

欧州・旧ソ連・アフリカにおける言語翻訳技術の研究  
開発等の動向調査

平成 25 年 7 月 31 日  
情報通信研究機構 欧州連携センター

# 目次

はじめに（基本構想・調査方法） .....	1
報告書全体の概要（報告書第一部と第二部のまとめ） .....	1
第一部 欧州における機械翻訳及び音声技術の研究開発の動向 .....	5
はじめに .....	5
第一章 欧州における機械翻訳技術及び音声技術の利用状況 .....	5
アンケート調査：フランス .....	5
調査結果 .....	6
調査結果のまとめ .....	9
第二章 欧州における機械翻訳技術及び音声技術の研究開発動向 .....	10
第一節 欧州連合の第七次枠組計画と競争・技術革新枠組計画 .....	10
1. FP7 の作業プログラムにおける機械翻訳及び音声技術の研究開発支援動向 ..	10
2. FP7 における研究プロジェクト .....	12
3. CIP の作業プログラムにおける機械翻訳及び音声技術の研究開発支援動向 ..	12
4. CIP における研究プロジェクト .....	13
第二節 英国 .....	14
エジンバラ大学 .....	14
ヒアリング議事録 エジンバラ大学統計機械翻訳グループ .....	16
ヒアリング議事録 エジンバラ大学 言語・認知・計算研究院 .....	20
第三節 フランス .....	28
グルノーブル情報学研究所 機械翻訳・自動言語・音声処理研究グループ .....	28
CNRS LIMSI 音声言語処理研究グループ .....	29
ヴォカピアリサーチ .....	30
ヒアリング議事録 ヴォカピアリサーチ .....	30
第四節 ドイツ .....	37
ドイツ人工知能研究センター 言語技術研究所 .....	38
ヒアリング議事録 ドイツ人工知能研究センター .....	38
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 .....	48
アーヘン・ライン=ヴィストファーレン工科大学 人間言語技術・パターン認識 研究グループ .....	48
第五節 その他の欧州諸国 .....	49
イタリア：ブルーノ・ケスラー基金（FBK） .....	49

アイルランド：次世代ローカライゼーションセンター .....	50
第三章 欧州企業による機械翻訳技術及び音声技術の商用化動向 .....	51
英国 .....	51
フランス .....	52
ドイツ .....	53
その他の欧州諸国 .....	55
ハンガリー .....	55
オーストリア .....	55
第四章 その他の動向 .....	56
第一節 欧州における機械翻訳及び音声技術関連の話題となる出来事 .....	56
視察レポート：デジタル技術の祭典「フューチャー・アン・セーヌ」における機 械翻訳及び音声技術の開発動向 .....	56
第二節 欧州特許庁とグーグル社による多言語特許翻訳の動向 .....	63
欧州特許庁の概要 .....	63
欧州特許庁と世界知的所有権機関 .....	64
グーグル社との提携関係の進展 .....	65
第三節 欧州におけるクラウドソース翻訳及び翻訳支援技術の研究開発の動向 ..	69
A) EU の動向 .....	69
B) 英国の動向 .....	69
C) アイルランドの動向 .....	71
D) その他 .....	72
第一部のまとめ（欧州の市場、技術、研究体制） .....	72
第二部 旧ソ連及びアフリカ諸国における機械翻訳技術の研究開発等の動向 .....	75
第一章 旧ソ連及びアフリカにおける機械翻訳技術及び音声技術の実用化の状況 ..	76
第一節 オンライン無料機械翻訳・音声技術サービス .....	76
第二節 機械翻訳・音声技術ソフトウェア .....	80
第二章 旧ソ連及びアフリカ諸国における言語処理に関する優れた研究機関の概要	
88	
第一節 旧ソ連諸国 .....	88
ウクライナ：ウクライナ人工知能研究院音声認識研究グループ .....	89
ウクライナ：トリデント・ソフトウェア .....	89
ウズベキスタン：ユニコン・UZ .....	90
ラトビア：ティルド .....	90

第二節 アフリカ諸国 .....	93
南アフリカ共和国 : Merake 研究院 人間言語技術研究所 .....	93
ケニア : ナイロビ大学 コンピューティング・情報学部 .....	94
ナイジェリア : オバフェミ・アウオロウオ大学 コンピューティング・インテリ ジェントシステム研究グループ .....	95
ヒアリング議事録 ナイジェリア オバフェミ・アウオロウオ大学 .....	95
第三章 旧ソ連及びアフリカ諸国の言語処理に関する研究プロジェクト例 .....	100
第一節 アゼルバイジャン・ディルマンクプロジェクト .....	100
第二節 FP7 MEDER プロジェクト .....	101
第三節 FP7 VOICES プロジェクト .....	102
第四章 旧ソ連及びアフリカ諸国において需要のある言語技術の分野及び日本語と の翻訳需要の可能性のある分野 .....	103
第五章 旧ソ連及びアフリカ諸国における無線インフラの状況 .....	105
旧ソ連諸国 .....	105
アフリカ諸国 .....	107
第二部のまとめ .....	108
補足資料 .....	111
FP7 及び CIP における言語技術の研究プロジェクトのリスト .....	111

## はじめに（基本構想・調査方法）

### 基本構想

独立行政法人情報通信研究機構（以下、NICTという。）では、世界の言語の障壁を超えることを目的とし、ネットワーク型言語翻訳の研究開発を行い、音声翻訳国際研究コンソーシアムU-STAR<sup>1</sup>（Universal Speech Translation Advanced Research）を通じて、各国の研究機関との研究協力を図っている。その上、iPhoneやアンドロイド携帯電話上で動作する多言語音声翻訳アプリケーションVoiceTraや、複数デバイス用多言語音声翻訳アプリケーションVoiceTra4U-Mなどを開発し、異なる言語間のコミュニケーションを実現させており、現在の対応言語はアジアやヨーロッパの主要言語等、約 23 言語に及んでいる<sup>2</sup>。

欧州は、伝統的に多数の国家が存立し、自ずと多言語の使用がコミュニケーションの課題となってきた地域であり、また、単一市場の形成を目指す欧州連合（以下、EU という。）では、常に言語の障壁への対応を迫られる事情があり、言語翻訳技術への関心が非常に大きい。さらに、NICT が U-STAR とともに言語翻訳を世界的に展開していく上で、欧州近隣の旧ソ連及びアフリカ諸国は有力な候補地となり得る。欧州系の言語としては既に、英語、フランス語、ドイツ語、イタリア語、オランダ語、スペイン語、デンマーク語、ハンガリー語、ポーランド語、ポルトガル語、ロシア語、トルコ語が対応済であるが、今後、旧ソ連及びアフリカにおいて需要が見込まれる言語の動向を把握していく必要がある。

このため、本調査では、1) 欧州における機械翻訳及び音声技術の研究開発の動向（第一部）として、機械翻訳技術と音声技術の利用状況、同技術の研究開発の動向（EU による開発支援動向及び各国の研究機関の概要、研究プロジェクトの事例）、欧州企業による商用化状況、そして、その他の動向として、フランスで開催されたデジタル技術の祭典「フューチャー・アン・セーヌ」における同技術の研究開発動向（仏独共同研究プログラム「クアエロ」を中心に）、欧州特許庁の特許翻訳とクラウドソース翻訳の動向、そして、2) 旧ソ連及びアフリカにおける機械翻訳及び音声技術の研究開発の動向として（第二部）、テキスト及び音声機械翻訳の実用化の動向、言語処理の優れた研究機関

---

<sup>1</sup> <http://www.ustar-consortium.com/>

<sup>2</sup> なお、VoiceTra は平成 25 年 3 月 31 日に運用終了予定。NICT は音声翻訳の技術を FEAT (<http://www.feat-ltd.jp/>) にライセンスし、平成 24 年 12 月 17 日から VoiceTra+ の名称で同機能のアプリケーションを開始。

の概要、需要のある翻訳の分野、日本語との翻訳の需要、無線インフラの状況を調査した。

全体の概要に関しては、本報告書の冒頭に収録した第一部と第二部のまとめを見ていただきたい。

### **調査方法**

- ・ インターネット及び刊行物を利用した情報収集
- ・ 関係者とのヒアリング
- ・ ICT イベントの視察
- ・ アンケート調査

これら4つの方法を調査トピックに合わせて利用し、調査を実施した。

本報告書では、情報を入手したウェブサイトの URL を参考のため注に載せているが、これらの記事はサイト運営者の都合で随時移動、修正、削除される可能性がある。従って、本報告書の発表後、注に記された URL から情報源となった記事にアクセスできないことがありうることを、ここで前もって注記しておきたい。

関係者のヒアリングでは、英国のエジンバラ大学言語・認知・計算研究院、同大学統計機械翻訳グループ、ドイツ人工知能センター言語技術研究所、フランスの民間開発組織ヴォカピアリサーチ、ナイジェリアのオバフェミ・アウオロウオ大学コンピューティング・インテリジェントシステムグループにインタビューを実施し、その議事録を本報告書に収録した。

### 調査支援組織情報

#### **ONOSO**

住所 : 2 Boulevard Anatole France, 92100, Boulogne-Billancourt, FRANCE

電話番号 : 01 46 03 06 53(フランス国外から: 0033 1 46 03 06 53)

メールアドレス : k.ono@onosofr

担当 : 小野 浩太郎

## 報告書全体の概要（報告書第一部と第二部のまとめ）

第一部では、欧州における機械翻訳技術と音声技術の研究開発及び同技術に関わる動向について記した。以上の調査結果を踏まえ、1) 市場、2) 技術、3) 研究体制という3つの観点から欧州の動向を簡単に記す。

1) 欧州は多数の国家から構成され、多数の言語が使用されている地域であるので、機械翻訳技術は自然に必要とされており、EUのFP7やCIP、そして、各国内で研究開発が積極的に促進されている。EUは言語の障壁を乗り越え、欧州単一市場創設のために、機械翻訳技術が必要であると考えており、FP7やCIPには言語技術に特化した公募枠がある（例：FP7 ICT-2009.2.2: 言語を基盤とするインタラクション、FP7 ICT-2011.4.2: 言語技術、FP7 ICT-2013.4.1: コンテンツ分析と言語技術、CIP 目標 5.1: 多言語ウェブのための機械翻訳、CIP 目標 6.1: オープン言語インフラストラクチャ）。

このような欧州の事情は、米国の研究機関等が同技術の研究開発を進める事情とは異なる。ヒアリング調査を行ったエジンバラ大学統計機械翻訳グループのフィリップ・コーエン教授によれば、米国では、軍事機関や諜報機関向けに機械翻訳技術や音声技術の研究開発が積極的に実施されている。このような傾向に対して、欧州は、米国よりも翻訳の市場規模が大きく、様々な分野、特に、ニュース、技術文書、ビジネス等で機械翻訳の需要があり、さらに、欧州議会や欧州委員会等のEU機関での機械翻訳の利用が進められている。従って、欧州では、軍事機関や諜報機関向けの研究開発は動機の1つに過ぎない。だが、欧州の研究開発機関（仏LIMSI等）も、米国防省のDARPA等の軍事機関や諜報機関向けの研究開発支援プログラムに参加しており、また、仏独共同研究開発プログラム「クアエロ」には仏政府の軍備総局も関与しており、軍事・諜報分野は欧州でも無視できない研究開発の動機の1つである。

特許翻訳分野では、欧州特許庁が米グーグル社と提携し、現在特許翻訳サービスが同庁のウェブサイト及びグーグルのプラットフォームで提供されている。現在、英語から、フランス語、ドイツ語、イタリア語、ポルトガル語、スペイン語、スウェーデン語、デンマーク語、オランダ語、フィンランド語、ハンガリー語、ノルウェー語、ポーランド語、中国語への翻訳、そして、これらの言語から英語への翻訳が可能である（現在14カ国語に対応している）。同庁の発表では、2014年末までに日本語、韓国語、ロシア語を含め、全32カ国語を対応させることを見込んでいる。だが、同庁が欧州の組織ではなく、米企業と提携したことが問題となり、この提携に対して否定的な反応をした欧州研究者もいたが、特許翻訳は翻訳の一つの需要に過ぎないとする研究者もいる。

フランスで行った機械翻訳技術と音声技術の利用状況のアンケート調査によれば、

PC でグーグル翻訳等のテキスト機械翻訳アプリケーションを使用している人の割合は高い（全体の 76%）。スマートフォンやタブレットでテキスト機械翻訳のモバイルアプリケーションを使用している人の割合は、調査対象者全体の 40%である。音声機械翻訳モバイルアプリケーションを利用している人は非常に少ない（全体の 1%）が、同アプリケーションに関心がある人の割合は全体の 45%に至る。また、音声機械翻訳以外の音声技術を使用するシステムやアプリケーション（音声転写アプリケーションや音声対話システム）を使用している人の数も少ない（全体の 14%）が、関心がある人は全体の 30%に昇る。このように、テキスト機械翻訳アプリケーションに比べると、音声翻訳を含める音声アプリケーションを実際に利用している人の割合は少ないが、関心がある人の割合が少ないわけではないので、技術の精度を高めるとともに、関心がある人の実際の利用を促すために、現在どのような技術やアプリケーションがあるのか、その内容を一般に周知させることも必要であると考えられる。

2) 技術に関して言えば、まず、エジンバラ大学の統計機械翻訳グループが開発した機械翻訳システムの MOSES の普及に注目すべきである。MOSES はオープンソースを理由に、欧州内外で、様々な研究機関（伊ブルーノ・ケスラー基金等）や民間組織（アジア・オンラインや SmartMATE）によって開発・利用されており、MOSES を利用する CAT (Computer Assisted Translation) ツール等を開発する FP7 プロジェクト (MATECAT や CASMACAT プロジェクト) も複数あり、また、MOSES の利用促進を目的とする FP7 プロジェクト (MOSESCORE プロジェクト) もあることから、今後も開発と利用が進められていくと考えられる。ついで、翻訳の精度の観点から、グーグル翻訳等の統計的手法だけに基づく機械翻訳システムへの批判がある。ヒアリング調査を行ったドイツ人工知能センター言語技術研究所のブズマン氏によれば、グーグル等の統計翻訳よりも、統計翻訳と規則（ルール）ベース翻訳を組み合わせた EuroMatrisPlus (FP7 プロジェクト) の翻訳システムの方が、翻訳の精度が高く、優れている。

クラウドソース翻訳に関しては、欧州委員会が「クラウドソーシング翻訳」と呼ばれる報告書を 2012 年 8 月に発表し、その概要と世界的な動向について報告しており、関心の高さが伺える上に、FP7 でも、クラウドソース翻訳に関わる CAT 技術が助成されている (MATECAT や CASMACAT プロジェクト)。また、欧州では、特にアイルランドのロゼッタ基金が、クラウドソース翻訳の活動に積極的である。

3) 研究体制に関しては、欧州各国には優れた研究機関があるとともに、これらの機関は FP7 のプロジェクト等を通して、しばしば共同で研究開発を実施していることが注目すべき点である。研究プロジェクトだけでなく、例えば、フランスの LIMISI とドイツの RWTH アーヘン大学とカールスルーエ技術研究院が共同で、多言語・マルチメディア

ィア技術研究院 (IMMI) という合同研究ユニットを創設している。また、META-NET という欧州の研究機関の提携を促進するアライアンスも設立されている (ドイツ人工知能センター言語技術研究所がコーディネーター)。機械翻訳技術は、必然的に他国の言語に関わる技術なので、他国の機関と提携して、研究開発を進めることが非常に重要である。従って、技術内容だけでなく、機械翻訳技術の研究提携の制度や提携活動の実行の面でも、FP7 等で他国との共同研究の経験が豊富な欧州から学ぶことは多いのではない。

第二部では、旧ソ連とアフリカ諸国における言語技術の実用化動向、研究開発動向、言語技術の需要及び無線インフラ状況を概観した。

実用化動向については、多くの企業によって、旧ソ連及びアフリカ諸国の言語のオンライン機械翻訳サービスが提供され、アプリケーションが開発されているが、その多くはグーグル社を筆頭に米企業であるが、欧州企業も少なからずある (英 SDL、ロシア・Yandex、ウクライナ・Online.ua、英 ATA、英 CAPITA、仏 Systran、仏ヴォカピアリサーチ、独 Languatec、オーストリア・Sonico、ハンガリー・Morphologic、ラトビア・ティルド、ウクライナ・トリデント)。対応言語に関しては、アラビア語に対応するサービスやアプリケーションが圧倒的に多いが、旧ソ連諸国であり、かつ EU 加盟国の言語、すなわち、ラトビア語、エストニア語、リトアニア語や、人口が多いウクライナ語に対応するサービスやアプリケーションも多くある。アラビア語以外のアフリカ言語に関しては、機械翻訳に対応し、実用化されている言語 (アフリカーンス語、ハウサ語、ソマリ語、スワヒリ語、ヨルバ語等) はあるものの、アプリケーションやサービスの数は限られる。

研究開発機関の動向に関しては、旧ソ連諸国では、ウクライナのトリデント社とラトビアのティルド社が欧州プロジェクトに参加しており、これらの国の言語に対応する機械翻訳技術の研究開発及び商用化を積極的に実施している。特に、ティルド社は多くの FP7 及び CIP のプロジェクトに参加しており、特にバルト三国の言語に関しては機械翻訳技術の分野で優れた研究機関の一つである。研究プロジェクトの動向に関しては、アゼルバイジャン政府が助成しているディルマンクプロジェクトが注目される。これは言語技術の包括的な研究プログラムであり、テキスト機械翻訳だけでなく、音声認識や音声翻訳システムの開発も実施している。アフリカ諸国では、南アフリカ共和国の Mareke 研究院が積極的に音声技術の開発を実施している。同国の公用語は 11 言語あり、政府機関向け等に音声技術の需要がある。また、ナイジェリアのオバフェミ・アウオロウオ大学では、現地の言語の 1 つであるヨルバ語を中心とする言語技術の研究開発を包括的に実施している。

言語技術と翻訳の需要に関しては、外国語学習（発音練習向け音声合成システム等）や観光の分野は一般的に需要があるが、国の政治や経済、言語利用状況によって需要分野は異なる。アゼルバイジャン等の小規模の国では、他国の情報収集が機械翻訳の研究開発の主な動機の1つである。また、旧ソ連諸国では、英語以外の欧米言語（仏語や独語）よりも、トルコ語やロシア語の方が需要がある場合がある（特に、アゼルバイジャン）。また、アフリカ諸国には非常に多くの言語があり、機械翻訳システムの前に、インターネットや携帯電話サービスをアフリカ言語で使用できるように、フォントやキーボード、ワープロソフト、音声合成技術の開発が必要とされる場合がある。農業国では、農業に関する情報を現地の言語で伝達する必要がある。

日本語との翻訳需要については、観光や外国語学習の他、アゼルバイジャン等の石油産出国では、ビジネス分野で機械翻訳の需要がある。また、日本のポップカルチャーの世界的な人気にも注目すれば、文化面で翻訳の需要の可能性はある。なお、ヒアリング調査を行ったアゼルバイジャンのディルマンクプロジェクトチームとナイジェリアのオバフェミ・アウォロウオ大学の研究員は、日本語へ機械翻訳システムを対応させる開発はまだ行っていないものの、日本の研究機関との提携開発には意欲的であった。

無線インフラ状況に関しては、旧ソ連及びアフリカ諸国の多くで携帯電話網が高い割合で人口をカバーしているが、3G 網に関しては国によって異なる。また、LTE 網の展開は、同地域でもすでに幾つかの国で開始されている（ウズベキスタン、エストニア、リトアニア、ラトビア、キルギスタン、アルメニア、ロシア、モルドバ、アゼルバイジャン、タジキスタン、カザフスタン、アンゴラ、ナミビア、タンザンニア、モーリシャス、南アフリカ共和国、ウガンダ）。

より詳しい情報については、報告書の本文を見ていただきたい。

## 第一部 欧州における機械翻訳及び音声技術の研究開発の動向

### はじめに

第一部では、機械翻訳技術及び音声技術の研究開発を中心に、EU 加盟国の動向について記す。第一章では、機械翻訳技術と音声技術の実際の利用状況を知るため、フランスでアンケート調査を実施し、その調査結果を記した。第二章では、同技術の研究開発動向を EU と欧州諸国の動向に分けて調査した。前者に関しては、特に第七次枠組計画（FP7）と競争・技術革新枠組計画（CIP）に研究開発動向（研究助成と研究プロジェクトの動向）を調査し、後者に関しては、英仏独の欧州主要国とイタリア、アイルランドにおける研究機関の研究体制と研究プロジェクトについて記した。第三章では、欧州企業による同技術の商用化動向を調べ、どのような製品が開発され、実際に商用化されているか記した。第四章では、その他の動向として、まず、フランスのデジタル技術の祭典「フューチャー・アン・セヌ」を視察し、同技術の開発動向を仏独共同開発プログラム「クアエロ」を中心に記し、ついで、欧州特許庁とグーグル社の特許翻訳分野での提携動向、最後に、欧州におけるクラウドソース翻訳及び CAT 技術（Computer Assisted Translation）の開発動向について記した。

インターネットや刊行物を通じた調査を補完するため、英国のエジンバラ大学言語・認知・計算研究院、同大学統計機械翻訳グループ、ドイツ人工知能センター言語技術研究所、フランスの民間開発組織ヴォカピアリサーチでヒアリング調査を実施し、その議事録を収録した。

### 第一章 欧州における機械翻訳技術及び音声技術の利用状況

#### アンケート調査：フランス

欧州のスマートフォンとタブレットのユーザーが、実際に機械翻訳技術（テキスト及び音声）と音声技術をどのように利用しているか、また、どのような技術やアプリケーションが実際に利用されているか明らかにするために、アンケート調査をフランスで実施した。

調査対象：スマートフォンか、タブレットを所有しているフランス在住者 50 名（性別、年齢、国籍を問わず）

## 調査結果

調査対象者：男性 27名：女性 23名  
年齢層：10代 6名：20代 18名：30代 20名：40代 4名：50代 1名：60歳以上 1名

### A) 音声機械翻訳アプリケーションの利用状況

1) スマートフォンか、タブレットで音声機械翻訳モバイルアプリケーションを使用している人の数： 2名（50名中）：約1%

2) スマートフォンか、タブレットで音声機械翻訳モバイルアプリケーションを使用している人の a) 主な使用目的（複数回答可）、b) 使用場所（複数回答可）、d) 使用しているアプリケーション、e) 対象言語（複数回答可）、f) 使用頻度

a) 勉強 1名：仕事 1名：旅行 0名：日常生活 1名

：その他 1名（外国語の発音練習）

b) 自宅 2名：学校 0名：仕事場 1名：通り 1名

：その他 0名

c) グーグル翻訳 1名：不明 1名

d) 対象言語：英語-仏語 1名：仏語-日本語 1名：仏語-ポルトガル語 1名

e) 1日に複数回 0名：1日に一回 0名：週に複数回 1名：週に一回 1名：一ヶ月に一回 0名：一ヶ月に一回以下 0名

3) 現在スマートフォンか、タブレットで音声機械翻訳のモバイルアプリケーションを使用していないが、この種のアプリケーションに関心がある人の数： 22名（48名中）：約45%

4) 現在スマートフォンか、タブレットで音声機械翻訳のモバイルアプリケーションを使用していないが、この種のアプリケーションに関心がある人の使用したい目的（複数回答可）：

勉強 7名：仕事 9名：旅行 15名：日常生活 21名

5) PC で音声機械翻訳アプリケーションを使用している人の数： 6名

6) PC で音声機械翻訳のアプリケーションを使用している人の使用目的  
勉強 1名：仕事 3名：日常生活 2名

## **B) テキスト機械翻訳アプリケーションの利用状況**

7) スマートフォンか、タブレットでテキスト機械翻訳モバイルアプリケーションを使用している人の数： 20名 (50名中) :40%

8) スマートフォンか、タブレットで、テキスト機械翻訳モバイルアプリケーションを使用している人の a) 主な使用目的 (複数回答可)、b) 使用場所 (複数回答可)、c) 使用しているアプリケーション (複数回答可)、d) 対象言語 (複数回答可)、e) 使用頻度

a) 勉強 7名：仕事 7名：旅行 4名：日常生活 12名

b) 自宅 15名：学校 3名：仕事場 4名：通り 1名：レストラン 1名：飛行機の乗り換え中：1名

c) グーグル翻訳 15名：iTranslate 2名：その他 1名

d) 対象言語：英語-仏語 16名：仏語-スロバキア語 1名： 仏語-日本語 1名： 仏語-スペイン語 1名： 英語-スロバキア語 1名

e) 1日に複数回 5名：1日に一回 2名：週に複数回 3名：週に一回 5名：一ヶ月に一回 1名：一ヶ月に一回以下 1名

9) 現在スマートフォンか、タブレットでテキスト機械翻訳のモバイルアプリケーションを使用していないが、この種のアプリケーションに関心がある人の数： 13名 (30名中) : 約43%

10) 現在スマートフォンか、タブレットでテキスト機械翻訳のモバイルアプリケーションを使用していないが、この種のアプリケーションに関心がある人の使用したい目的 (複数回答可) :

勉強 8名：仕事 7名：旅行 6名：日常生活 7名

11) PC でテキスト機械翻訳アプリケーションを使用している人の数： 38名 (50名

中) :76%

12) PC でテキスト機械翻訳アプリケーションを使用している人の使用目的 (複数回答可)

勉強 11名 : 仕事 15名 : 日常生活 18名

### **音声技術を利用するシステムやアプリケーションの利用状況**

13) スマートフォンか、タブレットで音声技術を利用するシステムやモバイルアプリケーションを使用している人の数: 7名 (50名中) : 約14%

14) スマートフォンか、タブレットで音声技術を利用するシステムやモバイルアプリケーションを使用している人の a) 主な使用目的、b) 使用場所 (複数回答可)、d) 使用しているアプリケーション (複数回答可)、e) 使用頻度

a) 音声対話システムによるスマートフォンの操作 4名 : ショートメッセージの作成  
2名 : インターネットサイト検索 1名

b) 自宅 7名 : 通り 1名

c) アップル社の Siri 3名 : iPhone の音声転写システム (ショートメッセージ向け) 2名 : アンドロイドの音声システム 1名

e) 1日に複数回 2名 : 1日に1回 0名 : 週に複数回 3名 : 週に1回 0名 : 1ヶ月に1回 3名 : 1ヶ月に1回以下 0名

15) 現在スマートフォンか、タブレットで音声技術を利用するシステムやモバイルアプリケーションを使用していない人のうちで、この種のアプリケーションに関心がある人の数: 13名 (43名中) : 約30%

16) 現在スマートフォンか、タブレットで音声技術を利用するシステムやモバイルアプリケーションを使用していないが、この種のアプリケーションに関心がある人の使用したい目的:

自動車内で GPS の操作 1名 : 勉強 1名 : インターネットショッピング 1名 : 日常生活 1名 : 旅行 1名 : スマートフォン上の地図アプリケーションの操作 1名 : 自動車内でのスマートフォンの操作 1名

17) PC で音声技術を利用するシステムやアプリケーションを使用している人の数： 2 名 (50 名中)

18) PC で音声技術を利用するシステムやアプリケーションを使用している人の使用目的 (複数回答可)

文章作成 1 名 : PC の操作 1 名

### 調査結果のまとめ

現在、フランスにおいて、音声機械翻訳モバイルアプリケーション、音声技術を使用するシステムやアプリケーションを使用している人の数は少ない (それぞれ全体の 1% と 14%)。だが、実際に使用してはいないものの、音声機械翻訳アプリケーションに関心がある人は全体の 45% であり、また同じく、実際に使用してはいないものの、音声技術を利用するシステムやアプリケーションに関心がある人は、全体の 30% である。テキスト機械翻訳モバイルアプリケーションを使用している人は、音声のアプリケーションに比べると割合が高い (全体の 40%)。PC で音声機械翻訳のアプリケーションを使用している人は少ない (全体の 12%)。PC でテキスト機械翻訳アプリケーションを使用している人の数は多い (全体の 76%)。PC で音声技術システムを使用している人の数は非常に少ない (全体の 1%)。

音声機械翻訳アプリケーションを現在使用していないが、関心がある人は、主に使用したい目的として、特に日常生活と旅行を挙げている。

テキスト機械翻訳モバイルアプリケーションに関しては、日常生活、勉強、仕事でよく使用されているが、レストランや空港 (飛行機の乗り換え中) という答えも見られた。場所は、自宅、学校、仕事場が多い。アプリケーションは Google 翻訳を使用している人がほとんどである。対象言語に関しては、仏語-英語がほとんどであり、調査地がフランスであることを考えれば、特に英語との翻訳需要が高いことが分かる。テキスト機械翻訳のモバイルアプリケーションを使用していない人のうちで、この種のアプリケーションに関心がある人の使用したい目的としては、勉強、仕事、日常生活、旅行がほぼ同じ割合で挙げられている。

音声技術を使用するシステムやアプリケーションの使用目的には、スマートフォンの操作、ショートメッセージの作成、インターネットサイトの検索と答えた人がいる。使

用場所は自宅の場合がほとんどである。現在音声技術を利用するシステムやモバイルアプリケーションを使用していない人のうちで、この種のアプリケーションに関心がある人の使用したい目的としては、自動車内で GPS の操作、勉強、インターネットショッピング、日常生活、旅行、スマートフォン上の地図アプリケーションの操作、自動車内でのスマートフォンの操作が挙げられている。

以上のように、音声機械翻訳音声を含めた音声技術を利用するシステムやアプリケーションは、テキスト機械翻訳アプリケーションと比べて、現在使用している人が少ない。だが、音声システムやアプリケーションに関心がある人は少ないわけではないので（音声機械翻訳アプリケーション：45%、その他の音声システム：30%）、技術の向上を高めるとともに、関心がある人の実際の利用を促すように、技術やアプリケーションの内容を一般に周知する必要があると考えられる。

## **第二章 欧州における機械翻訳技術及び音声技術の研究開発動向**

本章では、欧州における機械翻訳技術及び音声技術の研究開発動向を記す。まず、欧州連合（以下、EU と略す）の大型研究開発支援スキームである「第七次枠組計画」（以下、FP7 と略す）と、主に中小企業向け研究開発助成スキームである「競争・技術革新枠組計画」（以下、CIP と略す）における同技術の研究開発支援動向を記す。ついで、欧州諸国において、積極的に同分野で研究開発を実施している研究開発機関の概要と動向を記す。

### **第一節 欧州連合の第七次枠組計画と競争・技術革新枠組計画**

本節では、FP7 及びCIPにおける機械翻訳及び音声技術の研究開発支援動向及び研究事例について記す。両者のプログラムの研究事例については、欧州委員会のウェブサイト<sup>3</sup>を参考にした<sup>3</sup>。

#### **1. FP7 の作業プログラムにおける機械翻訳及び音声技術の研究開発支援動向**

FP7 は欧州の大型研究開発助成スキームで、2007 年から 2013 年までの 7 年間を対象とし、予算は約 505 億ユーロが見込まれている。2007 年より 2 年毎（最終年の 2013 年は別）に作業プログラムを作成し、公募の概要（公募する研究内容、採用基準、予算等）を定めている。2009-2010 年度作業プログラムより、EU内の言語の多様性がもたらす障害を越えるため、機械翻訳や音声技術に関する公募枠が独立し、重要視され始めている。以下に、作業プログラム毎に、機械翻訳技術及び音声入出力を利用する技術に関わる公

<sup>3</sup> [http://cordis.europa.eu/fp7/ict/language-tehnologies/portfolio\\_en.html](http://cordis.europa.eu/fp7/ict/language-tehnologies/portfolio_en.html)

募に関して記す。

**2007-2008 年度作業プログラム (WP2007-2008)**

WP2007-2008	
1)	
公募名	ICT-2007.2:認知システム・インタラクション・ロボティクス
対象技術	ロボット工学、認知システム、機械翻訳等
予算	1億9300万ユーロ（機械翻訳技術への支援に限られない）
2)	
公募名	ICT-2007.9:国際連携
対象技術	機械翻訳等、言語・音声技術（特にアラビア語を使用する地中海諸国との連携）
予算	1200万ユーロ（機械翻訳技術への支援に限られない）

**2009-2010 年度作業プログラム (WP2009-2010)**

WP2009-2010	
公募名	ICT-2009.2.2: 言語を基盤とするインタラクション
対象技術	機械翻訳技術等
予算	2600万ユーロ

**2011-2012 年度作業プログラム (WP2011-2012)**

WP2011-2012	
1)	
公募名	ICT-2011.4.1: デジタルコンテンツ及び言語に関する中小企業向け支援
対象技術	中小企業によるデジタルコンテンツ、言語リソースの共有促進
予算	3500万ユーロ（言語リソース共有の取り組みへの助成に限られない）
2)	
公募名	ICT-2011.4.2: 言語技術
対象技術	機械翻訳技術、音声対話システム技術等
予算	5000万ユーロ

## 2013 年度作業プログラム (WP2013)

WP2013	
1)	
公募名	ICT-2013.4.1: コンテンツ分析と言語技術
対象技術	視聴覚情報等のコンテンツ分析技術、高品質機械翻訳技術、音声対話技術
予算	2700 万ユーロ (機械翻訳と音声対話技術への助成に限られない)
2)	
公募名	ICT-2013.4.3:分析に関する中小企業向け支援
対象技術	中小企業によるオープンデータ、言語技術等の利用促進
予算	2000 万ユーロ

### 2. FP7 における研究プロジェクト

FP7 においては、非常に多くの言語技術に係る研究開発プロジェクトが実施されている。単なる研究開発だけでなく、同技術の分野で、欧州とアラビア語を利用する地中海諸国の連携を推進するプロジェクトもある (MEDERプロジェクト: 本報告書の第二部第三章第二節を参照のこと)。

FP7 の EuMatrixPlus プロジェクト、EU-bridge プロジェクト、Matecat プロジェクト、Casmacat プロジェクトについては、エジンバラ大学及びドイツ人工知能センターの研究者にヒアリング調査を実施し、詳しい動向についてお話を伺ったので、本報告書に収録したヒアリング議事録を参考にさせていただきたい (第一部第二章第二節英国、第三節ドイツ、第一部第四章第三節クラウドソース翻訳及び CAT 技術の開発動向)。

その他のプロジェクトについては、作業プログラム毎にリストを作成し、本報告書末に収録したので、そちらを参考にさせていただきたい。

### 3. CIP の作業プログラムにおける機械翻訳及び音声技術の研究開発支援動向

主に欧州の中小企業向けの研究開発助成スキームであるCIP<sup>4</sup>は2007年から2013年までの7年間を対象とし、予算は約36億ユーロが見込まれている。CIPは3つのプログラムから構成されるが、その一つがICTポリシー支援プログラム<sup>5</sup> (以下、ICT PSPと略す) である。ICT PSPでは、毎年作業プログラムが作成され、FP7の作業プログラムと同様

<sup>4</sup> <http://ec.europa.eu/cip/>

<sup>5</sup> [http://cordis.europa.eu/fp7/ict/language-technologies/ict-psp\\_en.html](http://cordis.europa.eu/fp7/ict/language-technologies/ict-psp_en.html)

に、公募の研究開発内容と採用基準等を示している。特に 2009 年度と 2010 年度の作業プログラムの「多言語ウェブ」という公募項目で、機械翻訳及び音声技術の公募が実施されている。同公募では、EU単一市場の形成を促進するために、ICTを基盤とする多言語間のコミュニケーションを強化することが目標とされている。

#### ICT PSP 2009 年度作業プログラム

ICT PSP 2009	テーマ 5 多言語ウェブ
公募名	目標 5.1: 多言語ウェブのための機械翻訳
対象技術	ウェブコンテンツ翻訳のための機械翻訳技術
予算	1400 万ユーロ (テーマ 5 全体の予算であり、目標 5.1 の予算にだけではない)

#### ICT PSP 2010 年度作業プログラム

ICT PSP 2010	テーマ 6 多言語ウェブ
予算	1600 万ユーロ (テーマ 6 全体)
1)	
公募名	目標 6.1: オープン言語インフラストラクチャ
対象技術	機械翻訳を利用するオンラインサービス提供の促進
2)	
公募名	目標 6.2: 多言語オンラインサービス
対象技術	目標 6.1 を補完し、現存する複数国にまたがるオンラインサービスの言語カバーを強化する。

#### 4. CIP における研究プロジェクト

CIPで実際に支援されている主な研究プロジェクトについては、リストを作成し、本報告書末に収録したので、そちらを参考にさせていただきたい（なお、2009 年度作業プログラムの機械翻訳の研究開発については、欧州委員会に専用のウェブサイトがある<sup>9)</sup>）。

6

[http://ec.europa.eu/information\\_society/apps/projects/index.cfm?obj\\_id=CIP-ICT-PSP.2009.5.1&menu=secondary](http://ec.europa.eu/information_society/apps/projects/index.cfm?obj_id=CIP-ICT-PSP.2009.5.1&menu=secondary)

## 第二節 英国

英国では、多くの大学研究機関で機械翻訳及び音声技術の研究開発<sup>7</sup>が実施されているが、特にスコットランド地方のエジンバラ大学<sup>8</sup>で同分野の研究が積極的に行われている。以下に、エジンバラ大学における研究開発動向を記す。また、同大学の研究員にヒアリング調査を実施した際の議事録も収録する。

### エジンバラ大学

エジンバラ大学では、統計機械翻訳グループで機械翻訳技術、言語・認知・計算研究院とスピーチ技術研究センターで音声技術の研究が実施されている。これらの組織は非常に多くの FP7 や CIP プロジェクトに参加し、また、多くのプロジェクトでコーディネーターを務めており、英国だけでなく、欧州、また世界的に見ても、言語技術分野の主要な研究機関である。

#### 1) 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<sup>9</sup>では、フィリップ・コーエン教授の下、主に統計的手法に基づく機械翻訳技術の研究開発が実施されている。同グループの人員は13名である（同組織のウェブサイトの情報）。同グループは機械翻訳技術に関わる多くの研究プロジェクトを実施しており<sup>10</sup>、本章第四節（ドイツ）で記すFP7 プロジェクトのEuroMatrixPlusにも参加している。本節では、同グループが開発した機械翻訳システムの「MOSES（モーゼス）」と関連プロジェクトについて記す。

### MOSES 及び関連プロジェクト

統計機械翻訳グループは、MOSES<sup>11</sup>というオープンソースの統計機械翻訳システムを開発し、供給していることで有名である。同システムは、2005年からフィリップ・コーエン教授の指導の下、ヒュウ・ホン氏により開発が開始され、現在に至る。このシステムはオープンソースなので、エジンバラ大学以外の研究者も利用でき、異なる研究組織と提携して、MOSESの開発を進めることができることを特徴とする。

MOSESを利用して、自らのニーズに適合する機械翻訳システムを開発することが可

---

<sup>7</sup> スピーチ技術に関しては、以下のサイトに研究機関がリストアップされているので、参考のこと。<http://www.ukspeech.org.uk/groups/>

<sup>8</sup> <http://www.ed.ac.uk/home>

<sup>9</sup> <http://www.statmt.org/ued/>

<sup>10</sup> <http://www.statmt.org/ued/?n=Public.HomePage>

<sup>11</sup> <http://www.statmt.org/moses/>

能である。例えば、アジア・オンライン社（シンガポール）<sup>12</sup>のアジア・オンラインや、英企業Capita Translation and Interpretingの SmartMate<sup>13</sup>という機械翻訳を必要に合わせてカスタマイズして提供するサービスにおいても、MOSESは利用されている<sup>14</sup>。なお、コーエン教授はアジア・オンラインの主任研究員を務めており、統計機械翻訳グループとアジア・オンラインの関係は非常に深い。さらに、MOSESは国際知的所有権機関<sup>15</sup>、欧州議会、欧州委員会でも利用が進められている。

MOSESは多くのFP7 プロジェクトにも関わっており、例えば、EuroMatrixPlusから支援を受けていた。また、FP7 のMATECAT<sup>16</sup>や、CASMACAT<sup>17</sup>というCAT技術の開発においても、MOSESは利用されている（両プロジェクトについては、第一部第四章第三節クラウドソーシング翻訳とCAT技術の研究開発動向において詳しく記す）。さらに、MOSESを普及させるために、MOSESCOREというMOSESの利用推進活動を目的とするFP7 プロジェクトも実施しており、同プロジェクトにはMOSESを開発したホン氏が関わっている。また、欧州とアラビア語圏の地中海諸国の機械翻訳技術分野での提携を促進するMEDERプロジェクトでも、MOSESが利用されている。CIPでも、MOSESを利用するLETMT!プロジェクトが実施されている<sup>18</sup>。以上のように、MOSESは今後もいっそう開発と利用が進められていくと考えられる。

#### FP7：MOSESCORE

略称	MOSESCORE
正式名称	拡大と開拓のためのMOSES オープンソース評価と提携支援
公募分野	ICT-2011.4.2
研究期間	2012年2月-2015年1月（36ヶ月）
全予算（EU 拠出金）	1250万ユーロ（1200万ユーロ）
プロジェクトコーディネーター	エジンバラ大学（英）
参加組織	プラハ・チャールズ大学（チェコ）、TAUS（蘭）、ブ

<sup>12</sup> <http://www.asiaonline.net/default.aspx>

<sup>13</sup> <https://www.smartmate.co/>

<http://www.capitatradingandinterpreting.com/smartmate/>

<sup>14</sup> <http://www.asiaonline.net/newsletters/201110.htm>

<sup>15</sup> [http://www.wipo.int/patentscope/en/news/pctdb/2011/news\\_0001.html](http://www.wipo.int/patentscope/en/news/pctdb/2011/news_0001.html)

<https://www3.wipo.int/patentscope/translate/translate.jsf>

<sup>16</sup> <http://www.matecat.com/>

<sup>17</sup> <http://www.casmacat.eu/>

<sup>18</sup> <http://project.letsmt.eu/>

	ルーノ・ケスラー基金（伊）、CAPITA（英）
ウェブサイト	<a href="http://www.statmt.org/mosescore/">http://www.statmt.org/mosescore/</a>
研究内容	MOSES の開発と利用を促進し、提携させるために、ワークショップや評価キャンペーン、機械翻訳マラソン等を実施する。

## ヒアリング議事録 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の研究開発動向を調査するため、同グループの責任者であるフィリップ・コーエン教授にヒアリング調査を行った。以下に、その議事録を収録する。

### **日程**

平成 25 年 6 月 14 日（木）

### **場所**

ONOSO 事務所（パリ・電話インタビュー）

### **先方（○）**

エジンバラ大学 統計機械翻訳グループ責任者：フィリップ・コーエン教授<sup>19</sup>

### **当方（△）**

ONOSO 研究員：小野 浩太郎

### **調査目的**

欧州における機械翻訳及び音声技術に関する動向を把握するために、英国のエジンバラ大学 統計機械翻訳グループにヒアリング調査を実施した。同機関は、機械翻訳システム MOSES を開発しており、欧州において同分野の主要な研究機関の一つである。同研究グループ責任者のフィリップ・コーエン教授にお話を伺い、1) MOSES の概要、2) MOSES を利用する研究プロジェクトの実施状況、3) 欧州における言語技術研究開発の特徴に質問した。

<sup>19</sup> コーエン教授は、仏シストラン社、アジア・オンライン、Research Triangle Institute でもコンサルティング活動を行っている。

## ヒアリングの概要

### 1) MOSESの概要

(△) あなたのグループでは、どの言語を対象に研究を実施しているか。

(○) 翻訳対象言語に関しては、研究者の関心や研究プロジェクトによって異なる。欧州プロジェクトでは、ドイツ語、フランス語、スペイン語、チェコ語、ロシア語等であり、米国のプロジェクトでは、中国語、アラビア語、日本語等である。

(△) あなたを中心とする統計機械翻訳グループがMOSES<sup>20</sup>を開発したと聞いた。他の機械翻訳システムと比べて、MOSESの長所とは何か。

(○) MOSES はオープンソースなので、そこから得られる利益が多い。例えば、各人が多くのデータベースでトレーニングできるので、性能を向上させることができる。

(△) なぜ、あなたはMOSESをオープンソースで発表したのか。オープンソースの方が有利な点が多いのか。

(○) その通り。オープンソースにより、他の研究組織と提携し、プロセスを共有して、研究を実施することが可能になる。例えば、我々はソースコードを共有し、米MIT、伊ブルーノ・ケスラー基金、チェコのプラハにある大学等、幾つかの研究グループと提携して研究を実施している。研究者が新しい要素を開発した場合には、彼らはMOSESにそれを統合することを望む。また、複数の組織で利用するので、互いに改善点をチェックし、テストすることが可能である。

(△) オープンソースの特に有利な点は、他の組織と提携して、研究開発することが可能になることであるか。

(○) その通り。

(△) MOSESは統計的手法に基づくのか。

(○) その通り。だが、当初は統計的手法だけに基づいていたが、開発が進み、文法等の他の要素と結びつけることが可能になった。

(△) MOSESはFP7プロジェクトEuroMatrixPlusから支援されたと聞いた。これは、MOSESが同プロジェクトにおいて改善されたという意味か。

(○) その通り。私が2005年にエジンバラ大学に来た時、統計機械翻訳グループでは別の機械翻訳システムを開発していたが、私の学生のホン(Hieu Hoang)<sup>21</sup>が全く新しい機械翻訳システムの開発を開始した。MOSESの開発に当たっては、欧州プロジェク

---

<sup>20</sup> <http://www.statmt.org/moses/?n=Moses.Overview>

<sup>21</sup> ウェブサイトの情報によれば、ホン氏は、現在エジンバラ大学統計機械翻訳グループの研究員であり、MOSESのツールを開発している。同氏は、アジア・オンラインでも働いていたことがあり、FP7のMOSESCOREに関わっている。

<http://statmt.org/~s0565741/>

トのEuroMatrix<sup>22</sup> (FP6 : 2006年9月 - 2009年9月)、EuroMatrixPlus<sup>23</sup> (FP7 : 2009年3月 - 2012年4月) 等から主な資金を供給されたが、当初は他の研究組織からも資金を得ていた<sup>24</sup>。

## 2) MOSESを利用する研究プロジェクトの実施状況 : FP7プロジェクト MATECAT及びCASMACAT

(△) FP7プロジェクトのMATECATとCASMACATでは、MOSESを利用し、CATツールを開発していると聞いた。これらのプロジェクトの概要について教えていただきたい。

(○) これらのプロジェクトは、主にプロの翻訳家を支援することを目的とする。現在、多くの翻訳サービス企業は機械翻訳を利用している。機械翻訳は完璧ではないので、翻訳者は、機械翻訳の後、翻訳された文章を修正する等、後編集を行わなければならない。これら二つのプロジェクトは後編集の作業を支援するためのツール、そして、翻訳者がタスクに適応させることが可能なシステムを開発している。特に、オープンソースのインターフェイスを開発しており、様々なツールや機能が統合されている。例えば、翻訳者が翻訳した文章を学習する機能があり、すぐに学習した文章が次の翻訳作業に反映される。以上の他にも、どのように翻訳者を支援するかという問題にも取り組んでいる。一言で言えば、インタラクティブな機械翻訳システムを開発することを目指している。

(△) これら二つのプロジェクトの違いは何か。

(○) これらはCATツールを開発する点では類似する。主な違いは、使用用途に由来する。MATCATは、翻訳会社が普段利用する機械翻訳向けのツールを開発することを目的としているので、利用スピードが速く、ユーザーが利用しやすいツールやシステム等を開発している。CASMACATでは、翻訳者の動作に注目し、よりインタラクティブな翻訳システムの研究開発を実施している。このため、我々は翻訳者が翻訳作業を行う時、どのような動作を行っているかについて研究し、どのように翻訳者を支援できるかを研究している。例えば、認知科学を踏まえ、翻訳作業中の目の動きについて研究している。

(△) これらの研究プロジェクトはクラウドソース翻訳を支援することにもつながるか。

(○) その通り。これらのプロジェクトはプロの翻訳者を支援することを目的としてお

---

<sup>22</sup> <http://www.euromatrix.net/>

<sup>23</sup> <http://www.euromatrixplus.net/>

<sup>24</sup> ウェブサイトの情報によると、チェコ共和国のチャールズ大学、伊ブルーノ・ケスラー基金、独RWTH・アーヘン大学、米メリーランド大学、米MIT、米DARPA、米NSF、米国防省からも資金を供給されている。

り、クラウドソース翻訳の支援が主な目的ではないが、それも可能である。開発されたツールをブラウザやクラウドソース翻訳のプラットフォームに装備することも可能である。

### 3) 欧州における言語技術研究開発の特徴

(△) 現在、言語技術の分野においてはグーグル等の米企業が世界を席巻していると思われる。例えば、欧州特許庁は特許翻訳に関してグーグルと提携することを決定している。この提携に対して、どのような反応が欧州の組織からあったのか。

(○) 不平を言う研究者がいたが、それは欧州にも欧州特許庁に対して機械翻訳サービスを提供できる類似する企業があるのに、特許庁は米国の企業と契約したからである。だが、他方で、他の国際機関、例えば、国際知的所有権機関<sup>25</sup>や欧州議会、欧州委員会の翻訳を所管する組織でも、MOSESに基づく機械翻訳システムを実際に利用しているか、利用しようとしており、MOSESを巡って様々な動きがある。

(△) あなたのお考えでは、米企業に比べて、欧州における言語技術分野の研究開発の強みや有利点は何か。

(○) 研究の質に関して言えば、あまり大きな違いはないと思う。だが、重要なことは欧州には機械翻訳の市場があることである。EU 公式言語は 23 カ国語にも昇る。従って、製品のローカライゼーションも含め、欧州には翻訳の非常に大きな需要があり、米国の市場よりも大きい。米国では、機械翻訳の研究は諜報機関向けに行われることが多い。つまり、機械翻訳は外国語の資料やウェブコンテンツを翻訳することに利用され、製品の情報等の取得が狙いである。従って、欧州と米国の違いは市場規模と種類が異なるだけであると思う。例えば、米国はアフガニスタンやパキスタンの言語の機械翻訳を必要としているが、欧州ではより精度の高いフランス語、スペイン語、ドイツ語、チェコ語、ポーランド語、ロシア語の機械翻訳の開発を目指している。

## 2) 言語・認知・計算研究院及びスピーチ技術研究センター

エジンバラ大学言語・認知・計算研究院<sup>26</sup>とスピーチ技術研究センター<sup>27</sup>は、音声技術の研究開発を実施している。前者はスティーブ・レネル教授が責任者を務めているが、同氏は後者のスタッフでもあり、両組織の関係は非常に深い。言語・認知・計算研究院の研究分野は、自然言語処理と計算言語学、音声言語処理、対話とマルチモーダルイン

<sup>25</sup> [http://www.wipo.int/patentscope/en/news/pctdb/2011/news\\_0001.html](http://www.wipo.int/patentscope/en/news/pctdb/2011/news_0001.html)  
<https://www3.wipo.int/patentscope/translate/translate.jsf>

<sup>26</sup> <http://www.ilcc.inf.ed.ac.uk/>

<sup>27</sup> <http://www.cstr.ed.ac.uk/>

タラクション、情報抽出・検索・プレゼンテーション、人間の認知の計算理論、教育・補助技術である<sup>28</sup>。スピーチ技術研究センターでは、言語・認知・計算研究院と同じく、音声認識、音声合成、音声信号処理、情報アクセス、マルチモーダルインターフェイス、音声対話システムの研究開発を実施しているが、より学際的な研究を実施している。

両組織は、同大学統計機械翻訳研究グループと同じく、多くのFP7プロジェクトに参加している。例えば、EMIME<sup>29</sup>、InEvent<sup>30</sup>、LISTA<sup>31</sup>、SCALE<sup>32</sup>、SIMPLE4ALL<sup>33</sup>、SSPNET<sup>34</sup>、Eu-Bridgeに参加している（Eu-Bridgeについては、両組織にヒアリングを行った際の議事録を参考のこと）。

また、エジンバラ大学の音声技術研究者は日本の研究機関とも関係が深い。特に、名古屋工業大学がFP7のEMIMEプロジェクトに参加しており、また、Udialogueというプロジェクト<sup>35</sup>をエジンバラ大学の研究者と実施している（Udialogueプロジェクトは、日本の独立行政法人科学技術振興機構における戦略的創造研究推進事業CRESTから助成されている）。また、日本の国立情報学研究所の山岸順一氏がスピーチ技術研究センターへ出向している。

エジンバラ大学言語・認知・研究院及びスピーチ技術研究センターにおける音声技術の研究開発動向を調べるためにヒアリング調査を実施した。以下に、その議事録を収録する。

#### ヒアリング議事録 エジンバラ大学 言語・認知・計算研究院

#### **日時**

2013年5月15日（水）午後12時30分～午後3時

#### **場所**

エジンバラ大学（英国・スコットランド地方）

#### **先方（O）**

エジンバラ大学 言語・認知・計算研究院所長 スティーブ・レネル氏

統計機械翻訳グループ ベイリー・ハドー氏

---

<sup>28</sup> より詳しくは以下のサイトを参考のこと。 <http://www.ilcc.inf.ed.ac.uk/research>

<sup>29</sup> <http://www.emime.org/>

<sup>30</sup> <http://www.inevent-project.eu/>

<sup>31</sup> <http://www.listening-talker.org/>

<sup>32</sup> <http://www.scale.uni-saarland.de/cms/>

<sup>33</sup> <http://simple4all.org/>

<sup>34</sup> <http://sspnet.eu/>

<sup>35</sup> <http://www.udialogue.org/ja/>

スピーチ技術研究センター 山岸順一氏 (NII から出向中)

## 当方 (△)

情報通信研究機構 欧州連携センター長 菱沼宏之

## ヒアリングの目的

欧州における機械翻訳及びスピーチ技術に関する動向を把握するために、英国のエジンバラ大学でヒアリング調査を実施した。同機関は、FP7 の EuBridge プロジェクトにも参加しており、欧州において同分野の主要な研究機関の一つである。同大学の言語・認知・計算研究院所長スティーブ・レネル氏、統計翻訳グループ研究員ベイリー・ハドリー氏、スピーチ技術センターへ日本の国立情報学研究所より出向中の山岸順一氏にお話を伺い、1) 同大学の研究体制の概要、2) 言語・認知・計算研究院における研究開発状況、3) EuBridge の概要、4) MOSES の概要、5) 日本との提携関係等について質問した。また、現地で無線 (3G) の伝送速度を計測したので、その数値も記す。

## ヒアリングの概要

### 1) エジンバラ大学における研究開発体制について

(△) エジンバラ大学 言語・認知・計算研究院の概要について教えて頂きたい。また、人員規模を教えて頂きたい。合わせて、スピーチ技術研究センターや統計機械翻訳グループについても教えて頂きたい。

(○：レネル氏) エジンバラ大学には、理工系 (Science)、人文系、医学系の 3 系統があり、合計 1 万 5000 人から 2 万人程度と思うが正確ではない。理工系の下に情報学 (Informatics) があり、約 500 名。うち言語・認知・計算研究院 (ILCC) に 125 名弱。内訳は、教授 25 名、研究員 40 人、ポスドク 50 名。言語・認知・計算研究院は垂直的なヒエラルキー構造をしている。逆に、スピーチ技術研究センター (CSTR) は、理工系 (情報学) と人文系 (言語学) にまたがった横断的な組織。また、統計機械翻訳グループは、関係する研究者が集まる組織で、やはり垂直的な仕組みとは異なる。

(△) NICT にも、6 つの垂直構造的な研究所 (Institute) と、4 つの組織横断的なセンター (Center) があり、非常に仕組みが似ている。英文の言い方まで同じである。

次に予算について教えて頂きたい。

(○) 言語・認知・計算研究院の年間予算は 700 万ポンド。これは、研究員及びポスドクの給与を含む。教授の給与は大学の人件費として直接支出されている。予算の源は、欧州委員会 (FP7 等) が最も多く、他に、英国政府 (英国研究会議等)、民間資金、

篤志家、米国政府、スイス政府、日本の科学技術振興機構（JST）の CREST などがある。CREST では名古屋工業大学と協力している。

## 2) 言語・認知・計算研究院（ILCC）における研究開発状況について

(△) 言語・認知・計算研究院の研究について教えて頂きたい。

(○：レネル氏) 自然言語処理と計算機による言語学、音声（スピーチ）言語処理、対話・マルチモーダル対話（仮想アバター・ジェスチャーなど）、情報抽出・検索・提示、人間の認知の計算理論、教育・（高齢者・障がい者）支援技術を研究している。

プロジェクトとしては、例えば、リアルタイムでストーリーを検知していく研究を行っている。これは、ツイッターでツイートされていく情報を、自然言語のアルゴリズムを用いて分析していくもので、流れていく話を追跡し、分析していく。これにより、新聞社の報道とツイッターとの先後関係・相互関係などが分かる。この分析は、地震や火山噴火の時にも使える。

また、欧州研究会議（ERC）と関係しているが、言語と視線やしぐさを組み合わせる研究を行っている。

(△) NICT でも同様の研究を行っている。

(○) 医療分野では、精神的にふさぎ込んでいる患者に対して、仮想エージェントにより、より良く眠れるようにしたり、目から言葉を伝えたりする研究も行っている。

教育分野でも幾つかプロジェクトを実施しており、自然言語対話を通して、単純な電子工学についての知識を教えるシステム、仮想文字とデジタルオブジェクトを利用して自閉症の子供の対話能力を向上させるシステム、また、大学の授業を共有できるソーシャルメディアの開発を行っている。

後により詳しく説明するが、我々は MOSES を開発、利用して、機械翻訳技術の開発を実施している。特に、我々は人間による通常の翻訳を支援するため、半自動インタフェースや、視覚認知、すなわち、目の動きなどを活用する研究も行っている。

スピーチ技術に関しては、英国国内の大きなプロジェクトとして、ナチュラルスピーチ技術の基礎研究を実施している。エジンバラ大学、シェフィールド大学、ケンブリッジ大学と協力して、音声認知及び音声合成のコア技術を開発している。音声認知に関しては、例えば、リッチコンテンツから成るメタデータを、個人に合ったものとする研究を実施している。問題は、どんな環境、どんな分野にも普遍的に対応するシステムを開発することである。研究成果は、例えば、放送事業者がアクセントや言語構造などを正確にするために使われる。音声合成に関しては、少数の訓練されたデータを用いて、異なる新しい音声へと変換し合成するシステムを開発している。例えば、

子供の会話を、言語を補うことにより、大人の会話に変えることが可能になる（音声デモを実演）。

我々はオープンソースソフトウェアでの開発に力を入れている。エジンバラ大学が開発したフェスティバル音声合成システム（HTS）<sup>36</sup>は統計的手法を利用する音声センテンス・システムであり、名古屋工業大学と協力している。また、スピアウト企業を3つ起業しており、音声グラフィックのアニメーションで、唇が動くものなどを開発している。シスコ、グーグル、EADS（仏エアバスの親企業）、東芝（ケンブリッジに研究所がある）などと協力している。

MOSES もオープンソースが特徴。オープンソースは色々な人々を結び付けるのが特徴である。MOSES は、新しい研究機関と民間企業のネットワークであるとともに、政府・非政府組織・EU も関係しており、多様性があり、グーグルやマイクロソフト等の機械翻訳の開発とは異なる。また、例えば、グーグル翻訳は特にオンライン翻訳に力を入れているが、機械翻訳分野はこれに留まらず、様々な部門、そして、様々な利用方法が可能であり、インターネットユーザーだけでなく、研究者、コンテンツサービスプロバイダー、人間の通訳者なども関心を有しており、非常に大きな分野である。機械翻訳には多くの言語のために多くのリソースを要求するので、グーグルやマイクロソフト等が全ての言語に対して同じように開発を進めるか分からない。特に欧州には多くの言語があり、欧州主要言語にはグーグルやマイクロソフトも力を入れているが、リソースの少ない欧州言語（チェコ語、デンマーク語、ギリシャ語等）に関しては、これらの企業の研究活動は明確ではない（ので、我々の出番がある）。MOSES は様々な研究プロジェクトで利用されており、欧州委員会でも翻訳・通訳に使用されている。

- (△) 音声認識・音声合成の研究対象言語は、欧州言語が多いのか。
- (○) 音声認識としては、英語が良く使われる。特に、欧州内のネイティブでない人向けの英語である。研究は、英語をはじめとした欧州言語に集中しており、欧州議会で使われたりする。音声合成では、文字を分析して、自動的に音声を作成する。
- (△) 音声認識は、音声から文字データへのインプット、音声合成は、文字データから音声へのアウトプットと捉えていると考えて良いか。
- (○) その通り。また、ドイツ人工知能研究センター（DFKI）の扱う EuroMatrix、EuroMatrixPlus は機械自動翻訳。欧州公用語 22 言語と 22 言語との翻訳の組合せは約 500 通りにもなる。どの言語が良くなるかは、どれだけのパラレル・データをコーパ

---

<sup>36</sup> <http://www.cstr.ed.ac.uk/projects/festival/>

スに収集できるかに依る。

(△) 当方の VoiceTra+や U-STAR も、アジアの言語を含めて、20 数か国語と 20 数か国語との組み合わせで 500 通り以上になる。

(○) 欧州言語以外でも、インドのウルドゥー語、アラビア語、中国語を研究している。また、国立情報学研究所 (NII)、シェフィールド大学、名古屋工業大学と協力している。EU の FP7 からは名古屋工業大学もファンディングを受けており、EMIME プロジェクトと呼ばれ、発生源の音声・言語 (例：日本語) から、相手先の音声・言語 (例：フランス語) に変換する。

なお、NICT の堀氏と協力しており、2013 年 2 月に 3 日間ほど来てもらった。また、同研究室のユー・チェン・ウー氏には 5 月 13 日 (月) 頃まで 3 か月ほど来てもらっていた。産業技術総合研究所 (AIST) ともポッドキャストで協力している。

### 3) EuBridgeについて

(○：レネル氏) EuBridge は、スピーチ (音声) 翻訳、スピーチから文字への翻訳を研究している。例えば、テレビ放送の字幕、自動講義録、欧州議会、融合通信 (TV 会議等) などに用いられる。なお同プロジェクトでは、音声合成技術に関する研究は実施していない。

Eu-Bridgeで開発を実施しているTEDトークスを紹介したい<sup>37</sup>。これは話されていく英語の単語を認識し、自動的にフランス語のセンテンスを作って字幕として表示していくことを可能にする (デモを実演)。これはNICTの堀氏のグループの技術と協力して組み合わせたもの。

(△) 翻訳されたのは分かりやすいフランス語に思える。しかし、英語の会話の前にフランス語への変換が行われるなど、少し変換のタイミングが早いようであるが。

(○) これはデモ用に予め作成しているため。なお、フランス人に言わせると自然なフランス語ではないそうだが、正確であり十分に実用的。

(△) EuBridge におけるエジンバラ大学の位置付けはどこにあるか。

(○) スピーチ (音声) 認識と機械翻訳とを組み合わせる点にある。機械翻訳では、シンタクティックで基本的なアプローチをしている。また、リソースが少ない言語についてもパラレルのコーパスを有している。音声認識に関しては、脳のネットワークをモデルとして活用する点がユニーク。

(△) ウルドゥー語、アラビア語、中国語の他に、欧州以外の言語は研究しているか。

---

<sup>37</sup> <https://wit3.fbk.eu/>

- (○) 音声合成に関しては、アフリカ・ナイジェリアのイボ語とマレー語を少し研究している。ナイジェリアから、研究者が同大学に来ていたのだ。
- (△) **EuBridge** の利点は何か。
- (○) 欧州ベースである点が挙げられる。
- (△) **EuBridge** は PC ベースかモバイルベースか。
- (○) 主として PC ベース。
- (△) 誰が研究結果を利用することを想定しているのか。EU 機関での利用を考えているか。
- (○) EU 機関はターゲットの一つであるが、一般の人でも利用できるだろう。
- (△) クラウドソースの翻訳は研究しているか。
- (○) あまり研究していないが、クラウドは面白く、訓練データの評価に利用している。

#### 4) MOSESについて

(○: ハドー氏) **MOSES** は、最新の技術を用いた統計ベースのモデル。単語（フレーズ）ベースと文章（センテンス）ベースで分析している。DIY（Do It Yourself）で、（グーグル等の）商用サービスと同じ技術を使用している。研究機関、製造業者、大学等の大量のデータを使用する。Asia Online、Kantan MT は良く知っている。バンコク・ベースであろう DIY SMT はあまり詳しくないが、英国の SmartMate は知っている。

**MOSES** は、オープンソースのライセンス（LGPL）であり、これを利用して人々が商用サービスを構築できる。2012 年から EU の EP7 の下で開発が支援されている。

- (△) **MOSES** のようなオープンソースのうまく行っている点と困難な点はどうか。
- (○) **MOSES** では、誤りが発生したり、うまくいかなかったりすれば、そのバージョンを取り除く。オープンソースなので混沌としているというのではなく、安定している。安定的な良いプラットフォームを目指している。多くの人が使うのでその圧力もある。

**MOSES** は、統計機械翻訳グループ（SMTG）だけが開発したとは言わないが、同グループが開発した主要なものである。フィリップ・コーエン教授が主催して、LGPL をフォーマライズした。

- (△) **MOSES** はクラウドソースと関係するか。
- (○) クラウドとは特に関係しない。**MOSES** は 2 ~5 人がコードを開発した、オープンソースのもの。

- (△) オープンソースにした理由は何か。EU が助成する条件だったのか。
- (○) フィリップ教授がオープンソースにした。EU は、オープンとすることに積極的だが、強制はしない。ただ、オープンにした方が現実的であり、相互運用性もある。また、オープンにすることにより知的所有権の問題を解決できた。

#### 5) 日本との協力について

- (△) 日本との協力関係について教えていただきたい。
- (○：山岸氏) 国立情報学研究所 (NII) とは MoU を締結し、ロボット等の分野で協力している。東京大学とは幸村 (こうむら) 教授と協力している。現在、日本から合計 5 人の研究者が来ている。特許庁からは間宮氏がサバティカル (研究休暇) で研究に来ている。
- (△) NII の坂内前理事長が本年 4 月に NICT の理事長に就任したところ。NICT で機械翻訳の研究にますます力が入ると考えている。
- (○：山岸氏) エジンバラ大学が U-STAR に支援・協力するのは易しいのではないか。ASR (音声認識) サーバは産総研のポッドキャスト経由で、また、フェスティバルという TTS (音声合成) のサーバが使えるはず。
- (○：レネル氏) 堀氏と話してみよう。
- (○：山岸氏) フェスティバルというサーバでは、英語、スペイン語、ドイツ語が使える。

#### 6) その他

- (△) グーグルをはじめとする米国の企業との競合・協力関係はどうか。
- (○：レネル氏) 米国企業にない点としては、例えば MOSES は、多くのステークホルダーから成り立っているし、様々な言語から成り立っている。学术界・商用開発者・多言語コンテンツプロバイダー、オンライン翻訳利用者、翻訳家といった多様性が強みであるし、また、①英語・スペイン語・ドイツ語・中国語など、リソースが豊富な言語、②チェコ語・デンマーク語・ギリシャ語など、二番手だが欧州連合の正式言語、③ゲール語・バスク語など、リソースが十分でない言語なども幅広く対象としている。  
なお、機械統計翻訳グループのトップであるフィリップ・コーエン教授が、グーグルと欧州特許庁との関係で名前が出てくるくらいであり、競合関係というよりは協力関係にあると思う。

#### 7) 無線 (3G) の速度 (XTREME LABS SPEEDTESTによる計測)

ロンドン市内 下り 平均 4,915Kbps、最大 5,181Kbps  
 上り 平均 43Kbps、最大 47Kbps

VoiceTra+翻訳時間 約 20～30 秒

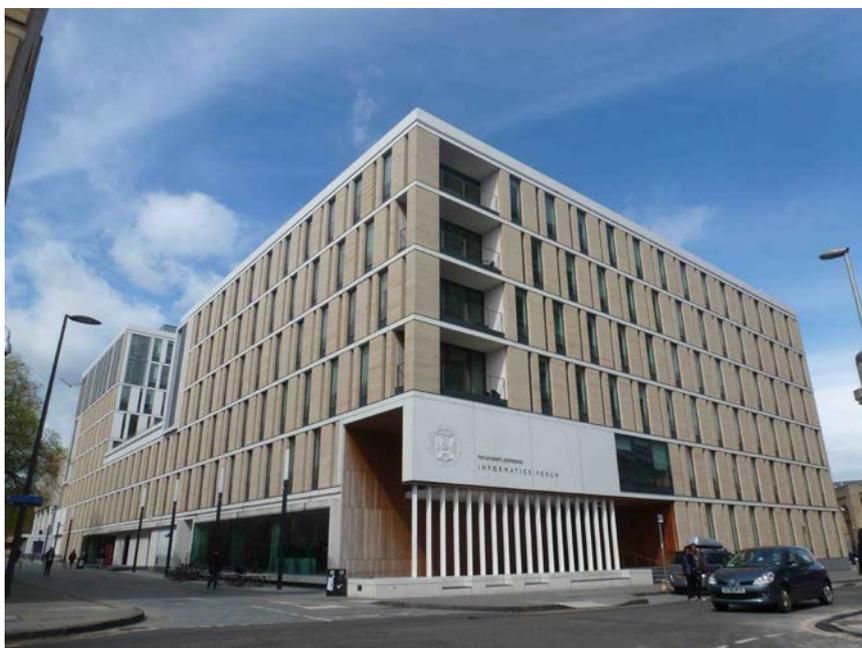
エジンバラ大学構内 下り 平均 326Kbps、最大 355Kbps  
 上り 平均 94Kbps、最大 106Kbps

VoiceTra+翻訳時間 約 5 秒

### 参考情報

略称	EuBridge
正式名称	異なる言語の間にかける橋
公募分野	ICT-2011.4.2
研究期間	2012年2月-2015年1月(36ヶ月)
全予算(EU 拠出金)	1051万ユーロ(787万ユーロ)
プロジェクトコーディネーター	カールスルーエ技術研究院(独)
参加組織	エジンバラ大学(英)、アルカテル(仏)、RWTHアーンヘン大学(独)、レッド・ビーメディア(英)、香港科学技術大学(中)、Pervoice(伊)、モバイル技術(独)、ポーランド日本情報工科大学(ポーランド)、ブルーノ・ケスラー基金(伊)、ACCIPIOプロジェクト(独)
ウェブサイト	<a href="http://www.eu-bridge.eu/">http://www.eu-bridge.eu/</a>
研究内容	様々なデジタル環境に <b>適応する</b> 音声翻訳技術の研究開発

## 写真（エジンバラ大学情報学フォーラム：ヒアリング実施場所の概観）



### 第三節 フランス

フランスでは、グルノーブル情報学研究所と国立科学研究センターの機械学・工学のための情報学研究所（LIMSI）で、機械翻訳及び音声技術研究開発が積極的に実施されている。以下に両組織の研究動向について記す。LIMSIはヴォカピアリサーチという民間企業と提携して、研究成果を商用化している。同企業にヒアリング調査を実施した際の議事録も合わせて収録する。

#### グルノーブル情報学研究所 機械翻訳・自動言語・音声処理研究グループ

グルノーブル情報学研究所<sup>38</sup>では情報学に関わる様々な研究活動が行われているが、機械翻訳・言語と音声自動処理研究グループ（GETALP）<sup>39</sup>で、機械翻訳と音声技術の研究開発が実施されている。

GETALPにおける機械翻訳分野の研究開発の歴史は長く、前身機関においては、1961年に研究に着手された。同グループの責任者はブザシエール・ローラン氏である。研究テーマは、機械翻訳とCAT、自然言語処理とプラットフォーム、言語データリソース収集と構築、情報システムローカライゼーション、音声・話者・音・地域語自動認識、

<sup>38</sup> <http://www.liglab.fr/spip.php?article107>

<sup>39</sup> <http://www.getalp.org/xwiki/bin/view/Main/>

知覚環境内における音分析とインタラクションである。

同グループは多くの研究プロジェクトを実施しているが、現在はFP7等の欧州プロジェクトよりも、フランス国内のプロジェクトを多く実施している<sup>40</sup>。機械翻訳技術関連の研究開発の他、言語技術をハンディキャップ支援へ応用するプロジェクト（HIS、Sweet-Home、TELMA）や、東南アジアの言語向け辞書製作プロジェクト（MOTAMOT）、オンライン辞書プロジェクト（Papillon）というプロジェクトがある。Papillonプロジェクトにおいては<sup>41</sup>、欧州言語だけでなく、アラビア語、エストニア語、ハウサ語、マレー語、日本語、中国語、インドネシア語、韓国語、タイ語、ベトナム語等、アフリカ言語、旧ソ連国言語、アジア言語も開発対象になっている。

国際提携関係に関しては、米国、ドイツ、ルーマニア、英国の研究機関とも提携しているが、特にアジア諸国との繋がりが深いことが特徴である。例えば、カンボジア、インド、マレーシア、タイ、ベトナム、日本の研究機関や企業と提携したことがある。日本に関しては、特に東京大学、京都大学、国立情報学研究所と提携して、研究プロジェクトを実施したことがある。

#### CNRS LIMSI 音声言語処理研究グループ

フランスの国立科学研究センター（CNRS）に設置されたLIMSI（機械学・工学のための情報学研究所）<sup>42</sup>には、機械学関係の他に、視聴覚技術を開発する研究グループが多く設置されている（聴覚・音響研究グループ、インタラクションのためのアーキテクチャとモデル研究グループ、情報・記述及び手話言語研究グループ、認知・知覚・利用研究グループ、RV&Z VENISE<sup>43</sup>（仮想空間・増大された実空間）研究グループ、音声処理研究グループ）。特に、音声処理研究グループで、音声技術に関する多くの研究開発（特に音声認識、言語及び話者特定、話者の感情状態の特定、視聴覚資料の構造化、音声機械翻訳、音声対話システムに関わる研究）が実施されている。

同研究グループの責任者は、ジャン＝リュック・ゴヴァン氏<sup>44</sup>である。研究テーマは、音響・語彙のモデル化、言語モデル化、放送番組の転写、会話転写、話者特定、感情認識、言語特定、聴覚分類、音声理解、音声対話システム、音声機械翻訳に係る非常に多くの研究プロジェクトを実施している<sup>45</sup>。現在は、FP7等の欧州プロジェクト

<sup>40</sup> <http://www.getalp.org/xwiki/bin/view/Projects/>

以上のサイトに現在進行中のプロジェクトがリストアップされている。

<sup>41</sup> <http://www.papillon-dictionary.org/papillon/Home.po>

<sup>42</sup> <http://www.limsi.fr/>

<sup>43</sup> <http://www.limsi.fr/Scientifique/tlp/>

<sup>44</sup> <http://www.limsi.fr/~gouvain/>

<sup>45</sup> <http://www.limsi.fr/tlp/projects.html>

に参加していないが、米国の研究開発助成プログラム「iARPA (Intelligence Advanced Research Projects Activity)」<sup>46</sup> (インテリジェントサービスに利用できる技術の開発支援プログラム)に参加している。多くのフランス国内のプロジェクトに関わっており、フランス国防省向けに音声認識技術の開発も実施している。また、ドイツのRWTHアーヘン大学とカールスルーエ技術研究院と、多言語・マルチメディア技術研究院<sup>47</sup>という合同研究グループを設立して、ドイツの研究機関と提携して機械翻訳及び音声技術の研究開発を実施している。なお、日本のNICTも参加しているU-STARに、2012年から参加している。

### ヴォカピアリサーチ

CNRS の LIMSI 音声処理研究グループは、仏研究開発企業ヴォカピアリサーチと緊密に提携し、研究開発を実施しており、ヴォカピアリサーチは LIMSI の開発成果を商用化させている。フランスにおける音声技術の研究開発動向を知るため、ヴォカピアリサーチにヒアリング調査を実施した。その議事録を以下に収録する。

### ヒアリング議事録 ヴォカピアリサーチ

#### **日程**

平成 25 年 6 月 10 日 午前 11 時～午後 12 時 30 分

#### **場所**

先方事務所 (フランス・パリ郊外南部・サッカレー)

#### **先方 (○)**

- ・ ヴォカピアリサーチ 最高経営責任者：ベルナール・プルー氏
- ・ CNRS LIMSI 音声言語処理グループ責任者 (ヴォカピアリサーチ 科学コンサルタント)：ジャン=リュック・ゴーヴァン氏

#### **当方 (△)**

- ・ NICT 欧州連携センター長：菱沼 宏之
- ・ ONOSO 研究員：小野 浩太郎

---

以上のサイトに現在進行中のプロジェクトがリストアップされている。

<sup>46</sup> <http://www.iarpa.gov/index.html>

<sup>47</sup> <http://www.immi-labs.org/?lang=en>

## 調査目的

欧州における機械翻訳及びスピーチ技術に関する動向を把握するために、フランスの民間研究開発組織ヴォカピアリサーチ<sup>48</sup>でヒアリング調査を実施した。同企業は、特にスピーチ技術の研究開発及び商用化を行っており、仏政府が助成しているICT大規模研究開発支援プログラム「クアエロ」に参加している。同企業の最高経営責任者（CEO）ベルナル・プルー氏と、同企業と技術提携しているCNRS<sup>49</sup> LIMSIS<sup>50</sup>の音声言語処理グループ責任者 ジャン=リュック・ゴーヴァン氏<sup>51</sup>にお話を伺い、1) ヴォカピアリサーチ社の概要、2) 同社が参加しているクアエロプログラムの動向、3) 言語技術分野における欧州の研究開発の強み、4) U-STARについて主に質問した。議事録末に、5) 現地の3G回線速度の計測結果を記入した。

## ヒアリングの概要

### 1) 先方組織の概要について

(△) ヴォカピアリサーチがあるサッカレーには、様々な研究組織が集まっているようである。サッカレーは先端工業地域のようなものであろうか。

(○：プルー氏) その通り。近くにはフランスで有名な理系エリート校、フランス理工科大学校（エコール・ポリテクニク）やパリ南大学、HEC（ビジネススクール）もある。現在さらに研究機関や企業が集まりつつあり、十年後には、特に ICT 部門の大規模な最先端工業地域となるだろう。

(△) ヴォカピアリサーチの概要について教えていただきたい。

(○) ヴォカピアリサーチは研究開発企業で、スピーチ技術分野のソフトウェア・アプリケーションを製品開発している。我々は、CNRS の LIMSIS と密接に技術提携しており、同機関と一緒に国際プロジェクトや国内プロジェクトに参加している。なお、LIMSIS の研究施設がヴォカピアリサーチのすぐ近くにある。我々のような小規模の企業にとって、優れた研究機関と提携することは非常に重要である。

(△) いつ、どのような経緯であなたの組織は設立されたのか。

(○) ヴォカピアリサーチは、私が 2000 年に創立した。それ以前、私はヴェクシス社という他の研究企業でスピーチ技術の開発等に携わっていた。この企業には、私が LIMSIS で博士課程を終了した後、1980 年に入社していた。だが、私とヴェクシス社の社

<sup>48</sup> <http://www.vocapia.com/index.html>

<sup>49</sup> CNRS はフランス国立科学研究センターのこと。

<sup>50</sup> LIMSIS は、「機械学・工学のための情報学研究所」のこと。 <http://www.limsi.fr/~gauvain/>

<sup>51</sup> ゴーヴァン氏は、CNRS LIMSIS 音声言語処理グループ責任者、ヴォカピアリサーチの科学コンサルタントとしての活動とともに、仏独合同の研究機関、情報多言語・マルチメディア研究院（IMMI）でも活動している。 <http://www.immi-labs.org/?lang=en>

長は異なる意見を持っており、私はヴェクシス内に、LIMSIと提携するヴェクシスリサーチという子会社を設立し、後に、会社名をヴェクシスと混同がないように、ヴォカピアリサーチと改名した。ヴェクシス社の社長が退職した際に、ヴェクシス社はベルタン社に売却され、ヴォカピアリサーチは独立することになった。現在、ベルタン社<sup>52</sup>は独自に音声認識技術を開発しているが、LIMSIと提携している我々の技術の精度には及ばないだろう。従って、LIMSIとの提携関係は、以前私がいた企業の頃から続いている。ちなみに、10年程前まで、私自身も研究活動を実施していたが、現在は企業経営に従事している。

(△) ヴォカピアという言葉には何か意味があるのか。

(○) 特にない。フランステレコムがオレンジという商標を持っているように、どの言語にも対応できる社名が必要だと私は考えた。

(△) 人員規模と総売上高について教えてほしい。

(○) 人員は10名以上いるが、秘書や事務員等はおらず、全てのスタッフは博士課程の学生であり、LIMSIから数名来ている。年間の総売上高は約150万ユーロである。

(△) 主な研究開発のテーマは何か。

(○) 特に多言語音声認識 (Speech-to-text) の研究を実施しているが、他にはオーディオ・セグメンテーション、話者特定、言語特定、字幕付与、音声-テキスト整列に関わる技術を開発している。

(△) どのような製品を商用化させているか。紹介していただきたい。

(○) 我々が開発しているソフトウェアは、主に2つの異なる音声データに向けてデザインされており、一方は放送向けであり、もう一方は電話会話向けである。これらは主に業務用であり、放送番組や電話会話の大量の音声データを転写するソフトウェアであり、音声認識等の技術を利用している。また、これらのソフトウェアを顧客のニーズ(例: 金融部門)に合わせて、カスタマイズするサービスも行っている。アプリケーションとしては、メディアモニタリング、電話会話転写、音声分析、自動字幕付与に関わるものを製品化している。なお、我々が開発している多くのアプリケーションは国防分野で使用されるものである。

我々はVoxSigma<sup>53</sup>というソフトウェアを販売しており、Linux向けの製品で、Windows向けではない。また、SaaSによりウェブサービスも提供している。なお、我々の製品は小売店等を通して販売されているわけではなく、ウェブ上で販売しているだけである。

(△) スピーチ技術の他、音声機械翻訳技術についても開発しているのか。

---

<sup>52</sup> <http://www.bertin.fr/>

<sup>53</sup> VoxSigma の技術参考情報 : <http://www.vocapia.com/voxsigma-speech-to-text.html>

(○) 音声機械翻訳技術についても、少しだが研究開発を行っている。この研究に関しては、フランスの国防省と契約しているので詳しくは話すことができないが、兵士が紛争地等で現地の人とコミュニケーションするのに、同技術は必要になる。

(△) どの言語を対象に研究を実施しているのか。

(○) 放送向けの音声認識に関しては、現在までのところ、アラビア語、オランダ語、英語、フランス語、フィンランド語、ドイツ語、ギリシア語、イタリア語、中国語（北京語）、ポーランド語、ポルトガル語、ルーマニア語、ロシア語、スペイン語を対象としている。近日中にブルガリア語、チェコ語、ウクライナ語に対応する予定であり、2013年末には、マルタ語を除く、全てのEU公式言語（23カ国語）に対応する予定である。英語に関しては、英国の英語と米国の英語に対応している。音声認識技術にとって、これら2つの英語はかなり異なる。またポルトガル語についても同様であり、ブラジルのポルトガル語は欧州のポルトガル語とかなり違う上に、ブラジルは市場が大きいので、重要である。日本語はまだ対応していないが可能である。

電話会話の音声認識については、対応言語数が少なく、7-8つの言語にしか対応していない。フランス語、英語、スペイン語、アラビア語等である。電話会話に関しては、使用できるコーパスが少なく、トレーニングが難しいからである。アラビア語に関してはレバノン国向けに対応させている。アラビア語は様々な国で使用されているので、国によって方言があり、電話会話に関しては国毎に対応する必要があり、放送向けのアラビア語のコーパスが使用できない。他には、小数の人々にしか使用されていない言語に関しても開発を進めているが、国防分野と関わるので、詳しくお話できない。

また、我々の音声認識技術は、仏ESTER（フランス語向け）<sup>54</sup>や蘭NBEST（オランダ語向け）や伊EVALITA（イタリア語向け）<sup>55</sup>という音声処理評価プログラムで高い評価を得ており、後者2つのプログラムでは、オランダやイタリアの研究機関が開発したソフトウェアよりも評価が高かった。

(△) CNRS LIMSI の他に、国内外に研究パートナーを持つか。

(○) 我々が参加しているクアエロプログラムには、ドイツのRWTHアヘン大学、カールスルーエ技術研究院もパートナーとして参加している。また、米国のDARPA（国防先端研究プロジェクト庁：Defense Advanced Research Projects Agency）<sup>56</sup>のプログラム、またIARPAプログラム<sup>57</sup>（Intelligence Advanced Research Projects Activity）と繋がりを持

<sup>54</sup> [http://www.afcp-parole.org/camp\\_eval\\_systemes\\_transcription/index.html](http://www.afcp-parole.org/camp_eval_systemes_transcription/index.html)

<sup>55</sup> <http://www.evalita.it/>

<sup>56</sup> <http://www.darpa.mil/>

<sup>57</sup> <http://www.iarpa.gov/index.html>

ち、IARPAプログラムのバベルプロジェクト<sup>58</sup>に参加している。

(△) EU の第 7 次枠組計画に参加しているか。

(○) まだ参加していない。

(△) 日本の研究機関の研究者からは、スピンオフ企業を設立することは簡単ではないと聞いている。もし助言があれば、いただきたい。

(○) 我々の企業は研究機関と強い繋がりを持つ企業であり、このような形態の研究開発企業としては、我々はフランスでパイオニアのような存在である。日本の状況が分からないと、助言するのは難しいが。

(△) 例えば、私（菱沼）が別の研究機関に話を聞いてきたドイツでは、大学、研究機関、企業が密接に連携しており、研究機関が大学と企業の橋渡しの役割を担っているそうである。つまり、大学は技術を商用化する術を知らないし、企業は大学の研究者とコンタクトを持つことが難しいが、研究機関が間に立ち、スピンオフ企業等の設立に関わることがある。

(○) 重要なことは、信頼関係をつくり、人間同士のコミュニケーションを取ることが重要であると思う。ちなみに、我々は CNRS の LIMSIS と技術提携しているが、収入の幾らかを同組織に払っている。こうして、Win-Win 関係を結ぶことにより、我々の間では様々な情報が滞ることなく伝達され、コミュニケーションがうまく取れている。

(△) 自らは商用化しない研究機関が、技術提携企業に対してその収入のうち幾らかを求めることについては、どのようにお考えか。

(○) 私はとてもいいことであると思う。

(△) LIMSIS には収入の何パーセントぐらいを支払っているのか。

(○) 残念ながら、割合を話すことはできない。

(△) 様々な教育機関、研究機関、企業が集まっているサッカーのような場所は、スピンオフ企業を設立するには適していると思う。

(○) その通りである。

(△) CNRS の LIMSIS はそれ自身では、技術を商用化することはしないのか。

(○) 否。そもそも商用化は彼らの仕事ではない。

## 2) クアエロの概要について

(△) ヴォカピアリサーチが参加しているクアエロプログラム<sup>59</sup>について教えていただ

<sup>58</sup> <http://www.iarpa.gov/Programs/ia/Babel/babel.html>

<sup>59</sup> クアエロプログラムは、仏独共同研究開発プログラムであり、特に多言語・マルチメディアコンテンツ自動分析に関わるアプリケーションを開発することを目的とする。特に、個人情報

きたい。クアエロは、欧州版グーグルを生み出すことが目的と言われ、2005年に発表され、2008年に開始されたが、発表当初ドイツの提携パートナーが同プログラムから離れていったと報道されている（2013年3月29日仏経済紙エコアの報道）。実際のところ、これはどのような事情であったのか。

(○) まず、この報道には誤りや誇張があることを指摘したい。クアエロは2005年にシラク仏大統領（当時）により発表された。報道では、欧州版グーグルを開発することを目的としていると言われたが、実際には、検索技術はクアエロの一部でしかない。また、ドイツの組織がクアエロから離れていったと言われているが、全部ではなく、2組織がクアエロに参加している。なお、離れていったドイツの組織は、クアエロとは別にテゼウス（THESEUS）<sup>60</sup>という別のプログラムをドイツで実施している。

(△) クアエロプログラムの予算について教えていただきたい。

(○) 全予算は5年間で約2億ユーロ、うち9900万ユーロを仏政府（フランス企業支援・イノベーション振興機構：OSEO）が拠出しており、独政府は助成してない。従って、クアエロは、ドイツの組織が参加するフランスのプロジェクトであると言える。このプロジェクトは、FP7等の欧州プロジェクトと比べると予算額が大規模である。このせいで、クアエロは2005年に発表されたが、欧州委員会が2008年に同プログラムの実施許可を出して実際に開始するまでに時間がかかってしまった。これは問題であった。

(△) なぜ欧州委員会の実施許可を待たねばならなかったのか。

(○) 政府が特定の民間企業に多額の助成金を拠出することは、市場競争を妨げる可能性があるため、欧州委員会の許可が必要であったという理由からである。

(△) クアエロプログラムでは、参加者は他の研究開発組織と提携して開発を実施しているのか。

(○) ヴォカピアリサーチはCNRS LIMSIと提携して研究開発を実施している。LIMSIに核となる技術を提供してもらい、それを我々が拡張する研究を行っている。だが、他の研究組織とは競争関係にある。クアエロでは、FP7のように、全参加組織が協力して研究開発を実施しているわけではなく、テーマ毎に、参加組織は他の参加組織と開発成

---

の保護に留意しながら、音声、文字、イメージ（写真等）、動画、音楽の自動処理アプリケーション（機械翻訳を含む）を開発している。仏独から全32組織（仏組織が主で、独からは2組織が参加）が参加し、大企業と中小企業を問わず産学を提携させて研究を実施させている。クアエロのウェブサイトの情報によれば、研究者の出身国は35カ国以上であり、参加研究者1000名以上である。全予算は2億ユーロで、うち9900万ユーロが仏政府（フランス企業支援・イノベーション振興機構：OSEO）から拠出されている（独政府からは助成されていない）。クアエロが発表された2005年には、欧州版グーグルを開発することが目的であると報道されていた。

<http://www.quaero.org/>

<sup>60</sup> <http://theseus.pt-dlr.de/en/>

果を比較し合うことになる。このような競争関係は研究開発を促進するのに重要である。例えば、クエアロでは、ドイツの研究組織も我々と同じくスピーチ技術に関する研究開発を実施している。このような協力・競争関係による研究開発は、米 DARPA と似ている。

(△) 研究成果の評価は誰が行っているのか。

(○) 主な評価機関は、仏政府の軍備総局と度量衡・実験国立研究所 (LNE)<sup>61</sup>である。

LNEは米標準・技術国立研究院 (NIST) と類似する機関である。なお米国では、DARPA とNISTが提携している。

(△) どのようなプロジェクトを実施しているのか。

(○) 我々はスピーチ技術に関する研究プロジェクトを実施している。対象となる言語は、欧州公式言語 23 言語である。

(△) 開発されたアプリケーションは、どのような形態で利用できるのか。

(○) PC やモバイル等、利用形態を問わない。

(△) 開発成果を商用化する予定であるか。

(○) その通り。実際にどれくらい売れるかはわからないが、商用化にチャレンジする予定である。

(△) フランスでは、パリ近郊で、2013年6月13日から23日まで、「Future en Seine」

(セーヌの未来) というデジタル技術の祭典が開催されると聞いたが、あなた方もそこへ参加するのか。

(○) その通り。クアエロプロジェクトは同イベントで会場にスタンドを持ち、開発成果のデモを行う予定である。

### 3) 欧州の言語技術研究開発の強みについて

(△) 言語技術については、グーグル等の米企業の勢力が強いと思うが、欧州の研究機関の強み、有利な点は何であるとお考えか。

(○: ゴーヴァン氏) 確かに、グーグルだけでなく、アップルやシスコも言語技術分野の研究開発を実施している。欧州の有利な点と言えば、欧州言語の研究開発に強いことである。重要なことは、欧州には多くの国があるので、言語技術が必要とされていることである。だが、技術的な観点から見れば、欧州が有利な点を持つわけではないと思う。

(△) 機械翻訳技術は欧州議会等の EU 機関で必要となると聞いたが。

(○) その通り。EU 機関では、翻訳料に多くの費用がかかっている。

---

<sup>61</sup> <http://www.lne.fr/>

#### 4) U-STARについて

(○：ゴーヴァン氏) 私はヴォカピアリサーチの科学コンサルタントであるが、本職は CNRS LIMSI 音声言語処理グループの責任者であり、NICT が主催している U-STAR に携わっている。LIMSI はフランスにおける U-STAR の責任者である。

(△：菱沼) 2012年6月のロンドンでの U-STAR の会合に参加していたのではないか。

(○) 参加していた。

(△) 我々は同じ会合に出席していたことになる。貴方をお見かけしたが、その時は声はかけなかった。

(○) (先方 VoiceTra+を操作しながら) VoiceTra+アプリケーションをフランスで使用すると、情報が日本とフランスを7回くらい行ったり来たりするので(フランスの話者→日本のサーバ→フランスのサーバ→日本のサーバ→フランスのサーバ→日本のサーバ→フランスの話者)、翻訳するのにとても時間がかかる。

(△) 我々の日本の研究者が現在フランスに専用サーバを設置して速度を改善しようとしており、私が現在場所探しを支援しているところである。

(○) このシステムは、日本で非常に中央制御化されているので、翻訳に非常に時間がかかるのが問題である。例えば、グーグル等は世界各地にサーバを置いており、翻訳するスピードが速い。また、翻訳のスピードを上げるには、音声データが取得されるのと同時に、そのデータを処理し始めること(ストリーミング)が重要である。話し終わってから処理を開始するのでは遅くなる。(注：U-STAR では一括送信と逐次送信のどちらの通信方式も選択でき、LIMISI は一括送信を採用している。なお、通信速度の遅い所では、逐次送信は途中で通信に失敗した場合の確認でさらに遅くなる可能性がある。)

なお、ヴォカピアリサーチでも、類似するアプリケーションを開発しているが、仏政府の軍備総局からこのアプリケーションを開発するために、資金を供給されており、商用化されていない。これは国防分野向けのアプリケーションである。

#### 5) 無線 (3G) の速度 (XTREME LABS SPEEDTESTによる計測)

ヴォカピアリサーチ近辺 (パリ郊外南部)

下り 平均 315Kbps、最大 1.437Kbps

上り 平均 296Kbps、最大 296Kbps

### 第四節 ドイツ

ドイツでも多くの研究機関で、機械翻訳及び音声技術に関する研究開発が実施されて

いる。以下に、ドイツ人工知能研究センター、カールスルーエ技術研究院、RWTH アーヘン大学の研究開発動向について記す。

#### ドイツ人工知能研究センター 言語技術研究所

ドイツ人工知能センターの言語技術研究所では、言語技術に関連する研究開発が積極的に実施されている。人員は、37名の研究者、17名のエンジニア、14名の大学の助手や客員研究者から構成される。同研究所の責任者は、ハンス・ウズコレイト氏<sup>62</sup>である。同研究機関は、META-NET (A Network of Excellence forgoing the Multilingual Europe Technology Alliance)<sup>63</sup>と呼ばれる欧州の言語技術の研究機関を提携させるアライアンスのコーディネーターを務めており、欧州の主要な研究機関の一つであると言える。主な研究テーマは、テキスト分析、機械翻訳、人間とロボットのインタラクションシステムである。研究プロジェクトに関しては、多くのFP7 プロジェクトに参加しており、EuroMatrixPlus<sup>64</sup>、QTLaunchpad<sup>65</sup>、T4ME NET<sup>66</sup> (META-NETを支援しているプロジェクト)、NiFTi<sup>67</sup> (人間とロボットのインタラクション) のコーディネーターを務め、その他、Excitement<sup>68</sup>、ALIZE<sup>69</sup>に参加している。欧州プロジェクトの他、多くのドイツ国内のプロジェクトにも参加し、またスピンオフ企業も3社設立している (ATTENSITY EUROPE<sup>70</sup>、ACROLINX'S<sup>71</sup>、YOCOY<sup>72</sup>)。YOCOY社は、中国語、英語、ドイツ語間のモバイル向け翻訳アプリ等を開発している。

ドイツ人工知能センター言語技術研究所における研究開発動向を知るために、ヒアリング調査を実施した。その議事録を以下に収録する。

#### ヒアリング議事録 ドイツ人工知能研究センター

##### **日程**

平成25年5月16日(木) 午後1時30分から午後3時30分

---

<sup>62</sup> <http://www.dfki.de/lt/card.php?id=201>

<sup>63</sup> <http://www.meta-net.eu/>

<sup>64</sup> <http://www.euromatrixplus.net/>

<sup>65</sup> <http://www.qt21.eu/launchpad/content/new-goal-quality-translation>

<sup>66</sup> <http://www.meta-net.eu/projects/t4me/>

<sup>67</sup> <http://www.nifti.eu/>

<sup>68</sup> <http://www.excitement-project.eu/>

<sup>69</sup> <http://www.aliz-e.org/>

<sup>70</sup> <http://www.attensity.com/home/>

<sup>71</sup> <http://www.acrolinx.de/>

<sup>72</sup> <http://www.yocoy.com/>

## 場所

ドイツ人工知能研究センター（ドイツ・ザールブリュッケン）

## 先方（○）

ドイツ人工知能研究センター・言語研究所 主任研究員 ステファン・ブズマン氏<sup>73</sup>

## 当方（△）

情報通信研究機構 欧州連携センター長：菱沼 宏之

## 調査目的

欧州における機械翻訳及びスピーチ技術に関する動向を把握するために、ドイツ人工知能研究センター・言語技術研究所でヒアリング調査を実施した。同機関は、FP7 の EuroMatrixPlus と META-NET のコーディネーターを務めており、欧州における同分野の主要な研究機関の一つである。同機関の主任研究員ステファン・ブズマン氏にお話を伺い、1) 同機関の概要、2) EuroMatrixPlus 及び 3) META-NET の概要、4) 欧州特許庁とグーグルの関係について主に質問した。また、5) 現地で無線（3G）の伝送速度も計測したので、参考情報として記す。

## ヒアリングの概要

- （1）ドイツ人工知能研究センターについて
- （△）（ヒアリング冒頭に NICT の概要を説明。VoiceTra+による音声翻訳を実演）
- （○）当方にも Yocoy（ヨコイ）<sup>74</sup>というスマートフォン上での翻訳を行うスピンオフ企業がある。主要対象分野は、旅行先の言語を知らない人向けに、タクシー、病院、買物、土産、レストラン、場所探しなど、20～40 の典型的な文章を入れている。
- （△）NICT のシステムもまずは観光分野を確立させた点では一緒であり、現在は他の分野に広げようとしている。なお、本日は1人なので実演できないが、U-STAR を使えば複数の人が iPhone を使って異なる言語で電話することができる。
- （○）スマートフォンでの音声翻訳の問題点は、雑音やエコーがあること。最近、SKYPE で一定程度の角度内であれば明確に音声を補足する技術が出たと聞いている。
- （△）ドイツ人工知能研究センター（DFKI）の概要を人数を含めて教えて頂きたい。
- （○）14 か国から、研究者が 37 名（うち、22 名がザールブリュッケン、15 名がベルリン）、エンジニアが 17 名、大学からの助手や客員研究者が 14 名いる。ドイツ人工

---

<sup>73</sup> [http://www.dfki.de/lt/  
http://www.dfki.de/lt/card.php?id=255](http://www.dfki.de/lt/http://www.dfki.de/lt/card.php?id=255)

<sup>74</sup> <http://www.yocoy.com/>

知能研究センターは、1988年に政府により設立され、まずカイザースロータン（本部）とザールブリュッケンに設立された。全体で10の研究所を有し、カイザースロータン、ザールブリュッケン、ベルリン、ブレーメンの4か所の規模が大きい。大学と協力しており、言語学や音声学から3～4人が来ている。

(△) ここザールラント大学の構内に研究機関が多いのはなぜか。貴方のいるドイツ人工知能センターだけでなく、ドイツの2大研究機関であるマックスプランク研究所とフラウンホーファー研究機構、さらには韓国科学技術研究院（KIST）まであり、全て独立した個別の建物に入居している。このようなことは、ドイツでは通常のことか。

(○) 大学構内に研究機関があると、研究成果であるアプリケーションを活かしたスピンオフ企業を設立しやすい。この点は政府が推奨しており、ドイツでは普通である。スピンオフ企業にとっても、低料金で場所を借りることができ、利益が一定額を超えた場合のみその一部納付すれば良いので、大学構内にあるのは有利である。大学構内では研究・教育・産業のシナジー効果があり、ドイツ人工知能研究センターは、産業界と研究者との架け橋となっている。基礎研究は政府やEUから資金が来る。応用研究はEUや産業界から資金が来る。プロトタイプを作り、ソリューションをカスタマイズしていく。

(△) 予算規模を教えてください。

(○) ドイツ科学技術省、ドイツ経済省、EU（実際は欧州委員会）、州政府、産業界等からの資金で、ザールブリュッケンの研究所の予算は年間440万ユーロ、ドイツ人工知能研究センター全体では3,600万ユーロの予算。資金源としては、85%が申請の必要な競争的資金で、10～15%は産業界からの固定的資金。1,000万ユーロ（27%）は、契約ベースで基礎研究に充てられる資金であるが、国際アドバイザリーボードの承認が必要。国際アドバイザリーボードは、助言だけでなく、申請された資金の承認も行う。EUからの資金は使用額の75%なので、残りの25%は産業界などから出してもらっている。

(△) 国際アドバイザリーボードは当方でも関心を持っている。次に、研究の概要を教えてください。

(○) 情報では、文字ベースでの概略抽出と音声質問・文字回答、多言語分野では文法・スタイルのチェックと翻訳、ナチュラルコミュニケーションでは話すロボットが救命に使われることもある。2002年の北イタリアの地震の際に歴史的建造物が崩壊して人間が中に入れずにロボットが入ったが、粉塵を巻き上げてしまい視界不良となることがある。入院中の子供と何週間も対話して遊ぶロボットもある。

(△) 翻訳の研究対象は欧州の言語のみか。

(○) 欧州の言語だけでなく、中国語や、以前は日本語も対象としていた。しかし、日本語は、統計的手法を使って文法ベースで研究しただけで、翻訳のスキーム全体は研究していない。JPSG（日本語句構造文法）で、辻井潤一先生に協力頂いていた。辻井先生はマンチェスター、スタンフォード、オスロといった色々な所で研究している。DELPH-IN というネットワークがあり、ドイツ人工知能研究センター、スタンフォード大学、NTT、東京大学、辻井研究室などが参加している。色々な文法を研究開発したが、日本語は文法だけを研究開発し、翻訳全体の研究開発はしていない。ドイツ語、英語、フランス語、イタリア語、スペイン語という強力な少数の言語と、ポーランド語、ロシア語、ブルガリア語といったソフトウェアが少ない言語を対象としていた。

(△) クラウドソース翻訳は研究しているか。

(○) EuroMatrixPlus のプロジェクトで研究している。だが、クラウドソースは、実際の翻訳よりも翻訳結果の評価のために使用している。ある人が話すことができるが文法などを正確には理解していない外国語について、グーグル翻訳は、文脈を適切に捉えておらず、良くない。しかし、コーパスをベースとすれば、色々な翻訳をみんなに評価してもらうことができる。みんなに翻訳してもらうよりも、どの翻訳が良いかをみんなに評価してもらう方が簡単（で効果的）である。DELPH-IN は、北京や上海ともつながりがあり、学生を交換留学させる契約もしている。主任教授の奥様は中国人であり、ここの学生にも中国人が多い。なお、留学生への教育は英語で行うが、2年くらい生活するためにはドイツ語が必要であり、留学生は習得していく。

## (2) EuroMatrixPlusについて

(△) EuroMatrixPlus の概要を教えてください。

(○) まず、音声でなく文字の翻訳であることを明確にしておきたい。統計ベース及び規則ベースの現在の翻訳技術にはそれぞれ難点があり、どう改善していくか、どう融合させて長所のみを取り入れたハイブリッド型にしていくかを研究した。異なるアプローチを組み合わせる利用者に提供する。約 500 ある欧州の言語間の翻訳の組み合わせでも、英語・ドイツ語間やフランス語・スペイン語間のようにその翻訳を対象とする技術の数が多いいものもあれば、デンマーク語・クロアチア語間やスロバキア語・リトアニア語間のようにその翻訳を対象とする技術の数が少ないものもある。評価するためのワークショップは毎年開催している。また、EuroMatrixPlus の枠組みの下で、欧州の機械翻訳のインフラとして MOSES というオープンソースの技術の開発が進展した。

(△) 昨日、英国エジンバラ大学で MOSES の関係者の話を聞いてきたところである。

ところで、EuroMatrixPlus の研究期間は終了したようだが、後継のプロジェクトがあるのか。

(○) 後継のプロジェクトはない。今後は、翻訳の誤りを分析して、少しずつ積み上げていくだけであり、センテンスが分からないときなど、より良くしていく。機械翻訳がなぜ失敗したのか統計的に見ていき、自動的に解決していく。誤りは少しの部分である。

ちなみに、QT Launchpad という、分野を特定し、高品質の翻訳を要求する、大きなプロジェクトがある。

(△) EuroMatrixPlus の予算の内訳はどのようなものか。

(○) 人・日で算定した研究者の給与、PC 等の設備、旅費、オーバーヘッド等である。

(△) EuroMatrixplus におけるドイツ人工知能研究センターの役割は何か。

(○) コーディネーターであり、EU とプロジェクト構成員との間に入る。自組織への研究予算を多くできるといったメリットがあるわけではないが、予算の配分を行うことができ、コーディネーター向けの特別な予算措置がある。

(△) 対象は欧州の言語のみか。

(○) 欧州の言語が中心であるが、ハイチ地震の際はアジアからのツイッターによるツイートが多くあり、統計翻訳のコーパスができたことがある。欧州の言語としては、主要言語だけでなく、リソースの少ない言語、例えばスロバキア語やブルガリア語などの中欧・東欧の言語にも力を入れている。英語・ブルガリア語間の翻訳を他の少数言語に活かしたり、英語・チェコ語間の翻訳と、チェコ語・スロバキア語間の翻訳を、英語・スロバキア語間の翻訳へ応用したりしている。

(△) EuroMatrixPlus の研究成果は、誰がどのように活用しているのか。

(○) ユーザは誰でも良い。ソフトウェアや科学論文にも活かされる。また、ソフトウェアとコーパスとをウェブサイトに公開している。活用手段として、PC はあるが、スマートフォンは想定していない。活用手段を想定したというより、技術を開発したつもりである。

(3) META、META-NET、T4ME NETについて

(△) META の特徴を教えてください。

(○) META とは、2013 年に再編成されたが、緩いつながりである。特段のゴールはなく、つながりを持たせることができる仕組みである。META SHARE プラットフォー

ム<sup>75</sup>という、メタデータを付ける目的のソフトウェアを技術開発する。情報やライセンス条件を設けており、データベースとしてはベルリンやエッセンなど3~4の異なるノードに置いている。

(△) 民間企業を含め多数が参加するので、知的財産権の問題が発生するのではないか。

(○) 知的財産権は META SHARE というソフトウェアに付与しているため、META に自由に参加することができても、知的財産権の問題は発生しない。

(△) META-NET について、META との違いを含めて、教えて頂きたい。

(○) META は誰でも加入できるアライアンスであるが、META - NET はそうではなく、全く異なる。META-NET は 34 か国の 60 の研究機関から成る。一方、META は参加者数が 727 と多く、欧州以外でも、中国が 6 機関、日本が 4 機関加盟している。日本からは、大日本印刷、Human Science、京都大学、産業技術総合研究所が参加している。META と META-NET の両者は、機械翻訳のアルゴリズムが目的というだけで、直接の関係はない。

META の参加者は META SHARE のデータや技術を自由に享受できる。利用者は META SHARE に対して代金を支払っても良いし、無料で使用しても良い。なお、META SHARE のライセンスは複雑であることから、ソフトウェアにライセンスを持たせた。

(△) META の後継はないのか。

(○) META の後継のプロジェクトはない。しかし、META では白書を公表している。なお、2014 年からホライゾン 2020 が始まるが、まだ確固とした具体的予定はない。

(△) T4ME NET について教えて頂きたい。

(○) T4ME NET とは EU のプロジェクトとしての名称であり、ブランド名としては一般的に META-NET と呼ばれている。META はこれとは異なり、関心があれば誰でも自由に参加できる緩い技術アライアンスであり、META-NET のようなプロジェクトのアライアンスではない。META は、参加者数が本日現在 727 と多く、参加者に対して特に何かを行うよう求めるものではないが、連絡先が分かるなど参加者がメリットを享受できる。META-NET は 2013 年までの予定で、現在進行中であり、EU からファンディングを受けている。

(4) 欧州特許庁とグーグルとの連携について

(△) 欧州特許庁とグーグルとの連携をどう捉えているか。

(○) 米国にサーバがあるグーグルに任せた場合、安全保障上の懸念がある。しかし、

---

<sup>75</sup> <http://www.meta-net.eu/meta-share>

ドイツ国内で、検索すること自体を「グーグルする」というくらいグーグルは普及しているのも事実である。また、欧州特許庁がグーグルに訓練させて良い翻訳にしていたのであろうが、全ての分野でグーグルが訓練していけるのであれば、EuroMatrixPlus は不要になりかねない。しかし、ドイツの人々はグーグル翻訳の品質に満足してはいない。グーグル翻訳では、動詞が分かれて別の場所に位置するドイツ語を、うまく英語等に翻訳できない。グーグルは大量のデータに物を言わせた統計翻訳であり、統計翻訳だけでなく規則（ルール）ベース翻訳もある EuroMatrixPlus になわれないと思う。EuroMatrixPlus は規則ベース翻訳なので、ドイツ語の動詞が分かれる点に対応できる。グーグル翻訳は、評価すると、一番ではないのである。また、例えば技術文書のように他人に渡したくない秘密文書の場合、グーグルの統計翻訳でなく、規則ベースで単語・単語ベースである EuroMatrixPlus が良い。多言語性というのも利点である。

米国は、国防総省の研究所 DARPA の資金によって機械翻訳を開発している。例えば CIA が 15 日間で外国の文書を調査する際に、言語が分からなければ理解できないので、統計翻訳によって早ければ翌日には翻訳してしまう。しかし、欧州は米国とは異なるユニークなものである。EU からファンディングされるということは、最終的に欧州市民にメリットがあるか、税金に見合うのかという観点で判断されることになる。オープンにし、高齢者の生活、健康、都市の将来といった分野に大きな関心が寄せられるものである。欧州議会と欧州員会による枠組みの中で研究開発している。

欧州特許庁がグーグルと翻訳で提携した点について、みんなは面白いとは思っていない。しかし、我々欧州の機械翻訳の研究者はグーグルとも共存していける。というのは、(グーグルが提携した) 特許分野は、機械翻訳の主要対象分野ではなく、翻訳の主要対象分野は、ニュース、技術文書、ビジネス等が中心なのである。例えば、ドイツ本社がスペイン子会社の従業員にドイツの規則を説明する際に、ドイツ語では分からないので、今は人手でスペイン語に翻訳しているのを、今後は機械で自動翻訳していくといったことがポイントである。翻訳が 100% 正確でなくても良く、誤りは練習により修正されていく。

#### (5) 無線 (3G) の速度 (XTREME LABS SPEEDTESTによる計測)

ザールブリュッケン中央駅前 下り 平均 303Kbps、最大 341Kbps

上り 平均 107Kbps、最大 107Kbps

VoiceTra+翻訳時間 約 10 秒

ドイツ人工知能研究センター付近 下り 平均 198Kbps、最大 347Kbps

上り 平均 80Kbps、最大 87Kbps  
 VoiceTra+翻訳時間 約 15 秒

## 参考資料

### FP7 プロジェクト

略称	EuroMatrixPlus
正式名称	欧州言語向け機械翻訳をユーザーへ
公募分野	ICT-2007.2.2
研究期間	2009年3月-2012年2月(36ヶ月)
全予算(EU 拠出金)	594万ユーロ(426万ユーロ)
プロジェクトコーディネーター	ドイツ人工知能研究センター
参加組織	プラハ・カレル大学(チェコ)、ジョン・ホプキンス大学(米)、ダブリン・シティ大学(アイルランド)、ブルーノケスラー基金(伊)、ルーシーソフトウェア・サービス(独)、CEET(チェコ)、ル・マン大学(仏)、エジンバラ大学(英)
ウェブサイト	<a href="http://www.euromatrixplus.net/">http://www.euromatrixplus.net/</a>
研究内容	機械翻訳技術の包括的な研究開発(EU公式言語向けの模範システムの開発、新しい方法の開発、評価、実用化)

### FP7 プロジェクト

略称	QTLAUNCHPAD
正式名称	精度の高い翻訳技術のための大規模アクションの準備と実施
公募分野	ICT-2011.4.1
研究期間	2012/7-2014/6(24ヶ月間)
全予算(EU 拠出金)	239万ユーロ(191万ユーロ)
プロジェクトコーディネーター	ドイツ人工知能センター
参加組織	ダブリンシティ大学(アイルランド)、アテネ情報・通

	信・知識技術研究・技術革新センター（ギリシア）、シーフィールド大学（英）
ウェブサイト	<a href="http://www.qt21.eu/launchpad/content/new-goal-quality-trans-lation">http://www.qt21.eu/launchpad/content/new-goal-quality-trans-lation</a>
研究内容	機械翻訳の精度を検査する新しいシステムを開発する。

#### FP7 プロジェクト

略称	T4ME NET
正式名称	多言語欧州情報社会のための技術
公募分野	ICT-2009.2.2
研究期間	2010年2月-2013年1月（36ヶ月）
全予算（EU 拠出金）	761万ユーロ（599万ユーロ）
プロジェクトコーディネーター	ドイツ人工知能研究センター
参加組織	ダブリン・シティ大学（アイルランド）、評価・言語資源分配庁（仏）、バルセロナメディア基金・ポンペウ・ファブラ大学（西）、RWTH アーヘン大学（独）、イタリア学術会議（伊）、ジョゼフ・ステファン研究院（スロベニア）、アルト大学（フィンランド）、アテナ情報通信知識技術研究イノベーションセンター（ギリシア）、ブルーノ・ケスラー基金（伊）、フランス国立科学研究センター・LIMSI（仏）、ユトレヒト大学（蘭）、プラハ・カレル大学（チェコ）
ウェブサイト	<a href="http://www.meta-net.eu/projects/t4me/">http://www.meta-net.eu/projects/t4me/</a>
研究内容	言語技術と他の技術（機械学習や認知システム）の研究開発機関を提携させる。META-NET を支援する。

#### 欧州言語技術アライアンス推進ネットワーク

略称	META-NET
正式名称	欧州多言語技術アライアンス推進ネットワーク・エクセレンス
予算	T4ME NET、CESAR <sup>76</sup> 、METANET4U <sup>77</sup> 、METANORD <sup>78</sup>

<sup>76</sup> <http://www.meta-net.eu/projects/cesar/summary>

	という 4 つの欧州プロジェクト（欧州委員会の拠出）から共同で資金を支給されている。
コーディネーター	ドイツ人工知能研究センター
参加組織	34 カ国から 60 の研究機関が参加
ウェブサイト	<a href="http://www.meta-net.eu/">http://www.meta-net.eu/</a>
活動内容	欧州の言語技術開発に関連する組織を提携させる。

### 写真（先方組織の外観）



ザールラント大学構内にある韓国科学技術研究院（KIST）ヨーロッパの施設



<sup>77</sup> <http://www.meta-net.eu/projects/METANET4U/>

<sup>78</sup> <http://www.meta-net.eu/projects/meta-nord/>

## 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

カールスルーエ技術研究院のインタラクティブシステム研究所では、人間の経験や人間同士及び人間と機械のコミュニケーションを簡便化することを目標に、音声認識、機械翻訳、音声合成、言語、視覚技術、人追跡と認識、マルチモーダル及びクロスモーダル知覚インターフェイス、スマートルーム、パーベイシブ計算をテーマに研究開発を実施している。責任者はアレックス・ヴァイベル氏<sup>79</sup>である。

研究開発に関しては、同研究所は、1991年という早い段階に、ヤヌス音声翻訳システム (Speech-toSpeech translation) という音声機械翻訳システムを開発している。現在進行中の研究プロジェクトとしては、FP7のEU-Bridge<sup>80</sup>にコーディネーターとして参加している。また仏独合同の研究開発プログラム「クアエロ」<sup>81</sup>に参加し、顔検知・認識・分析、マルチメディア人間特定、音声認識に関わる研究プロジェクトを実施している<sup>82</sup>。その他、自動音声認識と統計機械翻訳の組み合わせ、大学講義向けの翻訳システム<sup>83</sup>を開発している。また、過去には、C-STARやTC-STAR<sup>84</sup>という音声翻訳技術の研究開発を進める国際プロジェクトに参加している。

国際提携関係に関しては、米国のピッツバーグにあるカーネギーメロン大学コンピューターサイエンス学部とシリコンバレーに姉妹研究所を持つ。また、フランスのLIMSIとドイツのRWTHアーヘン大学と合同研究機関IMMIを設立し、提携して研究開発を進めている。ドイツのRWTHアーヘン大学、米国、フランスの研究機関と繋がりが深い。

## アーヘン・ライン=ヴィストファーレン工科大学 人間言語技術・パターン認識研究グループ

ドイツのRWTHアーヘン大学 (アーヘン・ライン=ヴィストファーレン工科大学)<sup>85</sup>では、人間言語技術・パターン認識研究グループで、音声認識、文字・音声機械翻訳、自

<sup>79</sup> [http://isl.anthropomatik.kit.edu/english/21\\_74.php](http://isl.anthropomatik.kit.edu/english/21_74.php)

<sup>80</sup> <http://www.eu-bridge.eu/>

<sup>81</sup> <http://www.quaero.org/>

クアエロプログラムに関しては、第一部第二章第三節フランスに収録したヴォカピアリサーチとのヒアリング議事録を参考のこと。

<sup>82</sup> <http://www.quaero.org/developpements-technologiques/>

<sup>83</sup> <http://isl.anthropomatik.kit.edu/english/1520.php>

<http://isl.anthropomatik.kit.edu/cmu-kit/english/1590.php>

<sup>84</sup> <http://www.tcstar.org/>

<sup>85</sup> <http://www-i6.informatik.rwth-aachen.de/web/Homepage/index.html>

然言語処理、音声対話システム、画像認識の研究開発が実施されている。

責任者は、イング・ハーマン・ネイ氏<sup>86</sup>である。最近の研究プロジェクトとしては、FP7 のEU-Bridgeに参加している。また、仏独研究開発プログラムであるクアエロでは、音声認識、画像分類、機械翻訳、手書きテキストの認識、音声機械翻訳のアプリケーションを開発している<sup>87</sup>。また、米国の国防省傘下のDARPA（国防先端研究プロジェクト庁）のBOLT（Broad Operational Language Translation）<sup>88</sup>プロジェクトに参加している。このプロジェクトでは、国防や国内治安の維持を目的に、中国語とアラビア語の諸方言を英語へと翻訳する技術の開発が目指される。特に、会話、Eメールテキスト、インスタントメッセージの翻訳が狙いとなる。また、自然言語処理、音声認識、機械翻訳等に関する多くのソフトウェアを開発している<sup>89</sup>。

## 第五節 その他の欧州諸国

本節では、いわゆる欧州主要国（英仏独）以外で、機械翻訳及び音声技術の研究開発を積極的に実施している研究機関について記す。

### イタリア：ブルーノ・ケスラー基金（FBK）

イタリアのFBK（ブルーノ・ケスラー基金）<sup>90</sup>は、トレント州に 2007 年に前身機関のトレント文化研究院を引き継いで設立された研究機関であり、人文科学、自然科学、技術開発分野で研究活動を行っている。FBKにはICT研究センターが設置されており、その人間言語技術ユニット<sup>91</sup>で、音声認識、機械翻訳、コンテンツ処理の研究開発が実施されている。責任者はマルチェロ・フェデリコ氏<sup>92</sup>である。

機械翻訳の研究開発に関しては、音声翻訳及びテキスト翻訳の研究を実施しており、エジンバラ大学が開発したMOSESの利用開発を積極的に行っている<sup>93</sup>。対象言語は、英語、イタリア語、スペイン語、中国語、アラビア語である。音声認識に関しては、FBKは 20 年来開発を行っており、研究の伝統がある<sup>94</sup>。電話会話の音声分析や視聴覚資料

---

<sup>86</sup> <http://www-i6.informatik.rwth-aachen.de/web/Staff/index.html>

<sup>87</sup> <http://www.quaero.org/developpements-technologiques/>

<sup>88</sup>

[http://www.darpa.mil/NewsEvents/Releases/2011/2011/04/19\\_DARPA\\_initiates\\_overarching\\_language\\_translation\\_research\\_Publishes\\_Broad\\_Agency\\_Announcement\\_for\\_Broad\\_Operational\\_Language\\_Translation\\_program.aspx](http://www.darpa.mil/NewsEvents/Releases/2011/2011/04/19_DARPA_initiates_overarching_language_translation_research_Publishes_Broad_Agency_Announcement_for_Broad_Operational_Language_Translation_program.aspx)

<sup>89</sup> <http://www-i6.informatik.rwth-aachen.de/web/Software/index.html>

<sup>90</sup> <http://www.fbk.eu/>

<sup>91</sup> <http://hlt.fbk.eu/en/home>

<sup>92</sup> <http://hlt.fbk.eu/people/federico>

<sup>93</sup> <http://hlt.fbk.eu/en/research/MT>

<sup>94</sup> <http://hlt.fbk.eu/en/research/ASR>

の音声転写アプリケーションを開発している。対象言語はイタリア語、英語、スペイン語である。研究プロジェクトに関しては、多くのFP7 のプロジェクト (Excitement、MateCat、Eu-Bridge、EuroMatrixPlus、Tosca-MP、PESCaDo、CoSyne、MosesCore、T4MNET、Dirha) に参加しており、非常に国際性が高い研究機関であると言える。MateCatプロジェクトとDirhaプロジェクトに関しては、コーディネーターを務めている。また、国内プロジェクトにも参加しており、2009 年よりトレント州の気象機関と提携し、トレント・メテオという天気予報向けの翻訳システムを開発している。

#### アイルランド：次世代ローカライゼーションセンター

アイルランドの次世代ローカライゼーションセンター (Centre for Next Generation Localisation : CNGL)<sup>95</sup>は、4つのアイルランドの大学 (トリニティ・カレッジ・ダブリン大学、ダブリンシティ大学、カレッジ・ダブリン大学、リムリック大学) と複数のアイルランドと諸外国の民間企業<sup>96</sup> (米国からはインテルやマイクロソフト、日本から大日本印刷) を提携させて研究開発を実施している。CNGLは、アイルランドの科学基金と民間企業のコアパートナーから資金 (4350 万ユーロ) を供給され、2007 年に設立された<sup>97</sup>。

CNGLは主にコンテンツに関わる技術の開発を実施しており、研究テーマの1つとして「翻訳・ローカライゼーション」<sup>98</sup>がある。同テーマの研究リーダーは、ダブリンシティ大学コンピューティング学教授のクン・リュ (Qun Liu) 氏である<sup>99</sup>。同テーマでは、機械翻訳、リソースパターン・ゲーム・メタデータの社会ローカライゼーション、言語技術の言語と社会への影響を研究対象とし、言語の障壁を取り除くことを目的としている<sup>100</sup>。

同テーマに関して、CNGLと提携する言語技術分野の研究センターは、ダブリンシティ大学コンピューティング学部の国立言語技術センター (機械翻訳、音声認識、音声合成)<sup>101</sup>、カレッジ・ダブリン大学のローカライゼーション研究センター (ローカライゼーションに関わる技術)<sup>102</sup>、MUSTERグループ<sup>103</sup> (音声技術)、創造言語システムグル

---

<sup>95</sup> <http://www.cngl.ie/>

<sup>96</sup> <http://www.cngl.ie/industry-commercialisation/our-partners/>

<sup>97</sup> <http://www.cngl.ie/about-us/>

<sup>98</sup> <http://www.cngl.ie/research/research-areas/translation-and-localisation/>

<sup>99</sup> <http://www.cngl.ie/people/theme-leaders/>

<sup>100</sup> <http://www.cngl.ie/research/research-areas/translation-and-localisation/>

<sup>101</sup> <http://www.nclt.dcu.ie/>

<sup>102</sup> <http://www.localisation.ie/>

<sup>103</sup> <http://muster.ucd.ie/>

ープ<sup>104</sup>である。

CNGLが関わっている研究プロジェクト（CNGLに参加している研究機関が参加しているプロジェクト）に関しては<sup>105</sup>、FP7のAbu-MaTran、CoSyne、EXPERT、KHERSMOI、LT-WEB、QTLanuchpad、EuroMatrixPlus、Panacea、PetaMedia、T4MENET、CIPのPlutoがあり、積極的に欧州プロジェクトに参加している。

また、同機関はクラウドソース翻訳に関わる研究開発も積極的に実施しており、スピンオフ組織として、クラウドソース翻訳を振興するロゼッタ基金を設立している（ロゼッタ基金に関しては、本報告書第一部第四章第三節を参考のこと）。

なお、KantanMT（利用者のニーズに合わせカスタマイズ可能な機械翻訳を提供するサービス）を提供しているXcelerator Machine Translation Solution社もCNGLの民間企業のパートナーである（同社は、ダブリンシティ大学内に設立されている）<sup>106</sup>。

### 第三章 欧州企業による機械翻訳技術及び音声技術の商用化動向

本章では、欧州企業（特にEU加盟国企業）による機械翻訳技術及び音声技術の商用化動向について記す。以下に、同分野の技術を商用化している企業の情報及び製品、サービスについて記す。機械翻訳に関連するソフトウェア等を商用化している企業については、対応言語についても記す。

#### 英国

企業名	SDL
企業種	ビジネスコンサルティング、翻訳、通訳の総合サービス（機械翻訳を含む）
企業本拠地	英国
サービス名	SDLFreeTranslation.com
サービス内容	オンラインテキスト機械翻訳サービス
ウェブサイト	<a href="http://www.freetranslation.com/">http://www.freetranslation.com/</a>
欧州言語	英語、独語、仏語等
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語、ハウサ語、ソマリ語

<sup>104</sup> <http://afflatus.ucd.ie/>

<sup>105</sup> <http://www.cngl.ie/research/affiliated-projects/>

<sup>106</sup> <http://www.kantanmt.com/aboutcompany.php>

企業名	ATA
企業種	ソフトウェア開発企業
企業本拠地	英国
開発製品名	AL-Mutarjim Al-Arabey v3,00 等
製品内容	アラビア語と英語向けテキスト機械翻訳ソフトウェア及び電子辞書
ウェブサイト	<a href="http://eshop.atasoft.com/eng/software.html">http://eshop.atasoft.com/eng/software.html</a>
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語

企業名	CAPITA
企業種	翻訳と通訳サービス（機械翻訳を含む）
企業本拠地	英国
サービス名	SmateMate
製品内容	ニーズに合わせてカスタマイズできるテキスト機械翻訳（MOSES を利用）
ウェブサイト	<a href="http://www.capitatranslationinterpreting.com/">http://www.capitatranslationinterpreting.com/</a> <a href="https://www.smartmate.co/">https://www.smartmate.co/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	ラトビア語、リトアニア語、エストニア語、ウクライナ語
アフリカ諸国の言語	アラビア語

## フランス

企業名	Systran
企業種	ソフトウェア開発企業
企業本拠地	フランス（創設は米国）
製品名	Systran、Systran Mobile、SystranLinks 等
製品内容	テキスト機械翻訳ソフトウェア（PC 及びモバイル向け）、オンラインテキスト機械翻訳サービス
ウェブサイト	<a href="http://www.systran.fr/">http://www.systran.fr/</a> 、 <a href="http://www.systranet.com/translate">http://www.systranet.com/translate</a> 、 <a href="http://www.systranlinks.com/#launch">http://www.systranlinks.com/#launch</a>

欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語

企業名	ソフティッシモ
企業本拠地	フランス
企業種	ソフトウェア開発企業
製品名	レヴェルソ・イントラネット
製品内容	パーソナライズされた機械翻訳
ウェブサイト	<a href="http://reverso.softissimo.com/fr/">http://reverso.softissimo.com/fr/</a>
欧州言語	英語、仏語、独語等、多数
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語、ハウサ語

企業名	ヴォカピアリサーチ
企業本拠地	フランス
企業種	ソフトウェア開発企業
製品名	VoxSigma (企業向け)
製品内容	音声認識、話者特定、音声転写ソフトウェア
ウェブサイト	<a href="http://www.vocapia.com/index.html">http://www.vocapia.com/index.html</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	ウクライナ語 (現在開発中)
アフリカ諸国の言語	アラビア語

## ドイツ

企業名	Linguatec
企業種	ソフトウェア開発企業
企業本拠地	ドイツ
1) 製品	ボイスリーダー
製品内容	音声合成ソフトウェア
2) 製品	ボイスプロ・エンタープライズ

製品内容	音声認識ソフトウェア（転写、字幕付け等）
3) 製品	パーソナルトランスレーター、コーポレートソリューション
製品内容	機械翻訳ソフトウェア
4) 製品	モバイル向けアプリケーション
製品内容	音声翻訳（旅行向け）、音声合成、音声認識アプリケーション
ウェブサイト	<a href="http://www.linguatec.de/">http://www.linguatec.de/</a>
欧州言語	英語、仏語、独語等（機械翻訳ソフトウェア及びモバイル旅行向け音声翻訳アプリケーション）
旧ソ連諸国の言語	アルメニア語、アゼルバイジャン語、ベラルーシ語、エストニア語、グルジア語、ラトビア語、リトアニア語、ウクライナ語（モバイル旅行向け音声翻訳アプリケーションのみ）
アフリカ諸国の言語	アフリカーンス語、アラビア語、スワヒリ語（モバイル旅行向け音声翻訳アプリケーション。アラビア語は音声認識及び音声合成ソフトウェアにも対応）

企業名	ルーシーソフトウェア・サービス
企業本拠地	ドイツ
企業種	ソフトウェア開発企業
製品名	ルーシーLT
製品内容	企業向けテキスト機械翻訳サービス
ウェブサイト	<a href="http://www.lucysoftware.com/english/home/">http://www.lucysoftware.com/english/home/</a>
旧ソ連諸国の言語	なし
アフリカ諸国の言語	なし

企業名	ヨコイ（ドイツ人工知能センター言語技術研究所のスピンオフ企業）
企業本拠地	ドイツ
企業種	ソフトウェア開発企業
製品名	ヨコイ・ランゲージ・ガイド
製品内容	英語、ドイツ語、中国語間の辞書・翻訳・学習モバイルアプリケーション

ウェブサイト	<a href="http://www.yocoy.com/">http://www.yocoy.com/</a>
旧ソ連諸国の言語	なし
アフリカ諸国の言語	なし

## その他の欧州諸国

### ハンガリー

企業名	MorphoLogic
企業種	ソフトウェア開発企業
企業本拠地	ハンガリー
製品名	MorphoWord
製品内容	機械翻訳ソフト
ウェブサイト	<a href="http://www.morphologic.hu/index.php">http://www.morphologic.hu/index.php</a>
欧州言語	英語とハンガリー語
旧ソ連諸国の言語	なし
アフリカ諸国の言語	なし

### オーストリア

企業名	Sonico
企業種	ソフトウェア開発企業
企業本拠地	オーストリア
製品名	iTranslate
製品内容	テキスト・音声自動翻訳アプリケーション（音声合成及び音声認識機能を含む）
ウェブサイト	<a href="http://www.itranslateapp.com/">http://www.itranslateapp.com/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	アゼルバイジャン語（テキスト翻訳のみ）、ベラルーシ語（テキスト翻訳のみ）、エストニア語（テキスト翻訳のみ）、グルジア語（テキスト翻訳のみ）、ラトビア語（テキスト翻訳のみ）、リトアニア語（テキスト翻訳のみ）、ウクライナ語（テキスト翻訳のみ）

アフリカ諸国の言語	アフリカーンス語（テキスト翻訳のみ）、アラビア語（音声認識及び音声合成あり）、スワヒリ語（テキスト翻訳のみ）
-----------	--

## 第四章 その他の動向

### 第一節 欧州における機械翻訳及び音声技術関連の話題となる出来事

#### 視察レポート：デジタル技術の祭典「フューチャー・アン・セーヌ」における機械翻訳及び音声技術の開発動向

フランスでは、平成25年6月13日から23日までデジタル技術の祭典、「Futur en Seine（フューチャー・アン・セーヌ：仏語の意味 未来のセーヌ）」<sup>107</sup>が開催された。同祭典を視察し、機械翻訳及び音声技術の開発動向を調査した。同祭典は、フランスの競争力拠点の1つ、キャップ・デジタル<sup>108</sup>により2009年に創設された。イル・ド・フランス州（首都パリ及びパリ近郊）とパリ市に支援され、毎年一度6月に開催されている。2013年は、通信事業者フランステレコム、ユビソフト（仏ビデオゲーム開発企業）、フランス郵便、セーヌ・サンドニ市が公式パートナーとして参加している。主なイベントは、パリ19区の芸術関連のイベント施設「サンカトル」で行われ、6月13-14日はステークホルダー向け、15-16日は子供、家族向けに会議等のイベントが開催された。約50の研究機関や大学組織、約620の民間企業（大規模企業グループ20社、その他は中小企業）等が参加しており、全部で120以上のデジタル技術に関わるプロジェクトのデモ等が行われ、サンカトル会場には多くのスタンドが立ち並んだ。機械翻訳及び音声技術に関しては、仏独共同研究プログラム「クアエロ」と仏ソフティッシモ社がスタンドを持ち、それぞれアプリケーションを紹介していた。

<sup>107</sup> <http://www.quaero.org/quaero-futur-en-seine-2013/>

フューチャー・アン・セーヌに関しては、日本フランス大使館が日本語でウェブサイトで紹介しているので、参考のこと。

<http://www.ambafrance-jp.org/article6618>

<sup>108</sup> <http://www.capdigital.com/cap-digital/>

#### a) サンカトル会場の様子



#### クアエロ

クアエロプログラム<sup>109</sup>は、仏独共同研究開発プログラムであり、特に多言語・マルチメディアコンテンツ自動分析に関わるアプリケーションを開発することを目的とする。特に、音声、文字、イメージ（写真等）、動画、音楽の自動処理アプリケーション（機械翻訳を含む）を開発している。仏独から全 32 組織（仏組織が主で、独からは 2 組織が参加）が参加し、大企業と中小企業を問わず産学を提携させ、研究を実施させている。クアエロのウェブサイトの情報によれば、研究者の出身国は 35 カ国以上であり、参加研究者 1000 名以上である。全予算は 2 億ユーロで、9900 万ユーロが仏政府（フランス企業支援・イノベーション振興機構：OSEO）から拠出されている（独政府からは助成されていない）<sup>110</sup>。

フューチャー・アン・セーヌのスタンドでは、クアエロプログラムに参加している仏シストラン社<sup>111</sup>とヴォカピアリサーチ社<sup>112</sup>に、同プログラムで開発したアプリケーションについて説明していただいた。

1) 仏シストラン社は、仏大手機械翻訳ソフトウェア開発企業であり、1968 年に世界に

<sup>109</sup> <http://www.quaero.org/>

<sup>110</sup> クアエロについては、ヴォカピアリサーチ社に行ったヒアリング議事録を参考のこと。

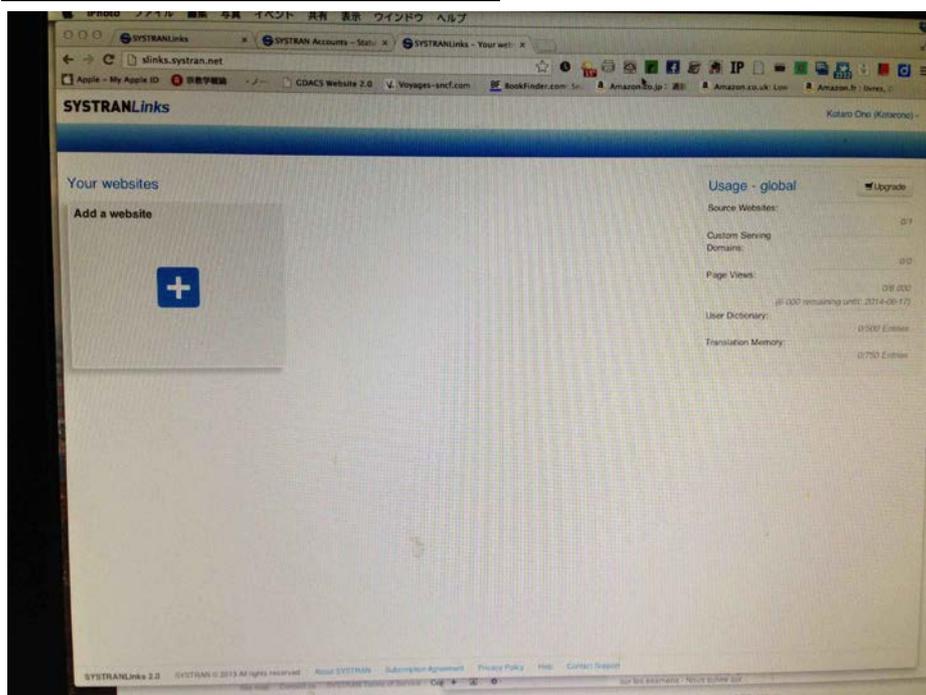
<sup>111</sup> <http://www.systransoft.com/>

<sup>112</sup> <http://www.vocapia.com/>

先駆けて創立された。同社には、シストラン・リンクス<sup>113</sup>という製品を紹介していた。この製品はウェブサイト翻訳向けのクラウド型オンライン機械翻訳サービスである。

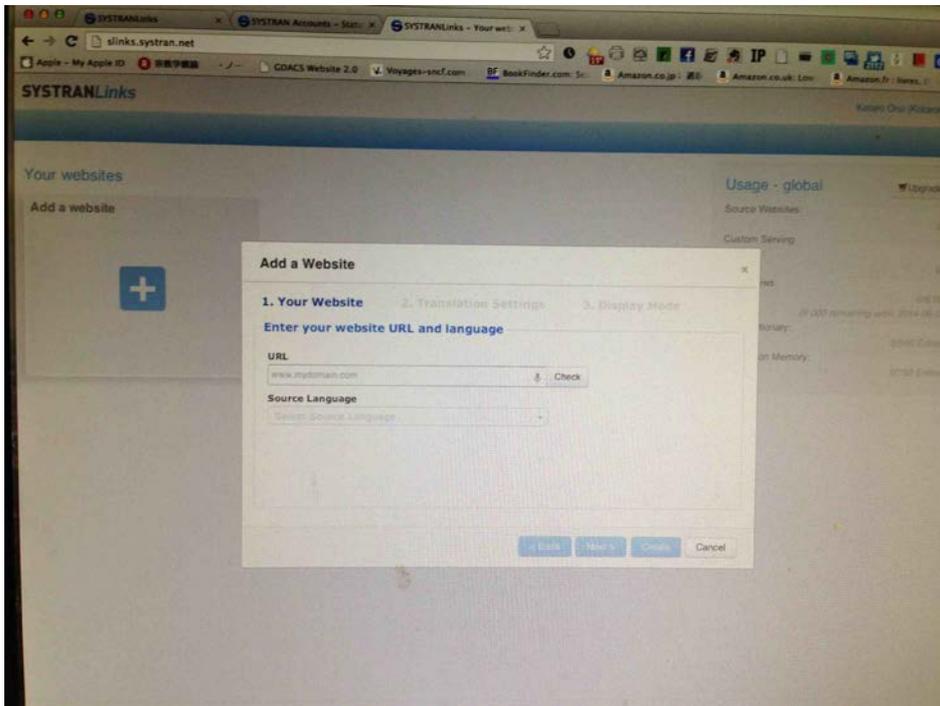
まず、利用者はシストラン・リンクスの専用ウェブサイトに行き、そこで翻訳したいウェブサイトの URL を入力すれば、そのサイトが翻訳される。翻訳後、機械翻訳が完全でない場合、そのウェブページ上で、ユーザーがその翻訳を後編集することも可能である。また、後編集を行う際に、他の人を招待して、複数の人で修正作業を行うことも可能である。同アプリケーションに利用登録すれば、1年間 6000 ページまでは無料で利用できるが、それ以上の場合には、使用料を払う。対応言語は、15 言語である（独語、英語、アラビア語、韓国語、中国語、スペイン語、フランス語、ギリシア語、イタリア語、日本語、オランダ語、ポーランド語、ポルトガル語、ロシア語、スウェーデン語）。

#### b) シストラン・リンクス トップ画面



<sup>113</sup> <http://www.systranlinks.com/#launch>

c) 翻訳対象のウェブサイトの URL を入力



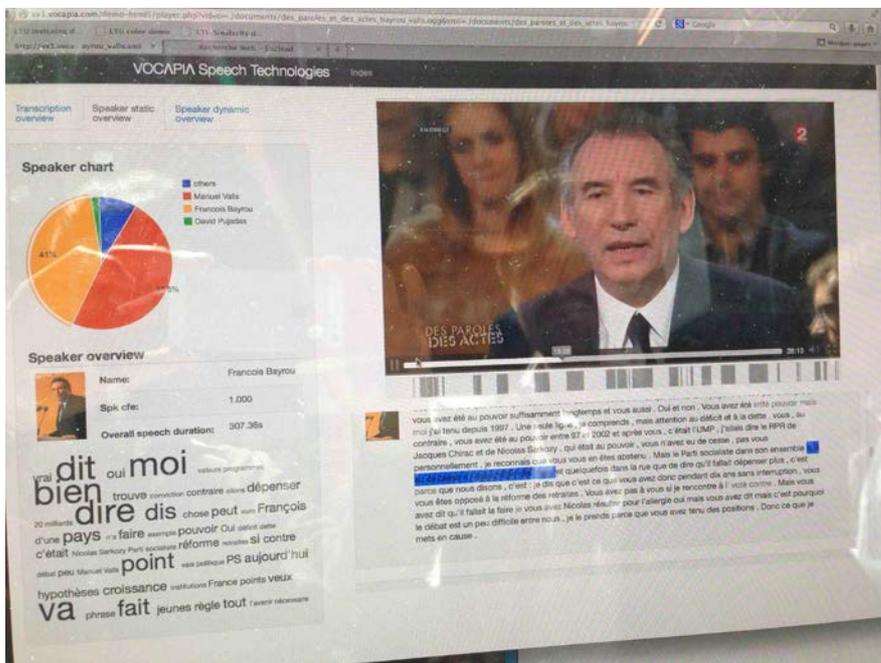
d) 後編集画面 (参考: 仏ル・モンド紙 仏語から日本語へ翻訳)



2) ヴォカピアリサーチには、同社のクラウド型 VoxSigma SaaS を利用するテレビ放送番組等の話者自動特定アプリケーションと自動転写アプリケーションを紹介していただいた。前者は、例えば、複数の政治家による討論番組で、各政治家の声を特定し、その人が話した瞬間をタイムコード上に記し、その人がいつ話したか、何分話したか知ることが可能になる。後者は、テレビ番組等で話している出演者の言葉を動画画面の下部に転写していくアプリケーションである。利用者は動画等を専用サイトにアップロードし、そのサイト上で、これら二つのアプリケーションを利用できる。

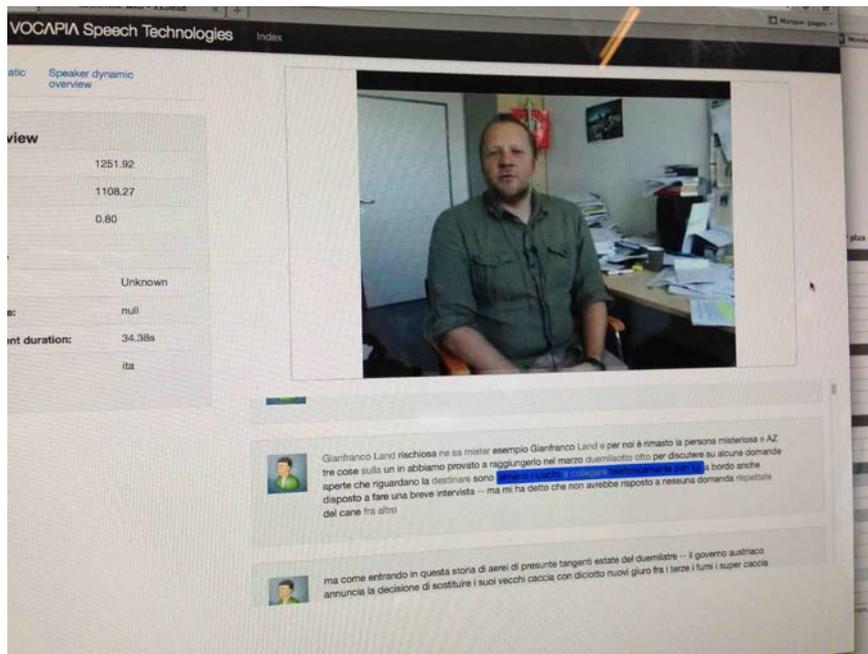
#### e) 話者自動特定及び自動転写アプリケーション

動画の下にタイムコードがあり、タイムコードの下に転写された文章が現れる。左上には、番組中に話者が話した時間の割合が現れる。



#### f) 自動転写アプリケーション

動画画面の下に転写された文章が現れる。



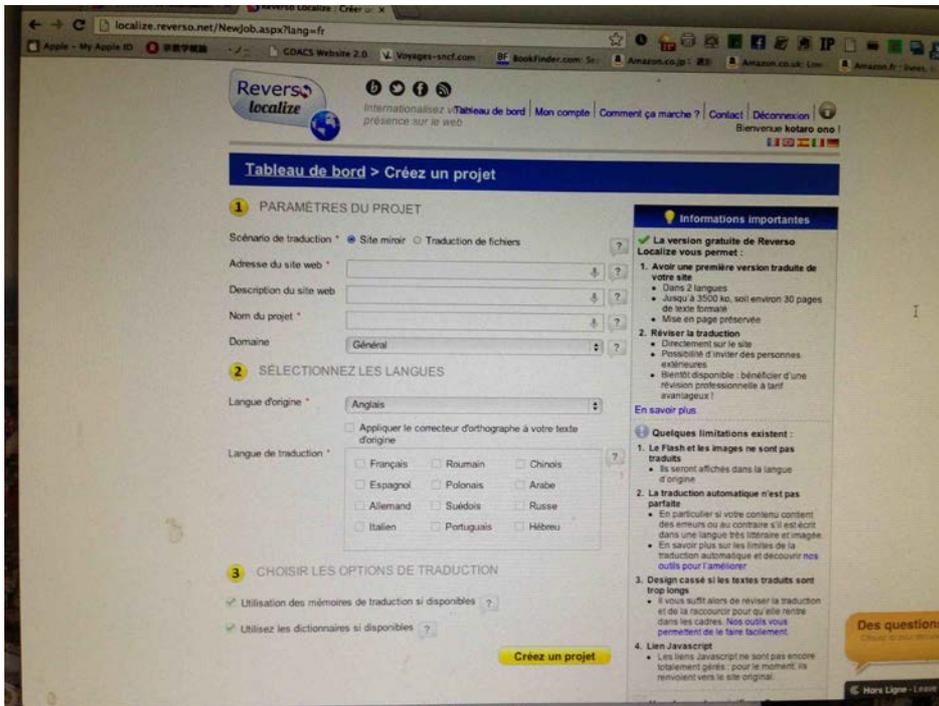
## ソフティッシモ

仏ソフティッシモ社には、レベルソ・ロカライズ<sup>114</sup>というクラウド型オンライン機械翻訳サービスを紹介していただいた。このサービスは、ワード等の文書やウェブサイトを翻訳するサービスであり、利用者は翻訳したい文書をアップロードし、それを専用のウェブサイト上で機械翻訳する。専用ウェブサイトで、ユーザーは翻訳された文章を自分で修正することも可能であるし、他の人を招待して複数名で後編集することも可能である。対応言語は 12 言語（英語、仏語、ドイツ語、スペイン語、イタリア語、ルーマニア語、ポーランド語、スウェーデン語、ポルトガル語、アラビア語、中国語、ロシア語、ヘブライ語）である。なお、同社はFlavius<sup>115</sup>というCIPプロジェクトに参加し、同サービスを開発した。

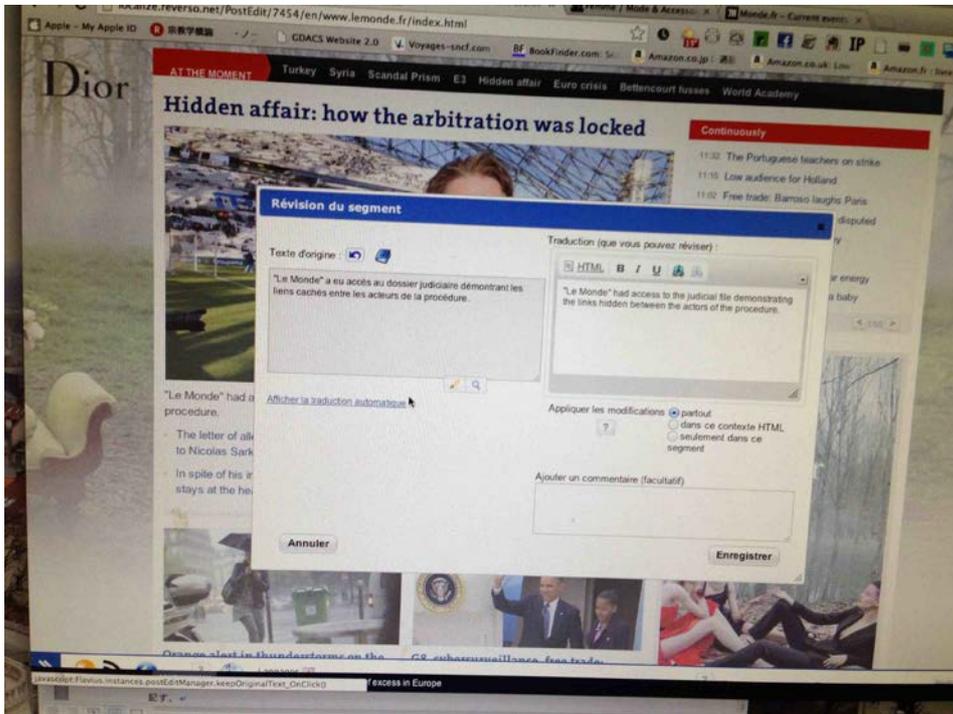
<sup>114</sup> <http://localize.reverso.net/?lang=fr>

<sup>115</sup> [http://ec.europa.eu/information\\_society/apps/projects/factsheet/index.cfm?project\\_ref=250528](http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=250528)

g) レベルソ・ロカライズのトップ画面



h) 後編集画面 (参考: 仏ル・モンド紙 仏語から英語へ翻訳)



参考：CIP プロジェクト

略称	Flavious
正式名称	インターネット及びユーザー作成サイトの外国語バージョン
公募分野	CIP-ICT-PSP.2009.5.3
研究期間	2010年4月-2012年9月（30ヶ月）
全予算（EU 拠出金）	370万ユーロ（185万ユーロ）
プロジェクトコーディネーター	ソフティッシモ（仏）
参加組織	ビデオポリス（ベルギー）、JFG ネットワークス（仏）、Qype（独）、アクロスシステム（独）、Language Weaver（ルーマニア）、Daedalus-data, Decisions and Language（スペイン）
ウェブサイト	<a href="http://www.project-flavius.eu/fr/en/">http://www.project-flavius.eu/fr/en/</a>

## 第二節 欧州特許庁とグーグル社による多言語特許翻訳の動向

本節では、欧州特許庁と米グーグル社による多言語特許翻訳の動向について記す。まず、欧州特許庁の概要について記し、ついで、グーグル社との提携関係の進展について記す。

### 欧州特許庁の概要

欧州特許庁（European Patent Office : EPO）<sup>116</sup>は、欧州諸国の統一的な特許管理を目的に、1977年に「欧州特許条約」に基づき設立された地域特許庁である。同条約は1973年に調印され、1977年に発効された。同庁は、欧州連合の機関でもなく、欧州委員会の下部組織でもないが、欧州委員会と密接な関係を持つ。特許取得出願者は、同庁に特許申請をし、所定の審査手続きを経れば、欧州特許条約締約国内全てで特許を付与される。同庁の設立により、各国で特許出願を申請する必要がなくなり、特許申請の手続きが簡素化されたが、最終的に特許は各締約国の決定で付与されるので、各国毎に特許を翻訳する必要があり、全締約国内で有効な特許を取得するには費用がかかる。条約締約国は、EU加盟国とともに、アルバニア、クロアチア、セルビアやトルコ等の国も含

<sup>116</sup> <http://www.epo.org/index.html>  
[http://ec.europa.eu/research/youngscientists/index\\_en.cfm?pg=epo](http://ec.europa.eu/research/youngscientists/index_en.cfm?pg=epo)

み、現在 38 カ国に及んでいる<sup>117</sup>。欧州特許庁への出願は、同庁の公式言語である英語、仏語、独語以外の言語でも可能であるが、その際には同 3 言語のどれかによる翻訳文の提出が必要となる。なお、同庁では商標や芸術作品の著作権等については取り扱っていない。

### 欧州特許庁の基本情報

公式言語	英語、仏語、独語
人員規模	7000 名 (30 各国以上の出身)
拠点	ミュンヘン (本部)、ハーグ、ベルリン、ウィーン、ブリュッセル
予算	2013 年度 20 億ユーロ (自己資金調達)

### 欧州特許庁と世界知的所有権機関

欧州特許庁は、特許を含めた知的所有権を所管する国際機関とも連携して活動しており、特に国際連合の一機関である世界知的所有権機関 (World Intellectual Property Organization : WIPO) と密接な関わりを持つ。同機関は、国際的な特許出願制度である特許協力条約 (Patent Cooperation Treaty : PCT)<sup>118</sup>の普及を促進している。特許協力条約は、自国内の特許庁に特許を出願するだけで、全締約国で出願した取り扱いを受けることを可能にする<sup>119</sup>。同条約には、現在 146 カ国が締約している<sup>120</sup>。

2012 年 5 月に、欧州特許庁と世界知的所有権機関は特許協力条約について協定を交わしている。欧州特許庁では、欧州特許条約だけでなく、特許協力条約に基づいて、特許出願を申請することもすでに可能であったが、2012 年 5 月の協定は、特許協力条約に基づく国際特許付与システム (特許申請、審査、付与のプロセス) を改善することを主な目的としていた<sup>121</sup>。

なお、世界知的所有権機関は、PATENTSCOPEと呼ばれるオンライン特許機械翻訳サービスを持ち、このサービスにはエジンバラ大学が開発したMOSESが利用されている<sup>122</sup>。翻訳可能な言語は、英語、仏語、独語、日本語 (ベータ版)、中国語である。だ

<sup>117</sup> 欧州特許条約締約国のリスト : <http://www.epo.org/about-us/organisation/member-states.html>

<sup>118</sup> <http://www.wipo.int/pct/en/>

<sup>119</sup> 国際特許出願手続きに関しては、日本特許庁のウェブサイトも参考のこと。

[http://www.jpo.go.jp/seido/s\\_tokkyo/kokusai1.htm](http://www.jpo.go.jp/seido/s_tokkyo/kokusai1.htm)

<sup>120</sup> 特許協力条約締約国のリスト :

[http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&treaty\\_id=6](http://www.wipo.int/treaties/en/ShowResults.jsp?lang=en&treaty_id=6)

<sup>121</sup> [http://www.wipo.int/pressroom/en/articles/2012/article\\_0008.html](http://www.wipo.int/pressroom/en/articles/2012/article_0008.html)

<sup>122</sup> [http://www.wipo.int/patentscope/en/news/pctdb/2011/news\\_0001.html](http://www.wipo.int/patentscope/en/news/pctdb/2011/news_0001.html)

<https://www3.wipo.int/patentscope/translate/translate.jsf>

が、同時にWIPOはグーグル翻訳も利用しており、複数のシステムを利用している<sup>123</sup>。

### グーグル社との提携関係の進展

現在、欧州特許庁と米グーグル社（Google）が特許検索の分野で連携を進めている。

### **背景**

2006年末、グーグル社はグーグル特許検索（Google Patents Search）というサービス名<sup>124</sup>で、米特許商標庁（United States Patents and Trademark Office：USPTO）に登録された特許と商標をオンラインで検索するサービスをベータ版で提供開始していた<sup>125</sup>。特許検索サービスに使用されている技術は、グーグル社が提供する書籍検索サービスと同じもので、同サービスにより、オンライン上で特許の文章や図案等を翻訳して閲覧することが可能になった。グーグル社はベータ版を発表した際に、今後特許検索の範囲を拡大していく方針を表明していた。

### **特許機械翻訳に関する欧州特許庁とグーグル社の提携動向**

2010年11月、欧州特許庁は特許翻訳の分野での提携に関して、グーグル社と合意に至った。欧州特許庁によれば、特許機械翻訳システムの導入により、各国の発明家、研究者等が自国の言語で特許を検索できるようになり、いわゆる言語の壁を破壊することができる。また、グーグル社にとっては、欧州特許庁より膨大な数の特許文書を受け取り、同社の機械翻訳技術を最適化することが可能になる。

2012年2月、同庁は、グーグル社による特許翻訳サービスの無料提供を開始した<sup>126</sup>。欧州特許庁は欧州各国の特許所管機関と提携して、グーグル社の機械翻訳エンジンのために、公式の特許文書、翻訳者が翻訳した特許文書等を同社に提供していた。これにより、特許検索者は、欧州特許庁のウェブサイト上のエスパスネット（Espacenet）、欧州出版サーバー（European publication server）で、特許を検索し、翻訳して閲覧することが可能になった<sup>127</sup>。2012年3月の時点では、英語から、フランス語、ドイツ語、イタ

<sup>123</sup> <http://patentscope.wipo.int/search/en/help/about.jsf>

<sup>124</sup> 特許検索サービスのページ

<https://www.google.com/?tbn=pts>

[http://www.google.com/advanced\\_patent\\_search](http://www.google.com/advanced_patent_search)

<sup>125</sup> <http://googleblog.blogspot.fr/2006/12/now-you-can-search-for-us-patents.html>

<sup>126</sup> 主に、欧州特許庁のウェブサイトや同庁が公表している「特許情報ニュース」、グーグル社の公式ブログ等を参考にした。

<http://www.epo.org/service-support/publications/patent-information/news/2013.html>

<http://googletranslate.blogspot.fr/2012/02/translating-patents-with-european.html>

<http://blog.epo.org/the-epo/patent-translate-service-goes-live/>

<sup>127</sup> 欧州特許庁の検索サイト

<http://www.epo.org/searching/free/patent-translate.html>

リア語、ポルトガル語、スペイン語、スウェーデン語への翻訳、そして、これらの欧州言語から英語への翻訳が可能だった。

2012年8月、グーグル社は2006年末に開始していた特許検索サービスを、同社の機械翻訳サービスに統合し、欧州特許庁の特許情報を取り入れた。これにより、特許検索者は、欧州特許庁のエスパスネット上だけでなく、同社のサービスプラットフォーム<sup>128</sup>で翻訳して閲覧すること、またPDFにダウンロードして保存することが可能になった<sup>129</sup>。グーグル社の特許検索サービスは、欧州特許庁の特許検索サービス（エスパスネット及び欧州出版サーバー）と連動している。また、同社は検索サービスを改善し、「先行技術発見（Prior Art Finder）」ボタンを設置した。これは同社の特許検索サービスと、同社の他のサービス、例えば、学術論文検索サービス、書籍検索サービス等を連動させることにより、ウェブ全体から、ある特許情報をその特許の関連情報とともに取得することを可能にする<sup>130</sup>。

2012年10月、機械翻訳の対応言語が拡大され、英語とデンマーク語、オランダ語、フィンランド語、ハンガリー語、ノルウェー語、ポーランド語の間で翻訳が可能になった（全13カ国語に対応）。

2012年12月には、英語と中国語の翻訳に対応し<sup>131</sup>、現在14カ国語に対応している。欧州特許庁の発表では、2014年末までに日本語、韓国語、ロシア語を含め、全32カ国語を対応させることを見込んでいる<sup>132</sup>。

### グーグルとの提携への反発

欧州特許庁は欧州特許の機械翻訳サービスの提供事業者として、米グーグル社と提携している。このような提携に対する欧州機関の反応はどのようなものであったのか。

我々の調査によれば、欧州特許庁とグーグル社との提携についての反応は研究者によって様々である。我々がヒアリング調査を実施したエジンバラ大学の言語・認知・計算研究院のレネル教授によれば、特にグーグル社と欧州の研究機関が競合しているわけではなく、むしろ協力関係にあるとも考えられる。これに対して、同じエジンバラ大学の統計機械翻訳グループのコーエン教授によれば、欧州にはグーグルと同じような特許翻訳サービスを提供できる研究機関があるのに、欧州特許庁がグーグルを選択したので、不満を言う研究者がいた。また、ドイツ人工知能研究センター言語技術研究所のブズマ

<sup>128</sup> グーグル特許翻訳のサービス画面：<https://www.google.com/?tbn=pts>

<sup>129</sup> <http://support.google.com/contact/bin/answer.py?hl=en&answer=2539193>

<sup>130</sup> <http://googleresearch.blogspot.fr/2012/08/>

<sup>131</sup> <http://www.epo.org/news-issues/news/2012/20121206.html>

<sup>132</sup> <http://www.epo.org/searching/free/patent-translate/faq.html>

ン研究員は、より批判的である。欧州の研究者は面白いとは感じておらず、また、米国にサーバを持つグーグルに任せただけの場合、安全保障上の懸念がある。だが、特許は機械翻訳の主要分野ではないので、それほど気にする事なく、欧州の研究機関はグーグル社と共存できるとブズマン氏は考えている。

### **欧州単一特許の導入と機械翻訳サービスの動向**

さて、以上の欧州特許庁とグーグル社の提携と平行して、欧州では欧州単一特許の導入が進められている。先ほども触れたように、すでに欧州では欧州特許条約に基づき、欧州特許で特許出願手続きを行えば、全条約締約国で出願するのと同じ取り扱いを受けることになるシステムが採用されていたが、最終的な特許付与は各国の所管となった。それゆえ、各国毎に特許登録経費を支払ったり、特許を翻訳しなければならず、そのための費用が必要であった。また、特許に関して訴訟が起こされた場合、出願者は同じ一つの特許に対して、複数国で訴訟に対応しなければならなかった。つまり、現在までの欧州特許は全締約国内において単一の法制度で付与されているわけではなく、本当の意味で単一の欧州特許ではなく、欧州特許庁は国内特許の束を与えるだけであると言われている。

特許保護の費用に関しては、欧州委員会の発表では、現在まで全EU加盟国で特許保護を得るには約 3 万 6000 ユーロ（主に各国での認可手続き料と翻訳料）が必要であったが、欧州単一特許制度の導入によって、12 年間の移行期間では 6500 ユーロ、移行後では 5000 ユーロにまで費用が削減される見込みである<sup>133</sup>。なお、米国では約 2000 ユーロ、中国では約 600 ユーロかかる。

さて、以上のような現状を改善するため、欧州では欧州単一特許の必要性が議論され、「特許パッケージ」<sup>134</sup>という名で法整備が進められてきた。欧州単一特許の導入はEU 単一市場の強化につながると考えられており、1960 年代から議論が開始され、2000 年には欧州委員会がEU規則を提案したが、制定には至らなかった。

特許パッケージは3つの要素、2つのEU法と統一裁判所の設立協定からなる<sup>135</sup>。

二つのEU規則: 2012 年 12 月、欧州特許導入に関する二つのEU法、「単一特許保護創造の領域における強化された提携を実施する規則」と「適用可能な翻訳調整に関して単一特許保護創造の領域における強化された提携を実施する規則」が承認された。両規則は

<sup>133</sup> [http://europa.eu/rapid/press-release\\_MEMO-12-970\\_en.htm?locale=en](http://europa.eu/rapid/press-release_MEMO-12-970_en.htm?locale=en)

<sup>134</sup> [http://europa.eu/rapid/press-release\\_MEMO-12-970\\_en.htm?locale=en](http://europa.eu/rapid/press-release_MEMO-12-970_en.htm?locale=en)

<sup>135</sup> [http://ec.europa.eu/internal\\_market/indprop/patent/index\\_en.htm#proposal](http://ec.europa.eu/internal_market/indprop/patent/index_en.htm#proposal)

2013年1月20日に発効されたが、実際に適用されるのは2014年1月1日からである<sup>136</sup>。こうして、同規則に反対していたイタリアとスペイン以外のEU 25加盟国で、欧州単一特許が導入されることが決定した。同特許は欧州特許庁によって付与され、単一の法制度で特許が保護されることになり、各国所管機関での認可と翻訳の必要性がなくなった。これにより、特許付与システムが単純化され、出願者が支払う経費が削減されることになる。

欧州特許統一裁判所：2013年2月には、欧州単一特許に関わる訴訟を一貫して審理する統一裁判所設立に係る国際協定に、EU加盟国が調印しており、各国で批准される予定である。同裁判所は各国で異なる特許の法制度の矛盾を解消し、特許訴訟の費用の削減に貢献して、欧州特許導入に必要不可欠な機関である<sup>137</sup>。統一裁判所は、第一審裁判所（パリの予定）、控訴院（ルクセンブルグの予定）、登録所から構成される<sup>138</sup>。

#### 欧州単一特許への異議

先にも触れたように、イタリアとスペインは欧州単一特許制度に反対しており、両国では同制度は採用されていない。なぜ、両国は反対したのか。

欧州単一特許請願は、従来の欧州特許と同じく、英語、フランス語、ドイツ語で可能であり、上記3言語以外の言語で特許を請願する場合には、従来のように、これらの言語のどれかによる翻訳文を提出する必要がある。

報道によれば<sup>139</sup>、イタリア、スペイン両国は英語でのみ請願可能であるべきであり、そうでなければ差別であり、英語だけの方が政治的に受け入れやすいと主張した。欧州委員会は、このような両国の主張を、英語、フランス語、ドイツ語による既存の欧州特許請願手続きに人々は慣れており、2009年の段階で、欧州特許の半分はフランス語とドイツ語であったという理由で退けている。なお、両国はいつでも後から欧州単一特許制度を採用することができ、欧州委員会は両国の参加を呼びかけている。

<sup>136</sup> <http://www.epo.org/law-practice/unitary.html>

<sup>137</sup>

<http://www.consilium.europa.eu/homepage/highlights/agreement-on-unified-patent-court-signed?lang=en>

<sup>138</sup> <http://www.epo.org/law-practice/unitary/patent-court.html>

<sup>139</sup> <http://www.zdnet.com/meps-give-green-light-to-common-patent-system-3040091853/>

<http://www.zdnet.com/european-commission-moves-to-free-up-patent-logjam-3040089421/>

<http://www.euractiv.com/innovation-enterprise/italy-spain-block-eu-wide-patent-news-499638>

### 第三節 欧州におけるクラウドソース翻訳及び翻訳支援技術の研究開発の動向

本節では、欧州におけるクラウドソース翻訳と翻訳支援技術の動向について記す。クラウドソース翻訳（クラウドソーシング翻訳とも言われる）とは、不特定多数の人にインターネットを通して翻訳業務を委託する雇用形態のことであるが、技術の面では、専用のプラットフォームや、翻訳者をコンピューターによって支援する CAT 技術（Computer Assisted Translation）が関わる。以下に、a) EU の取り組みを概観した後、b) 英国と c) アイルランドの動向について記す。

#### A) EU の動向

まず、EUの動向で注目すべきことは、欧州委員会が「クラウドソーシング翻訳」と呼ばれる報告書を 2012 年 8 月に発表していることである<sup>140</sup>。内容は同分野の現状を概観するもので、「クラウドソーシング」という言葉の定義から始め、クラウドソーシング翻訳の世界的な展開状況を俯瞰し、その有用性や問題点等について記されている。したがって、欧州委員会の政策方針を述べるものではないが、関心の強さが伺える。

また、FP7 でも、CAT 技術に関わる研究開発が実施されており、MATECAT や CASMACAT があるが、これらには特にエジンバラ大学が関わっており、同大学が開発した機械翻訳システムの MOSES が利用されている（MATCAT と CASMACAT については、B) 英国の動向で詳しく記す）。

#### B) 英国の動向

##### エジンバラ大学

FP7 のプロジェクトで見たように、英国のエジンバラ大学は積極的に CAT に関わる研究開発を実施しており、FP7 の MATECAT プロジェクトと CASMACAT プロジェクトに参加している。

##### FP7：MATECAT の概要

略称	MATECAT
正式名称	機械翻訳を利用するコンピューター支援翻訳の精度の向上
公募分野	ICT-2011.4.2
研究期間	2012 年 2 月-2015 年 1 月（36 ヶ月）

<sup>140</sup> 同報告書は欧州委員会のウェブサイト入手可能である。

<http://ec.europa.eu/dgs/translation/publications/studies/>

全予算 (EU 拠出金)	336 万ユーロ (265 万ユーロ)
プロジェクトコーディネーター	ブルーノ・ケスラー基金 (伊)
参加組織	エジンバラ大学 (英)、TRANSLATED (伊)、ル・マン大学 (仏)
ウェブサイト	<a href="http://www.mateAat.com/">http://www.mateAat.com/</a>

#### FP7 : CASMACAT の概要

略称	CASMACAT
正式名称	先端コンピューター補助翻訳のための認知分析と統計手法
公募分野	ICT-2011.4.2
研究期間	2011 年 11 月-2014 年 10 月 (36 ヶ月)
全予算 (EU 拠出金)	321 万ユーロ (250 万ユーロ)
プロジェクトコーディネーター	エジンバラ大学 (英)
参加組織	コペンハーゲンビジネススクール (デンマーク)、CELER ソリューションズ (西)、バレンシア工科大学 (西)
ウェブサイト	<a href="http://www.casmacat.eu/">http://www.casmacat.eu/</a>

ヒアリング調査を実施したエジンバラ大学の統計機械翻訳グループのコーエン教授によれば、これら二つのプロジェクトは、主にプロの翻訳家を支援するため、機械翻訳により翻訳した後に行う後編集 (post-edit) を支援するツールを開発することを目的とする。現在、機械翻訳は完璧ではないので、翻訳者は、機械翻訳により翻訳された文章を後で修正する必要があるが、これら二つのプロジェクトは、このような後編集の作業を支援するためのツールや、タスクに適応変化可能なシステムを開発している。特に、オープンソースのインターフェイスを開発しており、様々なツールや機能が統合され、学習機能を備えたインタラクティブな機械翻訳システムを開発することを目指している。2つのプロジェクトの主な違いは、MATECAT が、翻訳企業が普段利用する機械翻訳向けのツールの開発を目的としているので、利用スピードが速く、ユーザーが利用しやすいツールやシステム等の翻訳者支援システムを開発しているのに対して、CASMACAT では、翻訳者の動作に注目し、よりインタラクティブな翻訳システムの研

究開発を実施している。CASMACAT では、翻訳者が翻訳作業を行う時、どのような動作を行っているかについて研究しており、例えば、認知科学を踏まえ、翻訳作業中の目の動き等について研究している。これらの2つのプロジェクトで開発されたツールやシステムは、プロの翻訳者向けに開発を進めているが、クラウドソース翻訳にも利用でき、ツールをクラウドソース翻訳のプラットフォームに装備することも可能である。

## C) アイルランドの動向

### **ロゼッタ基金**

#### ロゼッタ基金の概要

アイルランドでは、ロゼッタ基金 (Rosetta Foundation)<sup>141</sup>がクラウドソーシング翻訳に関わる活動を積極的に展開している。同基金は、2009年9月にアイルランドのリムリック大学ローカライゼーション研究センター<sup>142</sup>と次世代ローカライゼーションセンター (CNGL)<sup>143</sup>のスピノフ組織として設立され、アイルランドのダブリンに本拠地を持つ。同基金は非営利組織であり、アイルランド政府の支援を受けている。リムリック大学ローカライゼーション研究センターの所長ラインハルト・シェーラー (Reinhard Schäler) 氏が同基金の創設者であり、現在会長を務めている。同基金は、可能な限り多くの人々へ情報伝達することを可能にすることを目的としている。母国語で情報へアクセスすることは人間の権利であると考え、言語差別に基づく知識のギャップを取り除くために活動している。

同基金は、マイナー言語が利用され、十分に母国語で情報や知識を入手することができなかった共同体へ翻訳サービスを提供するために、各地の非営利組織に、同基金に登録している翻訳ボランティアを紹介する活動をしている。同基金には、約4000人の翻訳者がボランティアとして登録しており、同基金の活動を支えている。言語としては、アブハジア語、アフリカーンス語、アルバニア語、アラビア語、アルメニア語、アラゴン語、アヴァル語、アイマラ語、バスク語、ベラルーシ語、ベンガリ語、ブルガリア語、カタラン語、中国語、クロアチア語、チェコ語、デンマーク語、オランダ語、英語、エストニア語、フィンランド語、フランス語、ガリシア語、ドイツ語、ギリシア語、グジャラート語、ハウサ語、ヘブライ語、ヒンディー語、ハンガリー語、インドネシア語、アイルランド語、イタリア語、日本語、ルワンダ語、韓国語、クルド語、ラテン語、ラトビア語、リトアニア語、マケドニア語、マレー語、モルジブ語、マルタ語、ノルウェ

<sup>141</sup> <http://www.therosettafoundation.org/>

<sup>142</sup> <http://www.localisation.ie/>

<sup>143</sup> <http://www.cngl.ie/people/theme-leaders/>

一語、オリエント語、ペルシア語、ポーランド語、ポルトガル語、パシュトー語、ルーマニア語、ルンディ語、ロシア語、スロバキア語、スロベニア語、スペイン語、スワヒリ語、スウェーデン語、タガログ語、タミル語、トルコ語、ウクライナ語、ウルドゥー語、バレンシア語、ベトナム語、ウェールズ語、ヨルバ語に対応できる翻訳者が登録している。

#### トロモンズ

2013年3月18日には、同基金はトロモンズ (Translation Commons : Trommons)<sup>144</sup>というオンラインプラットフォームを創設している<sup>145</sup>。このクラウドソース翻訳プラットフォームでは、翻訳してほしい文書等を持つ非営利組織がその文書をウェブサイトにタスクとして上げ、翻訳者はそのタスクにアクセスし、翻訳することができる。翻訳文書分野は様々であり、法律、教育、健康、経済等がある。同サイトには、SOLAS (Service Oriented Localisation Architecture Solution) というローカライゼーションプラットフォームが利用されており、これはリムリック大学ローカライゼーション研究センターによって開発された。SOLASは、ORM設計原則 (Open, Right, Minimalistic) に基づいて設計されており<sup>146</sup>、オープンライセンスであるGPLライセンスで利用でき、SOLASのウェブサイト上でダウンロードできる<sup>147</sup>。

#### D) その他

ドイツの人工知能研究センター言語技術研究所とエジンバラ大学言語・認知・計算研究院でおこなったヒアリング調査によれば、これらの研究機関はクラウドソース翻訳について直接研究をしているわけではないが、機械翻訳により翻訳された文章の評価に利用している。

### **第一部のまとめ (欧州の市場、技術、研究体制)**

第一部では、欧州における機械翻訳技術と音声技術の研究開発及び同技術に関わる動向について記した。以上の調査結果を踏まえ、1) 市場、2) 技術、3) 研究体制という3つの観点から欧州の動向を簡単に記す。

1) 欧州は多数の国家から構成され、多数の言語が使用されている地域であるので、機械翻訳技術は自然に必要とされており、EUのFP7やCIP、そして、各国内で研究開

---

<sup>144</sup> <http://trommons.org/>

<sup>145</sup> <http://www.therosettafoundation.org/blog/press-releases/>

<http://www.therosettafoundation.org/trommons/>

<sup>146</sup> <http://www.therosettafoundation.org/trommons/technology-resources/>

<sup>147</sup> <http://www.localisation.ie/solas/>

発が積極的に促進されている。EU は言語の障壁を乗り越え、欧州単一市場創設のために、機械翻訳技術が必要であると考えており、FP7 や CIP には言語技術に特化した公募枠がある（例：FP7 ICT-2009.2.2: 言語を基盤とするインタラクション、FP7 ICT-2011.4.2: 言語技術、FP7 ICT-2013.4.1: コンテンツ分析と言語技術、CIP 目標 5.1: 多言語ウェブのための機械翻訳、CIP 目標 6.1: オープン言語インフラストラクチャ）。

このような欧州の事情は、米国の研究機関等が同技術の研究開発を進める事情とは異なる。ヒアリング調査を行ったエジンバラ大学統計機械翻訳グループのフィリップ・コーエン教授によれば、米国では、軍事機関や諜報機関向けに機械翻訳技術や音声技術の研究開発が積極的に実施されている。このような傾向に対して、欧州は、米国よりも翻訳の市場規模が大きく、様々な分野、特に、ニュース、技術文書、ビジネス等で機械翻訳の需要があり、さらに、欧州議会や欧州委員会等の EU 機関での機械翻訳の利用が進められている。従って、欧州では、軍事機関や諜報機関向けの研究開発は動機の 1 つに過ぎない。だが、欧州の研究開発機関（仏 LIMSI 等）も、米国防省の DARPA 等の軍事機関や諜報機関向けの研究開発支援プログラムに参加しており、また、仏独共同研究開発プログラム「クアエロ」には仏政府の軍備総局も関与しており、軍事・諜報分野は欧州でも無視できない研究開発の動機の 1 つである。

特許翻訳分野では、欧州特許庁が欧州の組織ではなく、米グーグル社と提携したことが問題となり、この提携に対して否定的な反応をした欧州研究者もいたが、特許翻訳は翻訳の一つの需要に過ぎないと考える研究者もいる。

フランスで行った機械翻訳技術と音声技術の利用状況のアンケート調査によれば、PC でグーグル翻訳等のテキスト機械翻訳アプリケーションを使用している人の割合は高い（全体の 76%）。スマートフォンやタブレットでテキスト機械翻訳のモバイルアプリケーションを使用している人の割合は、調査対象者全体の 40%である。音声機械翻訳モバイルアプリケーションを利用している人は非常に少ない（全体の 1%）が、同アプリケーションに関心がある人の割合は全体の 45%に至る。また、音声機械翻訳以外の音声技術を使用するシステムやアプリケーション（音声転写アプリケーションや音声対話システム）を使用している人の数も少ない（全体の 14%）が、関心がある人は全体の 30%に昇る。このように、テキスト機械翻訳アプリケーションに比べると、音声翻訳を含める音声アプリケーションを実際に利用している人の割合は少ないが、関心がある人の割合が少ないわけではないので、技術の精度を高めるとともに、関心がある人の実際の利用を促すために、現在どのような技術やアプリケーションがあるのか、その内容を一般に周知させることも必要であると考えられる。

2) 技術に関して言えば、まず、エジンバラ大学の統計機械翻訳グループが開発した

機械翻訳システムの MOSES の普及に注目すべきである。MOSES はオープンソースを理由に、欧州内外で、様々な研究機関（伊ブルーノ・ケスラー基金等）や民間組織（アジア・オンラインや SmartMATE）によって開発・利用されており、MOSES を利用する CAT (Computer Assisted Translation) ツール等を開発する FP7 プロジェクト (MATECAT や CASMACAT プロジェクト) も複数あり、また、MOSES の利用促進を目的とする FP7 プロジェクト (MOSESCORE プロジェクト) もあることから、今後も開発と利用が進められていくと考えられる。ついで、翻訳の精度の観点から、グーグル翻訳等の統計的手法だけに基づく機械翻訳システムへの批判がある。ヒアリング調査を行ったドイツ人工知能センター言語技術研究所のブズマン氏によれば、グーグル等の統計翻訳よりも、統計翻訳と規則（ルール）ベース翻訳を組み合わせた EuroMatrisPlus (FP7 プロジェクト) の翻訳システムの方が、翻訳の精度が高く、優れている。

クラウドソース翻訳に関しては、欧州委員会が「クラウドソーシング翻訳」と呼ばれる報告書を 2012 年 8 月に発表し、その概要と世界的な動向について報告しており、関心の高さが伺える上に、FP7 でも、クラウドソース翻訳に関わる CAT 技術が助成されている。また、欧州では、特にアイルランドのロゼッタ基金が、クラウドソース翻訳の活動に積極的である。

3) 研究体制に関しては、欧州各国には優れた研究機関があるとともに、これらの機関は FP7 のプロジェクト等を通して、しばしば共同で研究開発を実施していることが注目すべき点である。研究プロジェクトだけでなく、例えば、フランスの LIMISI とドイツの RWTH アーヘン大学とカールスルーエ技術研究院が共同で、多言語・マルチメディア技術研究院 (IMMI) という合同研究ユニットを創設している。また、META-NET という欧州の研究機関の提携を促進するアライアンスも設立されている（ドイツ人工知能センター言語技術研究所がコーディネーター）。機械翻訳技術は、必然的に他国の言語に関わる技術なので、他国の機関と提携して、研究開発を進めることが非常に重要である。従って、技術内容だけでなく、機械翻訳技術の研究提携の制度や提携活動の実行の面でも、FP7 等で他国との共同研究の経験が豊富な欧州から学ぶことは多いのではないかと。

## 第二部 旧ソ連及びアフリカ諸国における機械翻訳技術の研究開発等の動向

第二部では、旧ソ連及びアフリカにおける機械翻訳及び音声技術の研究開発等に関する動向について記す。以下に、参考のため、上記の2地域で使用人口の多い上位5言語を表にした（人口数については、Ethnologueを参照にしたが、資料によって若干異なる場合がある<sup>148</sup>）。

### 旧ソ連地域

言語名	使用者人口
ウクライナ語	36,028,490
アゼルバイジャン語	24,237,550
ウズベク語	21,930,230
カザフスタン語	8,077,770
ベラルーシ語	7,818,960

### アフリカ地域

言語名	使用者人口
アラビア語	223,010,130
ハウサ語	24,988,000
アムハラ語	21,811,560
ヨルバ語	19,380,800
ソマリ語	16,559,722

特に、アラビア語の使用者が多いことが際立っているが、アラビア語には、国により多くの方言があるので、音声技術の開発にはこの点にも留意する必要がある。

第一章では、旧ソ連及びアフリカ諸国における機械翻訳及び音声技術の実用化の状況、第二章では、言語処理技術の優れた研究機関の概要、第三章では言語処理の研究プロジェクト例、第四章では同2地域で需要のある翻訳の分野及び日本語との翻訳の需要のある分野、第五章では無線インフラの状況について記した。また、研究開発動向を詳しく調査するために、ナイジェリアのオバフェミ・アウオロウォ大学の研究グループにヒア

<sup>148</sup> <http://www.ethnologue.com/>

リング調査を行い、その議事録を収録した。

## 第一章 旧ソ連及びアフリカにおける機械翻訳技術及び音声技術の実用化の状況

本章では、旧ソ連及びアフリカ諸国の言語を対象とする機械翻訳技術と音声技術の実用化状況について記す。以下に、上記の地域で利用されている言語に対応する機械翻訳技術と音声技術のサービスやアプリケーションを開発、もしくは提供している企業（企業国籍を問わず）の概要を記す。

### 第一節 オンライン無料機械翻訳・音声技術サービス

#### 概要

旧ソ連及びアフリカ諸国の言語を対象とするオンライン機械翻訳サービスを提供している企業には米国企業が多い。欧州の CIP プロジェクトの iTranslate4 はオンライン無料機械翻訳サービスを提供しているが、ウクライナの民間研究組織トリデントが同プロジェクトに参加し、ウクライナ語に対応している。ロシアやウクライナには、グーグルや Yahoo! のように、オンライン機械翻訳サービスを提供しているインターネットポータルサイトもある（Yandex と Online.ua）。また、通常の翻訳や通訳のサービスを提供している企業が、インターネット上でオンライン無料翻訳サービスも提供している場合もある（英 Capital、英 SDL）。

対応言語に関しては、アフリカ諸国で使用されている言語では、アラビア語に対応するサービスが圧倒的に多く、旧ソ連諸国で使用されている言語では、ウクライナ語やバルト三国の言語（ラトビア語、エストニア語、リトアニア語）に対応しているサービスが多い。ナイジェリアのオバフェミ・アウォロウォ大学は、英語を Yoruba 語に翻訳するオンライン無料機械翻訳サービスを提供している。

#### 米国

企業名	Star21.com
企業本拠地	米国
企業種	オンライン無料機械翻訳サービス
サービス名	Stars21
提供内容	オンライン無料テキスト翻訳サービス、オンライン無料辞書 (機械翻訳技術は複数の企業のものを使用している)
ウェブサイト	<a href="http://www.stars21.com/home.html">http://www.stars21.com/home.html</a>

欧州言語	英語、ドイツ語、フランス語等
旧ソ連諸国の言語	アルバニア語、ベラルーシ語、エストニア語、ラトビア語、リトアニア語、ウクライナ語
アフリカ諸国の言語	アラビア語、アフリカーンス語、スワヒリ語

企業名	Smart link Coporation
企業本拠地	米国
企業種	オンライン無料翻訳サービス
サービス名	imTranslator
提供内容	オンライン無料テキスト翻訳サービス、オンライン無料辞書、オンライン無料音声合成サービス等（機械翻訳技術は多数の企業のものを使用している）
ウェブサイト	<a href="http://imtranslator.net/translation/">http://imtranslator.net/translation/</a> （機械翻訳） <a href="http://text-to-speech.imtranslator.net/">http://text-to-speech.imtranslator.net/</a> （音声合成サービスは欧州主要言語、中国語、日本語、韓国語、ロシア語にのみ対応）
欧州言語	英語、ドイツ語、フランス語等
旧ソ連諸国の言語	ベラルーシ語、ラトビア語、リトアニア語
アフリカ諸国の言語	アラビア語、ハウサ語、スワヒリ語

企業名	WorldLingo
企業本拠地	米国
企業種	通訳・翻訳サービス
サービス名	Free Translation Online
提供内容	オンライン無料テキスト機械翻訳サービス（有料サービスもあり）
ウェブサイト	<a href="http://www.worldlingo.com/en/products_services/worldlingo_translator.html">http://www.worldlingo.com/en/products_services/worldlingo_translator.html</a>
欧州言語	英語、独語、仏語等
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語、ハウサ語、ソマリ語

企業名	BableFish
-----	-----------

企業本拠地	米国
企業種	翻訳サービス
サービス名	バベルフィッシュオンライン翻訳
提供内容	オンラインテキスト無料翻訳サービス（有料サービスあり）
ウェブサイト	<a href="http://altavista.babelfish.com/?ptype=ask-a-question">http://altavista.babelfish.com/?ptype=ask-a-question</a>
欧州言語	英語、独語、仏語等
旧ソ連諸国の言語	ウクライナ語
アフリカ諸国の言語	アラビア語

企業名	マイクロソフト
企業種	IT システム及びインターネットサービス提供企業
企業本拠地	米国
サービス名	Bing 翻訳
サービス	オンラインテキスト無料機械翻訳サービス
ウェブサイト	<a href="http://www.bing.com/translator">http://www.bing.com/translator</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	エストニア語、リトアニア語、ウクライナ語
アフリカ諸国の言語	アラビア語

### カナダ

企業名	IAC
企業種	オンライン辞書、翻訳サービス
企業本拠地	カナダ
サービス名	Dictionary.com
サービス内容	オンライン無料機械翻訳サービス
ウェブサイト	<a href="http://translate.reference.com/">http://translate.reference.com/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	ベラルーシ語、エストニア語、ラトビア語、リトアニア語、ウクライナ語
アフリカ諸国の言語	アフリカーンス語、アラビア語、スワヒリ語

### CIP プロジェクト : iTranslate4（より詳しくは次章第一節を参考のこと）

サービス名	iTranslate4
-------	-------------

サービス内容	オンライン無料機械翻訳サービス
プロジェクトコーディネーター	ハンガリー科学アカデミー言語学研究所
ウェブサイト	<a href="http://itranslate4.eu/en/">http://itranslate4.eu/en/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	カザフ語、ラトビア語、ウクライナ語
アフリカ諸国の言語	アラビア語

### 英国

企業名	SDL
企業種	ビジネスコンサルティング、翻訳、通訳サービス
企業本拠地	英国
サービス名	SDLFreeTranslation. com
サービス内容	オンラインテキスト無料機械翻訳サービス
ウェブサイト	<a href="http://www.freetranslation.com/">http://www.freetranslation.com/</a>
欧州言語	英語、独語、仏語等
旧ソ連諸国の言語	ウクライナ語、エストニア語、リトアニア語、
アフリカ諸国の言語	アラビア語、ハウサ語、ソマリ語

### ロシア

企業名	Yandex
企業種	オンラインポータルサイト提供サービス
企業本拠地	ロシア
サービス名	Yandex translate
サービス内容	オンラインテキスト無料機械翻訳サービス
ウェブサイト	<a href="http://translate.yandex.com/?lr=87">http://translate.yandex.com/?lr=87</a>
欧州言語	英語、独語、仏語等
旧ソ連諸国の言語	ベラルーシ語、ウクライナ語
アフリカ諸国の言語	なし

### ウクライナ

企業名	Online.ua
企業本拠地	ウクライナ

企業種	オンラインポータルサイト提供サービス等
サービス名	pereklad.Online
提供内容	オンラインテキスト無料機械翻訳（トリデント社の技術を利用）
ウェブサイト	<a href="http://pereklad.online.ua/">http://pereklad.online.ua/</a>
欧州言語	英語、独語、仏語等、多数
旧ソ連諸国の言語	アルバニア語、ベラルーシ語、リトアニア語、ラトビア語、ウクライナ語、エストニア語
アフリカ諸国の言語	アラビア語、アフリカーンス語、

### ナイジェリア

開発者名	オバフェミ・アウオロウオ大学
企業本拠地	ナイジェリア
機関種	大学研究開発組織
サービス名	English to Yoruba Translator
提供内容	オンラインテキスト無料機械翻訳
ウェブサイト	<a href="http://eng2yor.appspot.com/translate">http://eng2yor.appspot.com/translate</a>
対象言語	英語から Yoruba 語への翻訳

## 第二節 機械翻訳・音声技術ソフトウェア

### 概要

旧ソ連及びアフリカ諸国の言語を対象とする機械翻訳や音声技術のアプリケーションを開発している企業に関しても、グーグル社を筆頭に米国企業が多い。だが、欧州企業、例えば、英 ATA、英 CAPITA、仏 Systran、独 LINGUATEC、オーストリア iTranslate、ハンガリー MorPhoLogic 等の企業も、欧州言語の他、同地域の言語を対象に開発を進めている。音声機械翻訳アプリケーションについては、アフリカ諸国の言語の中では、アラビア語に対応するアプリケーションが多く、旧ソ連諸国については、ウクライナ語の他、ラトビア語、リトアニア語、ラトビア語に対応するアプリケーションが多い。また、ウクライナのトリデント社は、積極的に東欧や旧ソ連諸国内の言語を対象とする機械翻訳技術の開発を実施している。また、これらのソフトウェア企業がオンライン無料翻訳サービスを提供している場合もある。言語学習アプリケーションでは、発音練習のために音声合成技術を使用することがある。

## 米国

企業名	Google
企業本拠地	米国
企業種	インターネット検索サービス等
サービス名	Google Translate
提供内容	オンライン無料テキスト翻訳アプリケーション（PC・モバイル）、音声機械翻訳アプリケーション（機械翻訳、音声認識、音声合成。PC・モバイル）、音声検索モバイルアプリケーション（音声認識技術を使用）
ウェブサイト	<a href="https://translate.google.com">https://translate.google.com</a>
欧州言語	英語、ドイツ語、フランス語等、多数
旧ソ連諸国の言語	アルメニア語（テキストのみ）、アゼルバイジャン語（テキストのみ）、ベラルーシ語（テキストのみ）、エストニア語（テキストのみ）、グルジア語（テキストのみ）、ラトビア語（テキストのみ）、リトアニア語（テキストのみ）、ウクライナ語（テキストのみ）
アフリカ諸国の言語	アフリカーンス語（音声あり）、アラビア語（音声あり）、スワヒリ語（テキストのみ）

企業名	Ectaco
企業種	ソフトウェア開発企業
企業本拠地	米国
アプリケーション名	Ectaco Universal Translator for Android
内容	アンドロイド OS 向けテキスト機械翻訳モバイルアプリケーション（有料）
ウェブサイト	<a href="http://www.ectaco.com/ECTACO-UT-for-Android/">http://www.ectaco.com/ECTACO-UT-for-Android/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	アルメニア語、エストニア語、グルジア語、ベラルーシ語、カザフ語、ラトビア語、リトアニア語、ウズベク語、
アフリカ諸国の言語	アフリカーンス語、アラビア語、ハウサ語、ソマリ語、スワヒリ語、ズールー語

企業名	SayHi
企業種	ソフトウェア開発企業
企業本拠地	米国
アプリケーション名	SayHi
内容	AndroidOS 及び iPhone 向けテキスト・音声翻訳モバイルアプリケーション（機械翻訳、音声認識、音声合成技術の使用。欧州主要言語に関しては無料。幾つかのアフリカ言語、旧ソ連諸国の言語、アジア言語に関しては有料）
ウェブサイト	<a href="http://www.sayhitranslate.com/">http://www.sayhitranslate.com/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	アルバニア語、アルメニア語、アゼルバイジャン語、ベラルーシ語、エストニア語、ラトビア語、リトアニア語、ウクライナ語
アフリカ諸国の言語	アフリカーンス語、スワヒリ語、アラビア語

企業名	SpeechTrans
企業種	ソフトウェア開発企業
企業本拠地	米国
サービス名	SpeechTrans
提供内容	PC 向け音声転写ソフトウェア（音声認識）・音声機械翻訳ソフトウェア（機械翻訳、音声認識、音声合成）、AndroidOS 及び iPhone 向けテキスト・音声機械翻訳モバイルアプリケーション、音声翻訳電話アプリケーション
ウェブサイト	<a href="http://speechtrans.com/products/">http://speechtrans.com/products/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語（音声合成のみ）

企業名	LEC
企業種	ソフトウェア開発企業
企業本拠地	米国
サービス名	Translate

提供内容	テキスト機械翻訳ソフトウェア (PC 向け)
欧州言語	英語、仏語、独語等
ウェブサイト	<a href="http://www.lec.com/default.asp">http://www.lec.com/default.asp</a>
旧ソ連諸国の言語	ウクライナ語
アフリカ諸国の言語	アラビア語

企業名	Transparent Language
企業種	外国学習サービス提供企業
企業本拠地	米国
サービス名	Byki Mobile
サービス	外国語の学習向けモバイルアプリケーション (発音練習向けの音声合成)
ウェブサイト	<a href="http://www.transparent.com/">http://www.transparent.com/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	アルメニア語、アゼルバイジャン語、ベラルーシ語、エストニア語、カザフ語、ラトビア語、リトアニア語、トルクメニスタン語
アフリカ諸国の言語	アラビア語、スワヒリ語、ズールー語

企業名	Alterme
企業種	ソフトウェア開発企業
企業本拠地	米国
サービス名	Translator with voice
サービス	テキスト・音声機械翻訳アプリケーション
ウェブサイト	<a href="http://www.ifomia.com/apps/translator/index.html">http://www.ifomia.com/apps/translator/index.html</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	リトアニア語 (音声なし)、ウクライナ語 (音声なし)
アフリカ諸国の言語	アラビア語 (音声なし)

企業名	Unatezsoft
企業種	ソフトウェア開発企業
企業本拠地	米国

サービス	ヨルバ語・イボ語等の外国語学習アプリケーション（発音練習向け音声合成システムあり）
ウェブサイト	<a href="http://www.unatezesoft.com/">http://www.unatezesoft.com/</a>
欧州言語	仏語、スペイン語
旧ソ連諸国の言語	なし
アフリカ諸国の言語	ヨルバ語、イボ語

### カナダ

企業名	Pocketglow
企業種	ソフトウェア開発企業
企業本拠地	カナダ
サービス名	SpeakEasy Ukrainian
サービス	旅行に役立つフレーズを音声で出力するモバイルアプリケーション
ウェブサイト	<a href="http://pocketglow.com/speakeasyukrainian/">http://pocketglow.com/speakeasyukrainian/</a>
旧ソ連諸国の言語	ウクライナ語

### 英国

企業名	ATA
企業種	ソフトウェア開発企業
企業本拠地	英国
サービス名	AL-Mutarijim
提供内容	アラビア語と英語向けテキスト機械翻訳ソフトウェア
ウェブサイト	<a href="http://eshop.atasoft.com/eng/software.html">http://eshop.atasoft.com/eng/software.html</a>
アフリカ諸国の言語	アラビア語

企業名	CAPITA
企業種	翻訳と通訳サービス（機械翻訳を含む）
企業本拠地	英国
サービス名	SmatMate
サービス	テキスト機械翻訳
ウェブサイト	<a href="http://www.capitatranslationinterpreting.com/">http://www.capitatranslationinterpreting.com/</a>
欧州言語	英語、仏語、独語等

旧ソ連諸国の言語	アルバニア語、ラトビア語、リトアニア語、エストニア語、ウクライナ語
アフリカ諸国の言語	アラビア語

## フランス

企業名	Systran
企業種	ソフトウェア開発企業
企業本拠地	フランス
サービス名	Systran
提供内容	テキスト機械翻訳ソフトウェア（モバイルアプリケーションあり）及びオンラインテキスト機械翻訳サービス（オンライン無料サービスあり）
ウェブサイト	<a href="http://www.systran.fr/">http://www.systran.fr/</a> 、 <a href="http://www.systranet.com/translate">http://www.systranet.com/translate</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語

企業名	ヴォカピアリサーチ
企業種	ソフトウェア開発企業
企業本拠地	フランス
製品名	VoxSigma（企業向け）
製品内容	音声認識、話者特定、音声転写ソフトウェア
ウェブサイト	<a href="http://www.vocapia.com/index.html">http://www.vocapia.com/index.html</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	ウクライナ語（現在開発中）
アフリカ諸国の言語	アラビア語

## ドイツ

企業名	Linguatec
企業種	ソフトウェア開発企業
企業本拠地	ドイツ
製品	Travel Voice Translator
製品内容	旅行向けテキスト・音声機械翻訳モバイルアプリケーション

	(機械翻訳、音声認識、音声合成)
ウェブサイト	<a href="http://www.linguatec.de/">http://www.linguatec.de/</a>
欧州言語	英語、仏語、独語等
旧ソ連諸国の言語	アルメニア語 (テキストのみ)、アゼルバイジャン語 (テキストのみ)、ベラルーシ語 (テキストのみ)、エストニア語 (テキストのみ)、グルジア語 (テキストのみ)、ラトビア語 (テキストのみ)、リトアニア語 (テキストのみ)、ウクライナ語 (テキストのみ)
アフリカ諸国の言語	アフリカーンス語 (テキストのみ)、アラビア語 (音声あり)、スワヒリ語 (テキストのみ)

### オーストリア

企業名	Sonico
企業種	ソフトウェア開発企業
企業本拠地	オーストリア
サービス名	iTranslate
サービス	テキスト・音声機械翻訳アプリケーション (機械翻訳、音声認識、音声合成)
ウェブサイト	<a href="http://www.itranslateapp.com/">http://www.itranslateapp.com/</a>
欧州言語	英語、仏語、独語等、多数
旧ソ連諸国の言語	アゼルバイジャン語 (テキストのみ)、ベラルーシ語 (テキストのみ)、エストニア語 (テキストのみ)、グルジア語 (テキストのみ)、ラトビア語 (テキストのみ)、リトアニア語 (テキストのみ)、セルビア語 (テキストのみ)、ウクライナ語 (テキストのみ)
アフリカ諸国の言語	アフリカーンス語 (テキストのみ)、アラビア語、スワヒリ語 (テキストのみ)

### ハンガリー

企業名	MorphoLogic
企業種	ソフトウェア開発企業
企業本拠地	ハンガリー
製品名	Morpho Word

サービス	機械翻訳ソフトウェア (PC 向け)
ウェブサイト	<a href="http://www.morphologic.hu/index.php">http://www.morphologic.hu/index.php</a>
欧州言語	英語
旧ソ連諸国の言語	ウクライナ語
アフリカ諸国の言語	なし

### ラトビア

企業名	ティルド
企業種	ソフトウェア開発企業
企業本拠地	ラトビア
サービス名	機械翻訳
提供内容	機械翻訳モバイルアプリケーション
ウェブサイト	<a href="http://www.tilde.com/">http://www.tilde.com/</a> <a href="http://www.tilde.com/translator">http://www.tilde.com/translator</a>
欧州言語	英語
旧ソ連諸国の言語	ラトビア語、リトアニア語
アフリカ諸国の言語	なし

### ウクライナ

企業名	トリデント・ソフトウェア
企業種	ソフトウェア開発企業
企業本拠地	ウクライナ
サービス名	Pragma
提供内容	テキスト機械翻訳ソフトウェア (オンラインテキスト無料機械翻訳サービスも提供している)
ウェブサイト	<a href="http://translate.ua/us/pragma-6x">http://translate.ua/us/pragma-6x</a> <a href="http://online.translate.ua/">http://online.translate.ua/</a>
旧ソ連諸国の言語	アゼルバイジャン語、アルバニア語、アルメニア語、ベラルーシ語、グルジア語、リトアニア語、エストニア語、ラトビア語、ウクライナ語、カザフ語
アフリカ諸国の言語	アラビア語、スワヒリ語

## ガーナ

企業名	Nkyea Learnig Systems
企業種	ソフトウェア開発企業
企業本拠地	ガーナ
製品名	Nkyea
サービス	トウイ語・ハウサ語・スワヒリ語・ヨルバ語学習アプリケーション（発音練習向け音声合成システム）
ウェブサイト	<a href="http://www.nkyea.com/">http://www.nkyea.com/</a>
旧ソ連諸国の言語	なし
アフリカ諸国の言語	トウイ語、ハウサ語、スワヒリ語

## クウェート

企業名	Sakhr
企業種	ソフトウェア開発企業
製品 1	機械翻訳
製品内容	英語とアラビア語の機械翻訳ソフトウェア
製品 2	Sakhr S25
提供内容	アラビア語と英語向け音声翻訳モバイルアプリケーション
ウェブサイト	<a href="http://www.sakhr.com/tts.aspx">http://www.sakhr.com/tts.aspx</a>
旧ソ連諸国の言語	なし
アフリカ諸国の言語	アラビア語

## 第二章 旧ソ連及びアフリカ諸国における言語処理に関する優れた研究機関の概要

本章では、旧ソ連諸国及びアフリカ諸国における言語処理に関する優れた研究機関（民間研究機関も含む）について記す。

### 第一節 旧ソ連諸国

旧ソ連諸国で利用されている人口が多い言語は、ウクライナ語、アゼルバイジャン語、ウズベキスタン語である。アゼルバイジャン語の研究開発に関しては、次章で記すことにし、以下にウクライナとウズベキスタンの研究機関について記す。また、ラトビア語

を利用する人口は少ないものの（約 147 万人<sup>149</sup>）、ラトビアのティルド社は多くのFP7に参加している。

#### ウクライナ：ウクライナ人工知能研究院音声認識研究グループ

ウクライナ人工知能研究院<sup>150</sup>は、1991年に、ウクライナ教育・科学省と国立科学アカデミーの傘下にドネツクに創設された。同研究院では、人工知能を中心に様々なテーマで研究が実施されているが、言語技術に関しては、音声認識研究科<sup>151</sup>で研究開発が実施されている。

同研究科の主な研究テーマは、ロシア語とウクライナ語の音声認識システムの開発と、その応用研究である。現在まで開発成果としては、Voicedialing Phone（音声により電話番号を認識させ、発信するシステム）、Voice input formulas editor（数式を音声で入力するシステム）、ロシア語-英語音声コントロールフレーズブック（ロシア語の音声を英語に翻訳し、音声合成により英語の音声を出力する）等がある。

#### ウクライナ：トリデント・ソフトウェア

トリデント・ソフトウェア<sup>152</sup>は、機械翻訳システムの開発を行っているソフトウェア開発企業で、1998年にキエフに創設された。特に、ウクライナ語を中心とする翻訳技術の開発を実施しており、同社が開発したPragma 6.xは、英語、独語、仏語、ラトビア語、ロシア語、ウクライナ語、ポーランド語、カザフ語に対応している。オンライン無料機械翻訳サービス<sup>153</sup>も同社のウェブサイトを提供している。同社の製品は、ウクライナの大統領秘書室や閣僚官房、国立銀行等の政府機関や、産業や金融部門の民間企業、学校等で利用されている。また、トリデントは欧州のCIPプロジェクトでiTranslate4<sup>154</sup>にも参加しており、同プロジェクトはオンライン無料機械翻訳サービスを開発している。

参考：CIP プロジェクト

略称	iTranslate4
プロジェクト名	全欧州言語向けインターネット翻訳
プログラム分野	CIP-ICT-PSP.2009.5.1
全予算（EU 拠出金）	394 万ユーロ（197 万ユーロ）

<sup>149</sup> <http://www.ethnologue.com/language/lav>

<sup>150</sup> <http://iai.donetsk.ua/>

<sup>151</sup> [http://iai.dn.ua/general/frameset.php3?l=e&p=r&f=/general/index.php3?d=3\\*l=e](http://iai.dn.ua/general/frameset.php3?l=e&p=r&f=/general/index.php3?d=3*l=e)

<sup>152</sup> <http://translate.ua/us/company>

<sup>153</sup> <http://translate.ua/us/on-line>

<sup>154</sup> <http://itranslate4.eu/en/>

研究期間	2010/3- 2012/2 (24 ヶ月間)
ウェブサイト	<a href="http://itranslate4.eu/en/">http://itranslate4.eu/en/</a>
コーディネーター	ハンガリー科学アカデミー言語学研究所
参加組織	トリデント MT (トリデントのラトビア支社)、MorphoLogic (ハンガリー)、Systran (仏)、Linguattec (独)、pwn.pl (ポーランド)、SkyCode (ブルガリア)、Amebis (スロベニア)、Sunda (フィンランド)、Promt (ロシア)

#### ウズベキスタン：ユニコン・UZ

ユニコン・UZ<sup>155</sup>は、1992年にウズベキスタン共和国の通信省の命令により創設された電気通信部門の科学研究センターが前身機関であるが、現在は研究開発だけでなく、電気通信部門の市場調査、標準化等、総合的に電気通信部門の様々な研究活動を実施している。

言語処理技術に関しては、音声・言語情報化研究科<sup>156</sup>で実施されている。主に、CAT (Computer-aided translation)、音声認識、音声対話、機械翻訳技術の研究が実施されている。電子辞書 (ロシア語-ウズベク語-英語間) とウズベク語のスペルチェックシステムを開発している<sup>157</sup>。機械翻訳に関しては、日本語と中国語への対応も目指している。

#### ラトビア：ティルド

ティルド<sup>158</sup>は1991年に創設されたラトビアの民間研究開発組織で、主に利用者が少ない言語向けに機械翻訳や音声技術、スペルチェックシステム等の言語技術の研究開発を実施し、商用化を行っている。機械翻訳に関しては、統計ベースの手法と規則ベースの手法を統合して開発している。音声技術に関しては、音声合成技術を開発しており、テキストを読み上げるシステムを開発している。開発製品には、ラトビア語とリトアニア語向けのワープロソフト (音声出力、スペルチェック、電子辞書等が付属)<sup>159</sup>と機械翻訳モバイルアプリケーション (ラトビア語、英語、ロシア語に対応)<sup>160</sup>がある。

ティルドは、FP7 やCIP等の欧州プロジェクトに積極的に参加していることが特徴であり、欧州を中心に非常に多くの研究プロジェクトパートナーを持つ (米国からはマイ

<sup>155</sup> <http://unicon.uz/en/>

<sup>156</sup> [http://unicon.uz/en/about/etri/department\\_4.html](http://unicon.uz/en/about/etri/department_4.html)

<sup>157</sup> <http://unicon.uz/en/products-services/products>

<sup>158</sup> <http://www.tilde.com/>

<sup>159</sup> <http://www.tilde.com/products-and-services/language-products/translation-tools>

<sup>160</sup> <https://itunes.apple.com/jv/app/tilde-translator/id459067712?mt=8>

クrosoft) <sup>161</sup>。FP7 に関しては、TAAS、Accurat、METANORDのコーディネーターを務めている。また、CIPのLETSMT !、META-NORDのコーディネーターを務めている。

参考：ティルドが参加する FP7 プロジェクトと CIP プロジェクト

#### FP7

略称	TAAS
プロジェクト名	サービスとしてのターミノロジー
プログラム分野	ICT-2011.4.1
全予算 (EU 拠出金)	229 万ユーロ (182 万ユーロ)
研究期間	2012/6– 2014/5 (24 ヶ月間)
ウェブサイト	<a href="http://www.taas-project.eu/">http://www.taas-project.eu/</a>
コーディネーター	TILDE (ラトビア)
参加組織	ケルン応用科学大学 (独)、Kilgray (ハンガリー)、シェフィールド大学 (英)、TAUS (蘭)
研究内容	専門用語のデータの収集、共有、再利用を可能にするクラウド型のプラットフォームを開発する。専用プラットフォームには、CAT ツールを実装し、各種専門分野に対応する統計機械翻訳を提供する。

#### FP7

略称	ACCURAT
プロジェクト名	機械翻訳のリソースが少ない分野向け比較可能コーポラの分析と評価
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	355 万ユーロ (282 万ユーロ)
研究期間	2010/1 – 2012/6 (30 ヶ月間)
ウェブサイト	<a href="http://www accurat-project.eu/">http://www accurat-project.eu/</a>
コーディネーター	TILDE (ラトビア)
参加組織	シェフィールド大学 (英)、リード大学 (英)、アテネ情報通信・知識技術研究・技術革新センター (ギリシア)、ザグレブ大学 (クロアチア)、DFKI (独)、ルーマニアアカデミー・人

<sup>161</sup> <http://www.tilde.com/about-tilde/partnership>

	工知能研究院（ルーマニア）、Linguatec（独）、Zemanta（スロベニア）
研究内容	現在機械翻訳の精度はコーポラの規模に従い、言語リソースが小さい言語に関しては精度が落ちてしまう。このため、言語リソースが少ない言語向けにコーポラを開発する新しい方法、ツールを開発する。

### CIP

略称	METANORD
プロジェクト名	欧州バルト諸国・北欧オープン言語インフラストラクチャ
プログラム分野	CIP-ICT-PSP-2010-4
全予算（EU 拠出金）	450 万ユーロ（225 万ユーロ）
研究期間	2011/2- 2013/1（24 ヶ月間）
ウェブサイト	<a href="http://www.meta-nord.eu/">http://www.meta-nord.eu/</a>
コーディネーター	TILDE（ラトビア）
参加組織	コペンハーゲン大学（デンマーク）、タルチュ大学（エストニア）、ベルゲン大学（ノルウェー）、ヘルシンキ大学（フィンランド）、アイスランド大学（アイスランド）、リトアニア言語研究院（リトアニア）、ヨーテボリ大学（スウェーデン）
研究内容	バルト三国と北欧諸国の言語（デンマーク語、エストニア語、フィンランド語、アイスランド語、ラトビア語、リトアニア語、ノルウェー語、スウェーデン語）向けにオープン言語インフラストラクチャを設立する。そのため、これらの言語リソースを収集し、META-NET と提携して、リソースの共有を行う。

### CIP

略称	LETSMT !
プロジェクト名	トレーニングデータのオンライン共有とユーザーに適応する機械翻訳のためのプラットフォーム
公募分野	CIP ICT PSP. 2009. 5. 1: 多言語ウェブ: 多言語ウェブのための機械翻訳
全予算（EU 拠出金）	334 万ユーロ（167 万ユーロ）

研究期間	2010年3月-2012年8月（30ヶ月）
ウェブサイト	<a href="http://project.letsmt.eu/">http://project.letsmt.eu/</a>
コーディネーター	Tilde（ラトビア）
参加組織	エジンバラ大学（英）、ザグレブ大学（クロアチア）、コペンハーゲン大学（デンマーク）、ウプサラ大学（スウェーデン）、ZOOROBOTICS（蘭）、モラビア（チェコ）
研究内容	機械翻訳技術の精度は段々と向上しているものの、使用者が少ない言語（ラトビア語、リトアニア語、クロアチア語等）に関してはそうではない。これらの言語のために、データ共有と機械翻訳の構築のためのオンライン・提携プラットフォームを、MOSESを利用して開発する。

## 第二節 アフリカ諸国

本節では、アフリカ諸国における機械翻訳及び音声技術の研究開発を実施している研究開発機関について記す。音声技術の総合的な研究を実施している南アフリカのMeraka 研究院、スワヒリ語の機械翻訳技術等の研究を実施しているケニアのナイロビ大学、ヨルバ語を中心に研究を実施しているナイジェリアのオバフェミ・アウオロウオ大学について記す。また、アフリカ諸国の研究開発動向を知るため、オバフェミ・アウオロウオ大学の研究員にヒアリング調査を実施した。その議事録を収録する。

### 南アフリカ共和国：Merake 研究院 人間言語技術研究所

Merake研究院<sup>162</sup>は、2002年にムベキ大統領（当時）の主導で創設されたICT部門に特化した研究機関であり、研究者の育成、応用開発、最先端技術研究という3つの活動を行っている。同研究院では、スーパーコンピューター、地球観測向けICT、教育部門へのICTの応用開発、ハンディキャップ支援向けICTの研究開発活動等を実施している。言語技術に関しては、人間言語技術研究所<sup>163</sup>で研究が行われている。

同研究所では、言語技術を利用し、他の様々な技術を強化すること、そして、言語の障壁を打ち破ることを目標に置いている。特に後者の目標に関しては、南アフリカ共和国の公用語は11言語（ズールー語、コサ語、ベティ語、ツワナ語、ソト語、ツォンガ語、スワティ語、ヴェンダ語、ンデベレ語、アフリカーンス語、英語）であり、同技術は政府と企業が文書等を作成する際に役立つ、これにより特定の言語への歴史的な差別

<sup>162</sup> <http://www.meraka.org.za/about.htm>

<sup>163</sup> <http://www.meraka.org.za/humanLanguage.htm>

が解消されると考えられている。また、言語の多様性の他、オープンソースによるデータの共有、応用研究、国際共同研究の実施、若手研究者への研究環境の供与が研究方針として挙げられている。

研究テーマ<sup>164</sup>は、音声認識、音声合成、話者言語特定、イントネーションモデリング、発音モデリング、トピックモデリング、統計パターン認識である。研究プロジェクトとしては、多数の国内プロジェクトの他（11 公用語のリソース収集、目に障害がある人向けの筆記デバイスの開発、子供の声の音声認識及び音声合成システムの開発、言語学習アプリケーション、政府機関向けの 11 公用語に対応する音声技術を利用する電話システム<sup>165</sup>）、グーグル社のアンドロイド向けアプリケーションの開発や、FP7 プロジェクト VOICES（同プロジェクトに関しては、次章で詳述する）に参加している。

現在まで様々な製品を開発しており、Woefzela（アンドロイドOSを装備したスマートフォン向けモバイルアプリケーション。リソースが少ない言語の音声データを収集することを可能にする）<sup>166</sup>、DictionaryMaker（発音の電子辞書を作成するためのアプリケーション）<sup>167</sup>、Speect（音声合成システム）<sup>168</sup>、英語学習アプリケーション（特に、発音練習向け）がある。

提携パートナーとしては、南アフリカ共和国の諸大学、インド、オランダ、ベルギー、英国、米国と提携して研究開発を実施している。特に、同国の7つの大学と全国規模で人間言語技術ネットワークを形成し、同分野の研究開発を提携して実施している。

#### ケニア：ナイロビ大学 コンピューティング・情報学部

ナイロビ大学のコンピューティング・情報学部<sup>169</sup>では、様々なICT部門の研究が実施されている。言語技術に関しては、人間言語技術グループで、言語リソース、音声言語処理、機械翻訳等の研究開発が実施されている。研究プロジェクトとしては、1985 年以来、SALAMAプロジェクト<sup>170</sup>というスワヒリ語を対象とする言語技術の研究開発を、フィンランドのヘルシンキ大学とタンザニアのダルエスサラーム大学の研究組織と実施している。同プロジェクトでは、テキスト機械翻訳技術の開発を実施しており、スワヒリ語と英語間の翻訳システムを開発している。主に規則ベースに基づく手法を利用しているが、統計ベースの手法を組み合わせることも検討している。また、音声

<sup>164</sup> [http://www.csir.co.za/meraka/hlt/hlt\\_speech.html](http://www.csir.co.za/meraka/hlt/hlt_speech.html)

<sup>165</sup> <http://www.meraka.org.za/lwazi/about.php>

<sup>166</sup> <https://sites.google.com/site/woefzela/>

<sup>167</sup> <http://sourceforge.net/projects/dictionarymaker/>

<sup>168</sup> <http://www.meraka.org.za/lwazi/>

<sup>169</sup> <http://sci.uonbi.ac.ke/node/947>

<sup>170</sup> <http://www.njas.helsinki.fi/salama/index.html>

翻訳技術の開発も計画している。また、ANLocというカナダの組織IDRC (International Development Research Centre) <sup>171</sup>が助成するプロジェクトに参加している。このプロジェクトは、主にアフリカ諸言語へのローカライゼーション活動を促進するため、フォント、キーボードの開発から、ローカライゼーションツールやソフトウェア、スペルチェッカーの開発を実施している。

#### ナイジェリア：オバフェミ・アウオロウオ大学 コンピューティング・インテリジェントシステム研究グループ

オバフェミ・アウオロウオ大学のコンピューターサイエンス・工学部コンピューティング・インテリジェントシステム研究グループ<sup>172</sup>では、言語技術の包括的な研究が実施されている。人員は全部で研究者が18名である。研究テーマは、英語と特に Yoruba 語の言語技術の開発であり、音声合成、音声認識、機械翻訳、音声対話システム、言語学習システム、電子辞書の開発を行っている。現在、英語を Yoruba 語へと翻訳する機械翻訳システムをオンラインで提供している<sup>173</sup>。

より詳しい同組織の研究開発動向とアフリカにおける言語技術の研究開発について調査するために、同大学にヒアリング調査を実施した。以下に、その議事録を収録する。

#### ヒアリング議事録 ナイジェリア オバフェミ・アウオロウオ大学

##### **日程**

平成25年6月20日午後3時～午後4時

##### **場所**

NICT欧州連携センター事務所（フランス・パリ）<sup>174</sup>

##### **先方 (O)**

オバフェミ・アウオロウオ大学 コンピューターサイエンス・工学部 コンピューティング・インテリジェントシステム研究グループ<sup>175</sup>：アキナデ・グベンガ氏他、数名

##### **当方 (Δ)**

NICT 欧州連携センター長：菱沼 宏之

---

<sup>171</sup> <http://www.africanlocalisation.net/fr/node/42>

<sup>172</sup> <http://ifecisrg.org>

<sup>173</sup> <http://www.ifecisrg.org/yorubanumerals>

<sup>174</sup> インターネット電話サービス、スカイプを使用し、インタビューを実施した。

<sup>175</sup> 参考：<http://www.ifecisrg.org/>

ONOSO 研究員：小野 浩太郎

### **調査目的**

アフリカにおける機械翻訳及びスピーチ技術に関する動向を把握するために、ナイジェリアのオバフェミ・アウオロウオ大学 コンピューターサイエンス・工学部の研究者にヒアリング調査を実施した。同機関は、アフリカ言語の機械翻訳、音声認識、音声合成システム等に関連する技術の研究開発を包括的に進めており、アフリカにおける同分野の主要な研究機関の一つと考えられる。1) 同機関の概要、2) ナイジェリアで使用されている言語、3) 先方組織の言語技術の研究開発及び実用化状況、4) アフリカ諸国における言語技術の利用状況について主に質問した。

### **質問事項に対する回答**

#### **1) 先方の研究グループについて**

(△) あなたの研究グループの概要について教えていただきたい。

(○) 我々の研究グループは 2011 年に創立されたが、研究者たちはそれ以前より様々な技術開発を行っている。我々のグループには 18 名の研究者がおり、その内訳は博士号取得者 3 名、博士課程の学生 7 名、修士課程の学生 8 名である。

予算に関しては、2012 年度は 150 万米ドルであった。現在まで、この予算は大学からではなく、研究者の給料及びコンサルティング業務等によって拠出されている。その他、2000 米ドルをアフリカ言語技術イニシアティブから 2011 年にワークショップを開催するために得た。また、ナイジェリア政府から TETFUND という研究補助金 400 万ナイラ（約 2 万 5000 ドル）を 2011 年以来支給されている。

我々のグループでは言語に関わる様々な技術を開発しているが、一般的な関心は、コンピューターサイエンスによる近代的な技術とツールを使用して、アフリカ言語によりコンピューター部門の研究を行い、アプリケーションを開発することである。

研究パートナーに関しては、我々はナイジェリアのアフリカ言語技術イニシアティブ（ナイジェリア・イバダン）、アイルランドのヨーク制約コンピューターセンター（リチャード・ワラス博士）、英国のアストン大学（A.J.ボーモント博士と S.H.S ウォン博士）と提携している。

(△) コンサルティング業務に関しては、どのようなことを行っているのか。

(○) 言語技術を利用する新しいシステムやアプリケーションの開発に助言をしたりしている。例えば、教育分野での ICT ツール等の開発を例に挙げることができる。

(△) 民間企業と提携して開発を行っているか。

(○) 例えば、マイクロソフトのビスタに関しては、開発の際にコンサルティング業務

を行った。

## 2) ナイジェリアで利用されている言語について

(△) ナイジェリアの公用語は英語であると聞いた。英語も含め、言語の使用状況について教えていただきたい。

(○) ナイジェリアでは約 350 言語が使用されている。約 30%の国民が英語を話す、それ以外の人々はアフリカの言語しか話せない。アフリカ言語の中でも、特にハウサ語、ヨルバ語、イボ語が多くの人々により使用されている。使用者が最も多いのはハウサ語で約 40%、ヨルバ語は約 25%、イボ語は約 15%のである。また、全員ではないが、英語を話せる人々の多くは他のアフリカ言語も話すことができ、エリート層では2つの言語を使用できる人が多い。

(△) 英語が教育やビジネスで使用される言語であるか。

(○) その通り。なお、経済都市であるラゴスでは、中国語も使用されている。

(△) どのくらいの数の人々が中国語を話せるのか。

(○)それほど数は多くないと思う。

(△) 日本の企業がナイジェリアでビジネスをするには、機械翻訳技術は必要と思うか。

(○) 現在、日本のホンダ、トヨタ、ヨコハマ (タイヤ) 等の企業がナイジェリアで事業を行っており、彼らは機械翻訳を必要としていると思う。

(△) 現在、これらの日本企業は機械翻訳を利用していると思うか。

(○) 否。まだ利用していないと思う。だが、英語を話せない人が人口の 70%を占めるので、彼らと交流するには機械翻訳が必要となる。

## 3) 先方組織の言語技術の研究開発状況について

(△) あなた方は、現在どのような言語技術研究プロジェクトを実施しているか。幾つか紹介していただきたい。

(○) 以下に幾つか紹介しよう。

- ・ 機械翻訳：英語の文章をヨルバ語に自動翻訳する技術を開発している。この研究は研究者、言語学習者、言語教育者に言語リソース（英語—ヨルバ語）を供給し、彼らは自分の仮説をテストし、理論を正当化するためにそのリソースを利用できる。
- ・ アフリカ民話：アフリカの民話をコンピューターゲームと教育関連のソフトウェアアプリケーション向けに形式化するために、計算モデルを開発している。
- ・ 音声合成：この研究では、ヨルバ語テキストに対応する音を発生させるシステム

を開発する。このシステムを通して、自動的にヨルバ語テキストを音読することが可能になる。これはハンディキャップ支援とともに、言語教育に役立つ。このシステムはモバイルやマルチメディアプラットフォームでも利用可能である。この研究は自律分節音韻理論と言語計算理論に基づき、FST (Finite State Transducer) を利用するために設計されており、Festival TTSエンジン<sup>176</sup>を使って提供される。

- ・ 音声認識：ヨルバ語の母音を認識するための計算モデルを開発しており、そのため、fuzzy logicとneural networkという二つの計算技術を利用して開発を進めている<sup>177</sup>。
- ・ 音声区分復元：ヨルバ語のデジタルテキストに音声区分を復元させる。
- ・ 数字のヨルバ語テキストへの正規化：アラビア数字をヨルバ語の数の言語表現へどのように変換するかという問題に取り組み、そのプロセスのための計算モデルを開発した。計算モデルはオートマトンに基づく技術を利用し、数字をヨルバ語へ変化するためにデザインされた。このプロセスを達成するためのソフトウェアは、Pythonプログラミング言語<sup>178</sup> を使って実行された (<http://num2yor.appspot.com/>)。
- ・ マルチメディア多言語辞書プロジェクト (IFEDICO)

(△) アフリカ言語向けの機械翻訳サービスは、ナイジェリアで実用化されているか。

(○) 我々が開発した英語とヨルバ語の機械翻訳アプリケーションはすでに、誰にでも無料でオンライン上 (<http://eng2yor.appspot.com/translate>) で利用できる状態にある。

(△) 英語とヨルバ語の機械翻訳技術はモバイル向けにも開発されているか。

(○) アップル社のiOSとグーグル社のアンドロイドOS向けにも開発を開始したところである。

(△) 音声を利用する機械翻訳の研究を実施しているか。

(○) まだ実施していない。文字の翻訳技術だけである。

(△) あなた方はどの言語を研究対象にしているのか。

(○) デジタル辞書と機械翻訳の開発分野で、アフリカ言語に関しては、ヨルバ語、ハウサ語、イボ語を対象としており、欧州言語に関しては英語とフランス語である。機械

---

<sup>176</sup> <http://www.cstr.ed.ac.uk/projects/festival/>

<sup>177</sup> 参考論文：Akanbi, L. A. and Odejobi, O. A. (2011) Automatic recognition of oral vowels in tone language: experiments with fuzzy logic and artificial neural network models. Applied Soft Computing, Elsevier, vol. 11, no. 1, Pp 1467-1480, United Kingdom)

<sup>178</sup> <http://www.python.org/about/>  
<http://www.python.org/psf/about/>

翻訳に関しては、特に英語とヨルバ語の間での研究開発を進めている。

(△) ナイジェリア政府はあなた方の言語技術研究開発を支援しているか。

(○) その通り。我々の大学を通して研究開発を支援している。だが、我々の研究グループは政府から資金を支給されていない。

(△) あなた方は開発した技術を商用化しているか。

(○) 我々は研究機関なので、商用化はしていない。

(△) アフリカ内で言語技術に関する研究開発を実施している研究機関を提携させる試みはあるか。

(○) 否。例えば、アフリカ言語技術(AFLAT)、アフリカ言語技術イニシアティブ(ALTi)等の組織があり、アフリカ言語について研究開発を実施しているが、だが、アフリカ内で諸機関を提携させることは行っていない。なお、アフリカ言語技術イニシアティブ<sup>179</sup>は、マイクロソフト等の民間企業と提携し、言語技術に関わる様々な研究プロジェクトを実施したり、セミナーを開催したりしている。ハウサ語、ヨルバ語、イボ語を対象に研究を実施している。

#### 4) アフリカにおける言語技術の利用状況について

(△) アフリカにおける言語技術の利用状況について教えていただきたい。現在、ナイジェリアにはアフリカ言語向けの言語技術サービスやアプリケーションがあるか。

(○) テキストプロセッシング(スペルチェック)、携帯電話向け音声合成技術があり、特に携帯電話向けに事前に録音していたメッセージを再生するアプリケーションは普及している。

(△) 現在、アフリカではどのような言語技術の需要があるか。

(○) ビジネスに関しては、インターネットサービスをアフリカの現地の言葉で提供する必要がある。また、アフリカ言語で文章を書くためのワードプロセッサにも需要がある。医療に関しては、ヨルバの伝統医療において、意思決定のために言語技術の需要がある。農業に関しては、天気予報、市場の日、種まき季節、カレンダー等に関して、モバイル上で現地の言語により情報を提供する必要がある。

(△) 日本語とアフリカ諸言語の機械翻訳のニーズについては、どのようにお考えか。

(○) 我々の研究グループは、日本語とアフリカ諸言語を対象にした機械翻訳に関する研究を実施していない。日本語とナイジェリアで利用されている諸言語の間には音声表記法に関して、大きな違いがあり、幾つか難しい点もあると思うが、非常に関心の集ま

---

<sup>179</sup> <http://www.alt-i.org/index.php>

る研究になると我々と思う。もし日本の研究機関が我々のプロジェクトに興味を持つようであれば、我々としては大歓迎である。

(△) 将来的に、アフリカでは言語技術に基づくどのようなアプリケーションが必要とされるとお考えか。

(○) インターネットやウェブに基づくアプリケーションと携帯電話向けのアプリケーションを、アフリカ言語で利用できるようにする必要がある。

(△) ナイジェリアでは、3G 回線網は展開されているか。

(○) その通り。MTN、GLO、Etisalt、Zain 等の携帯電話通信事業者によって提供されている。

(△) 3G 回線の平均伝送速度はどのくらいか。

(○) 250-350Kbps であるが、接続が不安定な時がある。

(△) ナイジェリア以外では、どのアフリカ諸国において言語技術の研究開発が実施されているか。

(○) 南アフリカ共和国のマラケ研究院 (Meraka Institut)<sup>180</sup>で、言語技術の開発が盛んに実施されている。また、ケニアではスワヒリ語向け、エチオピアではアムハラ語向け、エジプト・チュニジア等の北アフリカ諸国ではアラビア語向けの言語技術の研究が実施されているそうである。

### **第三章 旧ソ連及びアフリカ諸国の言語処理に関する研究プロジェクト例**

本章では、旧ソ連及びアフリカ諸国における言語処理の研究プロジェクトの事例を記す。第一節では、アゼルバイジャンのディルマンクプロジェクト、第二節、第三節では FP7 の MEDER 及び VOICES プロジェクトについて記す。

#### **第一節 アゼルバイジャン・ディルマンクプロジェクト**

ディルマンクプロジェクト<sup>181</sup>は、アゼルバイジャンの包括的な言語技術研究プログラムであり、2005 年より政府が資金 (年間約 40 万ドル) を支給し (2014 年に終了予定)、機械翻訳、音声認識、音声合成、音声翻訳、電子辞書等の開発を実施している。人員は 15 名で、研究者は通信・情報技術省に雇用されている。

機械翻訳に関しては、2007 年にアゼルバイジャン語から英語に翻訳する機械翻訳システムを開発し、2008 年に、英語からアゼルバイジャン語に翻訳する機械翻訳システ

---

<sup>180</sup> <http://www.meraka.org.za/humanLanguage.htm>

<sup>181</sup> <http://www.dilmanc.az/en/node/8/>  
<http://www.dilmanc.az/en/project/about>

を開発している。また、2012年には統計的手法とルール（規則）ベース手法を組み合わせた機械翻訳システムを発表している。対応言語は、アゼルバイジャン語、英語、トルコ語、ロシア語である。音声認識技術に関しては、2007年に研究を開始し、現在音声認識システムを発表している。音声合成技術に関しては、2012年に音声合成システムを発表している。最終プロジェクトは、音声機械翻訳システム（英語とアゼルバイジャン語間・両方向）の予定である。

## 第二節 FP7 MEDER プロジェクト

FP7のMEDAR<sup>182</sup>は、アラビア語を対象に音声・言語技術の研究開発の分野で、欧州諸国（デンマーク、フランス、オランダ、英国）とアラビア語を使用する地中海諸国（ギリシア、レバノン、ヨルダン、パレスチナ自治区、エジプト、モロッコ）の研究組織を提携して、研究開発を実施することを目的としている。同プロジェクトは、同様の目的で、欧州委員会に助成されて、2003年から2005年に実施されていたNEMLARプロジェクトの後継プロジェクトである。

目標は、1) 技術発展の傾向、潜在的市場、提携可能性の予測に基づく提携活動ロードマップの作成、2) 機械翻訳向けのツールとリソースの開発サポート、3) BLARK (Basic language Resource Kit: 機械翻訳技術の研究に必要な最小限の言語リソースセット) のアップデート、4) アラビア語の人間言語技術分野の研究開発機関ネットワークを強化することである。同プロジェクトでは、特に英語とアラビア語間の機械翻訳システムの開発が実施されており、MOSESが利用されている。なお、MEDARが開発した機械翻訳パッケージは、ウェブサイトでダウンロード可能である<sup>183</sup>。

### MEDER プロジェクトの基本情報

略称	MEDER
正式名称	アラビア語を使用している地中海諸国との言語・音声技術分野における連携支援
公募分野	ICT-2007.9.1 国際連携
研究期間	2008年2月-2010年7月（30ヶ月）
全予算（EU 拠出金）	79万ユーロ（79万ユーロ：EUからほぼ全額の拠出）
プロジェクトコーディネーター	コペンハーゲン大学（デンマーク）

<sup>182</sup> [http://cordis.europa.eu/projects/rcn/85530\\_en.html](http://cordis.europa.eu/projects/rcn/85530_en.html)

<sup>183</sup> <http://www.medar.info/index.php>

参加組織	アテネ研究・技術革新センター（ギリシア）、原子エネルギー庁（仏）、フランス国立科学研究センター、ユトレヒト大学（蘭）、オープン・ユニバーシティ（英）、バラマンド大学（レバノン）、アマン大学（ヨルダン）、リヨン第二大学（仏）、ビルゼイト大学（パレスチナ自治区）、評価・言語資源配分庁（仏）、IBM ワールドトレードコーポレーション（エジプト）、SAKHR ソフトウェア（エジプト）、デジタルシステム開発工学カンパニー（エジプト）、モハメッド5世大学（モロッコ）、CNRS（仏）
ウェブサイト	<a href="http://www.medar.info/">http://www.medar.info/</a>

### 第三節 FP7 VOICES プロジェクト

FP7 の VOICES プロジェクトは、アフリカ諸国の都市部外も含めた社会・経済発展のために、モバイル ICT サービスの普及を促進することを目的とする。特に、コンテンツの強化と言語等の情報へのアクセスの障害を取り除くことが目指される。言語技術の開発に関しては、同プロジェクトでは、言語リソースの少ないアフリカ言語向けに音声合成及び音声認識の開発を進め、現地の言語で音声ベースの携帯電話サービスを利用可能にすることが狙われている。

問題は、アフリカの都市部外には、ラジオ放送網と第二世代携帯電話網は普及しているものの、インターネット接続が不可能な地域があり、またインターネットに接続できるとしても非常に通信料金が高価な場合があり、識字率も低いことがあることである。このため、VOICES プロジェクトでは、ラジオ、携帯電話、音声技術を結合し、様々な ICT サービスを開発し、実証実験を行っている。

ラジオ・マルシェ、タブレ、フォロバ・ビオンという3つのパイロットプロジェクトがマリで実施されている<sup>184</sup>。ラジオ・マルシェは農業事業者を支援するプロジェクトで、事業者が生産品の紹介、つまり広告をラジオを通して行うことを可能にするシステムを開発する。農業事業者は生産品の情報（価格等）を携帯電話で情報管理者に伝え、その情報を管理者がテキストとして専用のシステムに保存する。そして、そのテキスト情報は音声合成技術を使って音声出力されて、ラジオで伝播される。タブレは、あるグルー

<sup>184</sup> <http://mvoices.eu/node/80>

プがグループのメンバーにメッセージを伝達したい場合に利用される。メッセージ発信者がメッセージを専用のインターフェイスで録音すると、登録されたメンバーの携帯電話へとそのメッセージが一斉に伝達される。フォロバ・ビオンは市民報道向けのシステムであり、ある出来事の情報とその出来事に遭遇した人が携帯電話でその情報を話し、それが録音されて、一種の音声ブログが作られる。そこにラジオ局がアクセスし、ラジオでその情報が伝播される。

以上のように、VOICES プロジェクトでは最先端の ICT 技術の研究開発というよりはむしろ、旧来のラジオ、第二世代携帯電話、音声技術を組み合わせて、アフリカ諸国の社会・経済の発展を目指している。

### VOICES プロジェクトの基本情報

略称	VOICES
正式名称	社会発展のための音声に基づく共同体中心のモバイルサービス
公募分野	ICT-2009.9.1 国際連携
研究期間	2011年1月-2013年6月（30ヶ月）
全予算（EU 拠出金）	296万ユーロ（204万ユーロ）
プロジェクトコーディネーター	ERCIM（仏）
参加組織	CRS4（伊）、ポルトガルテレコム（ポルトガル）、TNO（蘭）、フランステレコム（仏）、VUA（蘭）、科学産業研究評議会（南アフリカ）、サヘル ECO（マリ）、電気通信他国家学校（セネガル）、ノースウェスト大学（南アフリカ）、ワールド・ワイド・ウェブ基金（スイス）、メイリュ基金（仏）
ウェブサイト	<a href="http://mvoices.eu/">http://mvoices.eu/</a>

## 第四章 旧ソ連及びアフリカ諸国において需要のある言語技術の分野及び日本語との翻訳需要の可能性のある分野

本章では、旧ソ連及びアフリカ諸国において需要がある言語技術の分野及び日本語との翻訳需要の可能性のある分野について記す。

まず、旧ソ連及びアフリカ諸国において需要のある言語技術の分野に関しては、外国

語学習や観光の分野は、どの国でも需要があるが、それ以外の分野では国により政治・社会・経済・言語使用状況が異なるので、需要分野は異なる。

例えば、旧ソ連諸国であるラトビア、エストニア、リトアニアの言語の使用者は少ないが、EU加盟国であるので、EU機関での翻訳が必要となり、また、FP7等の欧州プロジェクトに積極的に参加している機関もある（ラトビアのティルド社）。

アゼルバイジャンでは、国の人口が少ないため、他国の情報収集することに翻訳の需要があり、機械翻訳技術の研究開発を促進している側面がある。アゼルバイジャン語の報道サイトよりもトルコ語とロシア語の報道サイトの方が情報量が多く、また同国では欧米系の報道サイトを見るのが少ないので、同国のディルマンクプロジェクトでは、英語だけでなく、トルコ語とロシア語の言語（仏語や独語よりも）に対応する機械翻訳システムを開発している。また、同国は旧ソ連諸国で有数の石油産出国であるので、ビジネス分野でも機械翻訳は必要となる。

ナイジェリアもアゼルバイジャンのように石油産出国であるが、同国では英語が公用語であり、ビジネス分野では英語が使用されている。だが、英語を話せないナイジェリア人は人口の7割を占めるので、ビジネス分野でも業種によっては機械翻訳システムが必要になる可能性がある。ナイジェリアのオバフェミ・アウオロウオ大学では、英語をヨルバ語に翻訳するオンライン無料機械翻訳サービスを提供している。

南アフリカ共和国の公用語は11言語あり、Mareke 研究院は言語の多様性に対応するために、政府機関等に対して、音声技術等の言語技術を開発している。

アフリカ言語一般に言えることであるが、インターネットサービスや携帯電話向けのサービスをアフリカ言語で提供するため、ローカライゼーションを行う必要がある。このためには、機械翻訳システム以前の段階で、アフリカ言語向けのフォントやキーボード、ワープロソフトを開発する必要がある。ANLoc<sup>185</sup>というアフリカ言語のローカライゼーションを促進するプログラムもあり、カナダのIDRCが助成している。また、ローカライゼーション以外では、FP7のVOICESプロジェクトにおいては、インターネット接続環境が悪く、識字率が低い地域で、ラジオ、携帯電話、音声技術という既存の技術を組み合わせて、現地の言語で情報伝達を行う活動が実施されている。アフリカでは、言語技術は、ビジネス分野だけでなく、農業分野でも必要とされており、現地の言語で、天気予報、市場の日、種まき季節、カレンダー等に関して、現地の言語により情報を提供する必要がある。

ついで、旧ソ連及びアフリカ諸国の言語と日本語との翻訳需要に関してであるが、ヒ

---

<sup>185</sup> <http://www.africanlocalisation.net/fr/node/42>

アリング調査を行ったディルマンクプロジェクトのアゼルバイジャンの研究者によれば、需要は確かにあり、教育、研究、科学、観光、ビジネス等、様々な分野で両言語の機械翻訳の需要はある。また、アゼルバイジャンでは、日本文化を学ぶ人も多いそうである。確かに、マンガやアニメを始めとする世界各国での日本のポップカルチャーへの関心は高まりつつあり、同分野における翻訳の需要も増加していくと考えられる。同じくヒアリング調査を行ったナイジェリアのオバフェミ・アウォロウォ大学の研究者の話では、日本企業がナイジェリアでビジネスをするには、英語が公用語であるにしても、英語を話せない人々が人口の7割を占めるので、このような人々と交流するには機械翻訳技術が必要となる。

## 第五章 旧ソ連及びアフリカ諸国における無線インフラの状況

機械翻訳や音声技術等には、インターネットに接続して機能するネットワーク型のシステムがあり、これらの技術を使用するためには、必然的にインターネット接続が必要になる上に、アプリケーションを快適に使用するためには、LTE等の4G網の展開が不可欠である。以下に、旧ソ連諸国及びアフリカ諸国の無線インフラの展開状況について記す。アフリカに関しては国数が多いので、特に1000万人以上の人口の国についてのみ記した。参考資料としては、国際電気通信連合（ITU）が毎年発表している「World Telecommunication / ICT Indicators database 2013」（2013年6月17日発表）と、LTE網に関しては、GSA（Global Mobile Suppliers Association）のLTEレポート（2013年5月発表）を利用したが、国によってデータが不明な場合もある<sup>186</sup>。

### 旧ソ連諸国

まず、旧ソ連諸国では、全ての国で、携帯電話人口カバー率（2Gや3G等含める）は極めて高い。ついで、3G網以上でカバーされている人口の割合については、国によって様々である（ここで3G網以上とは、UMTS、HSPA、CDMA2000、EV-DO、モバイルWiMAX、LTEを含める）。

現在世界各国で整備が進むLTE網に関しては、旧ソ連諸国でもすでに商用化向けに回線網が展開されている。例えば、ウズベキスタン（2010年8月から）、エストニア（2010年12月から：2014年末までに全人口をカバーすることが見込まれている）<sup>187</sup>、リトアニア（2011年4月から）、ラトビア（2011年5月から）、キルギスタン（2011年12月

<sup>186</sup> <http://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx>

[http://www.gsacom.com/gsm\\_3g/info\\_papers.php4](http://www.gsacom.com/gsm_3g/info_papers.php4)

<sup>187</sup> <http://www.cellular-news.com/story/58976.php>

から)、アルメニア (2011 年 12 月から)、ロシア (2012 年 1 月から)、モルドバ (2012 年 4 月から)、アゼルバイジャン (2012 年 6 月)、タジキスタン (2012 年 10 月)、カザフスタン (2012 年 12 月) で、LTE網が商用向けに展開されており、他の国々でも実証実験が進められている (グルジア、ウクライナ)。

**旧ソ連諸国の無線インフラの状況 (人口数・携帯電話網人口カバー率・3G 網以上でカバーされている人口率)**

国名	人口数 (2012 年) : 単位 100 万人)	携帯電話網人口カバー率	3G 網以上でカバーされている人口の割合
ロシア	142	95% (2006 年)	不明
アゼルバイジャン共和国	9.4	100% (2011 年)	89.3% (2011 年)
アルメニア共和国	3.1	98.9% (2011 年)	97.4% (2011 年)
ウクライナ	44.9	99.9% (2011 年)	22% (2012 年)
ウズベキスタン	28	93% (2008 年)	不明
エストニア共和国	1.3	99% (2011 年)	97% (2011 年)
カザフスタン	16.3	95% (2010 年)	不明
ベラルーシ共和国	9.52	99.7% (2011 年)	90% (2011 年)
キルギス共和国	5.4	97% (2011 年)	45% (2012 年)
トルクメニスタン	5.1	不明	不明
タジキスタン共和国	7	不明	不明
グルジア共和国	4.3	99.1% (2011 年)	不明
モルドバ共和国	3.5	97.9% (2007 年)	61.9% (2009 年)
ラトビア共和国	2.2	98.8% (2007 年)	不明

リトアニア共和国	3.2	100% (2010年)	84% (2010年)

### アフリカ諸国

アフリカ諸国では、全ての国で、携帯電話人口カバー率は比較的高いが、50%か70%の国も多い。ついで、3G 網以上でカバーされている人口の割合については、国によって様々である（ここで3G 網以上とは、UMTS、HSPA、CDMA2000、EV-DO、モバイルWiMAX、LTE を含める）。LTE 網に関しては、アンゴラで2012年4月にアフリカ初で商用化向けにLTE 網が展開されている。その他、ナミビア（2012年5月）、タンザニア（2012年5月）、モーリシャス（2012年6月）、南アフリカ共和国（2012年10月）、ウガンダ（2012年10月）で展開されている。その他の国でも実証実験が行われているか、LTE 網の展開計画が発表されている（ボツワナ、赤道ギニア、エジプト、ガーナ、エチオピア、ケニア、リビア、ルワンダ、ナイジェリア、ジンバブエ、チュニジア）。

#### アフリカ諸国の無線インフラの状況（人口1000万人以上の国のみ）

国名	人口数（2012年）：単位100万人	携帯電話網人口カバー率	3G 網以上でカバーされている人口率
アルジェリア民主人民共和国	36.4	81.5% (2007年)	不明
アンゴラ共和国	20.1	40% (2006年)	不明
ウガンダ共和国	35.6	100% (2011年)	42% (2011年)
エジプト・アラブ共和国	83.9	99.70% (2011年)	91% (2011年)
エチオピア連邦民主共和国	86.5	73% (2012年)	40% (2011年)
ガーナ共和国	25.5	84% (2011年)	38% (2011年)
カメルーン共和国	20	58% (2006年)	不明

ギニア共和国	10.1	80% (2008年)	不明
ケニア共和国	42.7	89% (2011年)	不明
コートジボアール共和国	20	92% (2011年)	不明
コンゴ民主共和国	69	50% (2011年)	不明
ザンビア共和国	13	62% (2011年)	40% (2011年)
ジンバブエ共和国	13	72% (2011年)	15% (2011年)
スーダン共和国	45.7	86% (2011年)	42% (2011年)
セネガル共和国	13.1	90% (2011年)	不明
タンザニア連合共和国	47.6	85% (2011年)	不明
チャド共和国	11.8	75% (2011年)	不明
ナイジェリア連邦共和国	166.6	90% (2011年)	不明
ニジェール共和国	16.6	75% (2011年)	不明
ブルキナファソ	17.4	61% (2006年)	不明
マダガスカル共和国	21.9	23% (2006年)	不明
マラウイ共和国	15.8	85% (2011年)	不明
マリ共和国	16.3	20% (2006年)	不明
南アフリカ共和国	50	99% (2007年)	不明
モザンビーク共和国	24.4	不明	不明
モロッコ王国	32	99% (2011年)	75% (2011年)

## 第二部のまとめ

第二部では、旧ソ連とアフリカ諸国における言語技術の実用化動向、研究開発動向、

言語技術の需要及び無線インフラ状況を概観した。

実用化動向については、多くの企業によって、旧ソ連及びアフリカ諸国の言語のオンライン機械翻訳サービスが提供され、アプリケーションが開発されているが、その多くはグーグル社を筆頭に米企業であるが、欧州企業も少なからずある（英 SDL、ロシア・Yandex、ウクライナ・Online.ua、英 ATA、英 CAPITA、仏 Systran、仏ヴォカピアリサーチ、独 Linguattec、オーストリア・Sonico、ハンガリー・Morphologic、ラトビア・ティルド、ウクライナ・トリデント）。対応言語に関しては、アラビア語に対応するサービスやアプリケーションが圧倒的に多いが、EU の CIP プロジェクトで開発したサービス等（例：iTranslate4）もあり、旧ソ連諸国であり、かつ EU 加盟国の言語、すなわち、ラトビア語、エストニア語、リトアニア語や、人口が多いウクライナ語に対応するサービスやアプリケーションも多くある。

研究開発機関の動向に関しては、旧ソ連諸国では、ウクライナのトリデント社とラトビアのティルド社が欧州プロジェクトに参加しており、これらの国の言語に対応する機械翻訳技術の研究開発及び商用化を積極的に実施している。特に、ティルド社は多くの FP7 及び CIP のプロジェクトに参加しており、特にバルト三国の言語に関しては機械翻訳技術の分野で優れた研究機関の一つである。研究プロジェクトの動向に関しては、アゼルバイジャン政府が助成しているディルマンクプロジェクトが注目される。これは言語技術の包括的な研究プログラムであり、テキスト機械翻訳だけでなく、音声認識や音声翻訳システムの開発も実施している。アフリカ諸国では、南アフリカ共和国の Mareke 研究院が積極的に音声技術の開発を実施している。同国の公用語は 11 言語あり、政府機関向け等に音声技術の需要がある。また、ナイジェリアのオバフェミ・アウオロウオ大学では、現地の言語の 1 つであるヨルバ語を中心とする言語技術の研究開発を包括的に実施している。

言語技術と翻訳の需要に関しては、どの国でも外国語学習（発音練習向け音声合成システム等）や観光の分野は一般的に需要があるが、国の政治や経済、言語利用状況によって需要分野は異なる。アゼルバイジャン等の小規模の国では、他国の情報収集が機械翻訳の研究開発の主な動機の一つである。また、旧ソ連諸国では、英語以外の欧米言語（仏語や独語）よりも、トルコ語やロシア語の方が需要がある場合がある（特に、アゼルバイジャン）。また、アフリカ諸国には非常に多くの言語があり、機械翻訳システムの前に、インターネットや携帯電話サービスをアフリカ言語で使用できるように、フォントやキーボード、ワープロソフト、音声合成技術の開発が必要とされる場合がある。農業国では、農業に関する情報を現地の言語で伝達する必要がある。

日本語との翻訳需要については、観光や外国語学習の他、アゼルバイジャン等の石油

産出国では、ビジネス分野で機械翻訳の需要がある。また、日本のポップカルチャーの世界的な人気にも注目すれば、文化面で翻訳の需要の可能性もある。なお、ヒアリング調査を行ったナイジェリアのオバフェミ・アウォロウォ大学の研究員は、日本語へ機械翻訳システムを対応させる開発はまだ行っていないものの、日本の研究機関との提携開発には意欲的であった。

無線インフラ状況に関しては、旧ソ連及びアフリカ諸国の多くで携帯電話網が高い割合で人口をカバーしているが、3G 網に関しては国によって異なる。また、LTE 網の展開は、同地域でもすでに多くの国で開始されている（ウズベキスタン、エストニア、リトアニア、ラトビア、キルギスタン、アルメニア、ロシア、モルドバ、アゼルバイジャン、タジキスタン、カザフスタン、アンゴラ、ナミビア、タンザンニア、モーリシャス、南アフリカ共和国、ウガンダ）。

## 補足資料

### FP7 及び CIP における言語技術の研究プロジェクトのリスト

#### 2007-2008 年度作業プログラム (WP2007-2008)

略称	EuroMatrixPlus
正式名称	欧州言語向け機械翻訳をユーザーへ
公募分野	ICT-2007.2.2
全予算 (EU 拠出金)	594 万ユーロ (426 万ユーロ)
研究期間	2009 年 3 月-2012 年 2 月 (36 ヶ月)
ウェブサイト	<a href="http://www.euromatrixplus.net/">http://www.euromatrixplus.net/</a>
コーディネーター	ドイツ人工知能研究センター

略称	MEDER
正式名称	アラビア語を使用している地中海諸国との言語・音声技術分野における連携支援
公募分野	ICT-2007.9.1 国際連携
全予算 (EU 拠出金)	79 万ユーロ (79 万ユーロ : EU からほぼ全額の拠出)
研究期間	2008 年 2 月-2010 年 7 月 (30 ヶ月)
ウェブサイト	<a href="http://www.medar.info/">http://www.medar.info/</a>
コーディネーター	コペンハーゲン大学 (デンマーク)

略称	EMIME
プロジェクト名	モバイル環境における効果的な多言語インタラクション
プログラム分野	ICT-2007.2.1
全予算 (EU 拠出金)	432 万ユーロ (305 万ユーロ)
研究期間	2008/3 - 2011/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.emime.org/contact">http://www.emime.org/contact</a>
コーディネーター	エジンバラ大学 (英)

略称	CLASSIC
プロジェクト名	スピーチ会話のための適応システムにおける計算学習

プログラム分野	ICT-2007.2.1
全予算 (EU 拠出金)	476 万ユーロ (340 万ユーロ)
研究期間	2008/3 - 2011/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.classic-project.org/">http://www.classic-project.org/</a>
コーディネーター	ヘリオット・ワット大学 (英)

### WP2009-2010

略称	ACCURAT
プロジェクト名	機械翻訳のリソースが少ない分野向け比較可能コーポラの分析と評価
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	355 万ユーロ (282 万ユーロ)
研究期間	2010/1 - 2012/6 (30 ヶ月間)
ウェブサイト	<a href="http://www.panacea-lr.eu/">http://www.panacea-lr.eu/</a>
コーディネーター	ティルド SIA (ラトビア)

略称	MONNET
プロジェクト名	ネットワーク化された知識のための多言語オントロジー
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	320 万ユーロ (236 万ユーロ)
研究期間	2010/3 - 2013/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.monnet-project.eu/Monnet/Monnet/English?init=true">http://www.monnet-project.eu/Monnet/Monnet/English?init=true</a>
コーディネーター	アイルランド国立大学 (アイルランド)

略称	COSYNE
プロジェクト名	Wiki のための多言語コンテンツシンクロニゼーション
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	312 万ユーロ (235 万ユーロ)
研究期間	2010/3 - 2013/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.cosyne.eu/index.php/Main_Page">http://www.cosyne.eu/index.php/Main_Page</a>
コーディネーター	アムステルダム大学 (蘭)

略称	FAUST
プロジェクト名	ユーザーに適應する統計翻訳のためのフィードバック分析
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	376 万ユーロ (285 万ユーロ)
研究期間	2010/2 – 2013/1 (36 ヶ月間)
ウェブサイト	<a href="http://www.faust-fp7.eu/faust/">http://www.faust-fp7.eu/faust/</a>
コーディネーター	ケンブリッジ大学 (英)

略称	TTC
プロジェクト名	専門用語抽出、翻訳ツール、比較可能コーポラ
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	266 万ユーロ (202 万ユーロ)
研究期間	2010/1 – 2012/12 (36 ヶ月間)
ウェブサイト	<a href="http://www.ttc-project.eu/">http://www.ttc-project.eu/</a>
コーディネーター	ナント大学 (仏)

略称	PANACEA
プロジェクト名	人間言語技術のためのコスト削減言語リソース獲得及び自動的、標準化された注釈向けプラットフォーム
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	399 万ユーロ (268 万ユーロ)
研究期間	2010/1 – 2012/12 (36 ヶ月間)
ウェブサイト	<a href="http://www.panacea-lr.eu/">http://www.panacea-lr.eu/</a>
コーディネーター	ポンペイ・ファブラ大学 (スペイン)

略称	MOLTO
プロジェクト名	多言語オンライン翻訳
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	378 万ユーロ (297 万ユーロ)
研究期間	2010/3– 2013/5 (39 ヶ月間)
ウェブサイト	<a href="http://www.molto-project.eu/">http://www.molto-project.eu/</a>

コーディネーター	ヨーテボリ大学 (スウェーデン)
----------	------------------

略称	PRESEMT
プロジェクト名	パターン認識に基づく統計的に強化された MT
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	327 万ユーロ (250 万ユーロ)
研究期間	2010/1- 2012/12 (36 ヶ月間)
ウェブサイト	<a href="http://www.presemt.eu/">http://www.presemt.eu/</a>
コーディネーター	アテネ情報通信・知識技術の研究・技術革新センター (ギリシア)

略称	T4ME NET
プロジェクト名	多言語欧州情報社会のための技術
プログラム分野	ICT-2009.2.2
全予算 (EU 拠出金)	761 万ユーロ (599 万ユーロ)
研究期間	2010/2- 2013/1 (36 ヶ月間)
ウェブサイト	<a href="http://www.meta-net.eu/">http://www.meta-net.eu/</a>
コーディネーター	ドイツ人工知能センター

#### WP20011-2012

略称	MATECAT
正式名称	機械翻訳を利用するコンピューター支援翻訳の精度の向上
公募分野	ICT-2011.4.2
研究期間	2012 年 2 月-2015 年 1 月 (36 ヶ月)
全予算 (EU 拠出金)	336 万ユーロ (265 万ユーロ)
ウェブサイト	<a href="http://www.mateAat.com/">http://www.mateAat.com/</a>
コーディネーター	ブルーノ・ケスラー基金 (伊)

略称	CASMACAT
正式名称	先端コンピューター補助翻訳のための認知分析と統計手法
公募分野	ICT-2011.4.2
研究期間	2011 年 11 月-2014 年 10 月 (36 ヶ月)

全予算 (EU 拠出金)	321 万ユーロ (250 万ユーロ)
ウェブサイト	<a href="http://www.casmacat.eu/">http://www.casmacat.eu/</a>
コーディネーター	エジンバラ大学 (英)

略称	TAAS
プロジェクト名	サービスとしてのターミノロジー
プログラム分野	ICT-2011.4.1
全予算 (EU 拠出金)	229 万ユーロ (182 万ユーロ)
研究期間	2012/6-2014/5 (24 ヶ月間)
ウェブサイト	<a href="http://www.taas-project.eu/">http://www.taas-project.eu/</a>
コーディネーター	ティルド (ラトビア)

略称	X-LIKE
プロジェクト名	言語横断的知識抽出
プログラム分野	ICT-2011.4.2
全予算 (EU 拠出金)	474 万ユーロ (355 万ユーロ)
研究期間	2012/1-2014/12 (24 ヶ月間)
ウェブサイト	<a href="http://www.xlike.org/">http://www.xlike.org/</a>
コーディネーター	ヨゼフ・ステファン研究院 (スロベニア)

略称	EuBridge
正式名称	異なる言語の間にかける橋
公募分野	ICT-2011.4.2
研究期間	2012 年 2 月-2015 年 1 月 (36 ヶ月)
全予算 (EU 拠出金)	1051 万ユーロ (787 万ユーロ)
ウェブサイト	<a href="http://www.eu-bridge.eu/">http://www.eu-bridge.eu/</a>
コーディネーター	カールスルーエ技術研究院 (独)

略称	MOSESCORE
正式名称	拡大と開拓のための MOSES オープンソース評価と提携支援
公募分野	ICT-2011.4.2
研究期間	2012 年 2 月-2015 年 1 月 (36 ヶ月)

全予算 (EU 拠出金)	1250 万ユーロ (1200 万ユーロ)
ウェブサイト	<a href="http://www.statmt.org/mosescore/">http://www.statmt.org/mosescore/</a>
コーディネーター	エジンバラ大学 (英)

略称	SMESPIRE
プロジェクト名	デジタルコンテンツ及び言語環境に築かれた中小企業の欧州共同体
プログラム分野	ICT-2011.4.1
全予算 (EU 拠出金)	197 万ユーロ (179 万ユーロ)
研究期間	2012/5- 2014/4 (24 ヶ月間)
ウェブサイト	<a href="http://www.smespire.eu/index.php/project-overview">http://www.smespire.eu/index.php/project-overview</a>
コーディネーター	EPSILON (伊)

略称	QTLAUNCHPAD
プロジェクト名	精度の高い翻訳技術のための大規模アクションの準備と実施
プログラム分野	ICT-2011.4.1
全予算 (EU 拠出金)	239 万ユーロ (191 万ユーロ)
研究期間	2012/7- 2014/6 (24 ヶ月間)
ウェブサイト	<a href="http://www.qt21.eu/launchpad/content/new-goal-quality-translation">http://www.qt21.eu/launchpad/content/new-goal-quality-translation</a>
コーディネーター	ドイツ人工知能センター

略称	PORTDIAL
プロジェクト名	ポータブル多言語音声対話システムのための言語リソース
プログラム分野	ICT-2011.4.1
全予算 (EU 拠出金)	246 万ユーロ (187 万ユーロ)
研究期間	2012/7- 2014/6 (24 ヶ月間)
ウェブサイト	<a href="https://sites.google.com/site/portdial2/">https://sites.google.com/site/portdial2/</a>
コーディネーター	電気通信システム研究院 (ギリシア)

略称	DIRHA
プロジェクト名	ロバストアプリケーションのための遠距離スピーチインタラクション

プログラム分野	ICT-2011.4.2
全予算 (EU 拠出金)	484 万ユーロ (345 万ユーロ)
研究期間	2012/1- 2014/12 (24 ヶ月間)
ウェブサイト	<a href="http://dirha.fbk.eu/">http://dirha.fbk.eu/</a>
コーディネーター	ブルーノ・ケスラー基金 (伊)

略称	SIMPLE4ALL
プロジェクト名	適応学習を通して改善するスピーチ合成
プログラム分野	ICT-2011.4.2
全予算 (EU 拠出金)	399 万ユーロ (310 万ユーロ)
研究期間	2011/11- 2014/10 (36 ヶ月間)
ウェブサイト	<a href="http://simple4all.org/">http://simple4all.org/</a>
コーディネーター	エジンバラ大学 (英)

略称	TRANSLECTURES
プロジェクト名	ビデオ講義の転写と翻訳
プログラム分野	ICT-2011.4.2
全予算 (EU 拠出金)	499 万ユーロ (312 万ユーロ)
研究期間	2011/11- 2014/10 (36 ヶ月間)
ウェブサイト	<a href="http://www.translectures.eu/">http://www.translectures.eu/</a>
コーディネーター	バレンシア工科大学 (スペイン)

略称	LTCOMPASS
プロジェクト名	言語技術を研究から市場へ
プログラム分野	ICT-2011.4.2
全予算 (EU 拠出金)	327 万ユーロ (295 万ユーロ)
研究期間	2011/11- 2014/2 (39 ヶ月間)
ウェブサイト	<a href="http://cordis.europa.eu/projects/rcn/100698_en.html">http://cordis.europa.eu/projects/rcn/100698_en.html</a>
コーディネーター	INMARK ESTUDIOS (スペイン)

#### **CIP ICT 2009 年作業プログラム**

略称	FLAVIOUS
----	----------

プロジェクト名	インターネットとユーザー作製サイトの外国語バージョン
プログラム分野	CIP-ICT-PSP.2009.5.3
全予算 (EU 拠出金)	370 万ユーロ (185 万ユーロ)
研究期間	2010/4– 2012/9 (30 ヶ月間)
ウェブサイト	<a href="http://www.project-flavius.eu/fr/en/">http://www.project-flavius.eu/fr/en/</a>
コーディネーター	ソフティッシモ (仏)

略称	GALATEAS
プロジェクト名	自動翻訳と検索のエピソード分析のためのログ一般分析
プログラム分野	CIP-ICT-PSP.2009.5.1
全予算 (EU 拠出金)	370 万ユーロ (185 万ユーロ)
研究期間	2010/4– 2013/3 (30 ヶ月間)
ウェブサイト	<a href="http://www.galateas.eu/">http://www.galateas.eu/</a>
コーディネーター	XEROX (仏)

略称	iTranslate4
プロジェクト名	全欧州言語向けインターネット翻訳
プログラム分野	CIP-ICT-PSP.2009.5.1
全予算 (EU 拠出金)	394 万ユーロ (197 万ユーロ)
研究期間	2010/3– 2012/2 (24 ヶ月間)
ウェブサイト	<a href="http://itranslate4.eu/en/">http://itranslate4.eu/en/</a>
コーディネーター	ハンガリー科学アカデミー言語学研究所

略称	MultilingualWeb
プロジェクト名	多言語ウェブ向け標準と実践
プログラム分野	CIP-ICT-PSP.2009.5.2
全予算 (EU 拠出金)	41 万ユーロ (41 万ユーロ)
研究期間	2010/4– 2012/3 (24 ヶ月間)
ウェブサイト	<a href="http://www.multilingualweb.eu/">http://www.multilingualweb.eu/</a>
コーディネーター	GEIE (仏)

略称	Pluto
----	-------

プロジェクト名	特許言語翻訳オンライン
プログラム分野	CIP-ICT-PSP.2009.5.1
全予算 (EU 拠出金)	436 万ユーロ (218 万ユーロ)
研究期間	2010/4– 2013/3 (36 ヶ月間)
ウェブサイト	<a href="http://www.pluto-patenttranslation.eu/">http://www.pluto-patenttranslation.eu/</a>
コーディネーター	ダブリンシティ大学 (アイルランド)

略称	EASTIN-CL
プロジェクト名	生活補助サポート向けポータル内の言語横断・多様検索
プログラム分野	CIP-ICT-PSP.2009.5.1
全予算 (EU 拠出金)	215 万ユーロ (108 万ユーロ)
研究期間	2010/3– 2012/5 (26 ヶ月間)
ウェブサイト	<a href="http://www.eastin-cl.eu/">http://www.eastin-cl.eu/</a>
コーディネーター	LINGUATEC (独)

略称	ATLAS
プロジェクト名	言語に支援された CMS 向け応用技術
プログラム分野	CIP-ICT-PSP.2009.5.3
全予算 (EU 拠出金)	332 万ユーロ (166 万ユーロ)
研究期間	2010/3– 2013/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.atlasproject.eu/">http://www.atlasproject.eu/</a>
コーディネーター	テトラコムインタラクティブソリューション (ブルガリア)

### **CIP ICT 2010 年作業プログラム**

略称	CESAR
プロジェクト名	中央・南東欧州リソース
プログラム分野	CIP-ICT-PSP.2010.6.1
全予算 (EU 拠出金)	416 万ユーロ (208 万ユーロ)
研究期間	2011/2– 2013/1 (24 ヶ月間)
ウェブサイト	<a href="http://cesar.nytud.hu/">http://cesar.nytud.hu/</a>
コーディネーター	ハンガリー科学アカデミー言語学研究所

プロジェクト名称	Organic.Lingua
プロジェクト名	持続可能な農業・環境教育向け多言語ポータルの可能性の証明
プログラム分野	CIP-ICT-PSP.2010.6.2
全予算 (EU 拠出金)	350 万ユーロ (175 万ユーロ)
研究期間	2011/3- 2014/2 (36 ヶ月間)
ウェブサイト	<a href="http://www.multilingualweb.eu/">http://www.multilingualweb.eu/</a>
コーディネーター	アルカラ大学 (スペイン)

略称	PROMISLingua
プロジェクト名	欧州中小企業向けコンプライアンス支援のための操作可能・多言語インタラクティブサービス
プログラム分野	CIP-ICT-PSP.2010.6.2
全予算 (EU 拠出金)	528 万ユーロ (264 万ユーロ)
研究期間	2011/4- 2013/9 (30 ヶ月間)
ウェブサイト	<a href="http://www.promislingua.eu/en/">http://www.promislingua.eu/en/</a>
コーディネーター	INMARK ESTUDIOS (スペイン)

略称	SUMAT
プロジェクト名	機械翻訳による字幕付けのためのオンラインサービス
プログラム分野	CIP-ICT-PSP.2010.6.2
全予算 (EU 拠出金)	360 万ユーロ (180 万ユーロ)
研究期間	2011/4- 2014/3 (24 ヶ月間)
ウェブサイト	<a href="http://www.sumat-project.eu/">http://www.sumat-project.eu/</a>
コーディネーター	VICOMTECH (スペイン)