

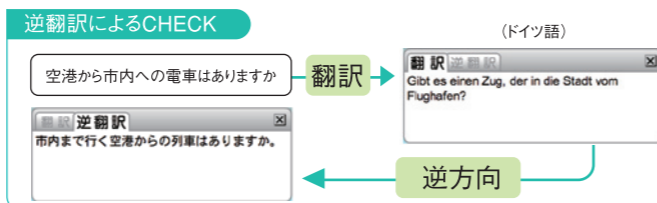
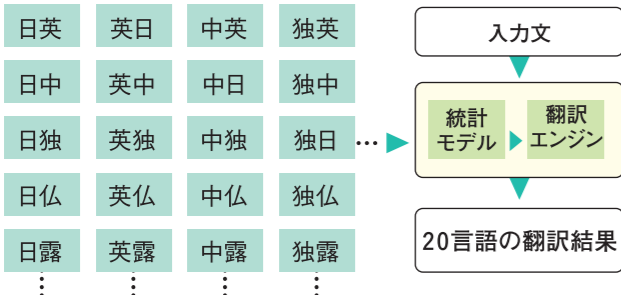
多言語翻訳

MASTARプロジェクト/言語翻訳グループでは、音声翻訳の翻訳部やテキスト翻訳に関して、基礎研究から実用化まで幅広く取り組んでいます。特に、世界の大多数の人口をカバーするという意味での多言語翻訳技術、及び、旅行、マニュアル、特許、新聞など様々な分野で高精度な翻訳を実現するための多分野翻訳技術を目指しています。

◎420 (21×20) 通りの組合せの翻訳を実現

構成

多言語対訳コーパス (21言語×20言語)



翻訳できる言語

| | | | |
|---------|----------|------------|--------|
| アラビア語 | タイ語 | 日本語 | ベトナム語 |
| イタリア語 | タガログ語 | ハンガリー語 | ポルトガル語 |
| インドネシア語 | 中国語(簡体字) | ヒンディ語 | マレー語 |
| 英語 | 中国語(繁体字) | ブラジルポルトガル語 | ロシア語 |
| オランダ語 | デンマーク語 | ガル語 | |
| スペイン語 | ドイツ語 | フランス語 | |

今後

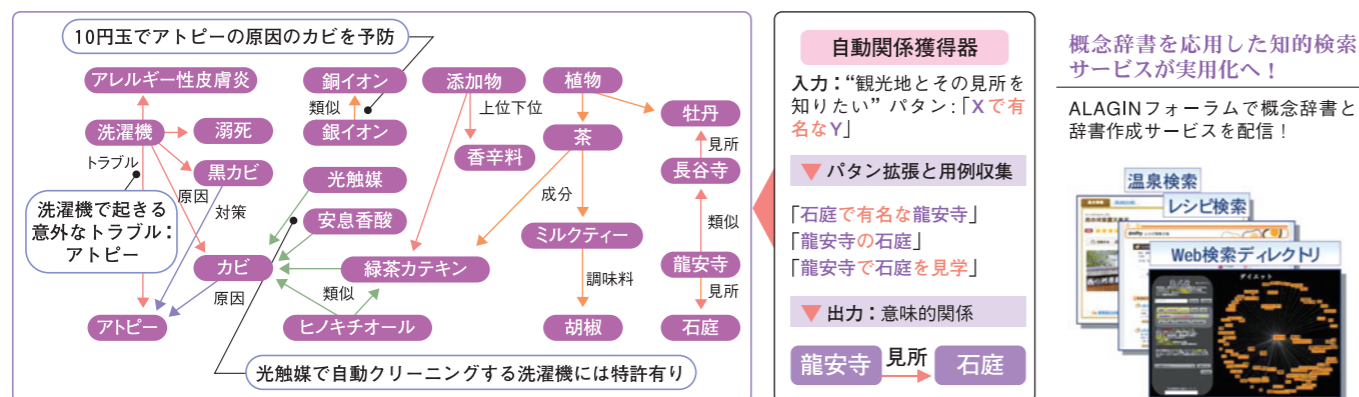
音声翻訳に適用してA-STAR(現在アジア8カ国が参加)による音声翻訳実験などで実証していきます。

概念辞書

言語に関する自動処理を行なおうとすれば、その第一の基礎はまず単語に関して様々な情報を記述した辞書ということになります。MASTARプロジェクトでは、様々な辞書を開発していますが、その内の一つ、概念辞書は、単語の間の様々な意味的な関係を含むもので、Webに自動知識獲得アルゴリズムを適用することによって作成されています。これは普通の辞書にはふくまれていないような単語間の生

きた意味的關係を含んでいます。こうした辞書、自動知識獲得アルゴリズムを用いることによって、意外でありながら有用な情報を発見する支援を行なう検索ディレクトリ、ユーザの意図に的確に答え、有用な情報をアドバイスするレシピ検索システムなど、実生活にフィットした様々な情報サービスが可能となります。

◎ユーザの要求にしたがって成長し続ける知の集積:概念辞書 多様な意味的關係をWebから自動獲得できる技術を開発



※普通の辞書にはない「生きた」関係。個人の知識を凌駕。現在、180万語/目標、250万語(広辞苑の10倍)



NICT 情報通信研究機構
National Institute of information and Communications Technology

新

ことばの壁を越えて



独立行政法人 情報通信研究機構
知識創成コミュニケーション研究センター
MASTAR プロジェクト
〒619-0289「けいはんな学研都市」京都府相楽郡精華町光台3丁目5番
TEL:0774-98-6300 FAX:0774-98-6955
HP:http://mstar.jp E-mail:info@mstar.jp

MASTAR
http://mstar.jp

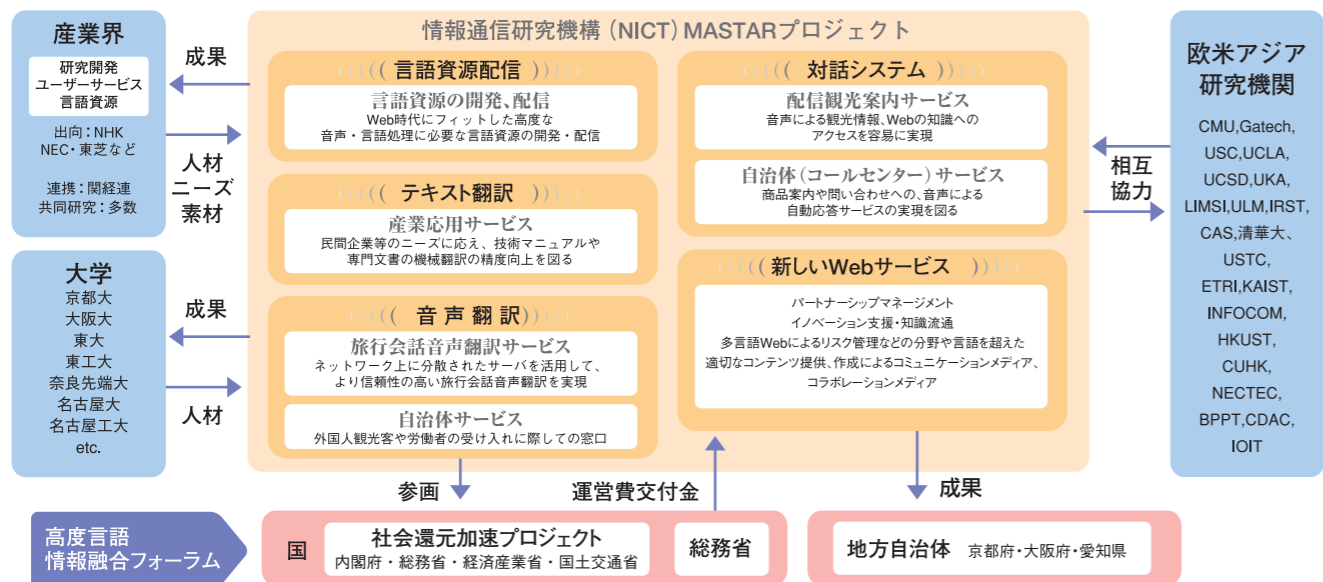
MASTARプロジェクトは、国内外の言語処理、機械翻訳、音声処理の研究者が集った音声・言語研究の世界的な研究拠点です。産官学が連携しながら横断的に多言語音声翻訳、多言語テキスト翻訳、多言語対話応答システム技術、多言語言語資源の研究開発を行います。これまで単一の機関では困難であった音声・言語資源の収集・蓄積を産学官連携で行い、これらを基に実用をにらんだ研究開発を進め、社会への還元を加速を目指します。

MASTARプロジェクトのねらい

1. NICTを音声・言語資源、技術に関する世界な拠点を構成する。
2. 産業界、社会とリンクした形で持続的に、音声・言語資源を蓄積、成長していく仕組みを作る。
3. 具体的には、
 - ① 総合科学技術会議の社会還元加速プロジェクトの一つに選定されたネットワーク音声翻訳に関する技術開発、社会実験、社会還元を行う。
 - ② 産業界と協力してマニュアルなどを対象としたWEB2.0型の機械翻訳サービスを行い、共通辞書、共通コーパスの蓄積、翻訳技術研究のポジティブ成長サイクルを確立する。
 - ③ ユニバーサルコミュニケーションの一環として、あらゆる利用者へ情報をお届けするための音声対話インタフェース技術の開発、社会実験、社会還元を行う。
 - ④ 世界的言語資源を構築し、社会に還元する。
4. 産官学が共同して研究を進めるオープン型研究開発機構を設置し、これにより、日本の情報処理を支える人材育成を行う。



MASTARプロジェクトの構成



高度言語情報融合フォーラム

Advanced Language Information Forum: ALAGIN

依然として、コミュニケーションには3つの壁がある

- 言語の壁：外国語で話されたり、書かれた情報は理解不能！機械と話が通じない！
- 情報の量の壁：Web上の大量のマルチメディア情報を消化できず本当に有用な情報が見つからない
- 情報の質の壁：Web上にはウソも一杯この3つの壁を乗り越えた時に、スーパー・コミュニケーションの時代が来る！どの言語の話者でも、信頼性が高く有用な情報だけを短時間、低コストで通信

フォーラムの狙い

- フォーラムを組織し
 - 3つの壁を乗り越える技術について議論
 - 産官学の協力体制を構築
 - NICTで構築した言語資源ツールの配信、共有、評価

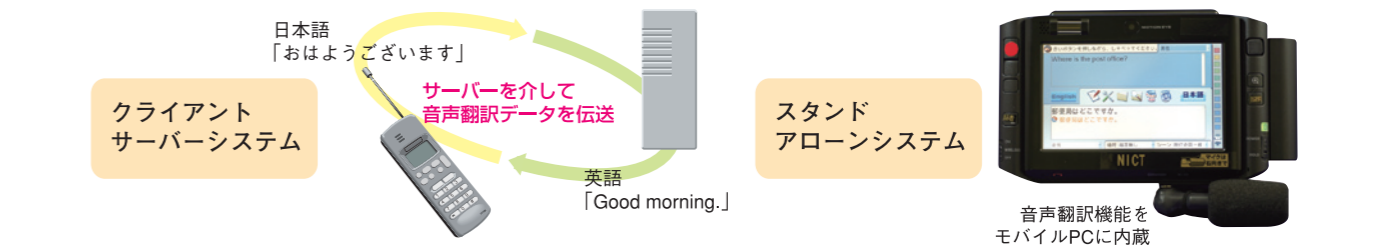
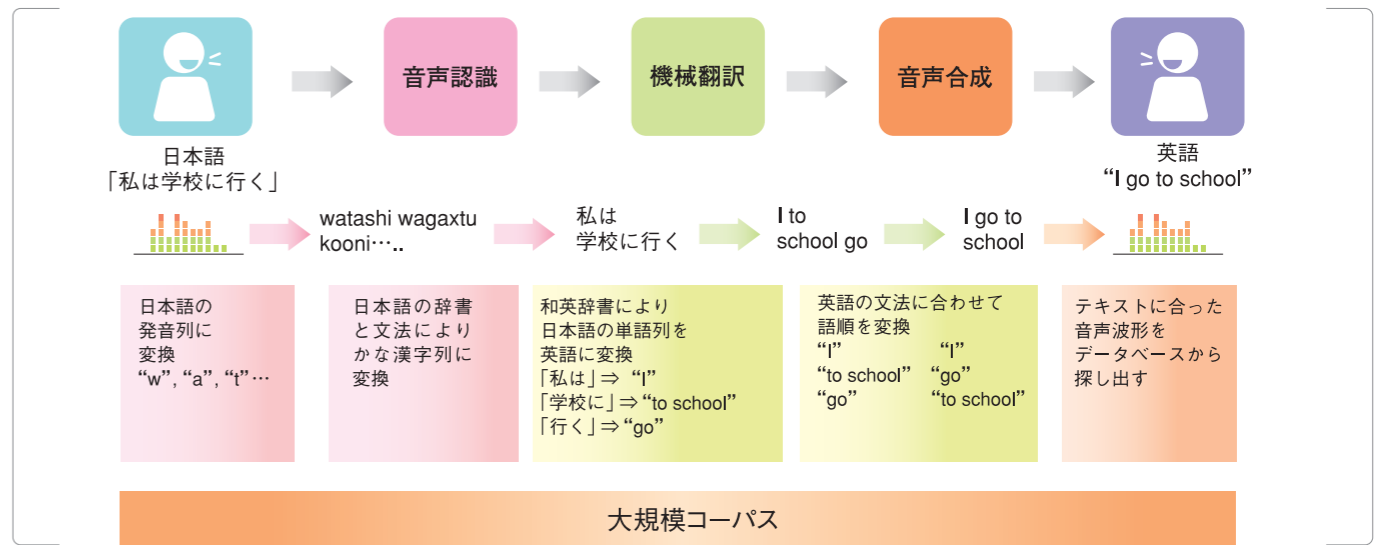
活動内容

- 普及促進部会：フォーラム成果の普及を促進し市場開拓を進める標準化、標準フォーマット、標準モジュールの提供
- 技術開発部会：関連分野の研究者・技術者に指針を与え、関連技術の研究開発を支援

高度言語情報融合フォーラム【ALAGIN】

人間同士あるいは人間と機械の「言葉の壁」、Web情報に内在する「量や質の壁」を克服する技術を対象に、言葉の「壁」を感じさせないコミュニケーション技術を普及・推進することを目的とし、産業界、大学、国などの教習を集結して戦略的に活動するフォーラムです。

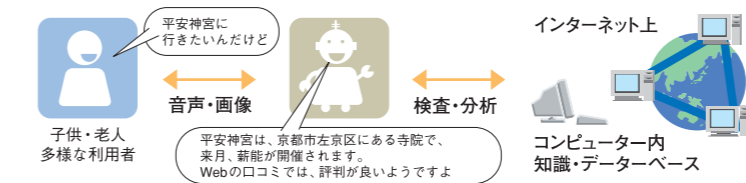
地球上には 6000以上の言語があるといわれています。言語の違いは、そのまま文化の違いに相当します。これらの言語を話す人々と直接コミュニケーションができればどんなにすばらしいことでしょうか。MASTARプロジェクトでは、自分の話した言葉を聞き取り（音声認識）、相手の言葉に翻訳し（機械翻訳）、その場で音声が出力される（音声合成）、言語の壁を越える究極のコミュニケーションシステム、音声翻訳システムの研究開発を進めています。



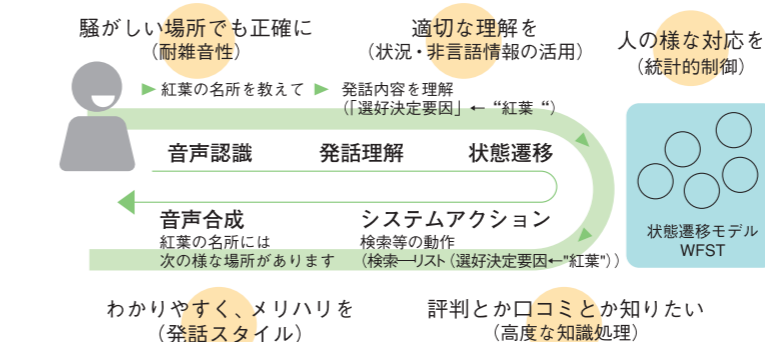
音声対話

だれもがストレスなく適切に情報を伝達できる情報通信システムの実現を目指して、ネットワーク端末とコミュニケーションするための音声解析技術や表情・身振り・手振りなどの言語以外の表現の認識技術、対話に必要な情報と推論のメカニズム等対話システムの研究開発を推進しています。

自然な表現でコンピューターと息のあったコミュニケーション



自然なコミュニケーションを表現するための対話技術・システム



マルチモーダルセンサーを利用した対話システム

- 自然な発話でネットワーク上の情報にアクセス
- 携帯回線・無線LANを利用
- クライアント・サーバ型システム



マルチモーダルセンサーを利用した対話システム

- 複数のセンサー情報 (音声、画像) を利用
- 音声、視線、動作によるコミュニケーションを実現

