NLPグループ

(1) ＳＦを用いた多言語機械翻訳に関する研究

ネットワーク時代となり、私たちは必要とする様々な情報を瞬時に手に入れることができるようになりました. しかしその情報が母国語以外で書かれているケースも珍しくありません. これらをいちいち手で翻訳していては時間がかかりすぎ効率が良いとは言えません. そのため機械翻訳の利用がなくてはならない時期が来ています. しかし現在の機械翻訳システムは翻訳の精度、品質がユーザの要求を十分に満たすレベルに到達していません. 特に不自然な訳文の出力がユーザの信頼を損なう原因のひとつになっています.

本研究では流暢な訳文を出力するために Super-Function(以下 SF)を用いた機械翻訳に着目します. SF とは、原言語と目標言語との対応を示す関数です. これらは対訳コーパスから作成されます. コーパスベースであるため流暢な訳文を抽出することができます. 図 1-1 に構築中のシステムを示します.

図 1-1:日英機械翻訳システム翻訳画面

(2) 株式予測と分析に関する言語生成システム

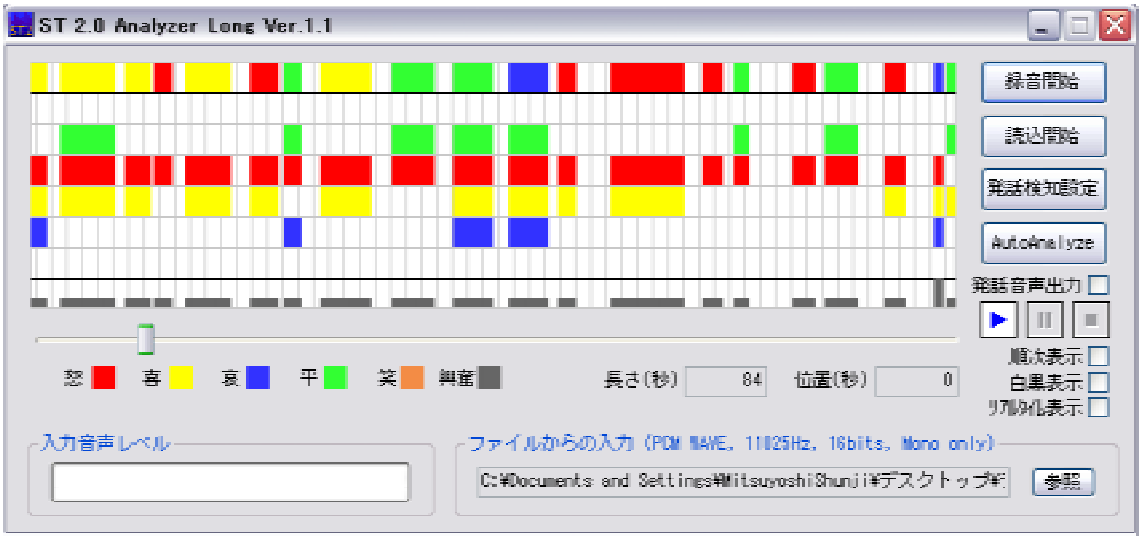
個人投資家に向けて、簡単な操作で自動的に株価の変動を言語の予測してくれるシステムを開発している. このシステムは、東証一部上場の全銘柄を対象として、株式評価モデルで株価の上昇、下落の方向性だけでなく、相場の転換点を予測して、説明の言語を生成する初めてのシステムである. 株価分析と予測では, データ間に潜む共通パターンを如何にして抽出すべきかが問題となる. そのため、データマイニングの多くの手法が提案されている. 本研究は、時系列データの回帰分析方法だけで市場予測を行っているが、他の手法を考慮していない. なお、データマイニングツールの選択、株式予測の結果に関する単語の対応等

について、今後検討を行う予定である.

(3) 中日機械翻訳における「把」文型の処理手法

中国語ではいくつかの漢字が結合して複合語となる場合が多い。特に「把」文型が含まれる文型が多い。「把」文型は変化が多く,意味も複雑であるため,構文解析における誤り,中日機械翻訳における曖昧性と深く関わっている。中日機械翻訳においては，「把」文型を正しく処理するのは極めて重要である。本稿では、中日機械翻訳に際し、「把」文型になりうる文字列を抽出して、「把」文型の自動翻訳手法を提案する。原言語の解析から目的言語の合成に至るまで、何段階の処理が必要とされる。まず、我々は大量の教科書、科学技術文献から「把」文型の例文を収集して構文解析し、「把」文型の翻訳ルールをまとめ、最後にこのような実例から抽出したルールを中日機械翻訳実験システムに組み込む。我々が構築した中語句「把」文型機械翻訳システムのインターフェイス画面を下図に示す.

(4) 韻律情報に基づく自然感情の認識とロボット、脳科学への応用



研究成果が市販された、韻律情報に基づく自然感情識別子 ST2.0＜(株)エイ・ジー・アイ＋日本 SGI(株)＞

リアルタイムに「怒り・喜び・悲しみ」の Feeling と「興奮（3 段階）・平常」の Emotion を識別する。

第一候補となる感情と、それに含まれる複合感情を時間軸に沿って同時に判定し、表示する。

fMRI 実験

解析風景

i. 研究背景

現在、米国では BCI（ブレイン・コピュータ・インターフェース）や BMI（ブレイン・マシン・インターフェース）が IT 産業に換わる主要ハイテクとして注目を浴びている。しかし、その実現には脳と機械が直接繋がるこれらの技術として、多くの問題（技術・倫理）がある。その一つが、人間の判断基準と記憶に大きな影響を及ぼす、感情＜Feeling＞と情動＜Emotion＞の定量化・測定である。

ii. 研究目的

そこで、人間がどのように感情を生起させ、物事の好き嫌いを判断するか？といったチューリング計算機科学の原点というべきテーマを中心に据え、ロボットや fMRI などでのシステム化から研究を進める。

このシステムの学際的活用により、不確実な主観評価に換わる脳と生理指標からの情動定量化基準を得る。

iii. 研究概要

人の音声からの感情（情動）の識別には、言語情報以外（非言語）での韻律情報が有効である．そこで、基本周波数や抑揚を利用した感情認識システムを構築し、ロボットや会話システムや BCI における非侵襲型脳イメージング技術との融合を目的とし、学際的な研究を実施する。(1)ロバストな韻律情報に基づく感情識別子を構築する．(2)会話システムとして(1)の識別子と音声認識を利用した、実際の携帯電話サービスで有効性を確認する．(3)会話型ロボットに(2)のシステムを利用し、実際の商品で有効性を確認する．(4)主観評価を使わない(1)の性能確認のために非侵襲型脳イメージング装置や生理指標との比較を行う。(5)(4)のシステムから情動研究の歴史的対立問題シャクター・キャノン・ランゲ論争の物理解釈を探る。

(5) 英作文支援システムの構築に関する研究

近年，国際化やインターネットの爆発的な普及に伴い，英語文書を作成する機会が増えている．特に，科学技術などの日本からの発信は英語で作成しなければならないが，ネイティブな英語文の作成は，多くの日本人にとって作業の負担が大きく困難である．そのため，英作文を支援するシステムの研究開発はますます重要となっている．

従来より，英作文を支援するシステムとして，機械翻訳システムが多く用いられてきた．機械翻訳とは，ある原言語を解析して別の目的言語で入力した言語とほぼ同じ内容の文章を生成する技術である．近代的な文法規則，さまざまな知識の整備によって研究開発が進められ，商用の翻訳システムも多数販売されている．しかし，さまざまな分野のテキストを高精度で翻訳できるシステムは未だ存在しない．特に論文，メール文などの情報発信のための英文は誤解やトラブルを防ぐために十分な品質を備えていなければならない．しかし，機械翻訳による一意的な訳文では不十分であることが多い．そこで我々は，日英の対訳用例コーパスを用いた英作文支援システムの構築を行った．本システムは，分野に応じた多様な表現文を検索し，利用者に提示することで英作文を支援する．また，用例文の編集や作成英文の訂正支援を行う．我々が構築した英作文支援システムのインターフェイス画面を図 1-2 に示す．

図 1-2：英作文支援システムの画面ダンプ図

(6) Chinese Semantic Analysis and its Application（中国語の意味解析とその応用）

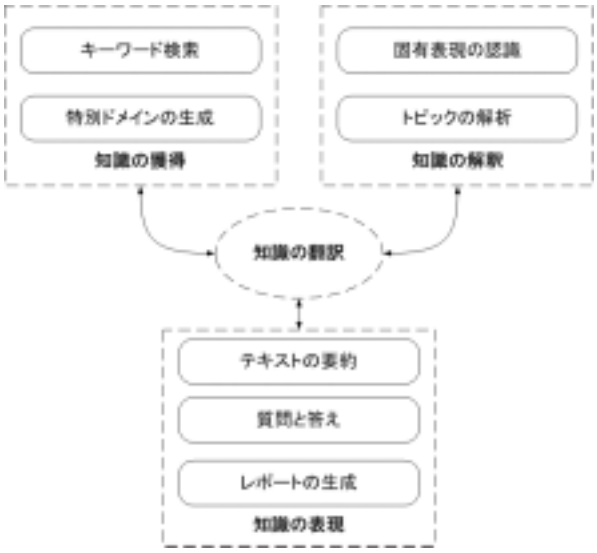
My research is on semantic analysis for Chinese. I have built a semantic analyzer that uses a maximum entropy classifier to determine the semantic relations between headwords and dependents. The system achieves an accuracy of just under 84%. The semantic relation tag set is a part of HowNet,

which is a prominent lexical dictionary in Chinese. Figure 1 shows an example of one annotated sentence. Currently, I am examining the uses of the semantic analyzer in emotion prediction for actors in sentences.

Figure 1-3, An example of one annotated sentence with semantic dependency relationships

(7) KANT システム

- KANT はどんなシステムですか？
  - 英語,日本語と中国語に関する多言語情報検索のシステムです。
    - 一つの言語を使って、色々な言語から情報を獲得できます。
- KANT の意味は何ですか？
  - Knowledge （知識）
  - Acquisition （獲得）
  - iNterpretation （解釈）
  - Translation （翻訳）
- KANT の知識はニュースから抽出した情報です。
- どうしてニュースを使うのですか？
  - ウェブで色々な国のニュースを簡単に見つけられます。
  - ニュースには色々な情報があります。
  - ニュースから文化をまなべます。
- システムは４つのモジュールがあります。



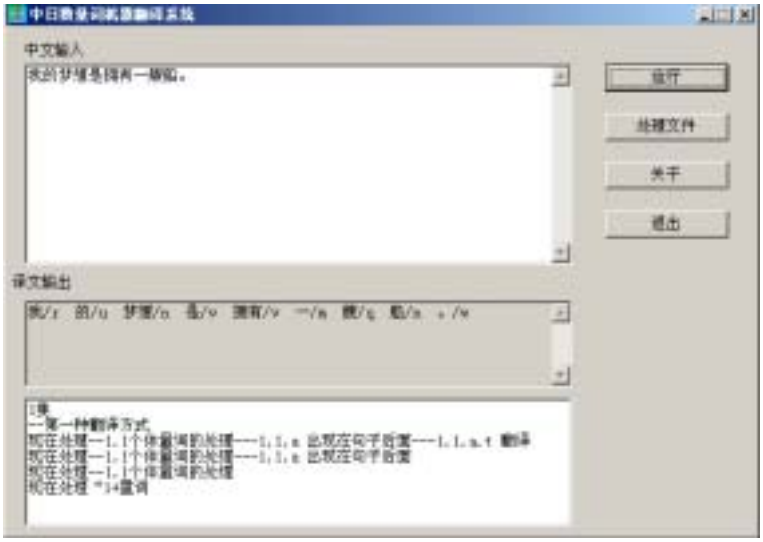
- 知識の獲得
  - キーワード検索
    - KANT のキーワード検索のアルゴリズムは標準のアルゴリズムより精度が高いです。
  - 特別ドメインの生成
    - 一つのドキュメントを与えて、ウェブ記事から特別ドメインを作成。
- 知識の解釈
  - 固有表現の認識
    - 人や場所や国などを抽出すします。
  - トピックの解析
    - 新しいトピックを見つけられます。
- 知識の翻訳
  - 機会翻訳
    - キーワードか文
  - 知識ネットワーク
- 知識の表現
  - テキストの要約
    - 要約したテキストは新しいニュースをブラウジングのために
  - レポットの生成
    - トピックを与えて、そのトピックについてレポートを生成します。
  - 質問と答え（QA）
    - 一つの言語で質問して、色々な言語から答えを貰えます。

(8) 機械翻訳における使役表現の翻訳規則について

日本語の使役表現のX（使役者）がY（被使役者）に／をVさせるにおいて，「させる」が動詞の未然形に下接する．中国語の使役表現はX（使役者）「叫」「 」「使」Y（被使役者）Vのような形で表され，使役詞「叫」「 」「使」と動詞がセットになって「させる」という意味になる．機械翻訳において，中国語の使役表現が「使役詞+動詞」で表現されるのを正しく認識できなければ，日本語に訳す時に大きな障害になる．本研究では，教科書及びホームページから大量の実例文を選出し，使役表現および関連情報を抽出し，その情報を分析し，使役表現の特徴などの検討によって，中日機械翻訳における使役表現の翻訳規則を提案する．

(9) 中日機械翻訳における数量詞の処理

中日機械翻訳における数量詞の処理は誤りが多い．本研究では,それらの文法特徴に基づき量詞を分類して処理する方法を提案する．まず,中日対訳コーパスから収集した数量詞の例文を形態素解析して,得られた量詞の種類と数量詞に修飾される名詞の語義特徴を統計する,そして異なる数量詞と出現する位置の異なりなどにより,機械翻訳における数量詞の翻訳規則を獲得する．これに基づいて構築した翻訳実験システムは2つのモジュールによって構成される,一つはこの数量詞を翻訳するかどうかを確認する．もう一つは,数量詞を翻訳する場合,翻訳形式を選定する．



中日機械翻訳における数量詞処理システム翻訳画面

オントロジに基づく古典資料の情報検索に関する研究

オントロジとは、人間が対象世界をどのように見ているかという、根元的な問題意識をもって、物事をその成り立ちから解き明かし、それをコンピュータと人間が理解を共有できるように、書き記したものである。オントロジ的とらえ方では、知識を構成する基本概念に立ち戻っての考察が行われることに特徴がある。情報検索に際して、ユーザ個別の要求、最も目的に適った検索や知識の発見を求める。本研究では、伝統的な情報検索システムが抱える問題点の解決策として、このオントロジを古典資料の情報検索に応用することを試みる。