

音声翻訳・同時通訳を実現する 音声言語処理技術

2012年2月3日(金)
(株)東芝 研究開発センター
住田一男



東芝グループは、持続可能な
地球の未来に貢献します。

音声言語処理に関する研究開発拠点

■ 東芝 欧州研究所(ケンブリッジ研究所、通信研究所)



ケンブリッジ
ブリストル

■ 東芝 中国研究開発センター



北京

■ 東芝 研究開発センター



川崎

東芝における音声言語処理技術

・ 音声合成技術

- 組み込み向け音声合成ミドルウェア(ToSpeak™)
 - コンパクトで高音質
- 音声合成クラウドサービス([ToSpeak Online](#))
 - 多様な声質・口調(短時間音声データによる話者適応)
 - 対応言語: 日本語、英語(アメリカ英語)、中国語(北京語)、スペイン語(アメリカ・スペイン語)、フランス語(カナダ・フランス語)の5ヶ国語対応

・ 音声認識技術

- コマンド音声認識から連続大語彙音声認識まで対応可能な柔軟性
- 組み込み・スタンドアロン動作からサーバー動作まで対応可能なスケーラビリティ
- 対応言語: 日本語、英語(アメリカ英語)、中国語
- 話者や環境雑音の変動に対してロバストな認識手法

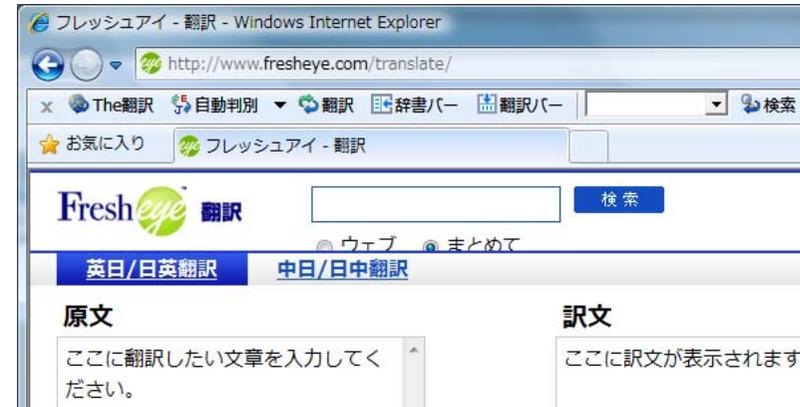
東芝における音声言語処理技術

機械翻訳技術

The翻訳™(パッケージソフト)



Webサービス



The翻訳エンタープライズ(サーバ)



Eiplaza™/MT(SaaS)



対応言語: 日・英・中双方向

機械翻訳とは

- ・ 自然言語処理の主要な応用の一つ

- 1940年代から研究がなされている

- ・ 主要な機械翻訳方式

- 規則ベース翻訳

- 原文を**解析**
- 原言語の構文木を目的言語の構文木に**変換**
- 目的言語の構文木から文を**生成**

} 規則で制御

- 用例ベース翻訳

- 原文と類似する翻訳例文を参照、訳を組み合わせる

- 訳例 「彼は学校へ行く」= “He goes to school.”

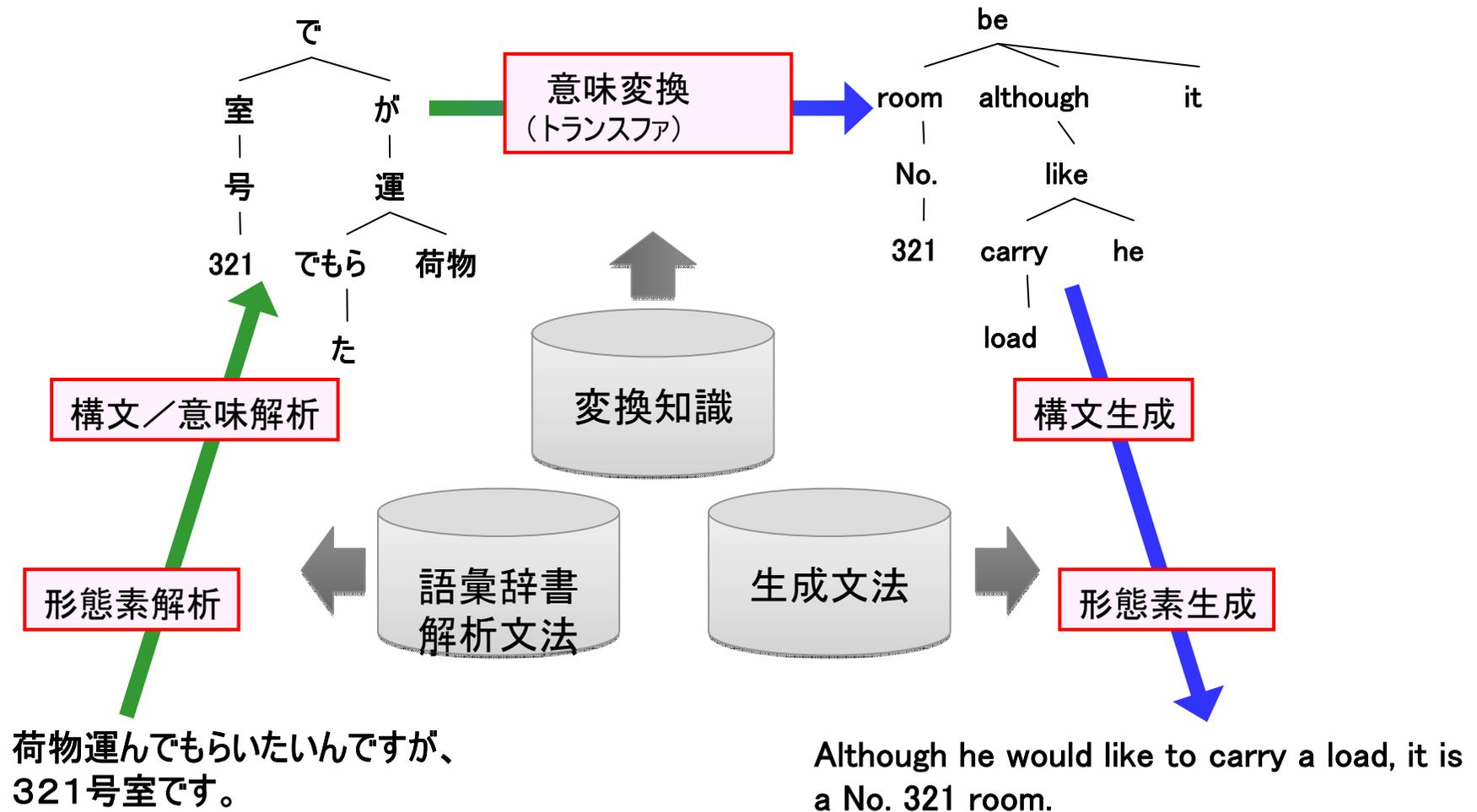
- 入力 「彼は会社へ行く」→ “He goes to an office.”

- 統計ベース翻訳

- 対訳コーパスから統計モデルを学習して変換

$$\hat{e} = \arg \max_e P(e | f) = \arg \max_e P(f | e)P(e)$$

規則ベース翻訳



高精度化の取り組み

・ 様々なユーザカスタマイズ機能

－ 翻訳メモリ辞書

E: Supplies are limited, so do not delay.

J: 数量に限りがありますのでお急ぎ下さい。

－ パタン翻訳辞書

E: A press on <A> starts the .

J: <A>を押すとが始まります。

－ 訳し分け辞書

に応じて=to suit; 実験結果; 決定する;

に応じて=in keeping with; データ; 温度(が)変化する;

－ ユーザ辞書 (単語・複合語)

－ 学習辞書 (訳語選択)

－ 特許前処理

－ セレクトコーパス辞書

単言語コーパスを利用した訳語選択

入力文

A spiky **landscape** shows **signs** of active **erosion** on the **surface** of Callisto.

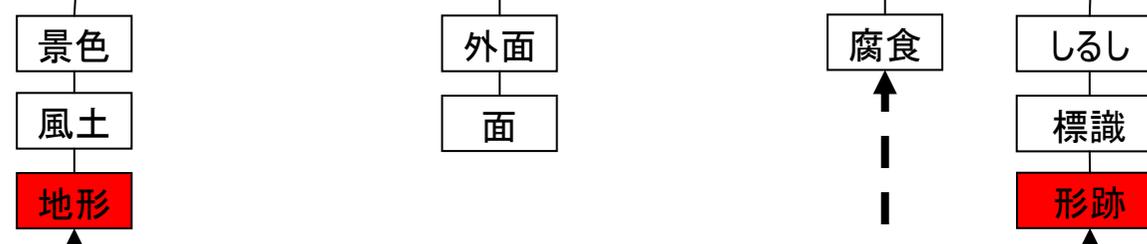


従来訳

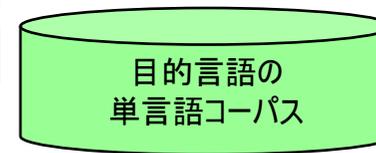
針**景観**な はカリストー**表面** 上の活発な **侵食** **を**示します。

訳語学習

学習後訳文



訳語選択



針**地形**な はカリストー**表面** 上の活発な **侵食** **形跡**ます。

翻訳精度の現状

- 一般用途(辞書登録なし)でも、斜め読みには有効.
- 利用者の英語能力により、評価が大きく分かれる.

Microsoft creates security ratings

マイクロソフト、セキュリティ格付けを作成

(IDG) -- Microsoft Corp. said it is adding a rating system to its security warnings to help customers take the appropriate steps when faced with a security threat.

(IDG) -- マイクロソフトは、セキュリティ脅威に直面した時顧客が適切なステップを取るのを助けるそのセキュリティ警告に、それが格付制度を加えていると言いました。

The company said security bulletins will be labeled with "critical," "moderate" or "low" severity ratings, and the bulletins will be sorted by type: client systems, Internet servers and internal servers.

会社は、セキュリティ公報がラベルを付けられるだろうと言いました「重大な」、「適度な」、「低い」厳しさ格付け、また、公報はタイプによってソートされるでしょう: クライアント・システム、インターネットサーバおよび内部サーバー

音声翻訳実現における課題

・ 話し言葉の認識

- フィラー(あの一、えーっと、…)への対応
- 言い淀み

・ 話し言葉の翻訳

(a) ロバストな解析、変換

- 音声認識誤りへの耐性
- 話し言葉固有の現象

(b) 原文の意味(意図)を保った、流暢な翻訳



話し言葉の難しさ(日本語を例に)

- ・ **非文法性**

助詞落ち、倒置、呼応表現欠如、文末省略、…

- ・ **口語表現**

「～しちゃう」、「～って」、「～んで」、…

- ・ **非流暢性**

言い淀み、言い直し、断片的発話、発話中断、…

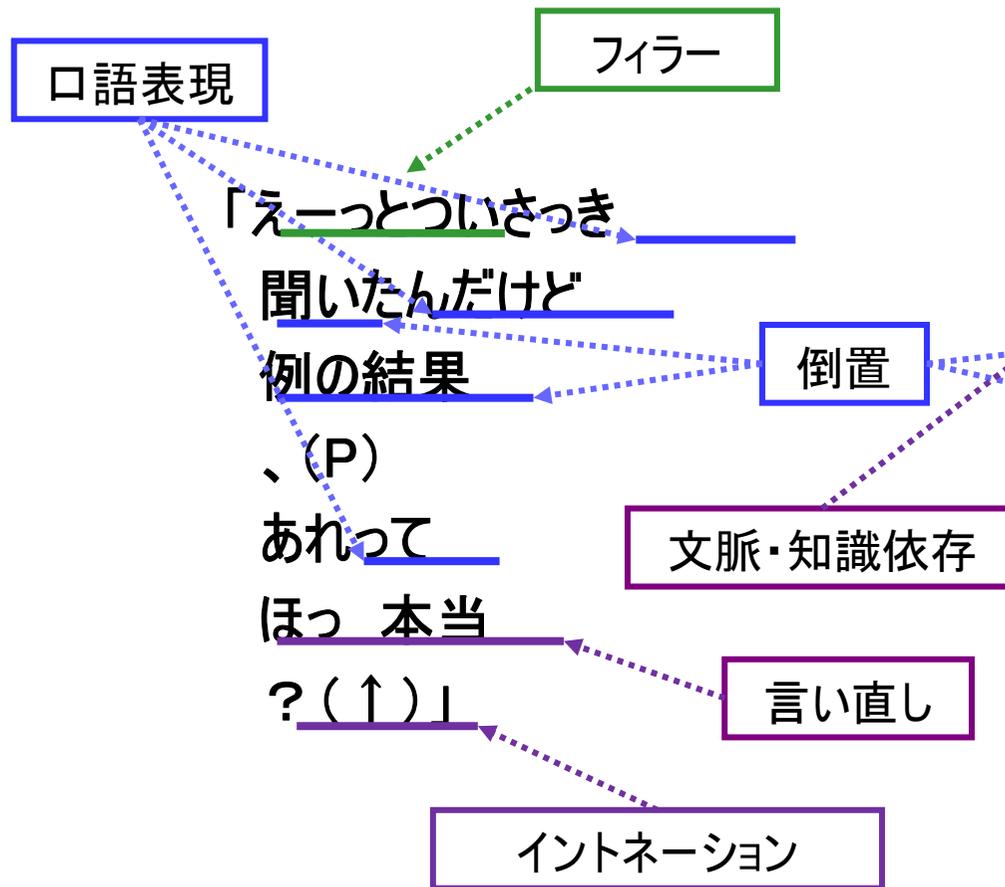
- ・ **文脈・知識依存性大**

省略、照応、状況・言語外知識依存、…

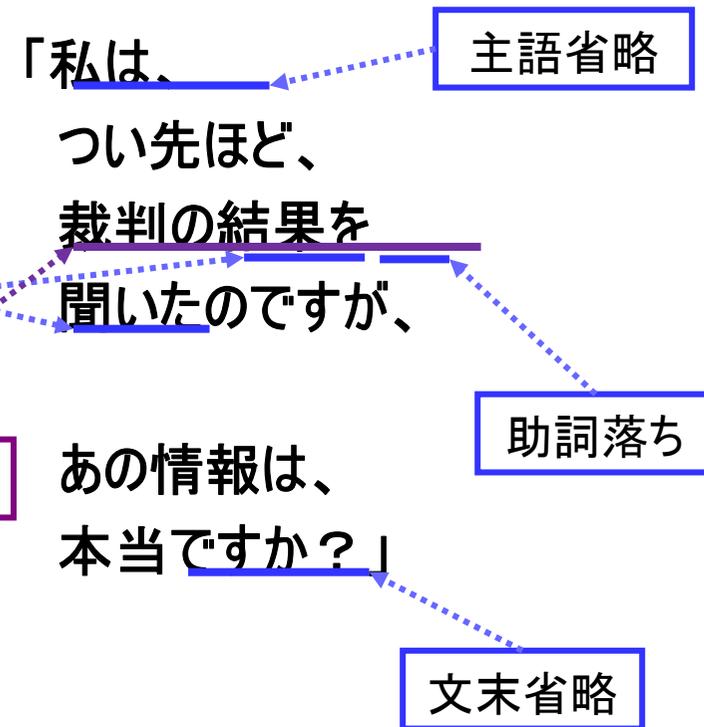
- ・ **無限の表現**

具体例

・ 話し言葉の例



・ 有りうる解釈の例



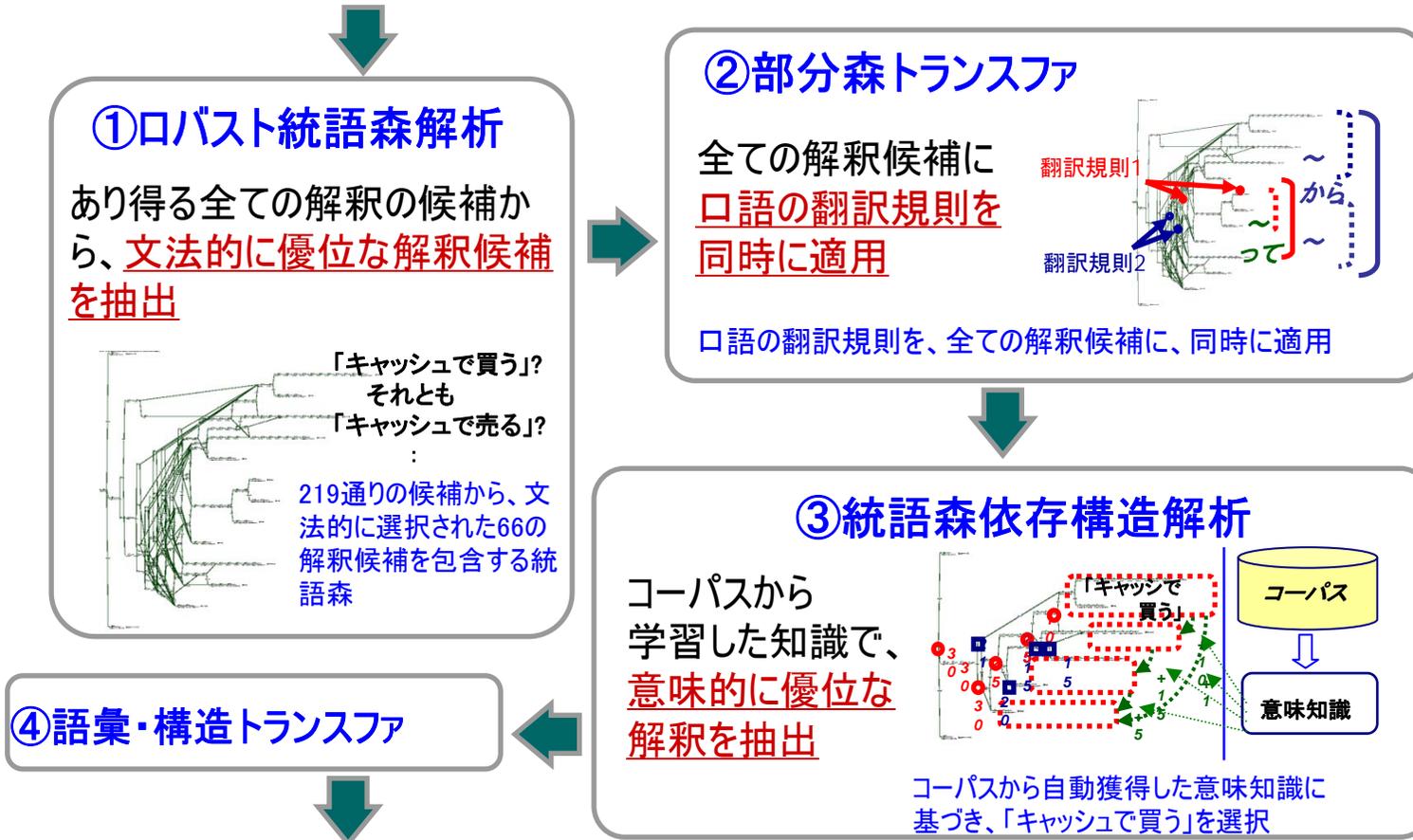
旅行会話向け音声翻訳アプリ

- ・ 機械解析の口語表現に対するロバスト化
- ・ 統計ベース翻訳の導入
- ・ 規則翻訳と統計翻訳のハイブリッド処理

話し言葉向け規則ベース翻訳

- ・ 話し言葉用日本語解析を独自開発
- ・ データ構造の最適化による小型化を実現

入力 = 「キャッシュで買うから300ドルで売って」



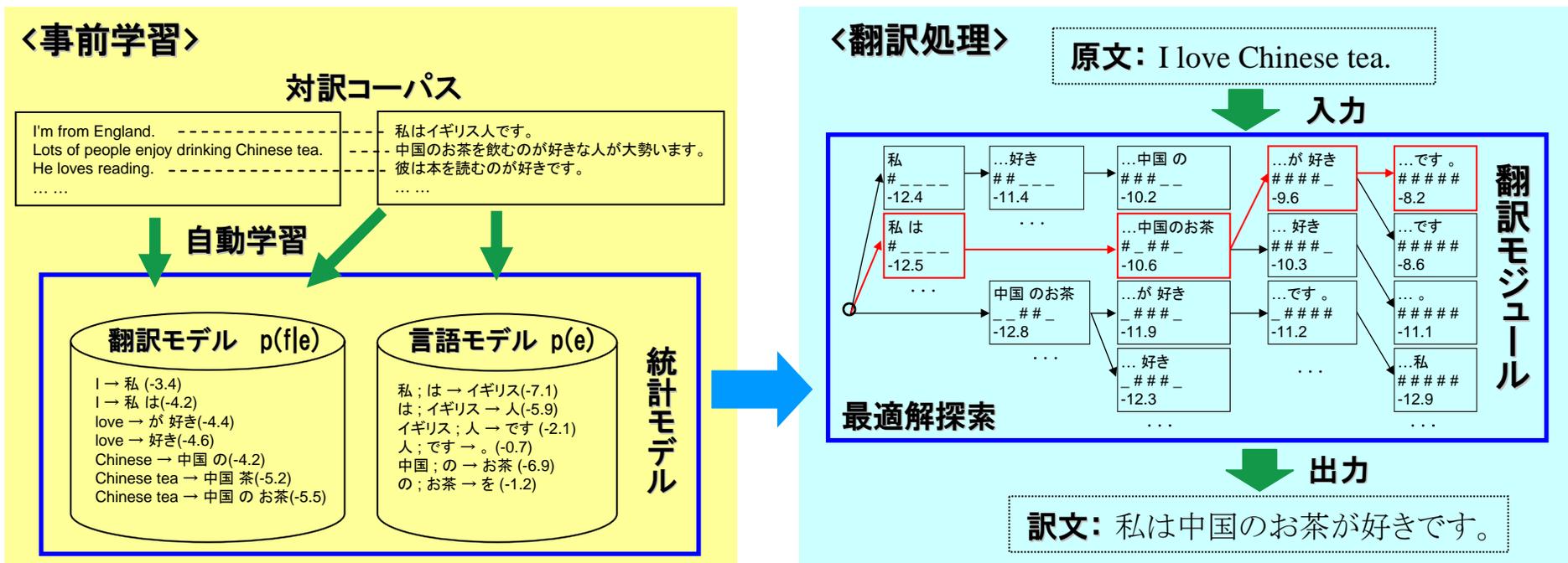
出力 = “I buy it for cash. Sell for 300 dollars.”

小型・高速な統計ベース翻訳

- 原文 f に対して最適な訳文 e を探し出す問題

$$\arg \max_e P(e | f) = \arg \max_e P(f | e)P(e)$$

- 効率的に候補枝刈りを行うアルゴリズムとデータ構造
 - スマートフォン上でも高速に動作するソフトウェアとして実現



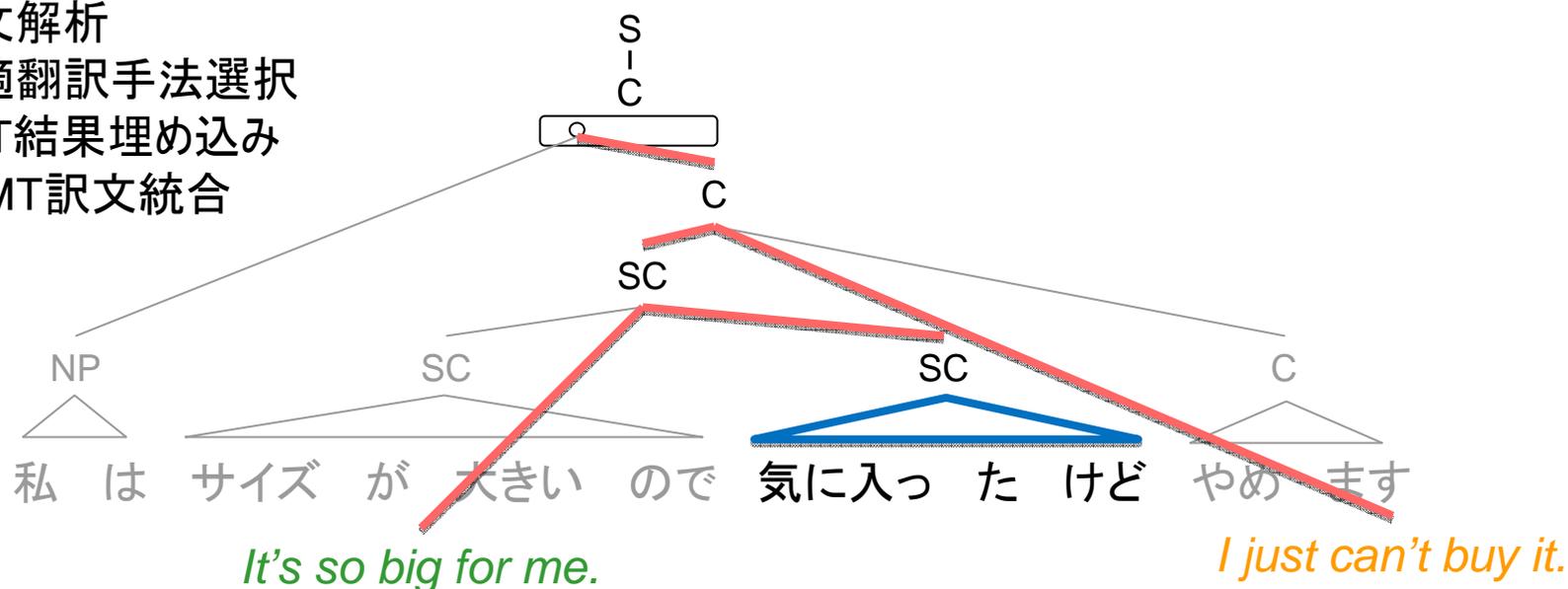
ハイブリッド翻訳

- ・ ハイブリッド=2つの相補的な手法を融合
 - 統計ベース翻訳(SMT)・・・大量の対訳コーパスに基づいて自然に翻訳
 - 規則ベース翻訳(RBMT)・・・文法規則に基づいて精密に翻訳

東芝ハイブリッド翻訳＝

統語構造に基づく文分割・訳文生成

- (1)構文解析
- (2)最適翻訳手法選択
- (3)SMT結果埋め込み
- (4)RBMT訳文統合



Transfer and generate by rules.

I like it, but it's so big for me, so I just can't buy it.

応用例「ポケット通訳™」

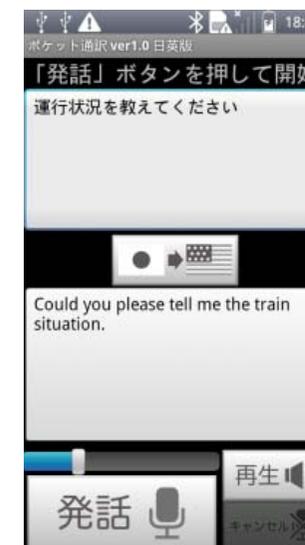
・ 旅行会話を対象とした音声翻訳アプリケーション

- 外国語に不安のあるユーザの、海外旅行支援が目的
- 端末内で、音声認識、翻訳、音声合成が全て動作
 - サーバとの通信不要
- 中国、北米旅行を想定して開発
 - Android Market公開版は、東日本大震災・復興関連語彙を強化



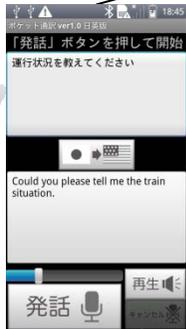
・ スタンドアロン型日英・日中・音声翻訳

- 周辺雑音など実環境に強い連続音声認識
- 話し言葉を自然に翻訳できる統計翻訳
- 自然で肉声に近い高音質な音声合成

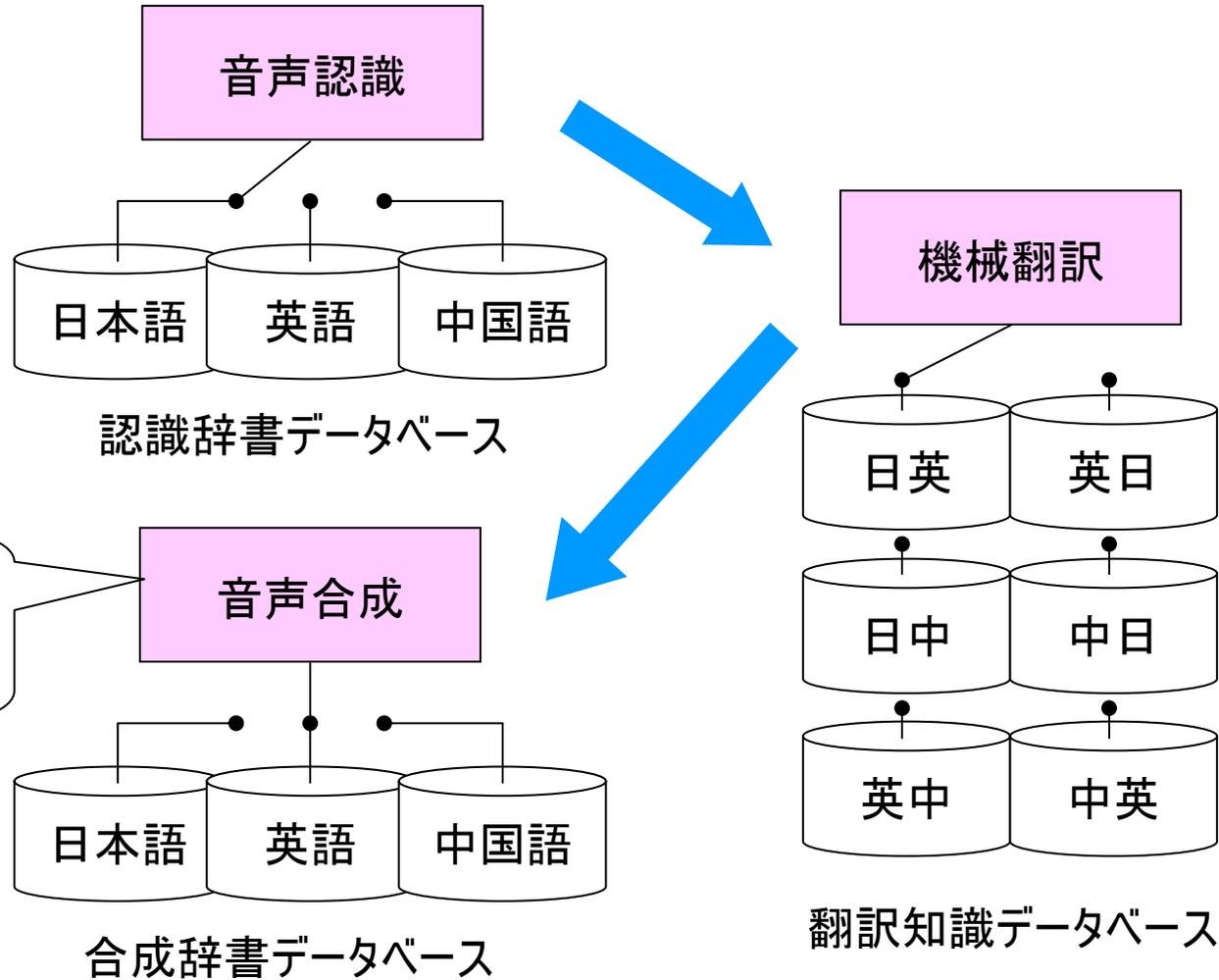


「ポケット通訳」の全体構成

運行状況を教えてください



Could you please tell me the train situation?



被験者評価実験(概要)

- 一方の被験者のみに「対話の目的」を伝え、音声通訳システムを使って、達成を試みさせる

Level-3-4	←レベル&目的ID
北京のバス路線図はどこで買えるか、 2箇所以上の場所を聞いて下さい。	←対話の目的
場所1: _____ 場所2: _____ : _____	←相手から得た情報の記入欄
OK <input type="checkbox"/> NG <input type="checkbox"/>	←目的の達成正否(被験者の自己評価)

- 対話のレベルを変えながら計26種のタスクを設定
 - 「挨拶」「謝意を伝える」「道順を尋ねる」
「観光先を推薦してもらおう」など

評価結果

・ 実験結果

- 「旅行における典型的なタスクの約7割を達成できる」性能

・ 課題の妥当性検証

- 被験者に提示した設問（日本語版アンケート抜粋(原文のまま)）

- 設問：課題の難しさ

- 今回提示した課題の内容は、自分が海外旅行をしたとき、または自国に来た海外旅行者の相手をするときなどに実際にありそうな事態と比較して、難しさについてどう感じたかを評価をしてください。また、評価理由やご意見などを自由に記入してください。
- 数値の意味：
5/難しい、4/やや難しい、3/ちょうど良い、2/やや簡単、1/簡単

- 3.4=「やや難しい」と「ちょうど良い」の中間

ビジネス利用可能な同時通訳に向けて

日常会話→対面業務、講演、会議等での利用

テレビ会議英語字幕という選択肢もあるが

- 音声認識の処理遅延の影響
- 音声認識率と英語スキルとの関係

逐次通訳

話者は一文ごと、あるいはある程度の長さで発話を区切り、通訳者がそれを訳す。
通訳が終わるまで話者は発話しない。

同時通訳

話者の発話を聞きながら、同時に訳していく。
逐次通訳と違い、会話が通訳のために中断されることはない。

デモ(対面業務向け同時通訳)

デモ

ビジネス利用可能な同時通訳に向けて

・ 音声認識

◆ 協調的発話 → 自由発話

- 複数文の連続発話
- フィラー(あの一、えーっと、等)の発生
- 口語表現
- 言い淀み(一文法単位内でのポーズ、言い直し、等)
- 抑揚の変化
- 発話中に発生する他発話の割込み(発話の分断、文脈の引き継ぎ)

◆ 専門用語への対応

◆ 方言、ノンネイティブ発話への対応

ビジネス利用可能な同時通訳に向けて

・ 漸進的口語翻訳

◆ 規則 × 統計のハイブリッド翻訳技術

⇒ 高精度化 + ドメイン適応性の向上
規則翻訳で培った翻訳知識の活用

◆ まとまりの検出による短単位訳出

⇒ 同時性の向上

◆ 文脈に依存する訳し分け

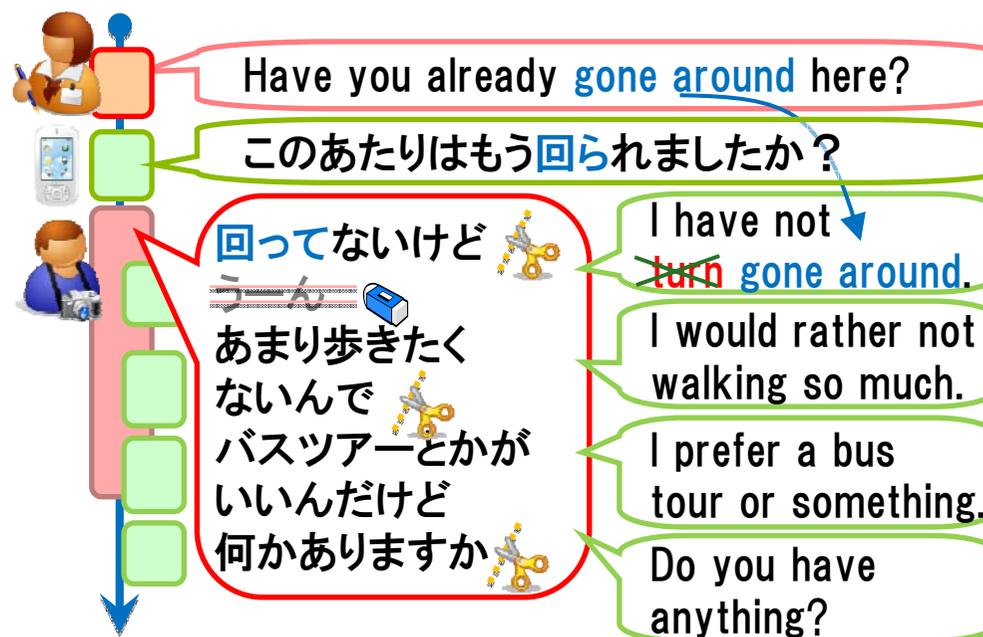
⇒ 高精度化

◆ 専門用語への対応

◆ 方言、ノンネイティブ

◆ 参照・直示

◆ 訳の分かりやすさ



ご清聴ありがとうございました

本資料掲載の商品、機能等の名称は、それぞれ各社が商標として使用している場合があります。