

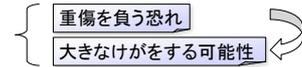
言い換え生成技術の 研究動向、課題、応用

藤田 篤
名古屋大学大学院 工学研究科
電子情報システム専攻

言い換えとは

■ 言い換え：換言、パラフレーズ、言い替え

同じ意味を持つ
異なる言語表現



ある表現の言い換え
を生成する行為

ひとは言い換える。それはなぜ？

■ [高塚, 99]

- 第二言語の学習者が自らの言語能力を補う
- 相手の理解を促進させるために自分の先行発話を言い換える
- 意味を確認するために相手の発話を言い換える

■ [国広, 00]

- メンツを傷つけることを避ける婉曲表現
- 仲間との連帯性を保つ集団語
- 相手によく理解してもらう

言い換え技術の重要性

■ 言語による情報伝達の円滑化

- 情報伝達ツール：言語（言語表現）
 - 分かりやすく言い換える（生成）
 - 同じことを言っていると理解する（認識）

■ 計算機による言い換え処理

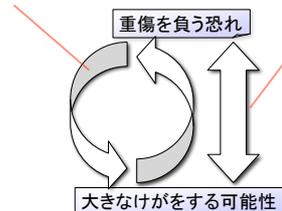
= 自然言語処理のコア技術

言い換え生成技術の 研究動向、課題、応用

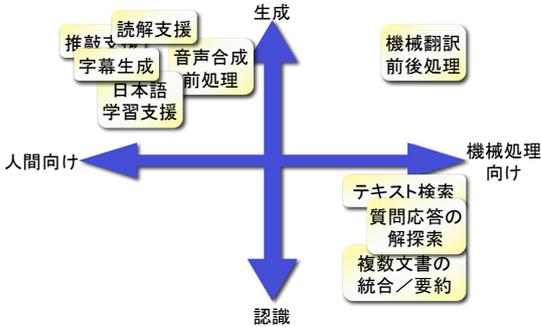
生成と認識

同じ内容を伝達する
異なる表現を**生成**

異なる表現が同じ内容を
伝達することを**認識**



言い換え技術の応用



2006.09.28

電気関係学会東海支部連合大会

7

機械処理向け言い換え認識

■ 複数の文書の要約生成 [Barzilay, 03]

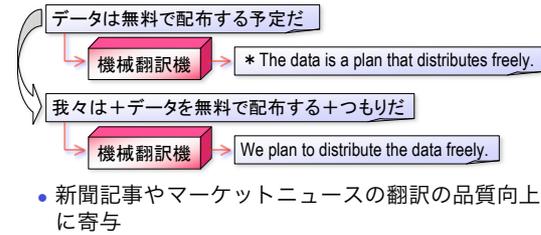


電気関係学会東海支部連合大会

8

機械処理向け言い換え生成

■ 機械翻訳の前編集 [Shirai+, 98] [吉見+, 00]



- 新聞記事やマーケットニュースの翻訳の品質向上に寄与

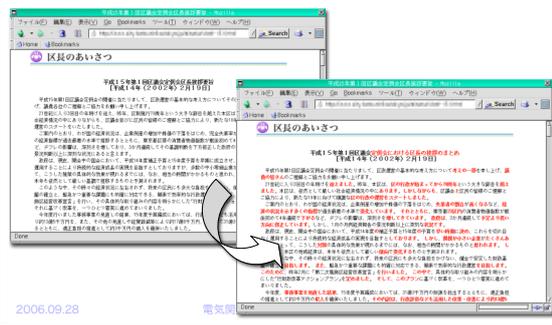
2006.09.28

電気関係学会東海支部連合大会

9

人間向け言い換え生成

■ テキスト簡単化による読解支援 [Inui+, 03]



2006.09.28

電気関係学会東海支部連合大会

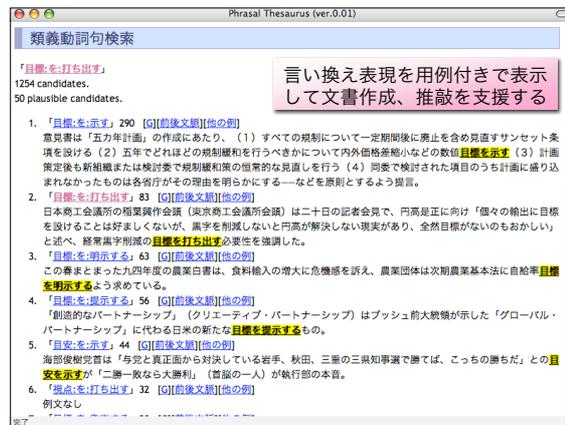
人間向け言い換え生成：いくつかの路線

- リライト：入力を自動的に書き換える
 - e.g. 文章の標準化 (新聞語彙, お薬DB, 翻訳)
- 情報発信支援：代替表現を提示する
 - e.g. 表現の発想や推敲の支援
- 情報受信支援：読解・解釈を促す情報を付加
 - e.g. 語釈文や類義語の提示

2006.09.28

電気関係学会東海支部連合大会

11



言い換え生成技術の研究動向、課題、応用

主要な研究路線

■ 生成

- 言い換え生成のフレームワーク
- 特定の種類の言い換えに着目した分析
- アプリケーション指向の変形規則の開発

■ 認識

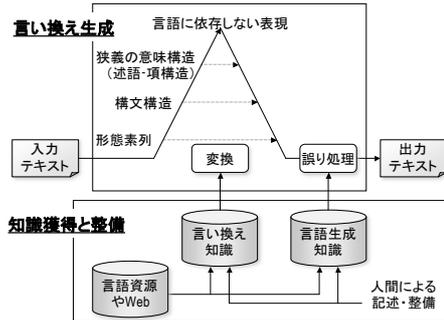
- 同義性判定 (複数文書要約、質問応答)
- 言い換え表現対の獲得 (言い換え生成、認識)

2006.09.28

電気関係学会東海支部連合大会

14

言い換え生成のフレームワークの典型



2006.09.28

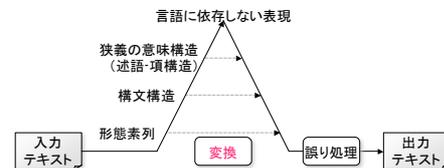
電気関係学会東海支部連合大会

15

トランスファ方式

■ 70～80年代に機械翻訳に用いられた方式

- 言い換える対象は入力の一部
- 構成性を仮定できるならば採用可能
 - 変換対応表は作れる、という信念
- どの抽象度で変形する？



2006.09.28

電気関係学会東海支部連合大会

16

語句の置換

■ 形態素列の情報のみに基づいて置換

[藤田+, 00] [Barzilay+, 01] [Shimohata+, 02]

海辺の丘陵に新しいマンションが完成した。
 海辺の高台に新しいマンションが完成した。

- 言い換えのための知識は個別的 (= 一般化困難)
 - 同義表現辞書を安価に構築することが課題

公算(N) ⇒ 可能性 逮捕する(V) ⇒ 捕まえる
 VB TO comfort PRP, burst into tears ⇒ cry
 ⇒ VB TO console PRP,

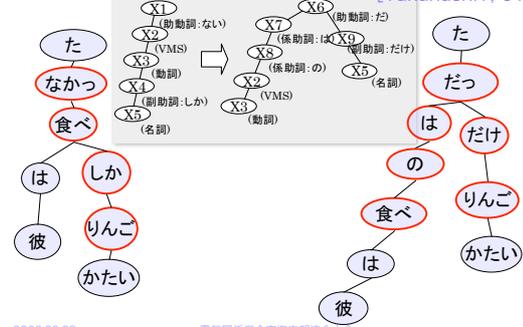
2006.09.28

電気関係学会東海支部連合大会

17

構文構造の変形

[Takahashi+, 01]



2006.09.28

電気関係学会東海支部連合大会

18

構文構造の変形

■ 様々な応用を考えたときに最低限必要な情報

- 実用レベルに円熟した解析技術に支えられている
- 自動獲得できる言い換え表現対の実装先
[鍛冶+, 01] [Takahashi+, 01] [Lin+, 01]
[Pang+, 03] [Shinyama+, 03]

今週当選したのは奈良県の男性でした。
今週は奈良県の男性が当選しました。

■ 論点

- どの程度の言い換えをカバーできるか?
- 変形規則の適用条件 (語に依存) をどう扱うか?

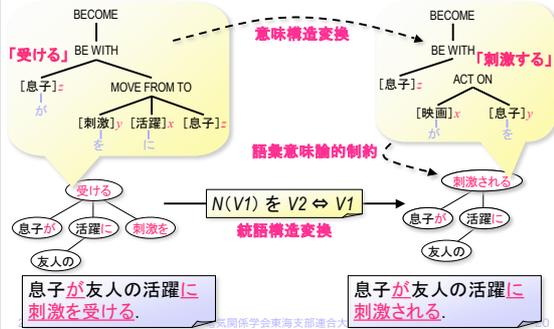
2006.09.28

電気関係学会東海支部連合大会

19

語彙意味情報に基づく統語トランスファ

[影山, 96] [Dorr, 97] [藤田+, 06]



語彙意味情報に基づく統語トランスファ

■ 言い換えの種類ごとに調査が必要

- 他の意味的な特性をとらえる枠組
 - 意味構造の変形・対応付け規則
 - 意味構造からの生成規則 (制約)

■ 語彙資源の構築

- 語の振る舞いや意味に関する情報

受ける(V) ⇔ [BECOME z [y MOVE FROM x TO z]]

- 語と語の関係の情報

刺激する(V) ⇔ 刺激(N)

暖かい(A) ⇔ 暖まる(V)

2006.09.28

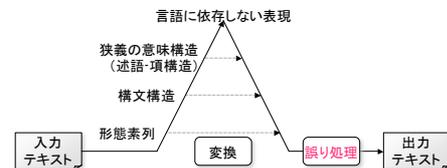
電気関係学会東海支部連合大会

21

誤り処理

■ 言い換えた箇所と周辺の表現のつながり方が不自然になる場合がある (変換誤り)

- 生成+ランキング/フィルタリング
 - 文法規則に基づく手法
 - 統計に基づく手法



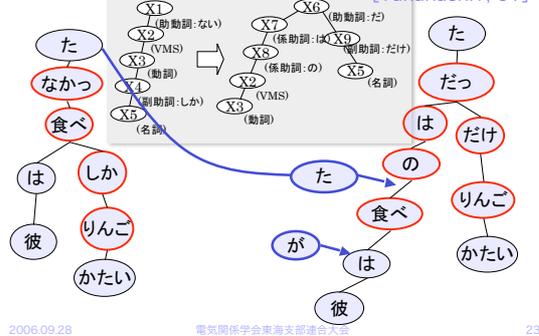
2006.09.28

電気関係学会東海支部連合大会

22

規則に基づく誤り修正処理の例

[Takahashi+, 01]



2006.09.28

電気関係学会東海支部連合大会

23

統計的言語モデルによる誤り検出

■ 生成された言い換え表現をふるいにかける

- e.g. 動詞と格要素の共起の誤り検出 [藤田+, 04]

言い換え候補表現	統計的に推定した 文法的適格性
傾向が強まる	4.7888e-05
賛成が半数を超す	1.0655e-07
こちらより下回る	8.9903e-08
判断を中心に据える	1.5676e-08
中心が買う	2.7903e-10

2006.09.28

電気関係学会東海支部連合大会

24

言い換え生成技術の研究動向、課題、応用

日本語の色々な言い換え

- 熾烈な視聴率争奪戦に進む**公算**が大きい。

激しい視聴率争奪戦に進む**可能性**が高い。

単語や句の置換
- 今週当選したのは**奈良**の男性**で**した。

今週は**奈良**の男性が**当選**しました。

構文の組み替え
- 息子が友人の活躍に**刺激**を受ける。

息子が友人の活躍に**刺激**される。

語の組合せ
- 古**びたお寺の年老いた住職を訪ねた。

古いお寺の年老いた住職を訪ねた。

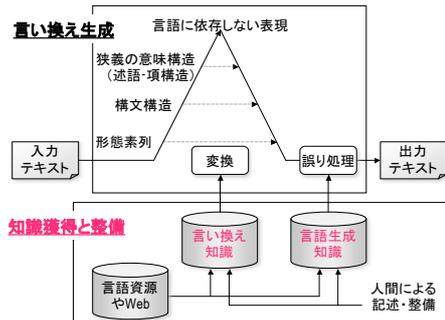
品詞交替

2006.09.28

電気関係学会東海支部連合大会

26

言い換え生成のフレームワークの典型



2006.09.28

電気関係学会東海支部連合大会

27

技術の確立に向けた課題 (まとめ1)

■ 高いカバレッジの変換知識ベースの構築

- 統語的言い換えについてはパターン記述が課題

 - VのはNだ ⇔ NがV
 - 言語学/統語論を取り入れればカバーできそう
- 語彙的言い換えについては表現対の収集が課題

 - 公算 ⇔ 可能性 | 手に負えない ⇔ 処理しきれない
 - 自動獲得だけでは不十分、分解・保守が鍵
- 語彙構成的言い換えは生産的だが未踏の現象



2006.09.28

電気関係学会東海支部連合大会

28

技術の確立に向けた課題 (まとめ2)

■ 語彙資源の構築

- 必要な情報

 - 語の振る舞いや意味に関する情報
 - 語と語の関係の情報
- 方法論の確立+実装

■ 曖昧性との付き合い

- 換喩なども含めて [Pustejovsky, 95]



2006.09.28

電気関係学

29

まとめ

言い換えという研究分野

■ 計算機科学と言語学の接点

- 実際は人間の認知、社会的背景にも関わる

■ 自然言語処理のコア技術

- 分かりやすく／処理しやすく言い換える（生成）
- 同じことを言っていると理解する（認識）

■ さらなる情報

- 名古屋大学→佐藤理史研究室→藤田

<http://sslab.nuee.nagoya-u.ac.jp/~fujita/paraphrase/>

2006.09.28

電気関係学会東海支部連合大会

31

Bibliography of paraphrasing

1. 会議 (conferences)
 - [IW, 2005] The 3rd International Workshop on Paraphrasing (IW2005) -, 2005.
 - [IW, 2003] The 2nd International Workshop on Paraphrasing: Paraphrase Acquisition and Applications (IW2003) -, 2003.
 - [NLP, 2001] The 1st Workshop on Automatic Paraphrasing: Theories and Applications -, 2001.
 - [IW, 2001] 言語処理学会10年次大会併設ワークショップ「言い換え/パラフレーズの自動化」-, 2001.
2. 解説 (survey)
 - [乾ら, 2004] 岩瀬太郎, 藤田真. 言い換え技術に関する研究動向. 自然言語処理, Vol. 11, No. 5, pp. 151-198, 2004.
 - [Hirst, 2003] Hirst, G. Paraphrasing Paraphrases. [IW, 2003], pp. Invited talk, 2003.
 - [乾, 2002] 岩瀬太郎. 言語表現を言い換える技術. 言語処理学会第10年次大会チュートリアル, pp. 1-21, 2002.
 - [佐藤, 2001] 佐藤理史. なぜ言い換え/パラフレーズを研究するのか. [IW, 2001], pp. 1-2, 2001.
 - [山本, 2001] 山本和英. 換言処理の現状と課題. [IW, 2001], pp. 93-96, 2001.
3. 言い換えという行為の分析 (socio-linguistic analysis)
 - [森口, 2001] 森口隆. 何のための「言い換え」か～読者の視点からのアプローチ. [IW, 2001], pp. 27-30, 2001.
 - [岡田, 2000] 岡田哲弥. 人はなぜ言葉を言い換えるか. 言語, Vol. 29, No. 10, pp. 20-25, 2000.
 - [川村, 2000] 川村英吾. 言語生成プロセスの一端～命題を言い換えるというコミュニケーション方略について～. 大阪府立工業高等学校研究紀要, Vol. 34, pp. 77-82, 2000.
 - [高塚, 1999] 高塚成彦. コミュニケーション方略としての「言い換え」～その前置内容と方法に関する基礎的研究～. 岡山大学教育研究紀要集, Vol. 第110号, pp. 1-12, 1999.

[Show/Hide meta]

言い換えるあれこれ / A classification of paraphrases

謝辞

ここで紹介した言い換える例は、奈良先端科学技術大学院大学情報科学研究所自然言語処理学講座（尚本研究室）において2003年夏に実施された「言い換える百選」プロジェクトにおいて集められたものをベースにしている。プロジェクトに参加して下さった方々に感謝の意を表す。

もくじ

- 動詞の言い換え
 - 【分詞文の言い換え】【文分割（連体節主節化）】【文分割（連用節・並列節の分割）】【接続表現の言い換え】
- 動詞の言い換え
 - 【否定表現の言い換え】【比較表現の言い換え】【類・使役の交換】【自他の動詞交換】【授受動詞の交換】【建置り/場所格交換】【出動詞の交換】【相互格の交換】【補文構文と格要素の交換】【句節動詞の言い換え】【授受の構文の言い換え】【修飾要素の交換（格の先の変更）】【数算符の移動】
- 文法カテゴリを養える言い換え
 - 【名詞と動詞（構態動詞構文の言い換え）】【動詞と形動詞】【名詞と形動詞】【ナ形動詞とイ形動詞】
- 主幹交換
 - 【名詞句】【句と節】【格と副詞句】
- 内容論の複合表現の言い換え
 - 【複合名詞の分解・構成】【「AのB」 ↔ 連体節】【複合動詞の分解・構成】【名詞後修飾の置換】
- 構態動詞/モダリティの言い換え
 - 【構態動詞置換】【取り立て動詞の移動】【動詞による特徴づけの削除】【伝達性モダリティ】【敬語表現の言い換え】【文体的変換】
- 内容論句の言い換え
 - 【名詞の言い換え】【動詞の言い換え】【形動詞の言い換え】【副詞の言い換え】
- 慣用表現の言い換え
 - 【慣用語】【表記の置換/換言】【換言】
- 日本語となくとも他の言語にある言い換え
 - 【いさかいをなすこと】

【編集】 すべて強調表示 大文字/小文字を区別

完了