

用例に基づく言い換え生成

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

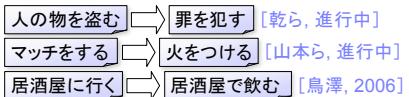
フレーザル・シソーラス

■語よりも大きな単位の表現の体系(化)

- 同義／類義関係(言い換え)



- 上位／下位、部分／全体、含意、因果などの関係



2006.09.15

NLP若手の会第1回シンポジウム

2

目標:「句」「節」の言い換え生成機構

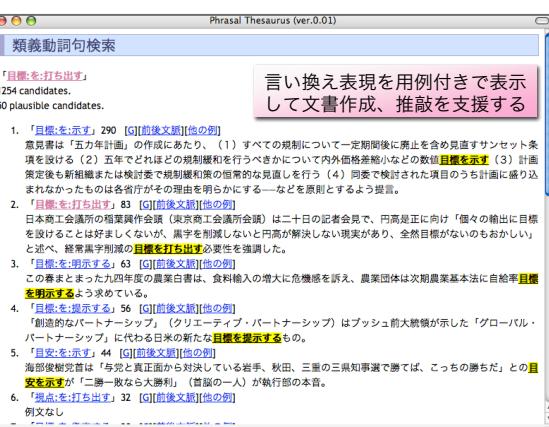
■仕様

- 入力:「句」「節」
 - 出力:入力表現の言い換え表現のリスト
- ねらい
- 頑健で言い換え生成機構を実現
 - 言い換えの基本処理単位についての検討
 - 意味的等価性の計算方法を検討
- ご利益
- 従来のシソーラスと同様、幅広いタスクに有用

2006.09.15

NLP若手の会第1回シンポジウム

3



設計上の論点

■提供したい機能

- 与えられた表現に対する言い換え表現のリストを生成する

■論点

- どの範囲の表現？
- 分類体系？距離空間？
- 辞書？生成機構？



2006.09.15

NLP若手の会第1回シンポジウム

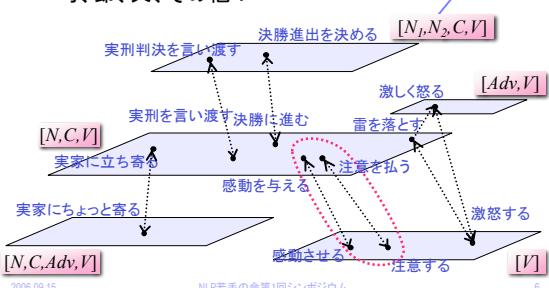
5

論点1:どの範囲の表現？

■語よりも大きな単位の表現

- 句、節、文、その他？

品詞に基づく表現の
パターン(以下「型」)



2006.09.15

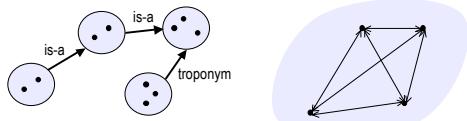
NLP若手の会第1回シンポジウム

6

論点2: 分類体系？距離空間？

■ 表現間の類似性の表現方法

- 事前に意味分類(クラス)を与えるか？
- 類似度を見積もって近傍を探索するか？



例: EDR、WordNet、FrameNet

2006.09.15

NLP若手の会第1回シンポジウム

7

論点3: 辞書？生成機構？

■ 辞書

- 語のシソーラスと同様の静的な辞書？

■ 生成機構

- 入力に対して類義表現を動的に生成する？

2006.09.15

NLP若手の会第1回シンポジウム

8

他の言い換え生成研究との比較

各種アプローチ	1.対象	2.分類	3.静動
シソーラスに基づく手法 (同分類の表現との置換)	語+成句	分類	静的
コーパスに基づく手法 (自動獲得したパターンの適用)	不定形 (※1)	分類/距離 (※2)	静的
変形規則に基づく手法 (規則/パターンと語彙的制約)	定型	—	動的
フレーザル・シソーラス	広義の動詞句	距離	動的

※1: 獲得の際に用いられる制約によって制限される

※2: クラスタリング(自動的に分類)する場合と
直接距離を計算する場合がある

2006.09.15

NLP若手の会第1回シンポジウム

9

なぜ動詞句？

■ 語を言い換える場合に生じる問題

- 語義曖昧性
- 文脈依存性
- 非構成的表現の誤認識

■ 上記の問題の軽減をねらって動詞句

- 自立的に言い換え可能なのでは？
= 言い換えの基本単位としてはどうか？

フレーザル・シソーラスでは
この仮説を検証する

2006.09.15

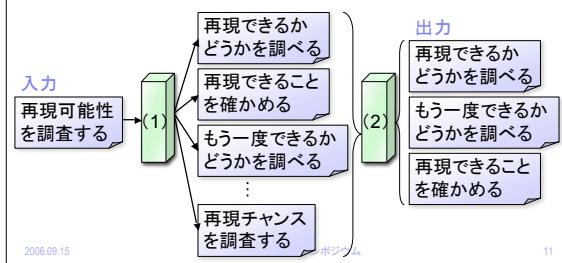
NLP若手の会第1回シンポジウム

10

研究課題

■ 2つの研究課題

- (1) 多様な候補表現の生成
- (2) 言い換え表現の順位付け



2006.09.15

11

研究課題1: 多様な候補表現の生成(1/2)

■ 複合語や節にも及ぶ



2006.09.15

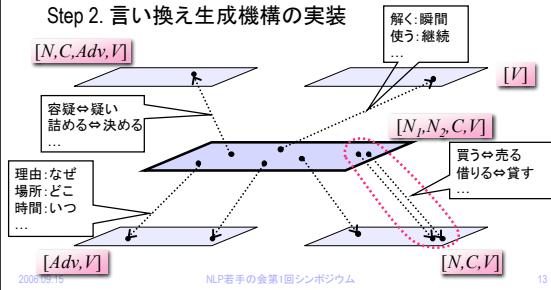
NLP若手の会第1回シンポジウム

12

研究課題1: 多様な候補表現の生成(2/2)

■ アプローチ

- Step 1. 入力を「型」で固定して事例収集
- Step 2. 言い換え生成機構の実装



研究課題2: 言い換え表現の順位付け(1/2)

■ 言い換え候補表現の「型」や長さに依存しない評価指標

● 入力と候補表現の意味的類似性

- 候補表現は入力と同じ意味を持っているか？
- どんな違いは許容でき、どんな違いは致命的か？

● 候補表現の文法的適格性

- 候補表現は受理できるものか？
- 候補表現は与えられた文脈で使えるか？
(文脈適合性)

2006.09.15 NLP若手の会第1回シンポジウム

14

研究課題2: 言い換え表現の順位付け(2/2)

■ 使える情報

- 意味的類似性:
 - 同じ型の表現:
構成語の類似度で近似
 - 異なる型の表現:
言い換え規則に当てはまれば類似度1
- 文法的適格性:
 - 表現そのもの:
コーパス中の出現頻度や出現確率で近似
(統計的言語モデル)
 - 文脈適合性:
ユーザ任せ

2006.09.15

NLP若手の会第1回シンポジウム

15

最初の調査

■ 「名詞+名詞+格助詞+動詞」 $[N_1, N_2, C, V]$

Step 1. 事例収集(214事例)

- 言い換え先の「型」と言い換え方
- 言い換え生成に必要となる語彙知識

言い換え先の「型」	用例数	言い換え方の例
名詞+格助詞+動詞	56	$[N_1, \text{が}, v(N_2)], [N_1, C, V]$
名詞+名詞+格助詞+動詞	48	$[N_1, N_2, C, V], [N_1, N_2, C, V]$
名詞+格助詞+名詞+格助詞+動詞	27	$[N_1, の, N_2, C, V], [N_2, の, N_1, C, V]$
節+名詞+格助詞+動詞	15	$[v(N_1), N_2, C, V]$
名詞+動詞	15	$[N_1, v(N_2)]$
形容動詞+名詞+格助詞+動詞	7	$[N_1, 的な, N_2, C, V]$
その他(18種類)	46	【一例】 $[N_1, ある, N_2, C, V]$

2006.09.15 NLP若手の会第1回シンポジウム 16

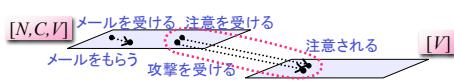
実現に向けて

■ 典型的でない部分もカバーすべき

- 未開拓、有用、チャレンジング(p.15)

■ 仮説

- ある表現の言い換え(類似表現)を探す
=ある表現と同じ「型」の類義表現を探す
+その類義表現の言い換えを真似て言い換える



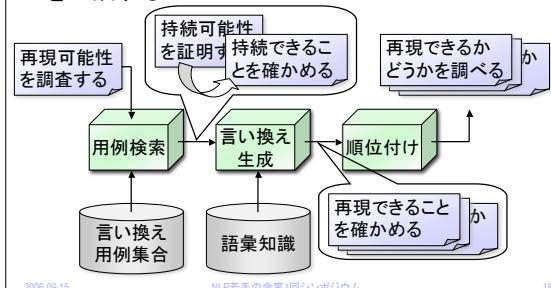
2006.09.15

用例に基づく言い換え生成で
多様性の確保と頑健な生成を目指す

17

用例に基づく言い換え生成

■ 人手で与えた言い換え用例に基づいて言い換えを生成する



実装状況

■ 評価実験に向けて

- 距離関数の定義についての考察
- 用例の増加に伴う精度向上の確認

■ 実装状況

- 類似する言い換え用例の検索
 - 共起する語のベクトル間の分布類似度
- 平行して語彙資源を構築中
 - 同義語辞書:名詞↔名詞、動詞↔動詞
 - 派生語:名詞↔動詞、動詞↔形容詞、名詞↔形容詞

2006.09.15

NLP若手の会第1回シンポジウム

19

今後の進め方

■ 知識獲得

- (動詞句よりもプリミティブな)語の関係の知識
 - 同義語、派生語、複合辞、慣用表現

■ 評価実験

- 用例検索に用いる距離関数についての考察
- 用例の増加に伴う精度向上の確認

■ 他の「型」を対象とした調査

- 「名詞+格助詞+動詞」
- 「名詞+格助詞+動詞+動詞」

2006.09.15

NLP若手の会第1回シンポジウム

20

