

# 用法基盤モデルに基づいた複合語形成の生産的パターンの抽出

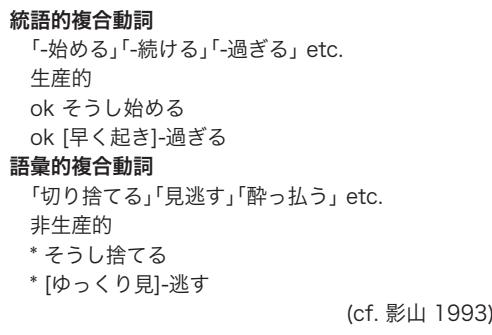
淺尾仁彦 (京都大学大学院)

## はじめに

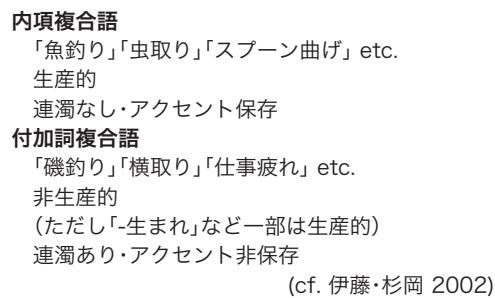
- ◆ 構文の集合としての形態論的知識
- ◆ 形態論における生産的パターンの明示的記述
- ◆ 生産的パターン学習の説明装置

## 先行研究

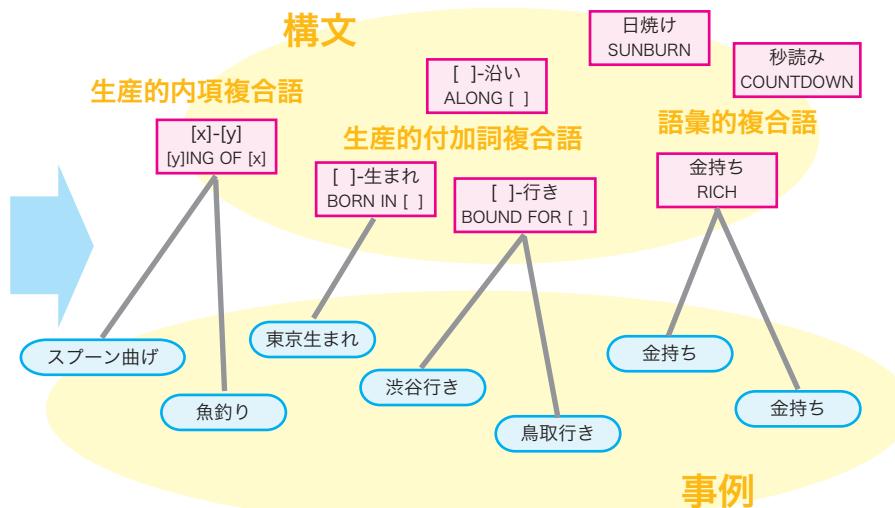
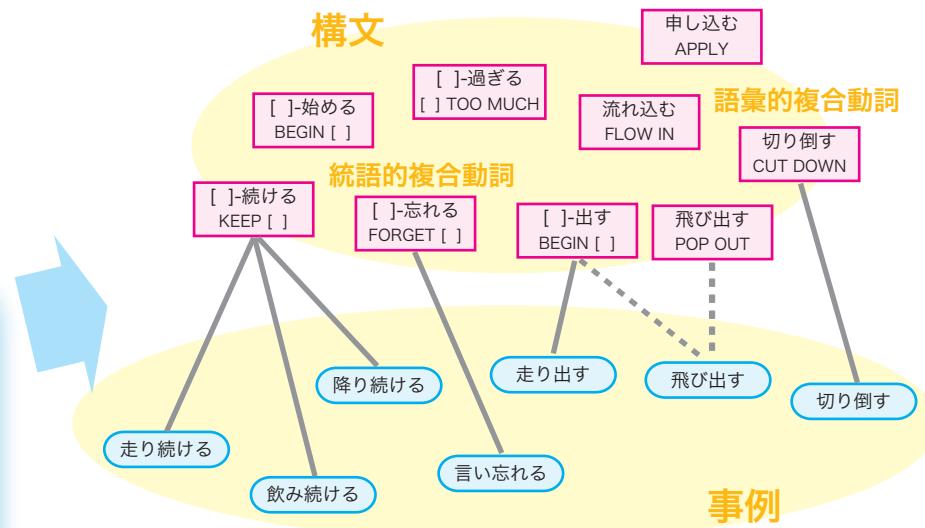
### 複合動詞



### 動詞由来複合語



## 構文による複合語の記述



## 構文の考え方

- ◆ 構文 = 意味と形式の慣習的結びつき  
(cf. Goldberg 1995, Booij 2006)
- ◆ 個々の言語経験は構文の実現事例
- ◆ 構文によって定着度は異なる
- ◆ 定着度の高い構文の実現事例は、理解・产出がしやすい
- ◆ あるいは、容認度が高い

## 構文の成立

- ◆ 構文は実際の事例に接することでボトムアップに成立する
- ◆ 最小一般化の原則: 構文はできるかぎり具体的に！  
(cf. Langacker 1990, 熊代 2003, Albright and Hayes 2002)
- ◆ 事例が増えれば増えるほど定着度が増す
- ◆ 先行研究の直観ベースの記述をシミュレーションによって再現できるか
- ◆ 構文の客観的認定
- ◆ さらに構文の学習の問題への示唆



## ■ シミュレーション 1

- ◆ 頻度の元データ:毎日新聞 95 年版
- ◆ MeCab + UniDic
- ◆ 複合動詞・動詞由来複合語 それぞれ 5 万語
- ◆ 各複合語は前項と後項のペアで表現される

### アルゴリズム

- モデルに複合語を一つずつ与える
- 既にその複合語を実現例とするような構文がある場合、その構文の定着度が 1 増加する
- その複合語を実現例とする構文がない場合、過去の複合語との共通性を抜き出した新しい構文が成立し、定着度初期値 1 が与えられる
- 上記 (ii), (iii) のプロセスでは、最も具体的な構文が優先される

- ◆ 統語的複合動詞の高い頻度直観に合う結果
- ◆ 語彙自体の区別以外も素性として取り込みたい

## ■ シミュレーション 2

- ◆ 頻度の元データ:Wikipedia 日本語版 (長谷部 2006)
- ◆ MeCab + UniDic
- ◆ 名詞・動詞・形容詞・形容動詞からなる複合語 10 万語
- ◆ 各語彙は「前項語彙、前項品詞、後項語彙、後項品詞、全品詞」の 5 つの素性で表現される

### 素性による表示

宝探し = {宝, 名詞, 探す, 動詞, 名詞}  
狭苦しい = {狭い, 形容詞, 苦しい, 形容詞, 形容詞}

- ◆ ある程度直観を反映する結果
- ◆ 統語的複合動詞と語彙の複合動詞を区別する能力などは不十分
- ◆ 結果を評価する手続きの必要性

## 文献

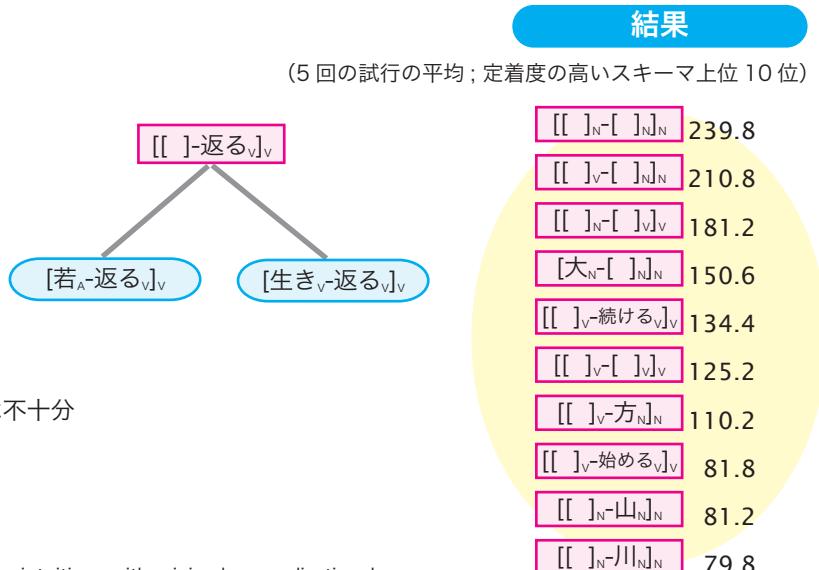
- Albright, Adam and Hayes, Bruce (2002). Modeling English past tense intuitions with minimal generalization. In Proceedings of the ACL-02 workshop on morphological and phonological learning, Vol. 6, pp.58-69. ACL.
- Booij, Geert (2005) Compounding and derivation: evidence for Construction Morphology. In Wolfgang U. Dressler, Franz Rainer, Dieter Kastovsky and Oskar Pfeiffer (eds.) Morphology and its demarcations. Amsterdam/Philadelphia: John Benjamins. 109-132.
- Goldberg, Adele E. (1995) Constructions: a construction grammar approach to argument structure. 長谷部陽一郎(2006) Wikipedia 日本語版をコーパスとして用いた言語研究の手法.『言語文化』9 (2), 373-403.
- 伊藤たかね・杉岡洋子 (2002)『語の仕組みと語形成』英語学モノグラフシリーズ 16 卷. 研究社.
- 影山太郎 (1993)『文法と語形成』ひつじ書房.
- 熊代文子 (2003) 認知音韻論. 吉村公宏(編)『認知音韻・形態論』pp.3-78. 大修館書店.
- Langacker, Ronald W. (1990) Concept, Image, and Symbol: the cognitive basis of grammar. Berlin/New York: Mouton de Gruyter.

## 結果

(5 回の試行の平均; 定着度の高いスキーマ上位 10 位)

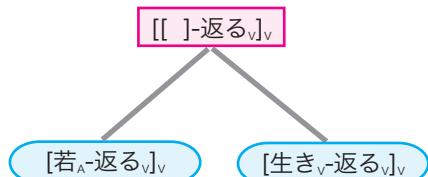


## 生産的付加詞複合語



## 結果

(5 回の試行の平均; 定着度の高いスキーマ上位 10 位)



## 謝辞

今回の研究にあたり、以下の方の  
ご助言・ご協力を得ました。どうも  
ありがとうございます。

- ・中川奈津子 (京都大学大学院)
- ・黒田航 (情報通信研究機構)