

# 格フレームと

## ノ型名詞句共起確率を

用いた

### 並列名詞句解析

東京工業大学  
吉橋 健治

## 背景 助詞トの曖昧性

助詞トの統語的機能

- 並列構造を作る
- 格要素になる

異質の構造

解析ミス

しかるべき 名詞と 地位を 与える

探照施設から 噴き出す 高度は ほぼ 同じ 150メートル。 <ト> PARA

欧州の 大砲と <ト> PARA 戦火を <ト> PARA 交えるまでに なっていた

人道支援の 医師と <ト> PARA 志に <ト> PARA 落ちる 人妻。

## 先行研究 格解析の研究は盛んだ。

用言と直前の格要素の組を単位とする格フレームの自動構築 河原(2001)  
 格フレーム辞書の漸次的自動構築 河原(2005)  
 大規模格フレームに基づく構文・格解析の統一的確率モデル 河原(2005)  
 日本語連体修飾節と被修飾名詞間の関係の解析 阿辺川(2005)  
 共起情報及び複数格の組み合わせを考慮した係り受け解析 阿辺川(2006)  
 動詞項構造辞書への大規模用例付与 平野(2006)

入力文

太郎が 食べた 弁当

格フレーム

食べる: 2  
 → ガ格: 人、学生、...  
 → ヲ格: 弁当、パン、...  
 → ニ格: 朝食、昼食、...  
 (対応なし)

## 先行研究 でも並列解析の研究は停滞している。

日本語技術文における並列構造 首藤(1986)

- 約1万文の日本語技術文を対象として並列構造を網羅的に抽出
- どのような型があるかを出現頻度もあわせて明らかにする

名詞句の並列の中で「NPとNP」は33.8% (一番多い)

長い日本語文における並列構造の推定 黒橋(1992)

文中の並列構造を類似した2つの文節列としてとらえる

- すべての文節対に対して、類似度を計算する
- 類似度の総和が最も大きい2つの文節列を並列構造の範囲とする

プログラミング言語は 2 2 0 2 2 2 0 0 2 0  
 問題解決の 2 0 2 4a 2 0 0 2 0  
 アルゴリズムを 2 2 4a 0 0 2 0  
 表現できる 0 0 0 2a 4a 0 2  
 a>記述力と 2 2 0 0 2a 0  
 計算機の 4 0 0 2 0  
 機能を 0 0 2 0  
 b>十分に 2a 0 2  
 駆動できる 0 2  
 枠組みが 0  
 必要である

## 着眼点 並列構造は、必ず類似しているのか？

道に落ちている百円玉は、いやでも眼につくが、錆びた折釘や路肩の雑草になると、無いも同然だ。  
 「箱男」安部公房

折れ釘、路肩、雑草の語彙的な類似性を考えるのはナンセンス

実際は類似性ではうまく解析できない例も数多く存在する

類似性に頼らない手法で並列解析の精度向上

ト格に関して、並列解析と格解析を統合してみる

## 新手法 格フレームに並列構造を取り入れる

その日付は 信玄と上杉謙信が激戦を交わした。 実行する勇氣と、退いて辛抱する勇氣とを兼ね備えた、 第4回の川中島の戦いの4カ月前だ。 真の勇氣ある者を目指してほしいのです。

従来の格フレーム

交わす ガ: 謙信 ヲ: 激戦

兼ね備える ガ: 者 ヲ: 勇氣

並列を取り入れた格フレーム

交わす <PARA>ガ: 信玄、謙信

兼ね備える <PARA>ヲ: 勇氣

### 新手法 名詞の共起情報に並列構造を取り入れる

点検と事故の因果関係

従来の共起データ抽出

事故: 因果関係

ひょっとして 点検: 因果関係 も?

並列を取り入れた共起データ抽出

<PARA>: 因果関係

事故: 因果関係

点検: 因果関係

### 新手法 解析step.1 助詞トがどこに結びつかを列挙する

信玄と上杉謙信が激戦を交わした川中島の戦いの4カ月前だ。

### 新手法 解析step.2 格フレーム辞書との対応付け

**並列を含む項構造**

以下の1~3の最大のものをスコアとする

- 1 用言(被修飾名詞)が<PARA>を格フレーム辞書(共起辞書)にもってれば、そのスコア
- 2 2と3は並列構造を展開して、それぞれの構造を対応付ける
- 3 並列を成す項の意味クラスが同じならスコアが大きくなるほう

格フレーム

<PARA>:ガ 交わす  
 激戦:ヲ

<PARA>:ノ 戦い  
 ヘルメット:ニ  
 身:ヲ 包む  
 防弾チョッキ:ニ  
 身:ヲ 包む

score(防弾チョッキ:ニ, 身:ヲ, 包む) > score(ヘルメット:ニ, 身:ヲ, 包む) > score(防弾チョッキ:ニ, 身:ヲ, 包む)

防弾チョッキ, ヘルメット → max(score(防弾チョッキ:ニ, 身:ヲ, 包む), score(ヘルメット:ニ, 身:ヲ, 包む))

### 新手法 解析step.2 格フレーム辞書との対応付け

**並列を含まない項構造**

ト格の名詞をあえてガ格スロットに入れてスコアが高ければそちらを採用

「~と(一緒に)~する」という用例は少ないから。

格フレーム

<人>:ガ ぶらぶらする  
 <場所>:ヲ

<人>:ト ぶらぶらする  
 <場所>:ヲ

<PARA>:ヲ ぶらぶらする

### 新手法 解析step.3 修飾語句の構造を列挙する

なんらかの制約条件やスコアで一つを選び出す 検討中

### 新手法 辞書構築とスコアの算出方法

- 1 生コーパスを構文解析する
- 2 助詞トを含む文を抽出する
- 3 助詞トの構造だけを人手で修正
- 4 用言 格要素構造//型名詞句を抽出する
- 5 用言 格要素構造のクラスタリング
- 6 入力構造と格フレーム//型名詞句共起対の対応付けのスコア

クラスタのスコア =  $\frac{\text{クラスタに含まれる用例数}}{\text{クラスタ中の用言を含む全用例数}}$  / 型名詞句 共起確率

$\max_{P \text{ クラスタ}} \text{類似度(入力構造, P)}$  語彙の類似度