

# 構文・述語項構造解析システム KNPの解析の流れと特徴

笹野 遼平 (東工大)

# KNPのイメージ？



# KNPのイメージ？

遅い

重い

古い



# KNPのイメージ？

遅い

重い

古い

ルールベースの係り受け解析器である



# KNPのイメージ？

遅い

重い

古い

ルールベースの係り受け解析器である

内部で何が行われているか分からない



# KNPのイメージ？

遅い

重い

古い

ルールベースの係り受け解析器である

内部で何が行われているか分からない

出力の読み方が分からない



# KNPのイメージ？

遅い

重い

古い

⇒ 係り受け解析器として見ると遅いし、重い、でも...

ルールベースの係り受け解析器である

内部で何が行われているか分からない

出力の読み方が分からない



# KNPのイメージ？

遅い

重い

古い

⇒ 係り受け解析器として見ると遅いし、重い、でも...

ルールベースの係り受け解析器である

⇒ 格解析、照応・省略解析なども行える**統合的解析器**

内部で何が行われているか分からない

出力の読み方が分からない





# KNPのイメージ？

遅い

重い

古い

⇒ 係り受け解析器として見ると遅いし、重い、でも...

ルールベースの係り受け解析器である

⇒ 格解析、照応・省略解析なども行える**統合的解析器**

内部で何が行われているか分からない

⇒ **ルール**と**統計情報**を用いて解析

出力の読み方が分からない



# KNPのイメージ？

遅い

重い

古い

⇒ 係り受け解析器として見ると遅いし、重い、でも...

ルールベースの係り受け解析器である

⇒ 格解析、照応・省略解析なども行える**統合的解析器**

内部で何が行われているか分からない

⇒ **ルール**と**統計情報**を用いて解析

出力の読み方が分からない

⇒ **最新版:KNP-4.1** **2013年12月公開** では  
分析用に可読性を重視した出力を追加



# 発表の背景と目的



## 背景

KNPが何をしているか知られていない

- 当初、ルールベースの構文解析器として公開されたが最新版では、格解析、照応・省略解析などを行うことが可能
- 構文解析のアルゴリズムなども当初とは大きく異なる
- 最新版では、複数の出力フォーマットが追加されている

# 発表の背景と目的



**背景** KNPが何をしているか知られていない

- 当初、ルールベースの構文解析器として公開されたが最新版では、格解析、照応・省略解析などを行うことが可能
- 構文解析のアルゴリズムなども当初とは大きく異なる
- 最新版では、複数の出力フォーマットが追加されている



KNPの解析の流れと特徴をまとめることで  
KNPに適したタスクを明らかにする

# もくじ

- 発表の背景と目的
- KNPが行う解析の概要
- KNPの出力フォーマット
- 述語項構造解析器としてのKNPの特徴
- まとめと今後の課題

# もくじ

- 発表の背景と目的
- **KNPが行う解析の概要**
  - 構文・格解析(デフォルト)
  - 照応解析(-anaphora)
- KNPの出力フォーマット
- 述語項構造解析器としてのKNPの特徴
- まとめと今後の課題

# 構文・格解析(デフォルト)の流れ

[河原ら'07]

1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む
2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

# 構文・格解析(デフォルト)の流れ

[河原ら'07]

1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む
2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

ルール + 統計情報



# 構文・格解析(デフォルト)の流れ

[河原ら'07]

1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む

– 係り先は活用や機能語等により絞り込み可能

e.g. 『用言の連体形』 ⇒ 係先は体言

『体言＋格助詞“を”』 ⇒ 係先は述語, etc.

# 構文・格解析(デフォルト)の流れ

[河原ら'07]

## 1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む

– 係り先は活用や機能語等により絞り込み可能

e.g. 『用言の連体形』 ⇒ 係先は体言

『体言＋格助詞“を”』 ⇒ 係先は述語, etc.

活躍した選手を  
監督が褒めた。

# 構文・格解析(デフォルト)の流れ

[河原ら'07]

## 1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む

– 係り先は活用や機能語等により絞り込み可能

e.g. 『用言の連体形』 ⇒ 係先は体言

『体言＋格助詞“を”』 ⇒ 係先は述語, etc.

係先は  
体言

活躍した選手を  
監督が褒めた。

# 構文・格解析(デフォルト)の流れ

[河原ら'07]

## 1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む

– 係り先は活用や機能語等により絞り込み可能

e.g. 『用言の連体形』 ⇒ 係先は体言

『体言＋格助詞“を”』 ⇒ 係先は述語, etc.

係先は  
体言

係先は  
述語

活躍した選手を  
監督が褒めた。

# 構文・格解析(デフォルト)の流れ

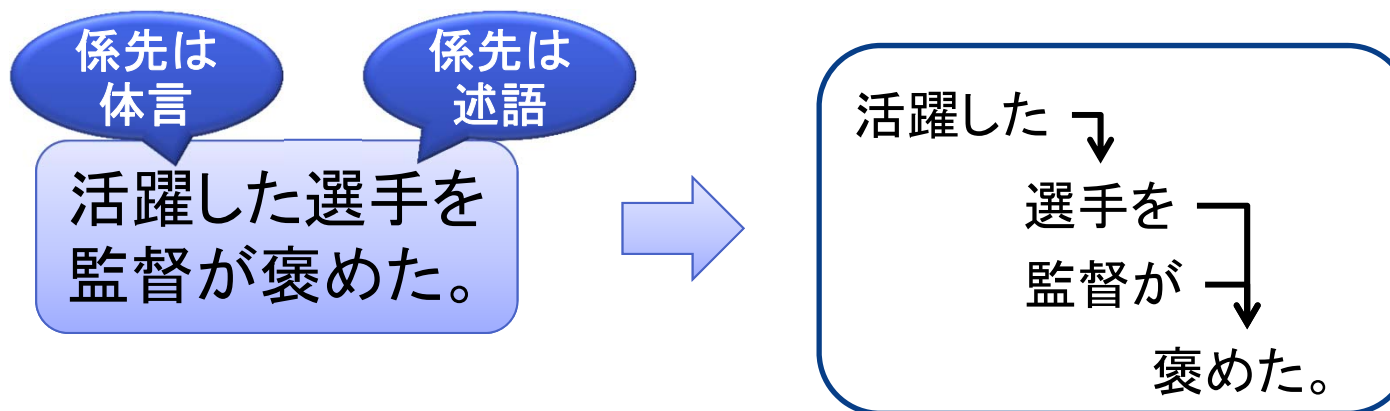
[河原ら'07]

## 1. 形態素の品詞・活用、機能語などの情報に基づき可能な構文・格構造候補を絞り込む

– 係り先は活用や機能語等により絞り込み可能

e.g. 『用言の連体形』 ⇒ 係先は体言

『体言＋格助詞“を”』 ⇒ 係先は述語, etc.

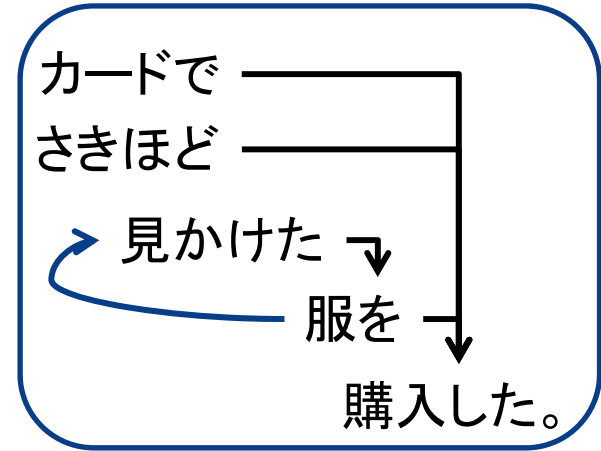
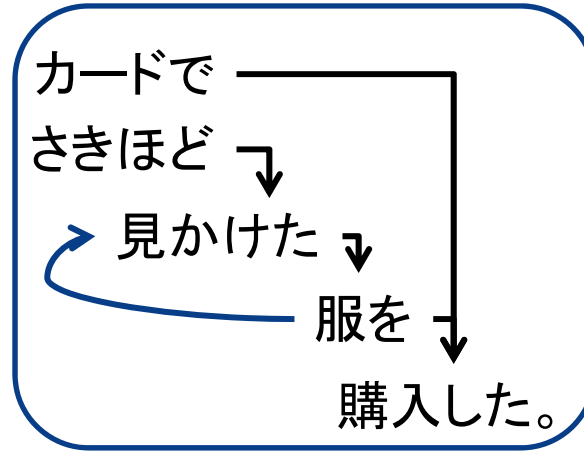
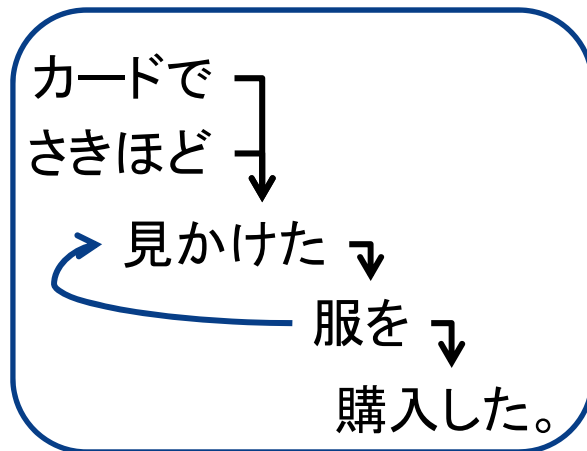


2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

カードでさきほど見かけた服を購入した。

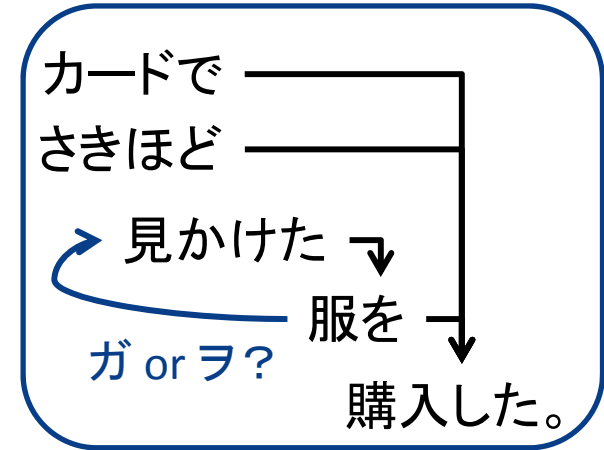
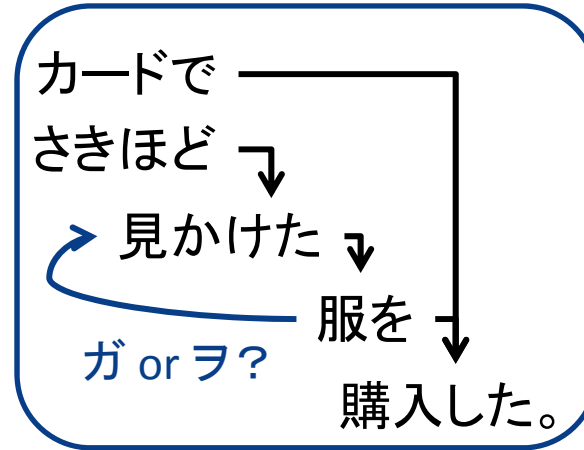
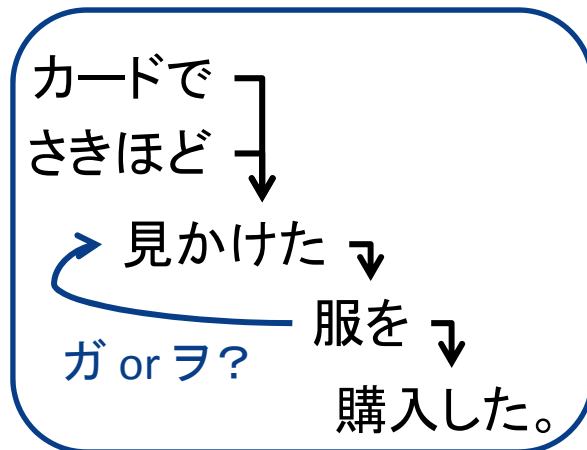
## 2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

カードでさきほど見かけた服を購入した。



## 2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

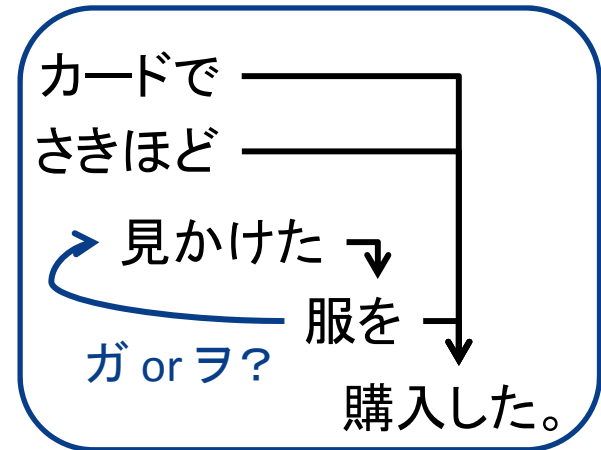
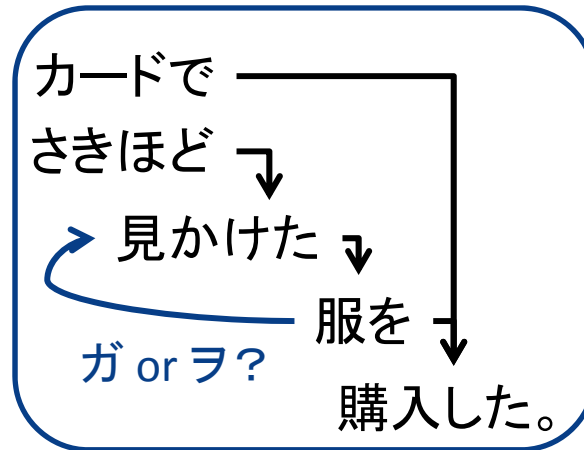
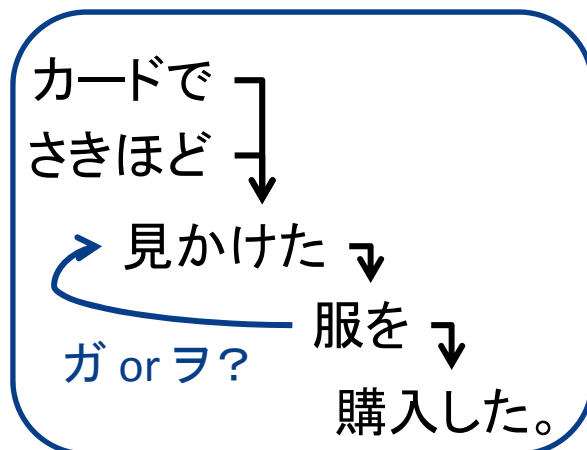
カードでさきほど見かけた服を購入した。





## 2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

カードでさきほど見かけた服を購入した。

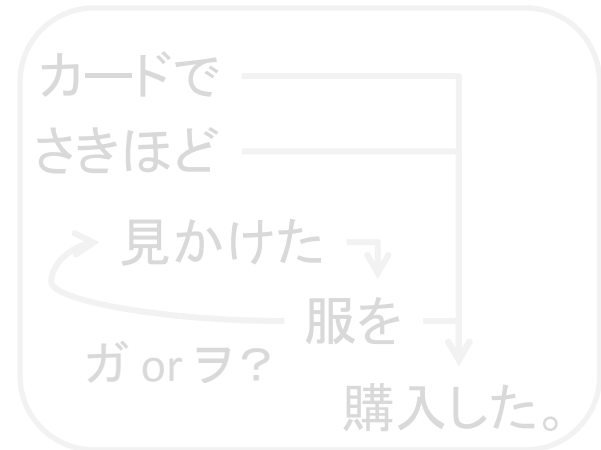
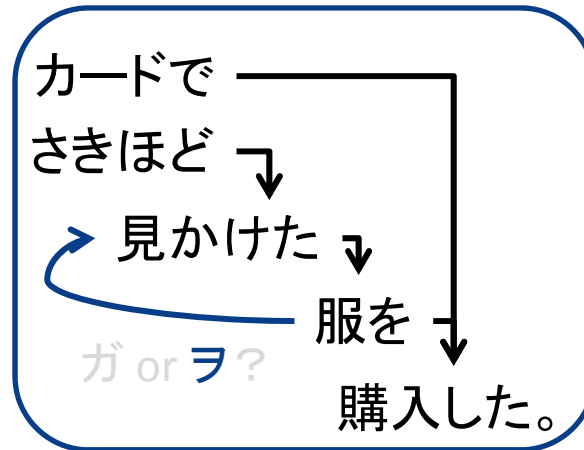
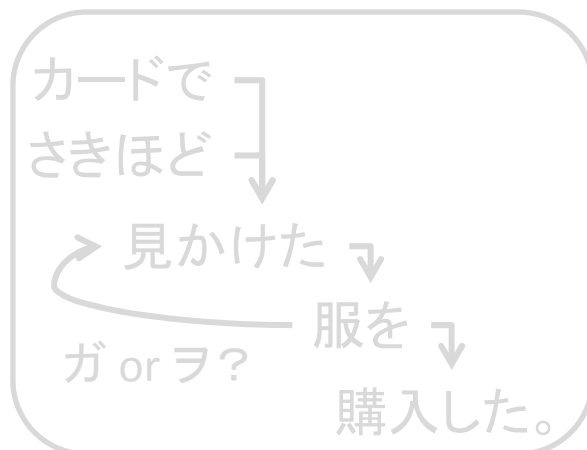


- 京大コーパスから得た統計情報
  - 時間表現 (e.g. さきほど) は近い用言に係りやすい
- 69億文から構築した格フレーム
  - カードで購入 >> カードで見かける
  - 服を見かける >> 服が見かける

生成確率  
を算出

## 2. 統計的情報に基づき各候補の生成確率を算出し、確率最大の構文・格構造を出力

カードでさきほど見かけた服を購入した。



- 京大コーパスから得た統計情報
  - 時間表現(e.g. さきほど)は近い用言に係りやすい
- 69億文から構築した格フレーム
  - カードで購入 >> カードで見かける
  - 服を見かける >> 服が見かける

生成確率  
を算出

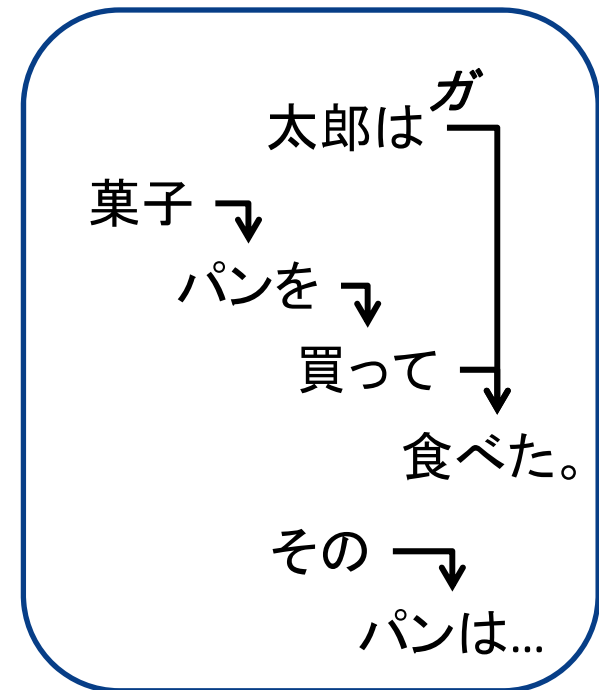
# 照応解析 (-anaphora)・固有表現解析 (-ne) を行う場合の解析の流れ

1. 構文・格解析(デフォルト)

2. CRF に基づく固有表現解析 (-ne)

3. ルールに基づく共参照解析 (-anaphora)

4. ゼロ照応解析も含む述語項構造解析 (-anaphora)  
by 大規模格フレームに基づく識別モデル [笹野ら'11]



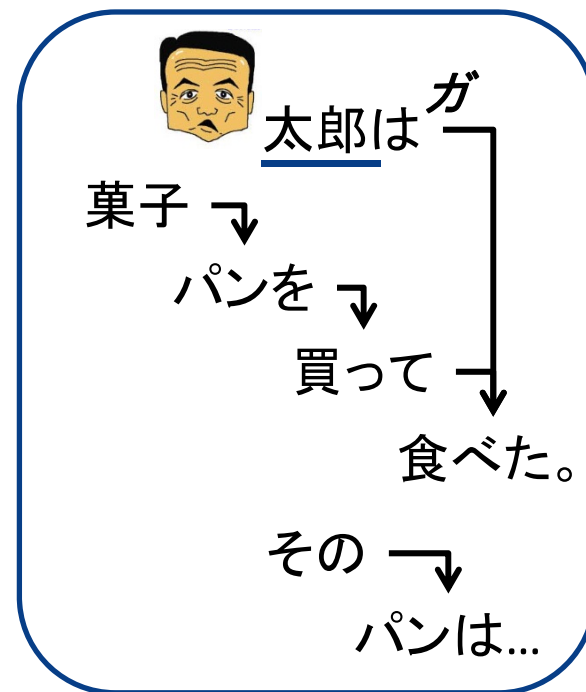
# 照応解析 (-anaphora)・固有表現解析 (-ne) を行う場合の解析の流れ

1. 構文・格解析(デフォルト)

2. CRF に基づく固有表現解析 (-ne)

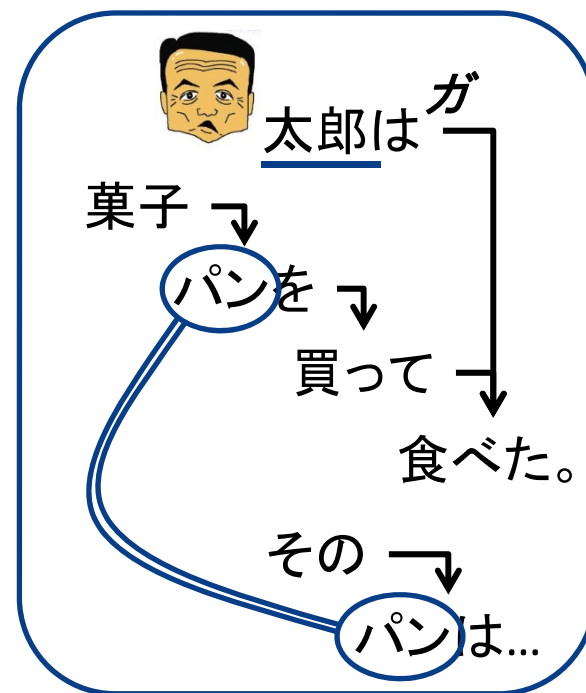
3. ルールに基づく共参照解析 (-anaphora)

4. ゼロ照応解析も含む述語項構造解析 (-anaphora)  
by 大規模格フレームに基づく識別モデル [笹野ら'11]



# 照応解析 (-anaphora)・固有表現解析 (-ne) を行う場合の解析の流れ

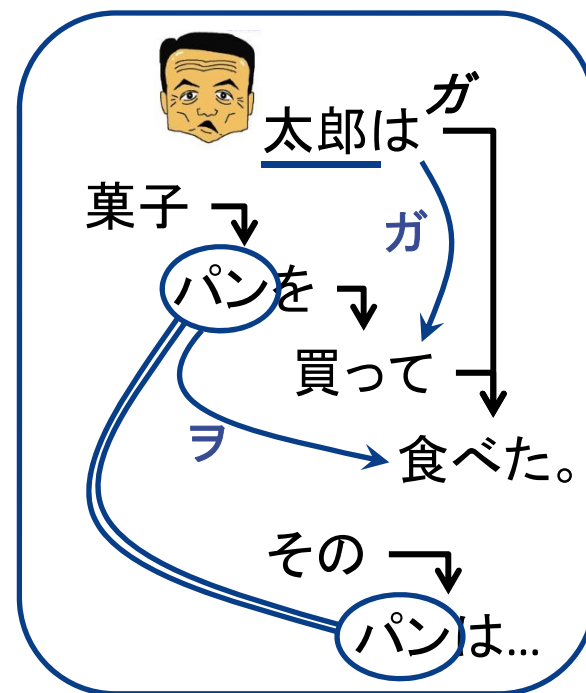
1. 構文・格解析(デフォルト)
2. CRF に基づく固有表現解析 (-ne)
3. ルールに基づく共参照解析 (-anaphora)



4. ゼロ照応解析も含む述語項構造解析 (-anaphora)  
by 大規模格フレームに基づく識別モデル [笹野ら'11]

# 照応解析 (-anaphora)・固有表現解析 (-ne) を行う場合の解析の流れ

1. 構文・格解析(デフォルト)
2. CRF に基づく固有表現解析 (-ne)
3. ルールに基づく共参照解析 (-anaphora)
4. ゼロ照応解析も含む述語項構造解析 (-anaphora)  
by 大規模格フレームに基づく識別モデル [笹野ら'11]



# 構文・格解析とゼロ照応解析の関係

手法	係り受け 関係	格解析 (非省略項)	格解析 (省略項)
構文・格解析 (デフォルト)[河原ら'07]	○	○	×
ゼロ照応解析 (-anaphora)[笹野ら'11]			

# 構文・格解析とゼロ照応解析の関係

手法	係り受け 関係	格解析 (非省略項)	格解析 (省略項)
構文・格解析 (デフォルト)[河原ら'07]	○	○	×
ゼロ照応解析 (-anaphora)[笹野ら'11]			○



# 構文・格解析とゼロ照応解析の関係

手法	係り受け 関係	格解析 (非省略項)	格解析 (省略項)
構文・格解析 (デフォルト)[河原ら'07]	○	○	×
ゼロ照応解析 (-anaphora)[笹野ら'11]	×		○

× 計算量の爆発を防ぐため係り受けはそのまま使用

# 構文・格解析とゼロ照応解析の関係

手法	係り受け 関係	格解析 (非省略項)	格解析 (省略項)
構文・格解析 (デフォルト)[河原ら'07]	○	○	×
ゼロ照応解析 (-anaphora)[笹野ら'11]	×	◎	○
	再解析しない	再解析	

- × 計算量の爆発を防ぐため係り受けはそのまま使用
- ◎ 格解析結果が改善する場合があるため再解析

# 構文・格解析とゼロ照応解析の関係

手法	係り受け 関係	格解析 (非省略項)	格解析 (省略項)
構文・格解析 (デフォルト)[河原ら'07]	○	○	×
ゼロ照応解析 (-anaphora)[笹野ら'11]	×	◎	○
	再解析しない	再解析	

× 計算量の爆発を防ぐため係り受けはそのまま使用

◎ 格解析結果が改善する必要があるため再解析

e.g. 太郎が戻るとさきほど見かけた女の子はいなくなっていた。

- 省略項を考慮しない場合: 『女の子ガ 見かけた』

- 省略項を考慮した場合: 『太郎ガ 女の子ヲ 見かけた』

# もくじ

- 発表の背景と目的
- KNPが行う解析の概要
- **KNPの出力フォーマット**
- 述語項構造解析器としてのKNPの特徴
- まとめと今後の課題

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)

女の子を ↘  
見かけた。

## 2. -tab: 詳細出力

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><正規化代表表記:女の子/おんなのこ><主辞代表表記:女の子/おんなのこ>

+ 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><正規化代表表記:女の子/おんなのこ><照応詞候補:女の子><解析格:ヲ><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人 ドメイン:家庭・暮らし" <代表表記:女の子/おんなのこ><カテゴリ:人><ドメイン:家庭・暮らし><正規化代表表記:女の子/おんなのこ><文頭><かな 漢字><名詞相当語><自立><内容語><タグ単位始><文節始><文節主辞><係:ヲ格>

ををを 助詞 9 格助詞 1 \* 0 \* 0 NIL <かな漢字><ひらがな><付属>

\* -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><主辞代表表記:見掛ける/みかける>

+ -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><用言代表表記:見掛ける/みかける><主題格:一人称優位><格関係0:ヲ:女の子><格解析結果:見掛ける/みかける:動1:ガ/U/-/-/-/-;ヲ/C/女の子/0/0/1;ニ/U/-/-/-/-;ト/U/-/-/-/-;デ/U/-/-/-/-;カラ/U/-/-/-/-;ヨリ/U/-/-/-/-;マデ/U/-/-/-/-;時間/U/-/-/-/-;外の関係/U/-/-/-/-;修飾/U/-/-/-/-;ノ/U/-/-/-/-;トスル/U/-/-/-/-;ツウジル/U/-/-/-/-;ニカギル又/U/-/-/-/-;ニオル/U/-/-/-/-;ニカギル/U/-/-/-/-;ニトル/U/-/-/-/-;ヲフクメル/U/-/-/-/-;ニツク/U/-/-/-/-><EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 タ形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

。。。特殊 1 句点 1 \* 0 \* 0 NIL <文末><英記号><記号><付属>

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

女の子を  
↓  
見かけた。

## 2. -tab: 詳細出力

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><正規化代表表記:女の子/おんなのこ><主辞代表表記:女の子/おんなのこ>

+ 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><正規化代表表記:女の子/おんなのこ><照応詞候補:女の子><解析格:ヲ><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人 ドメイン:家庭・暮らし" <代表表記:女の子/おんなのこ><カテゴリ:人><ドメイン:家庭・暮らし><正規化代表表記:女の子/おんなのこ><文頭><かな 漢字><名詞相当語><自立><内容語><タグ単位始><文節始><文節主辞><係:ヲ格>

ををを 助詞 9 格助詞 1 \* 0 \* 0 NIL <かな漢字><ひらがな><付属>

\* -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><主辞代表表記:見掛ける/みかける>

+ -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><用言代表表記:見掛ける/みかける><主題格:一人称優位><格関係0:ヲ:女の子><格解析結果:見掛ける/みかける:動1:ガ/U/-/-/-;ヲ/C/女の子/0/0/1;ニ/U/-/-/-;ト/U/-/-/-;デ/U/-/-/-;カラ/U/-/-/-;ヨリ/U/-/-/-;マデ/U/-/-/-;時間/U/-/-/-;外の関係/U/-/-/-;修飾/U/-/-/-;ノ/U/-/-/-;トスル/U/-/-/-;ツウジル/U/-/-/-;ニカギル又/U/-/-/-;ニオル/U/-/-/-;ニカギル/U/-/-/-;ニトル/U/-/-/-;ヲフクメル/U/-/-/-;ニツク/U/-/-/-><EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 タ形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

。。。特殊 1 句点 1 \* 0 \* 0 NIL <文末><英記号><記号><付属>

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

女の子を →  
見かけた。

## 2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><正規化代表表記:女の子/おんなのこ><主辞代表表記:女の子/おんなのこ>

+ 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><正規化代表表記:女の子/おんなのこ><照応詞候補:女の子><解析格:ヲ><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人 ドメイン:家庭・暮らし" <代表表記:女の子/おんなのこ><カテゴリ:人><ドメイン:家庭・暮らし><正規化代表表記:女の子/おんなのこ><文頭><かな 漢字><名詞相当語><自立><内容語><タグ単位始><文節始><文節主辞><係:ヲ格>

ををを 助詞 9 格助詞 1 \* 0 \* 0 NIL <かな漢字><ひらがな><付属>

\* -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><主辞代表表記:見掛ける/みかける>

+ -1D <文末><時制-過去><句点><用言:動><レベル:C><区切:5-5><ID:(文末)><係:文末><提題受:30><主節><格要素><連用要素><動態述語><正規化代表表記:見掛ける/みかける><用言代表表記:見掛ける/みかける><主題格:一人称優位><格関係0:ヲ:女の子><格解析結果:見掛ける/みかける:動1:ガ/U/-/-/-;ヲ/C/女の子/0/0/1;ニ/U/-/-/-;ト/U/-/-/-;デ/U/-/-/-;カラ/U/-/-/-;ヨリ/U/-/-/-;マデ/U/-/-/-;時間/U/-/-/-;外の関係/U/-/-/-;修飾/U/-/-/-;ノ/U/-/-/-;トスル/U/-/-/-;ツウジル/U/-/-/-;ニカギル又/U/-/-/-;ニオル/U/-/-/-;ニカギル/U/-/-/-;ニトル/U/-/-/-;ヲフクメル/U/-/-/-;ニツク/U/-/-/-><EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 タ形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

。。。特殊 1 句点 1 \* 0 \* 0 NIL <文末><英記号><記号><付属>

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

女の子を →  
見かけた。

## 2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><正規化代表表記:女の子/おんなのこ><主辞代表表記:女の子/おんなのこ>

+ 1D <文頭><SM-主体><SM-人><ヲ><助詞><体言><係:ヲ格><区切:0-0><格要素><連用要素><名詞項候補><先行詞候補><正規化代表表記:女の子/おんなのこ><照応詞候補:女の子><解析格:ヲ><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人 ドメイン:家庭・暮らし" 代表表記:女の子/おんなのこ <カテゴリ:人><ドメイン:家庭・暮らし> 内容語><タグ単位始

><文節始><文節主辞><係:ヲ格><助詞:ヲ><体言:名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人 ドメイン:家庭・暮らし" 代表表記:女の子/おんなのこ <カテゴリ:人><ドメイン:家庭・暮らし> 内容語><タグ単位始

ををを 助詞 9 格助詞 1 \* 0 \* 0 "代表表記:を/を カテゴリ:助詞" 代表表記:を/を <カテゴリ:助詞> 内容語><タグ単位始

\* -1D <文末><時制-過去><動詞 2 \* 0 母音動詞 1 夕形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

正規化代表表記:見掛ける/みかける 動詞 2 \* 0 母音動詞 1 夕形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

みかける:動1:ガ/U/-/-/-/-/;カ/U/-/-/-/-/;キ/U/-/-/-/-/;ク/U/-/-/-/-/;ケ/U/-/-/-/-/;コ/U/-/-/-/-/;カギル/U/-/-/-/-/;ニカギル/U/-/-/-/-/;ニトル/U/-/-/-/-/;ヲフクメル/U/-/-/-/-/;ニツク/U/-/-/-/-/;<EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 夕形 10 "代表表記:見掛ける/みかける" <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><表現文末><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞>

。。。特殊 1 句点 1 \* 0 \* 0 NIL <文末><英記号><記号><付属>

解析結果を手で分析することを  
想定した出力オプションを追加



# KNPの代表的な出力オプション

1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

## 2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

## 3. -simple: tab出力の簡易版 **New!**

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <体言><係:ヲ格>

+ 1D <体言><係:ヲ格><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人ドメイン:家庭・暮らし" <係:ヲ格>

ををを 助詞 9 格助詞 1 \* 0 \* 0 NIL

\* -1D <用言:動><係:文末>

+ -1D <用言:動><係:文末><EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 夕形 10 "代表表記:見掛ける/みかける"

。。。 特殊 1 句点 1 \* 0 \* 0 NIL

EOS

# KNPの代表的な出力オプション

## 1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

## 2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

## 3. -simple: tab出力の簡易版 **New!**



それなりに見やすく  
解析詳細も少し分かる

# S-ID:1 KNP:4.1-CF1.1 DATE:2013/03/12 SCORE:-5.16440

\* 1D <体言><係:ヲ格>

+ 1D <体言><係:ヲ格><EID:0><述語項構造:女の子/おんなのこ:名1>

女の子 おんなのこ 女の子 名詞 6 普通名詞 1 \* 0 \* 0 "代表表記:女の子/おんなのこ カテゴリ:人ドメイン:家庭・暮らし" <係:ヲ格>

ををを 助詞 9 格助詞 1 \* 0 \* 0 NIL

\* -1D <用言:動><係:文末>

+ -1D <用言:動><係:文末><EID:1><述語項構造:見掛ける/みかける:動1:ヲ/C/女の子/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 タ形 10 "代表表記:見掛ける/みかける"

。。。 特殊 1 句点 1 \* 0 \* 0 NIL

EOS

# KNPの代表的な出力オプション

1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

3. -simple: tab出力の簡易版 **New!**



それなりに見やすく  
解析詳細も少し分かる

# KNPの代表的な出力オプション

1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

3. -simple: tab出力の簡易版 **New!**



それなりに見やすく  
解析詳細も少し分かる

4. -td: 構文木 + 詳細出力 **New!**

# KNPの代表的な出力オプション

1. -tree: 構文木出力(デフォルト)



見やすい



詳細は分からない

2. -tab: 詳細出力



詳細は分かる



もの凄く見づらい

3. -simple: tab出力の簡易版 **New!**



それなりに見やすく  
解析詳細も少し分かる

4. -td: 構文木 + 詳細出力 **New!**



見やすく  
詳細も分かる

# テーブル表示ツール用出力(-td)

解析結果汎用表示ツール x

← → ↻ nlp.ist.i.kyoto-u.ac.jp/nl-resource/tabledisplay/index.cgi ☆ ☰

## テーブル表示ツール

表示するファイル:  選択されていません

エンコード:  UTF-8  EUC-JP

---

### 解析結果汎用表示ツールTableDisplay

**概要**

近年のNLP技術の目覚ましい発展や、マシンパワーの向上、さらにはWebなどに代表されるような大規模リソースへの容易なアクセスが可能となったことにより、研究対象がより複雑化、大規模化している。必然的に、それらを正確に扱う処理(プログラム)も複雑化、大規模化している。またシステムの複雑化により、実験結果を議論する際の手間が無視できないくらいに増えてきている。そこでこのような理由から、作業や研究の効率化のために、システムの解析結果を視覚的・直感的にわかりやすい形で表示できるようなツールの必要性を感じ、その構築を試みた。ツールは誰もが容易にアクセスでき、多人数での議論においても有用なものにするため、CGIで実現し、Webブラウザでアクセスすることを考えた。Webブラウザはどんな種類のマシンにも搭載されており、プラットフォームを気にする必要がないため、十分な汎用性・簡便性を確保できる。また、HTMLの機能を用いることにより、視覚的にもわかりやすい表示を実現可能である。

TableDisplayをダウンロード (187,530 bytes)

TableDisplayを試してみる → 上のフォームに表示したいファイルを指定してください。  
なおこのページもTableDisplayにより作成されています。

# テーブル表示ツール用出力 (-td)

解析結果汎用表示ツール x

nlp.ist.i.kyoto-u.ac.jp/nl-resource/tabledisplay/index.cgi

## テーブル表示ツール

表示するファイル:  選択されていません

エンコード:  UTF-8  EUC-JP

### 解析結果汎用表示ツールTableDisplay

近頃の世は技術の目覚ましい発展が、コンピュータの向上と共に世に知られだに代わられるようになった。

されており、プラットフォームを気にする必要がないため、十分な汎用性・簡便性を確保できる。また、HTMLの機能を用いることにより、視覚的にもわかりやすい表示を実現可能である。

[TableDisplayをダウンロード \(187,530 bytes\)](#)

TableDisplayを試してみる → 上のフォームに表示したいファイルを指定してください。  
なおこのページもTableDisplayにより作成されています。

```
$ echo "太郎が戻るとさきほど見かけた女の子はいなくなっていた。"  
| juman | knp -anaphora -ne -td > result.td
```



# テーブル表示ツール用出力(-td)

# S-ID:1 SCORE:-26.58452	EID	解析結果
太郎が――	0	PERSON:太郎
戻ると――	1	*[ガ:太郎]
さきほど――	2	
見かけた――	3	*ガ:太郎 [時間:さきほど] [ヲ:女の子]
女の子は――	4	*
いなくなっていた。	5	*[ガ:女の子]

# テーブル表示ツール用出力(-td)

# S-ID:1 SCORE:-26.58452	EID	解析結果
太郎が――	0	PERSON: 太郎
戻ると――	1	*[ガ: 太郎]
さきほど――	2	
見かけた――	3	*ガ: 太郎 [時間: さきほど] [ヲ: 女の子]
女の子は――	4	*

\* 4D <時制-過去><連体修飾><用言:動><係:連格><レベル:B><区切:0-5><ID:(動詞連体)><連体節><動態述語><正規化代表表記:見掛ける/みかける><主辞代表表記:見掛ける/みかける>

+ 4D <時制-過去><連体修飾><用言:動><係:連格><レベル:B><区切:0-5><ID:(動詞連体)><連体節><動態述語><正規化代表表記:見掛ける/みかける><用言代表表記:見掛ける/みかける><格関係2:時間:さきほど><格関係4:ガ:女の子><格解析結果:見掛ける/みかける:動7:ガ/N/女の子/4/0/1;ヲ/U/-/-/-/-;ニ/U/-/-/-/-;ト/U/-/-/-/-;デ/U/-/-/-/-;カラ/U/-/-/-/-;時間/C/さきほど/2/0/1;外の関係/U/-/-/-/-;修飾/U/-/-/-/-;ノ/U/-/-/-/-><EID:3><述語項構造:見掛ける/みかける:動1:ヲ/N/女の子/4;時間/C/さきほど/2;ガ/O/太郎/0>

見かけた みかけた 見かける 動詞 2 \* 0 母音動詞 1 夕形 10 代表表記:見掛ける/みかける <代表表記:見掛ける/みかける><正規化代表表記:見掛ける/みかける><かな漢字><活用語><自立><内容語><タグ単位始><文節始><文節主辞><係:連格>

# テーブル表示ツール用出力 (-td)

# S-ID:1 SCORE:-26.58452	EID	解析結果
太郎が――	0	PERSON:太郎
戻ると――	1	*[ガ:太郎]
さきほど――	2	
見かけた――	3	*ガ:太郎 [時間:さきほど] [ヲ:女の子]
女の子は――	4	*
いなくなっていた。	5	*[ガ:女の子]

# テーブル表示ツール用出力(-td)

# S-ID:1 SCORE:-26.58452	EID	解析結果
太郎が――	0	PERSON:太郎
戻ると――	1	*[ガ:太郎]
さきほど――	2	
見かけた――	3	*ガ:太郎 [時間:さきほど] [ヲ:女の子]
女の子は――	4	*

**Saliency score ranking**  
**4:女の子:2.000, 0:太郎:1.000, 2:さきほど:0.300**

**Predicate argument structure ranking**

**0:** -19.455 【見掛ける/みかける:動1】 ヲ:4:女の子 時間:2:さきほど ガ:0:太郎 ニ:×

**1:** -19.499 【見掛ける/みかける:動7】 ガ:4:女の子 時間:2:さきほど ヲ:× ニ:-

**2:** -20.641 【見掛ける/みかける:動1】 ヲ:4:女の子 時間:2:さきほど ガ:× ニ:×

**3:** -23.046 【見掛ける/みかける:動7】 ガ:4:女の子 時間:2:さきほど ヲ:0:太郎 ニ:-

**4:** -23.827 【見掛ける/みかける:動1】 ガ:4:女の子 時間:2:さきほど ヲ:× ニ:×

**5:** -25.081 【見掛ける/みかける:動3】 ガ:4:女の子 時間:2:さきほど ヲ:× ニ:×

**6:** -25.425 【見掛ける/みかける:動1】 ガ:4:女の子 時間:2:さきほど ヲ:0:太郎 ニ:×

**7:** -26.509 【見掛ける/みかける:動3】 ガ:4:女の子 時間:2:さきほど ヲ:0:太郎 ニ:×

**8:** -29.211 【見掛ける/みかける:動1】 ヲ:4:女の子 時間:2:さきほど ニ:0:太郎 ガ:×

**9:** -29.673 【見掛ける/みかける:動7】 ガ:4:女の子 時間:2:さきほど ニ:0:太郎 ヲ:×

# もくじ

- 発表の背景と目的
- KNPが行う解析の概要
- KNPの出力フォーマット
- **述語項構造解析器としてのKNPの特徴**
- まとめと今後の課題

# CaboCha + SynCha との主な違い

1. KNPは文全体を見て最適な構文・格構造を決定
  - ガーデンパス文等、**遠距離**、かつ、**語彙情報を考慮**する必要がある係り受けに対しても頑健

# S-ID:1 SCORE:-65.59176
太郎が—
電池を—
売った—
ラジオに—
おまけで—
つけた。

CaboCha : 係り受け解析器 <https://code.google.com/p/cabochoa/> [工藤ら]

SynCha : 述語項構造解析器 <https://www.cl.cs.titech.ac.jp/~ryu-i/syncha/> [飯田]

# CaboCha + SynCha との主な違い

1. KNPは文全体を見て最適な構文・格構造を決定
  - ガーデンパス文等、**遠距離**、かつ、**語彙情報を考慮**する必要がある係り受けに対しても頑健

# S-ID:1 SCORE:-65.59176
太郎が—
電池を—
売った—
ラジオに—
おまけで—
つけた。

- CaboChaはガーデンパス文の解析が苦手[浅原'11]
- この場合も「太郎」が「売った」に係ると解析してしまう

CaboCha : 係り受け解析器 <https://code.google.com/p/cabochoa/> [工藤ら]

SynCha : 述語項構造解析器 <https://www.cl.cs.titech.ac.jp/~ryu-i/syncha/> [飯田]

# CaboCha + SynCha との主な違い

1. KNPは文全体を見て最適な構文・格構造を決定
  - ガーデンパス文等、**遠距離**、かつ、**語彙情報を考慮**する必要がある係り受けに対しても頑健
2. KNPは格フレームに含まれる**全ての格**が解析対象
  - 双主構文『象は鼻が長い』の主題格も解析できる(ガ2格)
  - ただしゼロ照応解析の対象はガ、ヲ、ニ格のみ

CaboCha : 係り受け解析器 <https://code.google.com/p/cabochoa/> [工藤ら]

SynCha : 述語項構造解析器 <https://www.cl.cs.titech.ac.jp/~ryu-i/syncha/> [飯田]



# CaboCha + SynCha との主な違い

1. KNPは文全体を見て最適な構文・格構造を決定
  - ガーデンパス文等、**遠距離**、かつ、**語彙情報を考慮**する必要がある係り受けに対しても頑健
2. KNPは格フレームに含まれる**全ての格**が解析対象
  - 双主構文『象は鼻が長い』の主題格も解析できる(ガ2格)
  - ただしゼロ照応解析の対象はガ、ヲ、ニ格のみ
3. 受身文、使役文の場合、**出現格**を解析
  - SynCha は原形格を出力(cf. NAISTテキストコーパス)

CaboCha : 係り受け解析器 <https://code.google.com/p/cabochoa/> [工藤ら]

SynCha : 述語項構造解析器 <https://www.cl.cs.titech.ac.jp/~ryu-i/syncha/> [飯田]

# KNPは遅いか？ -解析時間の比較-

対象	新聞記事	Web
文書数	696	500
文数	9,284	15,415
文字数	404,007	397,047
CaboCha 0.66	5.16	4.57
SynCha 0.3 [n=2]	2,984	1,602
SynCha 0.3 [n=∞]	9,031	17,161
KNP 4.1 [談話解析なし]	5,905	5,405
KNP 4.1 [談話解析あり]	7,595	6,889

# KNPは遅いか？ -解析時間の比較-

対象	新聞記事	Web
文書数	696	500
文数	9,284	15,415
文字数	404,007	397,047
CaboCha 0.66	5.16	4.57
SynCha 0.3 [n=2]	2,984	1,602
SynCha 0.3 [n=∞]	9,031	17,161
KNP 4.1 [談話解析なし]	5,905	5,405
KNP 4.1 [談話解析あり]	7,595	6,889

KNPの解析速度は  
1秒あたり2~3文

# KNPは遅いか？ -解析時間の比較-

対象	新聞記事	Web
文書数	696	500
文数	9,284	15,415
文字数	404,007	397,047
CaboCha 0.66	5.16	4.57
SynCha 0.3 [n=2]	2,984	1,602
SynCha 0.3 [n=∞]	9,031	17,161
KNP 4.1 [談話解析なし]	5,905	5,405
KNP 4.1 [談話解析あり]	7,595	6,889

述語項構造解析器であるSynChaとは速度面で大きな違いはない



KNPの解析速度は1秒あたり2~3文

# KNPは遅いか？ -解析時間の比較-

対象	記事	Web
文書数	696	500
文数	284	15,415
文字数	404,007	397,047
CaboCha 0.66	5.16	4.57
SynCha 0.3 [n=2]	2,984	1,602
SynCha 0.3 [n=∞]	9,031	17,161
KNP 4.1 [談話解析なし]	5,905	5,405
KNP 4.1 [談話解析あり]	7,595	6,889

係り受け解析器であるCaboChaと比較すると圧倒的に遅い



述語項構造解析器であるSynChaとは速度面で大きな違いはない



KNPの解析速度は1秒あたり2~3文

# ゼロ照応解析の精度比較 (F値)

	ガ			ヲ			ニ		
	文内	文間	合計	文内	文間	合計	文内	文間	合計
Taira ら [4]	30.2	23.5	-	11.4	9.3	-	3.7	11.8	-
Imamura ら [3]	50.0	13.1	38.6	30.8	0.7	24.5	0.0	0.0	0.0
Iida ら [2]	-	-	34.6	-	-	-	-	-	-
Yoshikawa ら [5]	54.1	-	-	10.3	-	-	0.0	-	-
KNP 4.1	41.6	26.3	35.7	22.6	13.8	20.1	8.8	1.0	5.9

- NAISTテキストコーパスを用いて実験
  - 実験条件が完全に同じでないため単純な比較はできないものの、関連研究と同程度の精度
  - 文間のガ格・ヲ格ゼロ代名詞の解析精度が高い傾向

# ゼロ照応解析の精度比較(F値)

	ガ			ヲ			ニ		
	文内	文間	合計	文内	文間	合計	文内	文間	合計
Taira ら [4]	30.2	23.5	-	11.4	9.3	-	3.7	11.8	-
Imamura ら [3]	50.0	13.1	38.6	30.8	0.7	24.5	0.0	0.0	0.0
Iida ら [2]	-	-	34.6	-	-	-	-	-	-
Yoshikawa ら [5]	54.1	-	-	10.3	-	-	0.0	-	-
KNP 4.1	41.6	26.3	35.7	22.6	13.8	20.1	8.8	1.0	5.9

- NAISTテキストコーパスを用いて実験
  - 実験条件が完全に同じでないため単純な比較はできないものの、関連研究と同程度の精度
  - 文間のガ格・ヲ格ゼロ代名詞の解析精度が高い傾向

# ゼロ照応解析の精度比較(F値)

	ガ			ヲ			ニ		
	文内	文間	合計	文内	文間	合計	文内	文間	合計
Taira ら [4]	30.2	23.5	-	11.4	9.3	-	3.7	11.8	-
Imamura ら [3]	50.0	13.1	38.6	30.8	0.7	24.5	0.0	0.0	0.0
Iida ら [2]	-	-	34.6	-	-	-	-	-	-
Yoshikawa ら [5]	54.1	-	-	10.3	-	-	0.0	-	-
KNP 4.1	41.6	26.3	35.7	22.6	13.8	20.1	8.8	1.0	5.9

- NAISTテキストコーパスを用いて実験
  - 実験条件が完全に同じでないため単純な比較はできないものの、関連研究と同程度の精度
  - 文間のガ格・ヲ格ゼロ代名詞の解析精度が高い傾向

## 【参考】係り受けの精度比較

- CoNLL2009テストセット: MSTParser(2次) 92.44, KNP 93.14
- Web759文でのテスト: MSTParser(2次) 92.83, KNP 94.01



# まとめと今後の課題

- KNPとは

- 😊 KNPは格解析や照応・省略解析まで行う**統合的な解析器**
- 😊 **大規模格フレーム** (69億文から構築) を使用しており、遠距離、かつ、語彙情報知識が必要な係り受けも正しく扱える
- 😊 ただし、速度は**係り受け解析器と比べ遅い**ため、大規模な係り受け解析結果がほしい場合には不向き

- 今後の課題

- 原形格構造の出力に対応 (受動 ⇔ 能動変換)
- ゼロ照応解析の精度向上
- 高速化、特に格フレーム検索の高速化