

自然言語処理論 I

5. 文法3(素性構造と単一化)

1

実用的な構文解析

- 今までの例に挙げた文法は非常に単純
- 実用的な文法
 - いろいろな文に対応しなければならない
 - それだけ規則の数も増える
 - 規則を効率的に管理する必要がある

2

一致の例

- 英語における一致(agreement)
 - 数(単数形, 複数形)
 - 人称(1人称, 2人称, 3人称)
- 名詞句の例
 - ○ a desk
 - ○ the desks
 - × a desks
 - NP → det n という規則では、これらは全て許される

3

一致を正しく扱うには

- 文法の記号を細分化する
$$\text{NP}_{\text{singular}} \rightarrow \text{det}_{\text{singular}} \text{ n}_{\text{singular}}$$
$$\text{NP}_{\text{plural}} \rightarrow \text{det}_{\text{plural}} \text{ n}_{\text{plural}}$$
 - 問題点
 - 規則の数が組み合わせ的に増大する
 - 人称の一致も考慮すると...

$\text{n}_{\text{singular_1st}}$	$\text{n}_{\text{singular_2nd}}$	$\text{n}_{\text{singular_3rd}}$
$\text{n}_{\text{plural_1st}}$	$\text{n}_{\text{plural_2nd}}$	$\text{n}_{\text{plural_3rd}}$
- ※ 6個の規則が必要

4

素性による一致のチェック

- NP → det n
 - a desk
(number=singular) (number=singular)
 - × a desks
(number=singular) (number=plural)

- 素性(feature)
 - ex. number
- 値(value)
 - ex. singular, plural

5

素性構造(feature structure)

- 素性と値のペアの集合
- 単語や書き換え規則を素性構造で表現

theの素性構造

$$\left[\begin{array}{l} \text{CAT}=\text{det} \\ \text{NUM}=\{\text{singular, plural}\} \\ \text{PER}=\text{3rd} \\ \text{HEAD}=\text{the} \end{array} \right]$$

girlの素性構造

$$\left[\begin{array}{l} \text{CAT}=\text{n} \\ \text{NUM}=\text{singular} \\ \text{PER}=\text{3rd} \\ \text{HEAD}=\text{girl} \end{array} \right]$$

規則 NP→det n

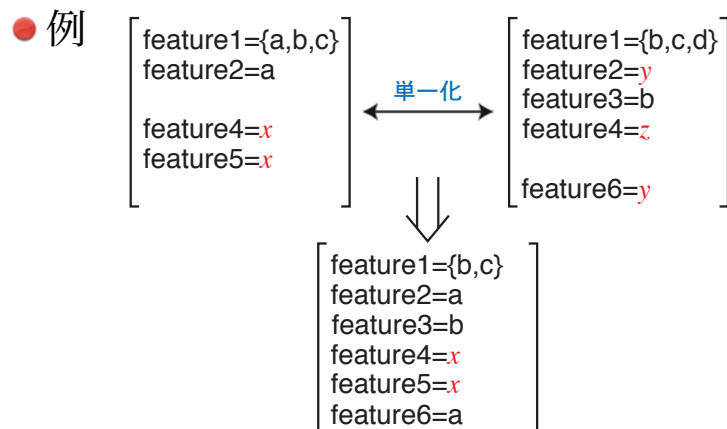
$$\left[\begin{array}{l} \text{CAT}=\text{NP} \\ \text{NUM}=\textit{n} \\ \text{PER}=\textit{p} \end{array} \right] \longrightarrow \left[\begin{array}{l} \text{CAT}=\text{det} \\ \text{NUM}=\textit{n} \\ \text{PER}=\textit{p} \end{array} \right] \left[\begin{array}{l} \text{CAT}=\text{n} \\ \text{NUM}=\textit{n} \\ \text{PER}=\textit{p} \end{array} \right]$$

斜字は変数を表す

6

単一化(unification)

- 2つの素性構造を素性と値に矛盾がないように1つの素性構造にまとめる操作



7

単一化

- 単一化できない場合
 - 同じ素性の値が異なる定数になるとき

■ 例1

$$\left[\text{NUM}=\text{singular} \right] \xleftrightarrow{\text{単一化}} \left[\text{NUM}=\text{plural} \right]$$

■ 例2

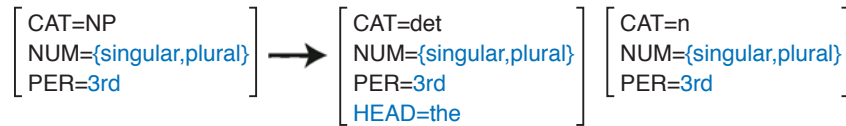
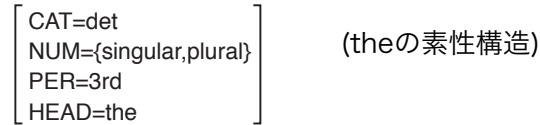
$$\left[\begin{array}{l} \text{feature1}=\textit{x} \\ \text{feature2}=\textit{x} \end{array} \right] \xleftrightarrow{\text{単一化}} \left[\begin{array}{l} \text{feature1}=a \\ \text{feature2}=b \end{array} \right]$$

8

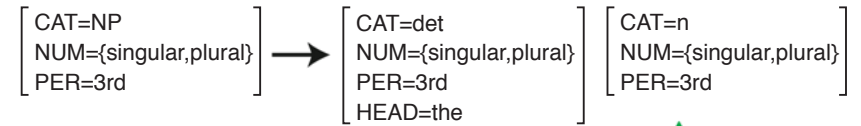
“the girl”の例



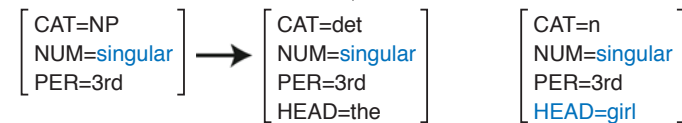
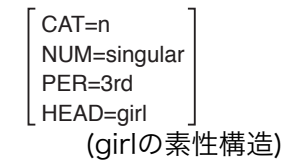
↑ 単一化 ↓



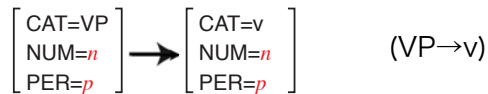
“the girl”の例



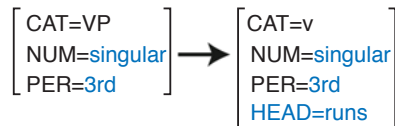
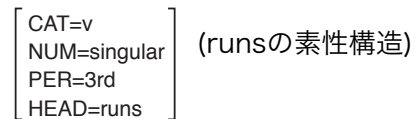
↑ 単一化 ↓



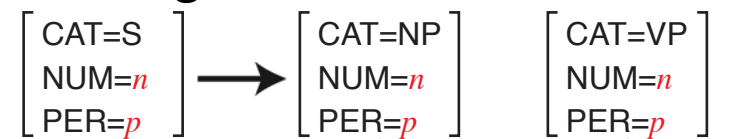
“runs”の例



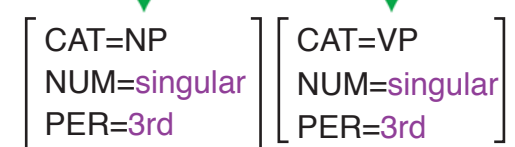
↑ 単一化 ↓



“the girl runs”の例



↑ 単一化 ↓



n=singular
p=3rd で単一化可能

基本的な素性

- 一致に関する素性
 - NUM,PER
- 動詞の形態に関する素性
 - VFORM
 - ◆ base(原形),ing(現在分詞),past(過去形)...
- 動詞の格要素支配に関する素性
 - SUBCAT
 - ◆ NP-NP(第3文型), NP-NP-NP(第4文型)

13

ここまでのおさらい

- 単語間の言語的な整合性をチェックしたい場合、素性を使うと良い
- 素性を使うことの利点
 - 文法の記号を細分化しなくてよい
 - 規則数は増加しない
- 整合性のチェックは素性構造の単一化によって実現されている
 - いろいろな素性を統一的に取り扱える

14

単一化文法

- unification grammar
- 素性構造間の単一化によって言語の文法的制約を表す文法
- 具体例
 - 語彙機能文法(LFG)
 - ◆ Lexical Functional Grammar
 - 一般化句構造文法(GPSG)
 - ◆ Generalized Phrase Structural Grammar
 - 主辞駆動句構造文法(HPSG)
 - ◆ Head-driven Phrase Structural Grammar

15

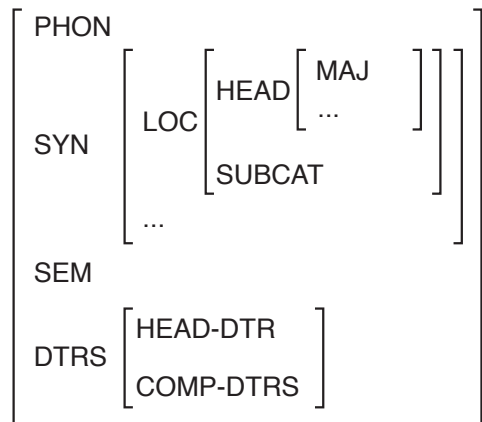
HPSGの概要

- 書き換え規則は使用しない
- 素性構造の単一化のみで解析を進める
 - 単一化に関する少数の原則に従う
- 単語は素性構造を持つ
 - 豊富な素性
 - 解析に必要な情報のほとんどが記述される

16

HPSGの素性構造

- 単語, 句, 文は全て素性構造で表現される
 - 入れ子構造を持つ



17

句の素性構造の生成

- 2つの素性構造A,Bを組み合わせ、上位の句の素性構造を作る
 - 例: saw + Sandy \Rightarrow saw Sandy (VP)
 - 規則は使わず、単一化によって作成
- ボトムアップに繰り返し、最終的に文全体の素性構造を得る

18

主辞(head)とは?

- 語彙的主辞(lexical head)
 - 句の中で最も中心的な役割を果たす単語
 - ◆ NP(det+n)の場合, 主辞はn(名詞)
 - ◆ VP(v+NP)の場合, 主辞はv(動詞)
- 句は主辞要素と補語要素から成る
 - 主辞要素(head daughter)
 - ◆ 語彙的主辞を含む要素
 - 補語要素(complement daughter)
 - ◆ それ以外の要素
 - 例: SがNPとVPから構成されるとき
VPが主辞要素、NPが補語要素

19

HPSGの主な素性

- PHON素性(音韻素性)
 - 句が支配する単語の列を保持する素性
- SYN素性(統語素性)
 - 句の構文に関する情報を保持する素性
 - SYN|LOC|HEAD素性
 - ◆ 語彙的主辞の情報を保持
 - SYN|LOC|SUBCAT素性
 - ◆ 語彙的主辞が下位範疇化する要素の情報を保持
 - ◆ 下位範疇化: 動詞が主語や目的語を取る
 - SYN|LOC|HEAD|MAJ素性
 - ◆ 語彙的主辞の品詞を保持する素性

20

HPSGにおける主な素性

- SEM素性(意味素性)
 - 句の意味に関する情報を保持する素性
- DTRS素性
 - DTRS|HEAD-DTR素性
 - ◆ 主辞要素の素性構造をそのまま保持
 - DTRS|COMP-DTRS素性
 - ◆ 補語要素の素性構造をそのまま保持

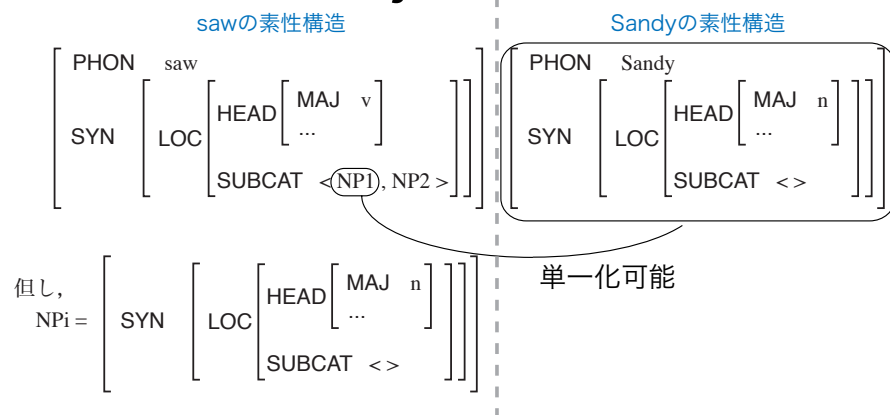
21

素性構造生成のための原理

- A,B 2つの素性構造を組み合わせる
- どちらが主辞要素になるか?
 - AのSYN|LOC|SUBCAT素性中の素性構造が、Bの素性構造全体と単一化可能であるとき
→Aが主辞要素、Bが補語要素

22

saw + Sandyの場合



※ sawが主辞要素、Sandyが補語要素

23

素性構造生成のための原理

- 新しく生成する素性構造の素性の値をどのように決めるか?
- SYN|LOC|HEAD素性
 - 主辞要素のHEAD素性をコピー
 - 主辞素性原理(head feature principle)
- SYN|LOC|SUBCAT素性
 - 主辞要素のSUBCAT素性から、補語要素を削除したものを与える
 - 下位範疇化原理(subcategorization principle)

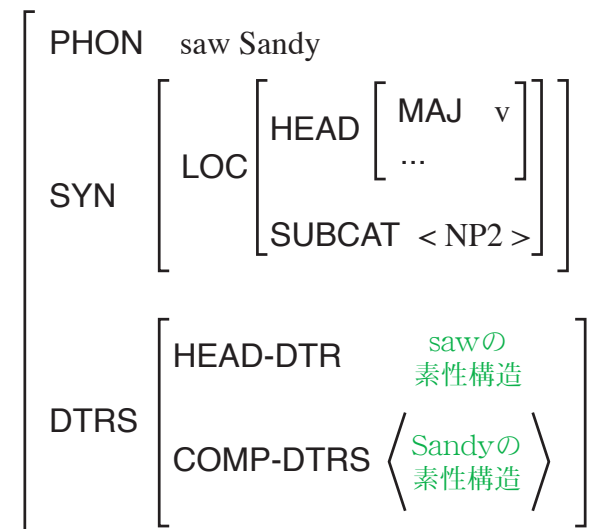
24

素性構造生成のための原理

- PHON素性
 - 構成要素のPHON素性の値を連結したものを与える
- DTRS | COMP-DTRS素性
 - 補語要素の素性構造のリストを与える
- DTRS | HEAD-DTR素性
 - 主辞要素の素性構造全体を与える

25

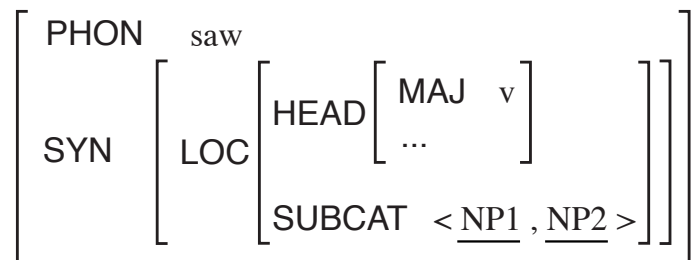
“saw Sandy”の素性構造



26

HPSG

- 主辞素性原理が重要
 - 句や文の構造は語彙的主辞によって決まる
 - sawの場合、目的語と主語を取ることがSUBCAT素性に明示されている



27