

# 自然言語処理入門

岸山 健 (31-187002)

Dec. 17, 2018

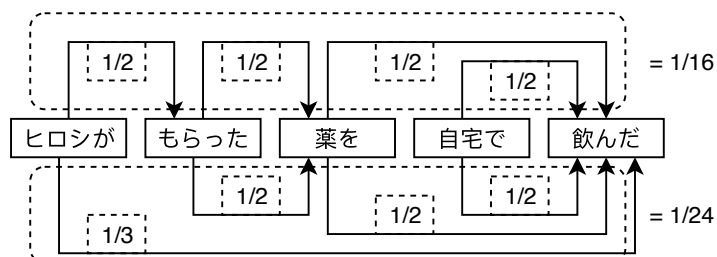
## 課題

- 以下の2つの文を、グラフに基づく手法で依存構造解析し、それぞれについて最も確率の高い依存構造を示せ
  - ヒロシが/もらった/薬を/自宅で/飲んだ
  - ヒロシは/もらった/薬を/自宅で/飲んだ
- 文節の依存関係の確率は以下とする
  - $P(1|\text{格助詞を機能要素とする文節, それより右 } n \text{ 番目の用言を主要素とする文節}) = 1/(n+1)$
  - $P(1|\text{係助詞ハを機能要素とする文節, 入力末から } n \text{ 番目の用言を主要素とする文節}) = 1/(n+1)$
  - $P(1|\text{機能要素が連体形である文節, それより右 } n \text{ 番目の体言を主要素とする文節}) = 1/(n+1)$
  - $P(1|\text{上記以外の関係にある文節対}) = 1/10$

## 1 回答

与えられた例文1において、「ヒロシが」は「もらった」と「飲んだ」の両方に依存しうる。下の図は例文のチャンクとそれぞれの依存先、そしてそれらの確率を示したものである。例文の場合、「ヒロシが」が「もらった」に依存する確率と「飲んだ」にかかる確率の差がそのまま各構造の尤もらしさを反映される。前者の確率の積は $1/16$ になるのに対し、後者は $1/24$ となる。従って「ヒロシが」が「もらった」に依存する構造が尤もらしいということが分かる。また、「薬を自宅で飲んだ」という構造は「薬を」「自宅で飲んだ」と「薬を自宅で」「飲んだ」という構造がとられるが、4つ目のルールにより後者は排除される。

- ヒロシが/もらった/薬を/自宅で/飲んだ



他方、主語の助詞が「ヒロシは」のように係助詞の「ハ」が与えられていた場合の結果は異なる。2つ目のルールにより、最初に出てきた用言「もらった」のコストは入力末から2番目であるため、その確率は $1/3$ となる。しかし同じルールを最後の用言である「飲んだ」に適用した場合、その確率は $1/2$ となる。したがって、

前者の確率の積は  $1/16$  になるのに対し、後者は  $1/24$  となる。従って「ヒロシは」が「飲んだ」に依存する構造が尤もらしいということが分かる。

- ヒロシは/もらった/薬を/自宅で/飲んだ

