

# 格構造解析による自動要約生成文の冗長性の除去

橋本哲弥<sup>\*1</sup>, 三輪誠<sup>\*2</sup>, 鶴岡慶雅<sup>\*1</sup>, 近山隆<sup>\*1</sup>  
<sup>1</sup>東京大学 近山・鶴岡研究室 <sup>2</sup>マンチェスター大学

## ■自動要約

膨大な電子テキストから必要な情報を抽出

◇一般的な自動要約システム

- ①各文の重要度を単語の出現頻度から計算し、重要文を抽出
- ②抽出された文を圧縮

## ■文圧縮の課題

既存の文圧縮手法<sup>[2]</sup>は、表層格のみに基づく原文-要約文の組から学習した削除規則によって、除去する句を決定

お腹が空いていたので、1000円でおやつを買った。

↓  
お腹が空いていたので、おやつを買った。  
→所持金に焦点を置いた話でなければ文法的・意味的に正しい

意味関係(深層格)の違いを考慮しておらず、誤った圧縮を行ってしまう可能性有り

鍵がかかっていたので、合鍵で扉を開けた。

↓  
鍵がかかっていたので、扉を開けた。  
→文法的には正しいが、意味的には不自然

深層格を考慮することで文法的且つ意味的に正しい文圧縮を行う

## ■文圧縮の手法

雑音のある通信路モデル(Noisy-Channel Model)

$$P(s/t) \propto P(t/s) \cdot P(s)$$

原文tから要約文sが出力される確率を、sを拡張したときにtが生成される確率とsの生じる確率(文法的に正しいかなど)から計算

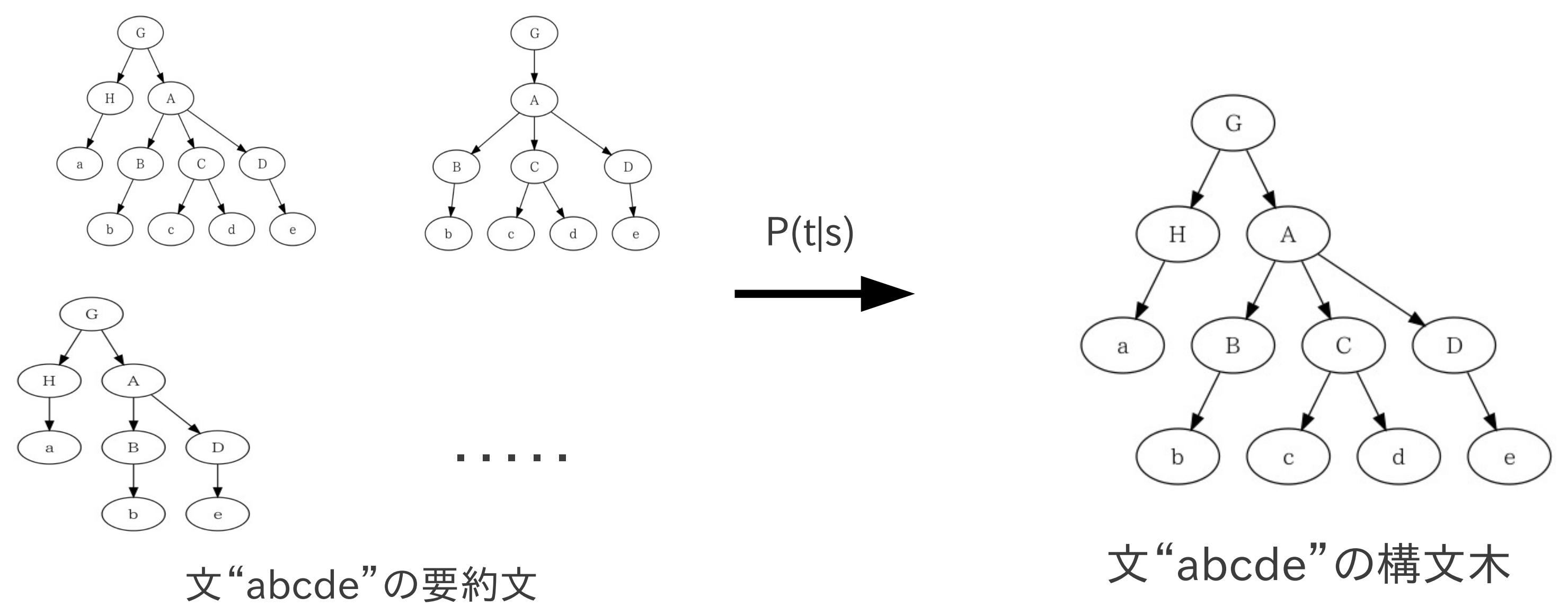
原文 t: "a b c d e" (長さ L = 5, a~eはtに含まれる句)

↓ tの各要素を削除するかしないか

要約文候補 s: "a", "a b", "a c", ... (2<sup>L</sup>通り)

P(t|s): sに含まれていないtの要素をsに挿入する際の隣接する句同士の条件付き確率から計算

P(s): sの生じる確率



## ■提案手法: 深層格を考慮した文圧縮

文法的、表層的な関係である表層格に加えて意味的な関係である深層格を利用し、文の意味を考慮した文圧縮を行う

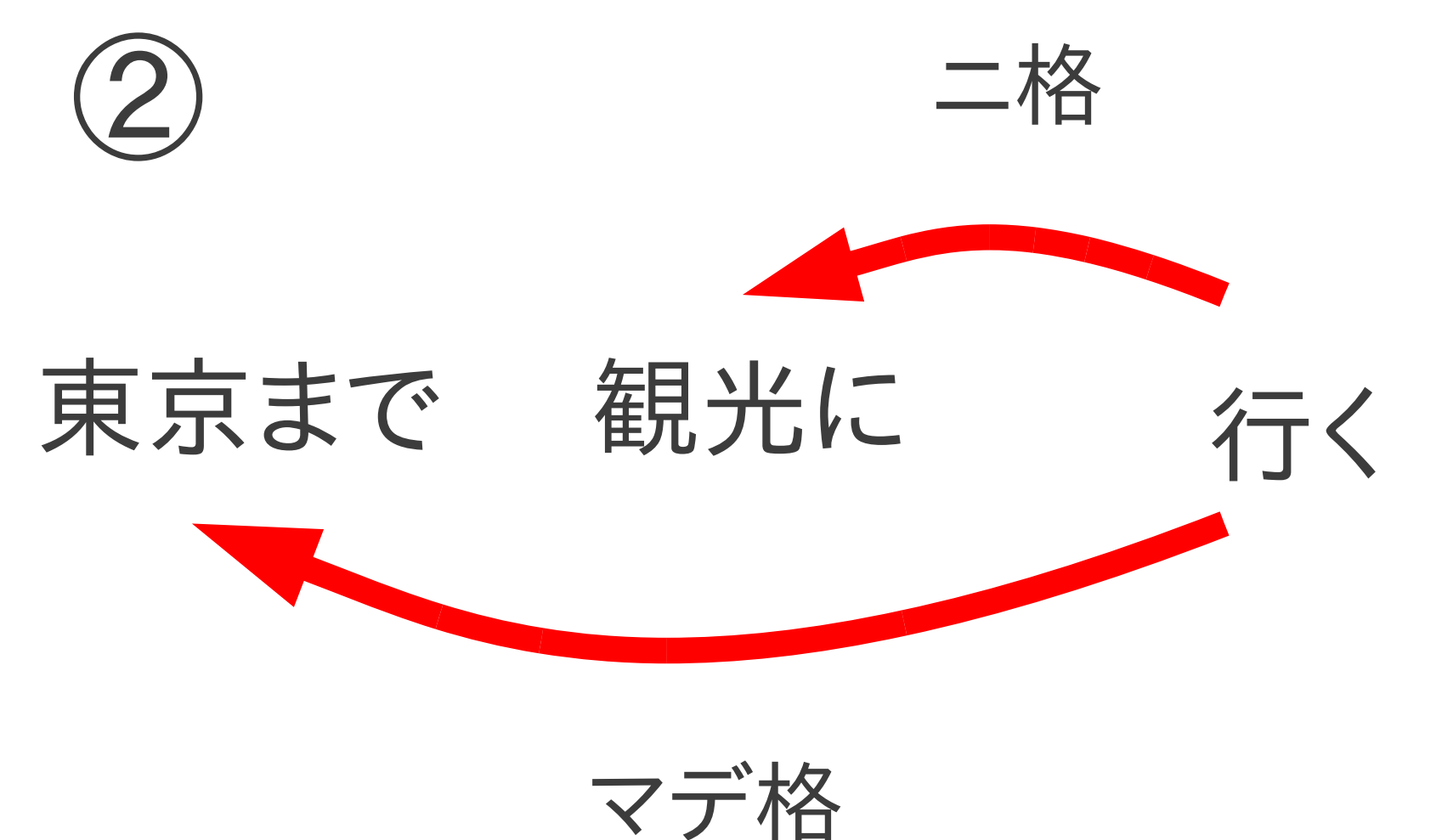
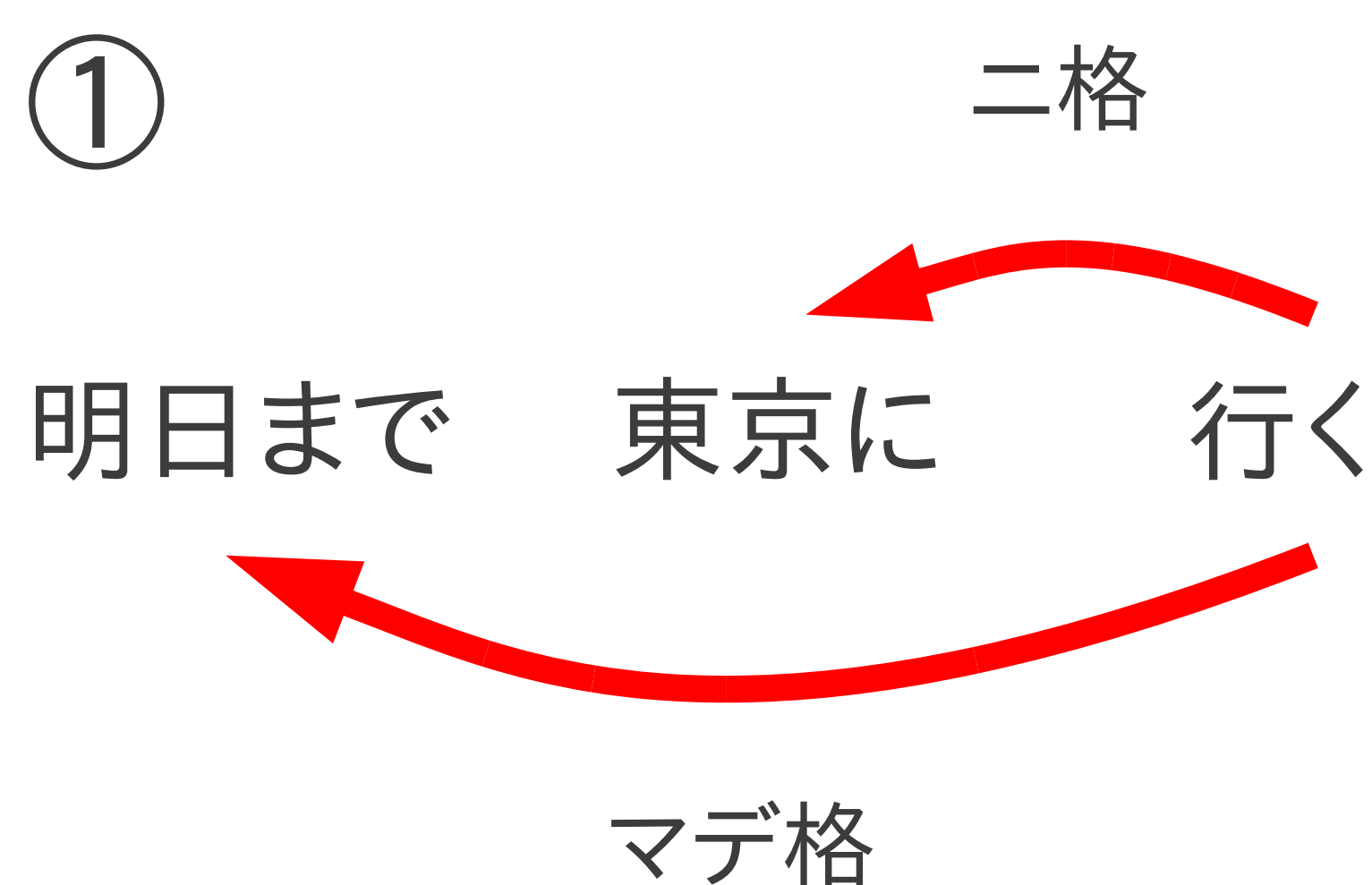
### ■日本語における表層格

ガ格、ヲ格、カラ格  
ト格、デ格、ニ格  
ヨリ格、ヘ格、マデ格  
無格、時間格、複合辞

### ■深層格

チャールズ・フィルモアによって提唱されたもの

- ・動作主格(agent)
- ・対象格(object)
- ・経験者格(experiencer)
- ・道具格(instrument)
- ・源泉格(source)
- ・目標格(goal)
- ・場所格(location)
- ・時間格(time)



- ①では「明日まで」は時間格で、「東京に」は場所格
  - ②では「東京まで」は場所格・対象格で、「観光に」は目的格
- 深層格は表層格に対して1対1対応をしていないので、表層格とは別に判別する必要がある

深層格を付与した原文-要約文の組を学習し、tの構文木において、Sに含まれていないtの葉ノードが省略できる任意格である確率の計算を含めて、P(t|s)を計算

動詞の取る表層格と、名詞の持つ場所や時間などの属性から、雑音のある通信路モデルにおけるP(t|s)で拡張される句が、削除しても意味の通じる任意格かを推定する。

## ■まとめ

- ・雑音のある通信路モデルによって、構文木から要約文を確率的に決定
- ・深層格によって文法的正しさに加えて、意味的正しさを考慮

## ■参考文献

- [1]H. Jing, "Sentence Reduction for Automatic Text Summarization", Proc. 6th ANLP, pp.310-315, 2000.
- [2]K.Knight and D.Marcu "Statistics-Based Summarization - Step one: Sentence Compression", Proc. 17th AAAI, pp.703-710, 2000.
- [3]K.Yoshikawa, R.Iida, T.Hirao, and M.Okumura, "Sentence Compression with Semantic Role Constraints", Proc. ACL, pp.349-353, 2012