

指定指示・代行指示を区別した 指示連体詞の照応解析

井之上 直也[†] 飯田 龍^{††} 乾 健太郎[†] 松本 裕治[†]

[†]奈良先端科学技術大学院大学
{naoya-i, inui, matsu}@is.naist.jp

^{††}東京工業大学
ryu-i@cl.cs.titech.ac.jp

言語処理学会 第15回年次大会 2009年3月4日

照応解析とは

照応関係

先行詞

照応詞

麻生首相が鳥取を訪問した。首相は記者会見で国政について語った。

- ▶ ある言語表現が文章内の別の表現を指示するとき、これら2つは**照応関係**にある、という (首相と麻生首相)
 - **照応詞** ... 指示する側の表現 (首相)
 - **先行詞** ... 指示される側の表現 (麻生首相)
- ▶ **照応解析** ... **照応関係**を特定する処理
 - 機械翻訳や情報抽出等の応用分野で必要となる要素技術

直接照応と間接照応

書道の腕前を見せてもらった。腕前は抜群だった。

初めて彼の料理を食べた。腕前は抜群だった。

- ▶ どちらの指示となるかは先行文脈に依存
 - **曖昧性を解消する必要がある**
- ▶ 本研究の対象は、照応詞がこの/その/あのを含む場合

図書館で資料を入手した。このデータは重要だ。

「この/その + NP」で照応 → **指定指示（直接照応）**

5年間、水質調査を行った。このデータは重要だ。

「この/その」で照応 → **代行指示（間接照応）**

指定指示、代行指示をまとめて指示関係と呼ぶ ③

指示連体詞の照応解析に必要な2つのタスク

▶ 先行詞同定

- ・与えられた照応詞に対して**先行詞**を同定する

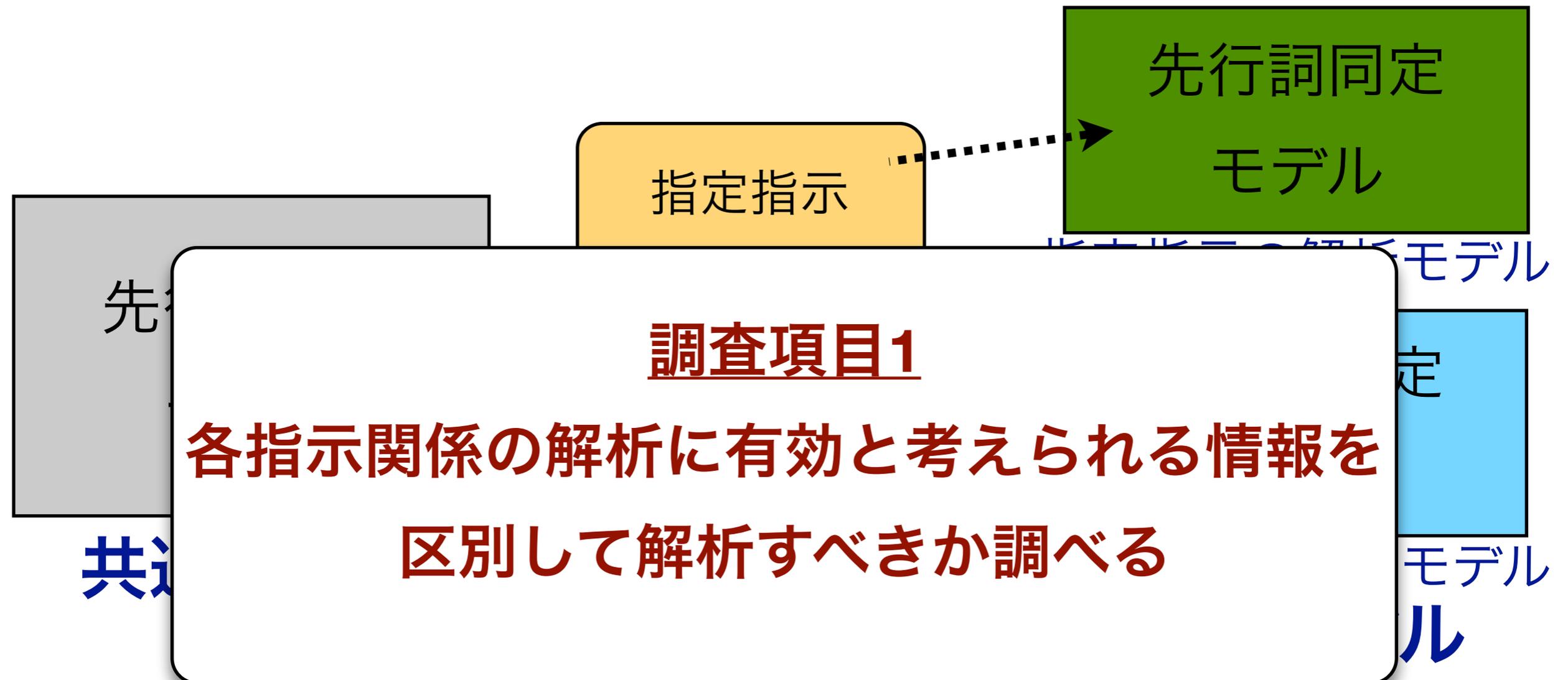
▶ 指示関係分類

- ・与えられた照応詞が**指定指示(直接照応)**か**代行指示(間接照応)**か分類する

本研究の焦点

機械学習を用いた照応解析の枠組みにおいて、
2つのタスクをどのように組み合わせるべきか
(解析順序など) 調査する

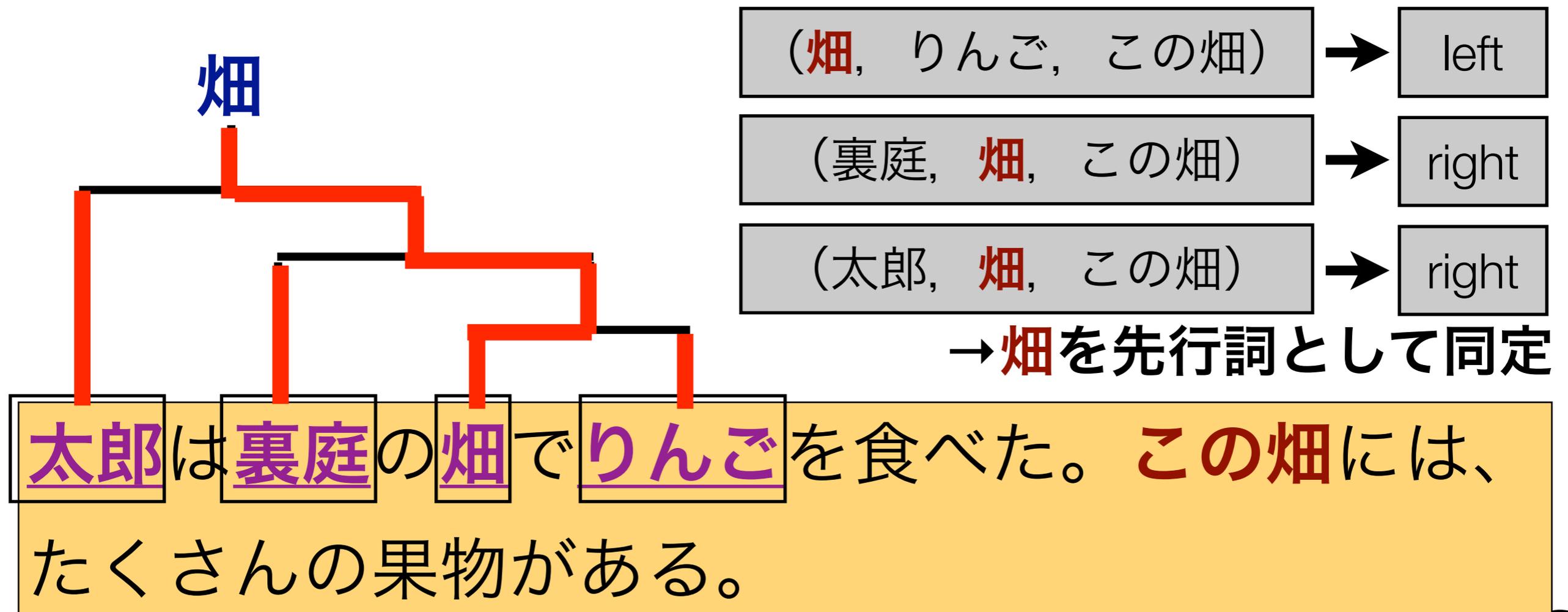
考えられる先行詞同定の手法



- ▶ 指定指示と代行指示のそれぞれの解析に有効と思われる情報を**全て使う（共通）** or **分けて使う（独立）**

先行詞同定に用いる枠組み

- ▶ 先行詞の同定にはトーナメントモデル（飯田ら,04）の枠組みを用いる
- ▶ 先行詞候補間の勝ち抜き戦を行い、勝者を先行詞とする



指示連体詞の照応解析に必要な2つのタスク

▶ 先行詞同定

- 与えられた照応詞に対して**先行詞**を同定する

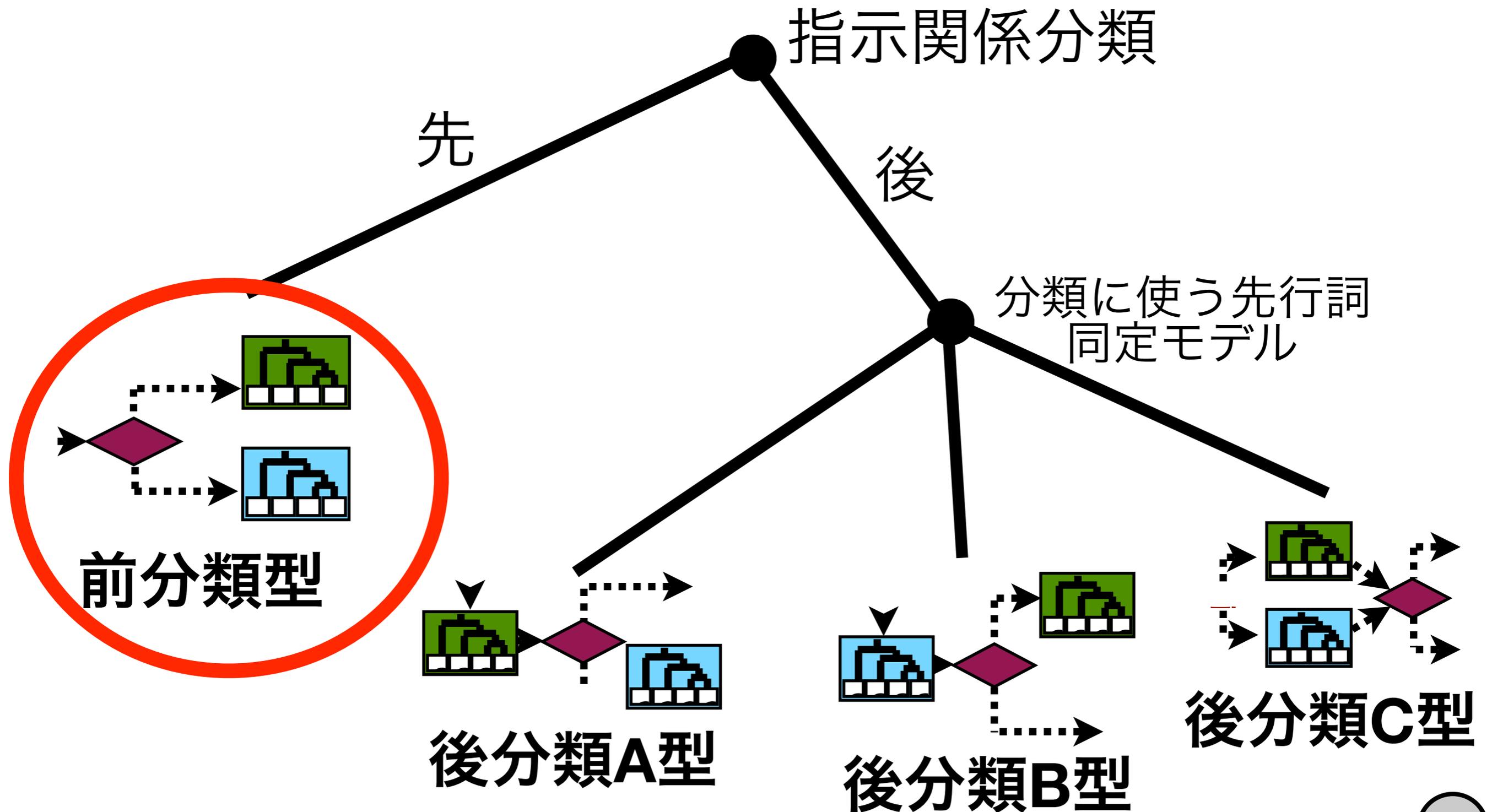
▶ 指示関係分類

- 与えられた照応詞が**指定指示(直接照応)**か**代行指示(間接照応)**か分類する

本研究の焦点

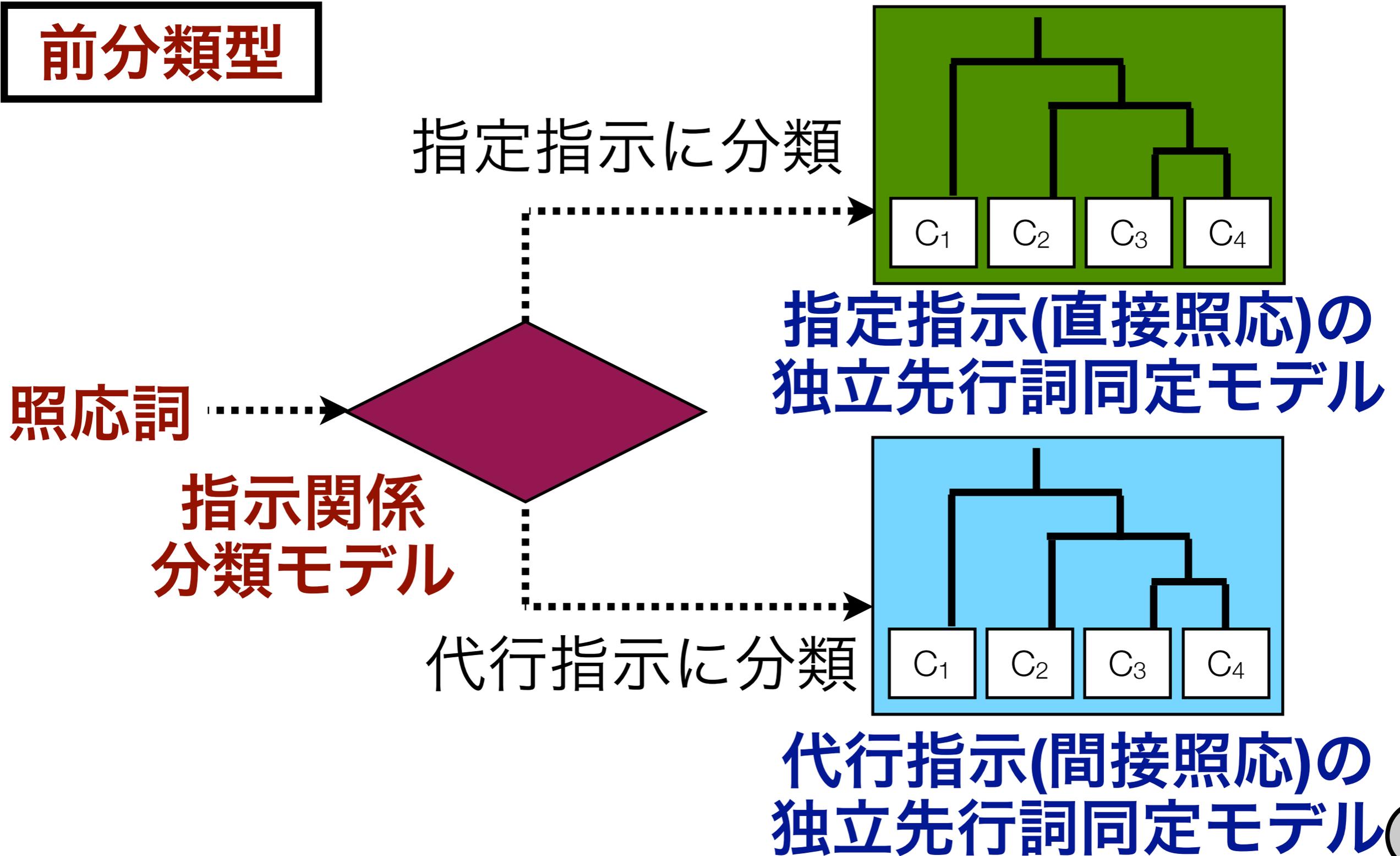
機械学習を用いた照応解析の枠組みにおいて、
2つのタスクをどのように組み合わせるべきか
(解析順序など) 調査する

比較する4つの指示関係分類モデル

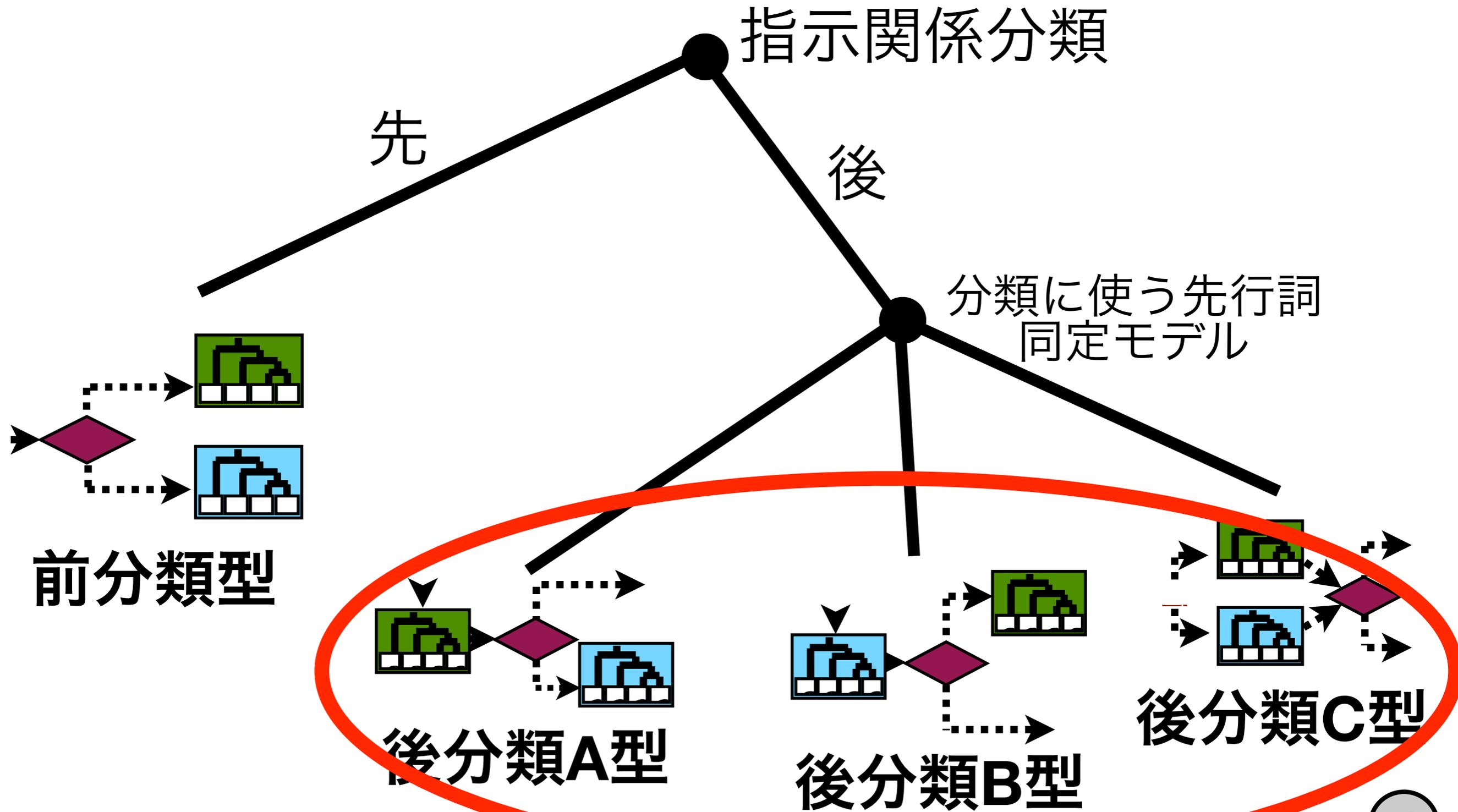


1. あらかじめ指示関係を分類してから先行詞同定

前分類型



比較する4つの指示関係分類モデル



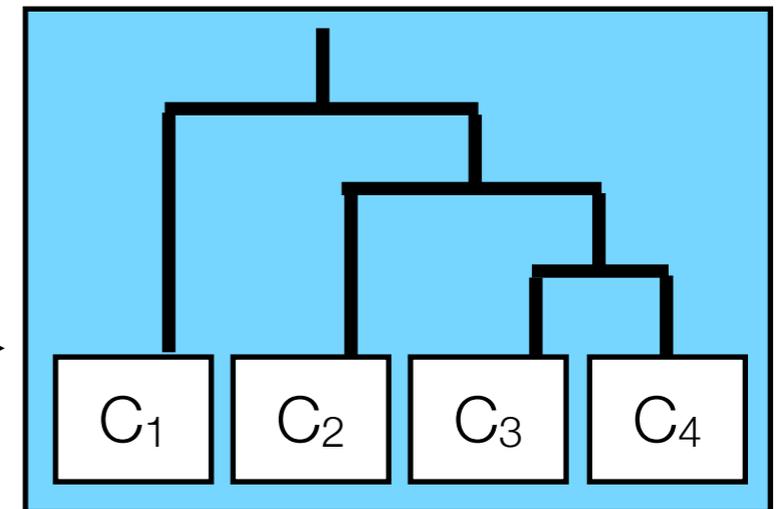
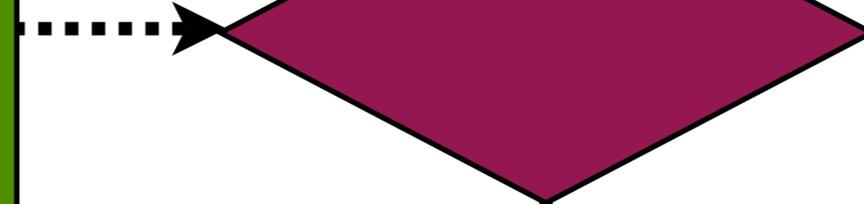
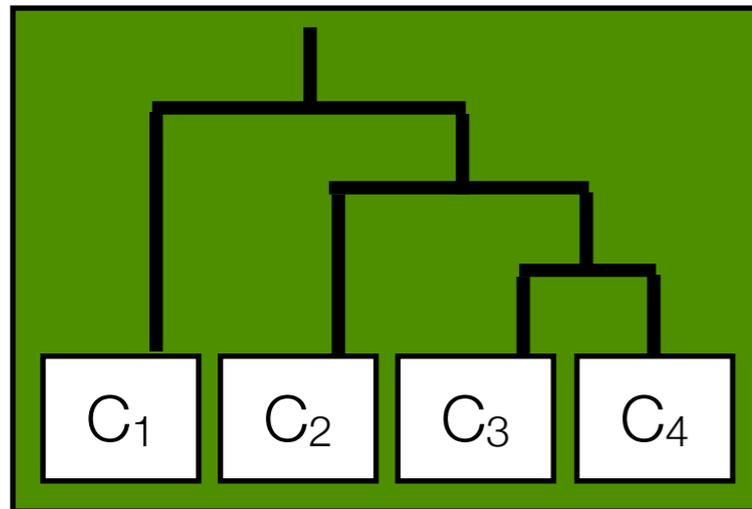
2. 先行詞同定の結果を用いて指示関係分類

後分類A型

照応詞

指定指示に分類

指示関係
分類モデル



指定指示(直接照応)の
独立モデル

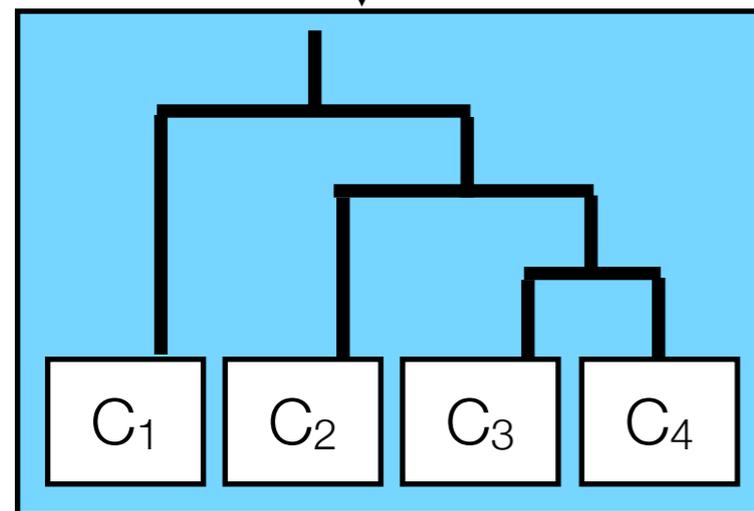
代行指示に分類

代行指示(間接照応)の
独立モデル

3. 先行詞同定の結果を用いて指示関係分類

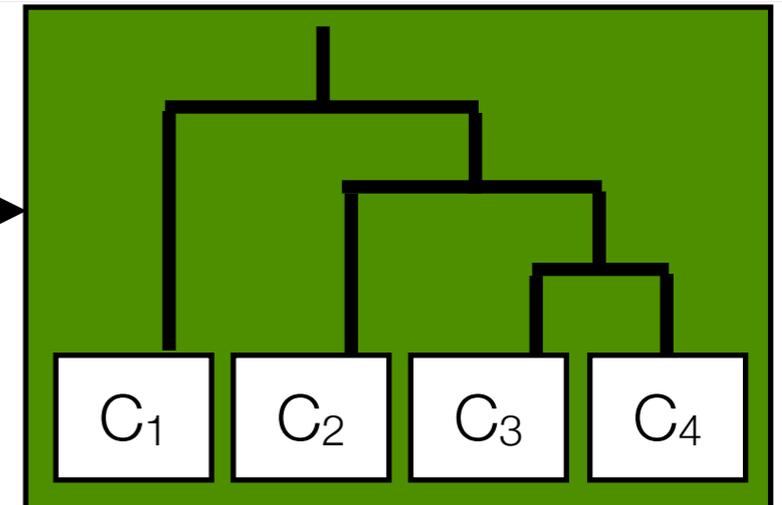
後分類B型

照応詞



代行指示(間接照応)の
独立モデル

指定指示に分類

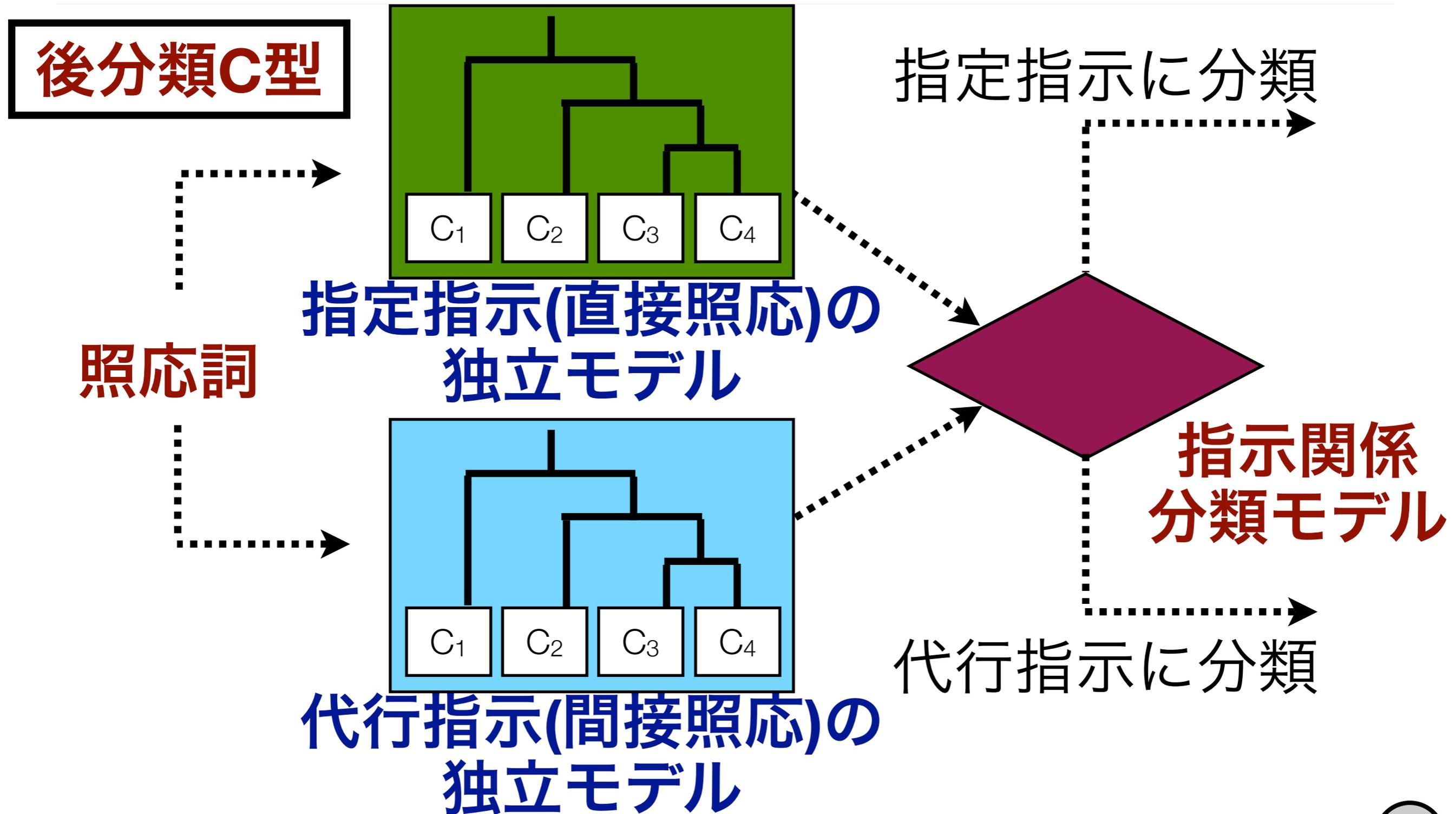


指定指示(直接照応)の
独立モデル

指示関係
分類モデル

代行指示に分類

4. 先行詞同定の結果を用いて指示関係分類



指示関係分類に使う情報の違い

- ▶ 前分類型 (指示関係分類 → 先行詞同定)

昨日、太郎の家の近くの図書館で資料を手に入れた。このデータは機械的に処理される。

調査項目2

どのような方法で先行文脈情報をエンコードするのが
指示関係分類に有効か調べる

昨日、太郎の家の近くの図書館で資料を手に入れた。このデータは機械的に処理される。

- 同定した先行詞

評価実験

- ▶ NAISTテキストコーパス（飯田ら, 07）の報道記事869記事
 - 今回新たにタグ付与を行った指示連体詞**1,363**
（10分割交差検定）
 - 指定指示(直接照応):**553**, 代行指示(間接照応):**810**
 - 先行詞が述語となる場合も対象とする
 - 今回は外界指示は対象としない
- ▶ 分類器にはSupport Vector Machine（Vapnik, 95）を使用
 - 線形カーネル、実装はSVM^{light}

実験に用いた素性 - 先行詞同定

▶ 語彙情報

- ・ 主辞、品詞、固有名詞の種類

▶ 統語的情報、位置情報

- ・ 格助詞、照応詞との文距離、先行詞候補間の文距離

▶ 意味的類似度の情報...独立モデルは指定指示(直接照応)のみ

- ・ 文字列一致、分布類似度、同義表現か
- ・ 照応詞の下位語か

▶ 共起尺度...独立モデルは代行指示(間接照応)のみ

- ・ 「(先行詞)の(照応詞)」のパターンで共起する強さを表す自己相互情報量 (PMI)

実験に用いた素性 - 指示関係分類

▶ 前分類型 (指示関係分類 → 先行詞同定)

• 先行文脈に各指示関係の先行詞の特徴を持つ候補があるか

- 照応詞の下位語、同義表現があるか
- 照応詞との類似度、「(先行詞) の (照応詞)」のPMI

• 照応詞の語彙、統語情報

- 照応詞の主辞、品詞、指示連体詞の種類
- 照応詞の格助詞

▶ 後分類型 (先行詞同定 → 指示関係分類)

• 同定した先行詞の情報

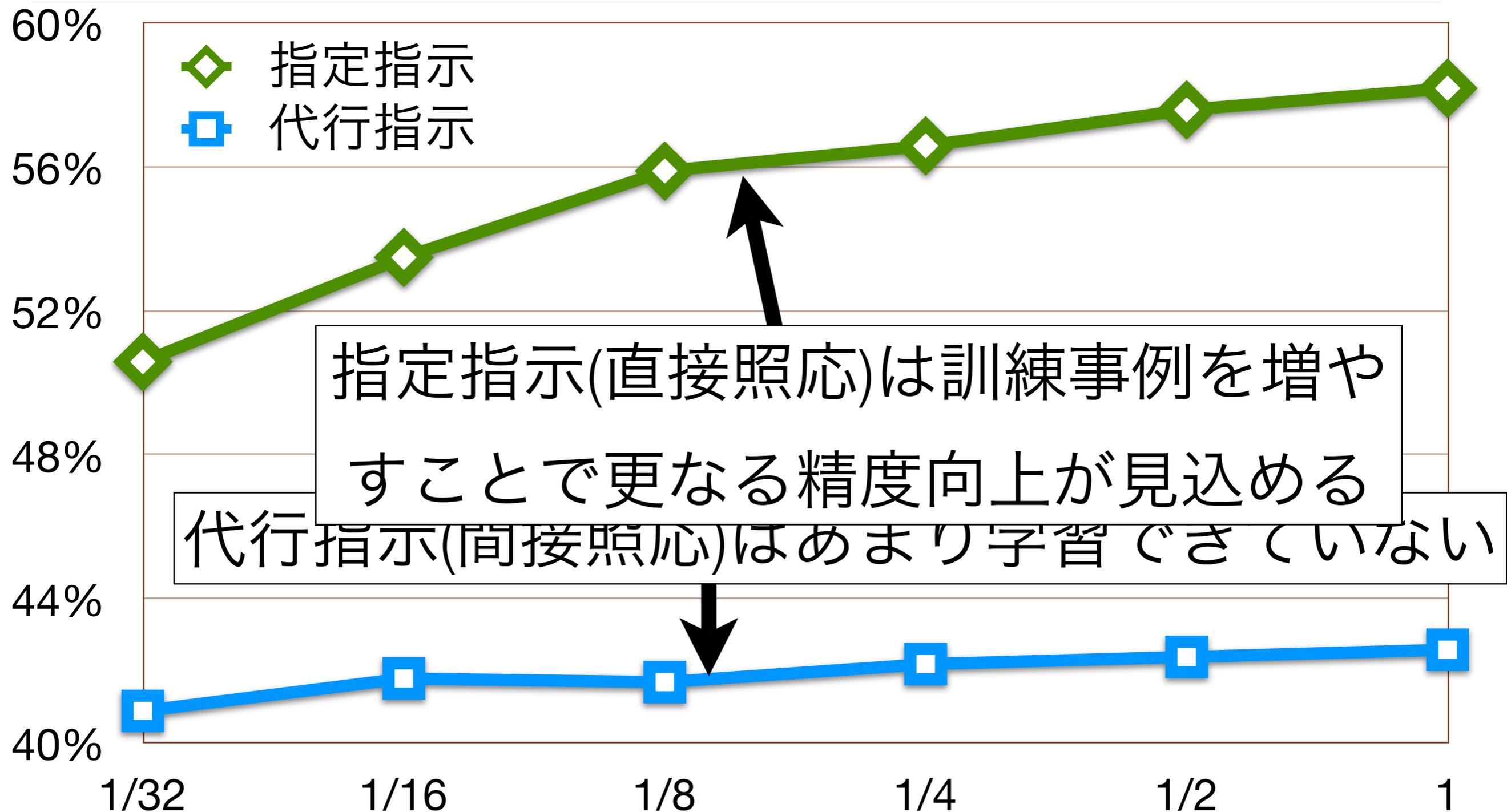
- 利用した先行詞同定モデルの素性をそのまま使う

実験結果 - 先行詞同定の精度

	指定指示 (直接照応)	代行指示 (間接照応)	全体
共通	59.5% (329/553)	28.8% (233/810)	41.2% (562/1,363)
独立	58.2% (322/553)	42.6% (345/810)	48.9% (667/1,363)

- ▶ 独立モデルの方が精度よく解析できたことから、同定に用いる情報を分けて使うほうがよいことを確認

独立先行詞同定モデルの学習曲線



先行詞同定の主な誤り

▶ 指定指示(直接照応)

映画の知識はないが『フランケンシュタイン』
ぐらいいは知っている。この映画は最高だ。

- 文字列一致しているが先行詞ではないものを誤って同定
→ 意味的類似性だけでは正しく解析できない

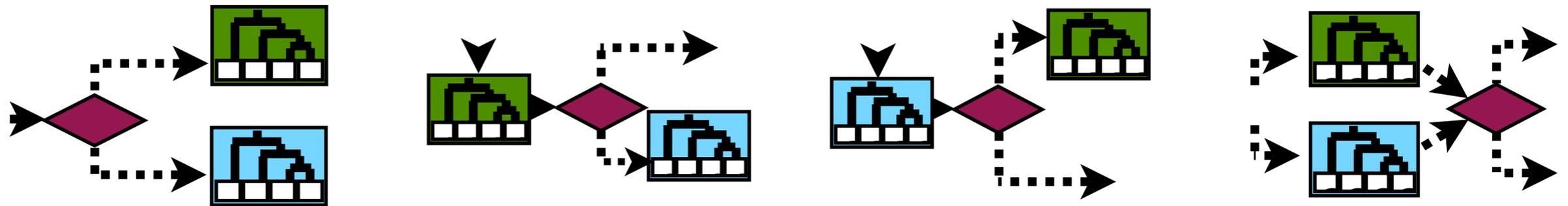
▶ 代行指示(間接照応)

日川は花園では勝てない。
今回もそのジंकスを破れなかった。

- 先行詞が述語の場合、誤って他の名詞句候補を先行詞と同定
→ 述語の先行詞を解析する手がかりが不十分。先行詞が述語となる事例は代行指示の照応の45.3%を占めるため、重要な問題

実験結果 - 指示関係分類の精度

	前分類型	後分類A型	後分類B型	後分類C型
精度	85.8% (1,170/1,363)	86.2% (1,175/1,363)	73.9% (1,007/1,363)	83.7% (1,141/1,363)
上限	85.8% (1,170/1,363)	86.9% (1,184/1,363)	84.4% (1,151/1,363)	90.3% (1,231/1,363)

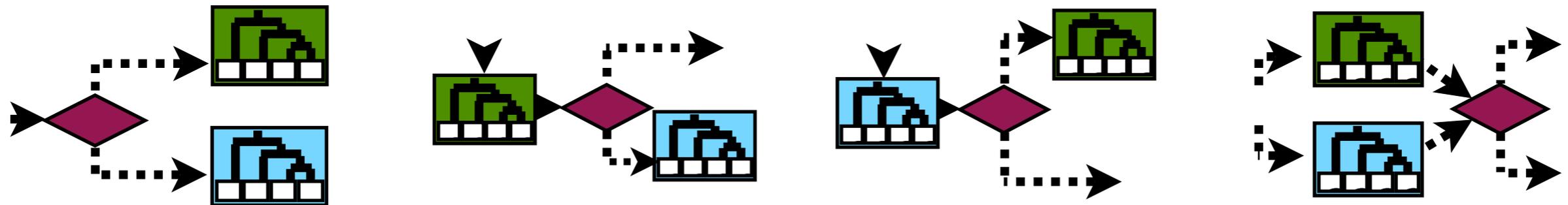


※上限 ... 後分類型には正解の先行詞を与えた場合の精度

- ▶ **同定した先行詞を先行文脈情報としてエンコード**することが、指示関係分類に役立つ

実験結果 - 照応解析全体の正解率

前分類型	後分類A型	後分類B型	後分類C型
45.7% (623/1,363)	46.4% (633/1,363)	36.5% (497/1,363)	45.1% (615/1,363)



正解率 = $\frac{\text{先行詞、指示関係ともに正しく解析できた数}}{\text{指示連体詞の総数}}$

- ▶ 照応解析全体として見ても、**同定した先行詞を先行文脈情報としてエンコード**するのが有効だと確認できた

関連研究

▶ 指定指示(直接照応)

- 名詞句共参照解析におおむね対応。
- 評価型タスクなどを通して盛んに研究されている (Soonら, 01; Ngら, 02; Yangら, 03; Beanら, 04)

▶ 代行指示(間接照応)

- Bridging Reference (Clark, 77) におおむね対応。
- 具体的な解析の試み (Poesioら, 04)

▶ **多くが指示関係を独立に扱っているが、実際の解析では解析対象が指定指示(直接照応)か、代行指示(間接照応)か不明**

▶ **指示関係の分類問題を焦点とする研究は少ない (本研究の貢献)**

まとめと今後の課題

- ▶ 指示連体詞の照応解析を**先行詞同定**と**指示関係分類**に分け、2つの調査を行った
- ▶ **調査項目1) 先行詞同定と指示関係の区別**
 - **指定指示(直接照応)と代行指示(間接照応)で解析に用いる情報を区別**し同定を行うほうがよいことを確認した
- ▶ **調査項目2) 指示関係分類に有効な先行文脈情報**
 - **同定した先行詞を先行文脈情報としてエンコード**することにより、指示関係分類に有効に働くことが分かった
- ▶ **今後の課題**
 - 代行指示(間接照応)の先行詞同定に有効な手掛かりを吟味
 - 外界指示を含めた指示関係分類を検討する



述語の先行詞を捉える手がかり

日川は花園では勝てない。

今回もそのジंकスを破れなかった。

- ▶ (NP (PRED X) Y)のような統語構造を持つ(X, Y)の対を獲得しておき、述語照応の手がかりとして利用する
 - e.g.) ...絶対に勝てないジंकスがあり...
 - → (勝てない, ジंकス) を獲得