

# 日本語事実性解析課題の経験的分析

成田 和弥<sup>†</sup>, 水野 淳太<sup>††</sup>, 乾 健太郎<sup>†</sup>

<sup>†</sup>東北大学 <sup>††</sup>奈良先端科学技術大学院大学

# はじめに：事実性とは

- 事実性 (Factuality) :

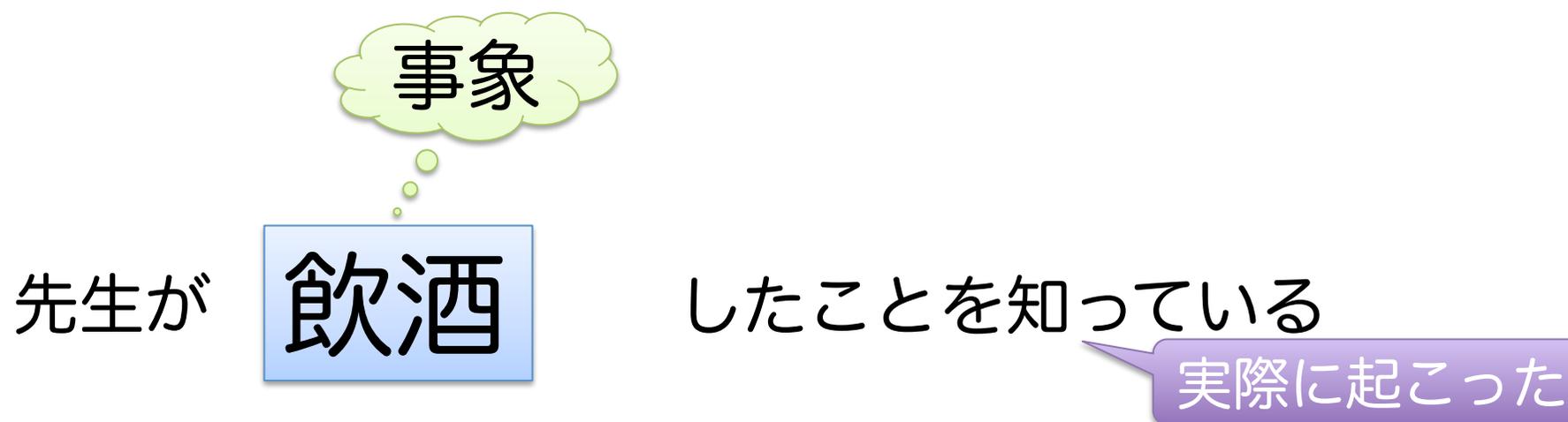
文中のある事象が実際に起こったことなのか、  
あるいは起こる可能性を述べただけなのか、 に関する情報



# はじめに：事実性とは

- 事実性 (Factuality) :

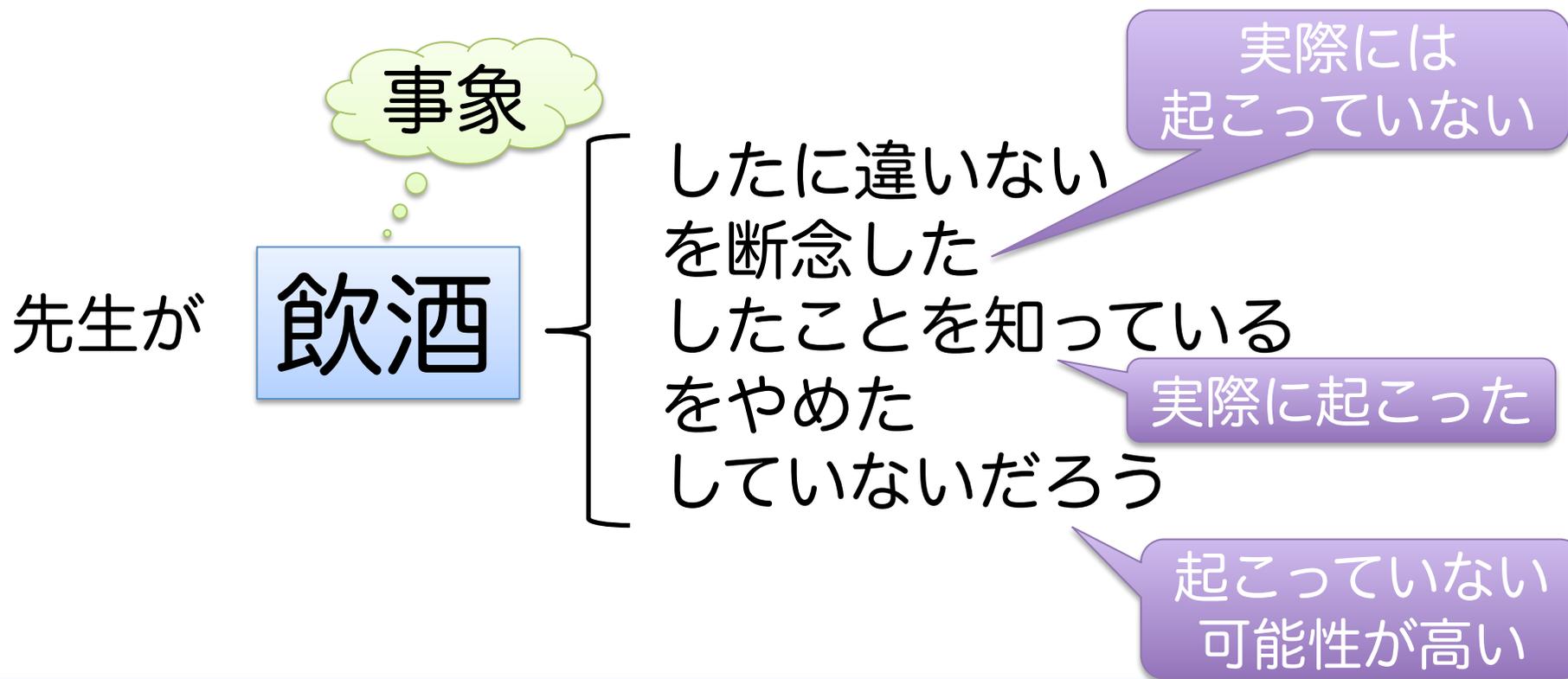
文中のある事象が実際に起こったことなのか、  
あるいは起こる可能性を述べただけなのか、 に関する情報



# はじめに：事実性とは

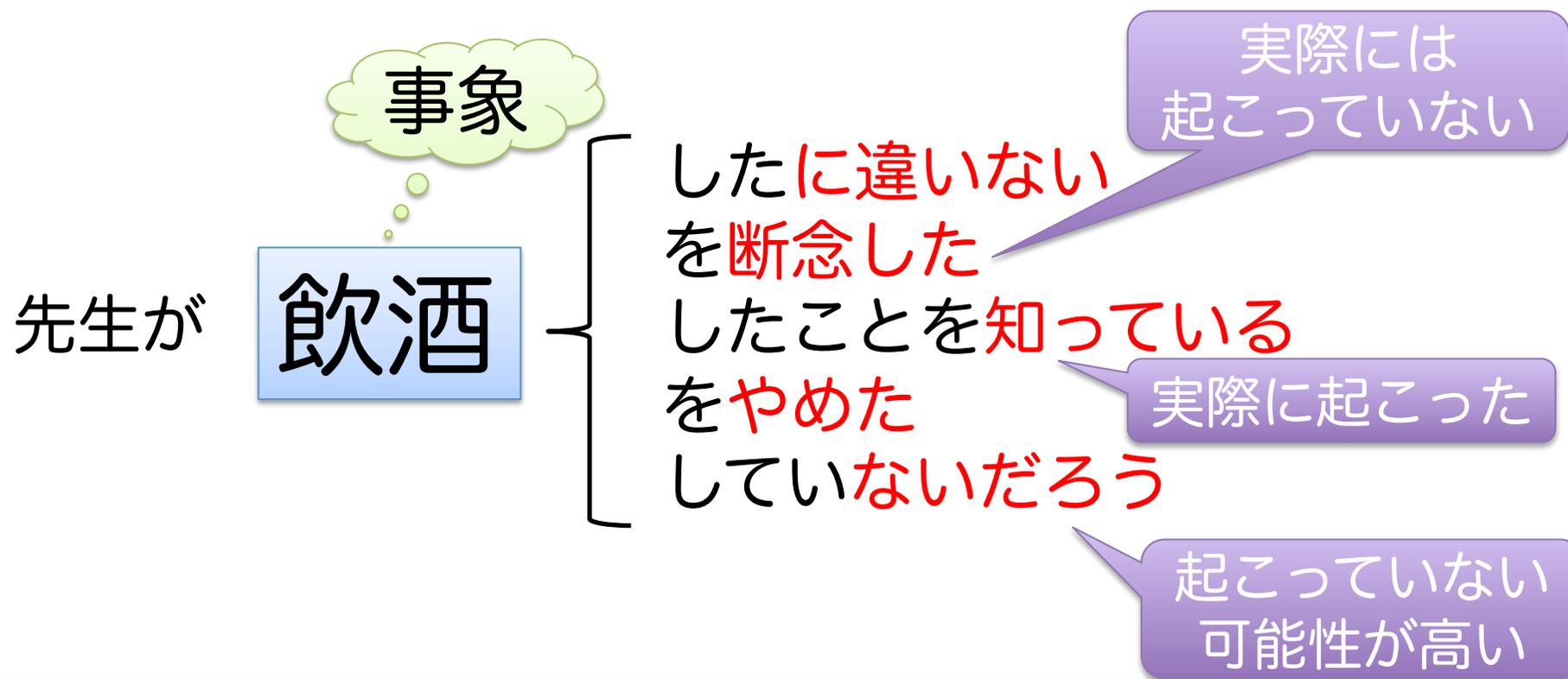
- 事実性 (Factuality) :

文中のある事象が実際に起こったことなのか、  
あるいは起こる可能性を述べただけなのか、 に関する情報



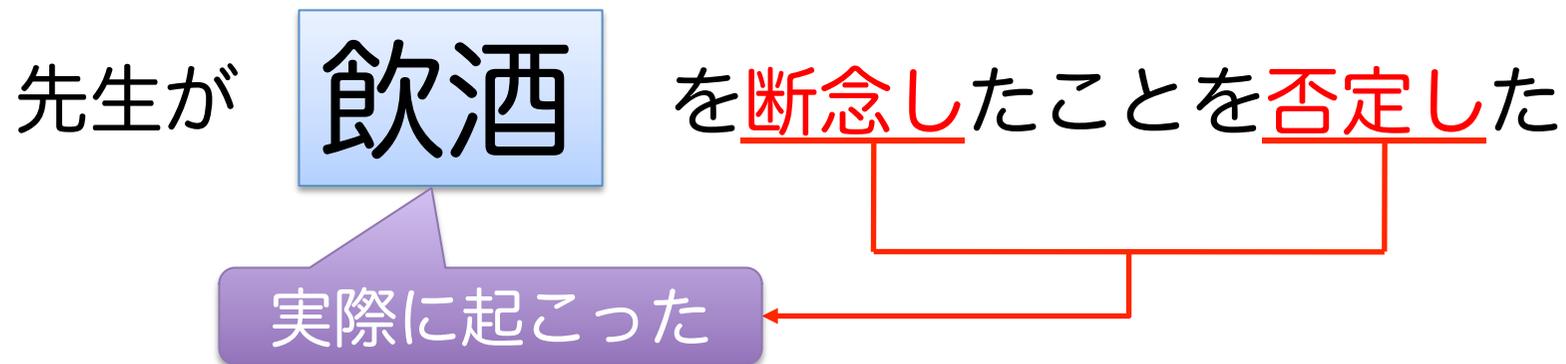
# はじめに：事実性解析の手がかり

事実性解析に対する手がかりとなる表現の存在



# はじめに：事実性解析の手がかり（組み合わせ）

事実性解析に対する手がかりとなる表現の存在



手がかり表現（語彙知識）の組み合わせ問題としてとらえる

# はじめに：本研究の目的

- 事実性解析 = 語彙知識の組み合わせによる問題

事実性に影響を与える語彙知識に  
焦点を当て、分析を行う

- 語彙知識の組み合わせに基づく，日本語事実性解析器を構築  
→ 実験・誤り分析

# 目次

- はじめに
  - 事実性とは
  - 本研究の目的
- **関連研究**
- **日本語事実性解析器**
  - 事実性の定義
  - 解析モデル
  - 使用する語彙知識
- 実験
- 誤り分析
- おわりに

## 関連研究：機械学習に基づく手法

- 条件付確率場を利用した手法[江口+ 10]
  - 態度表明者，態度，真偽判断などの6項目からなる拡張モダリティに対する解析
  - 事象間の依存関係を考慮
- 最大エントロピーモデルを用いた手法[松吉+ 11]
  - 拡張モダリティに対する解析を行い，主に態度に関して分析

これらの手法は，語彙知識に関する分析という目的には適用しづらい

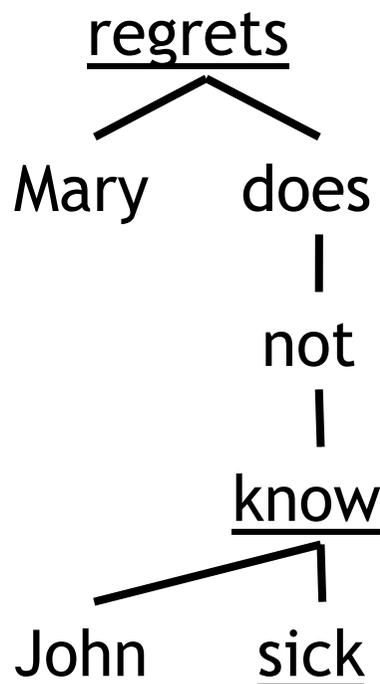
## 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

## 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

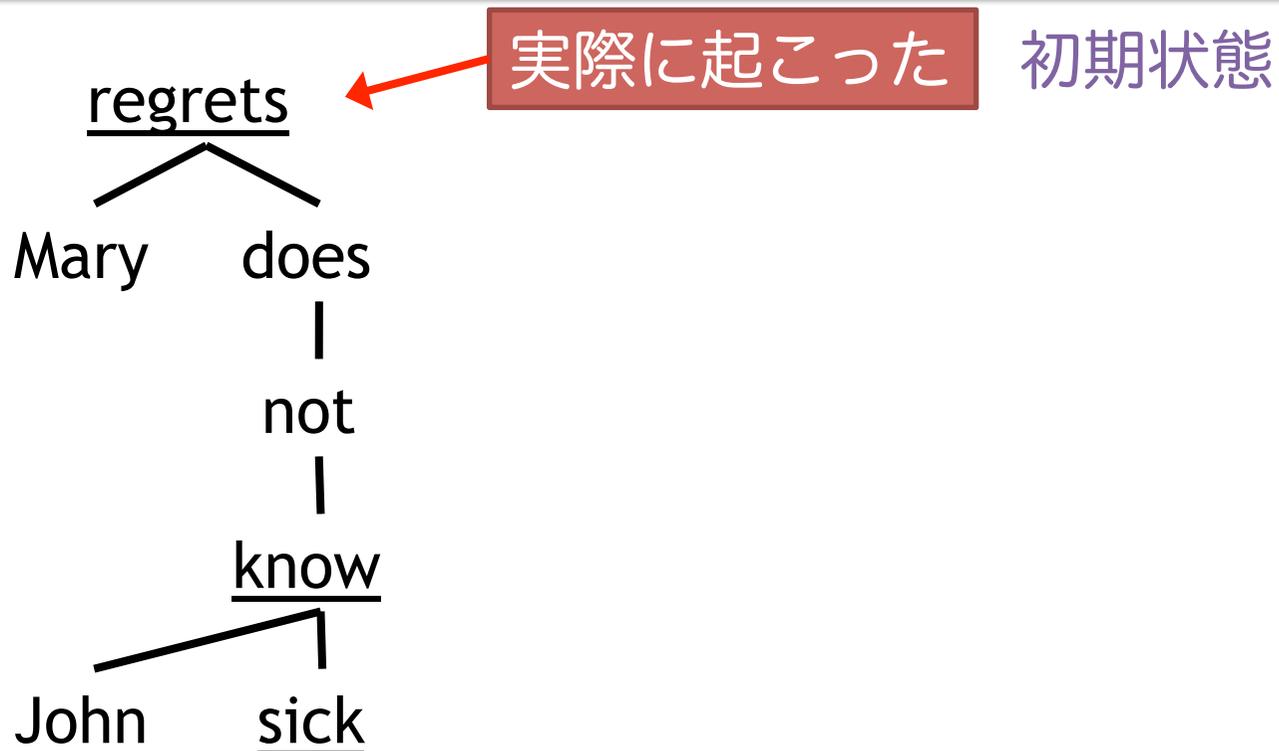
Mary regrets that John does not know he is sick.



## 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

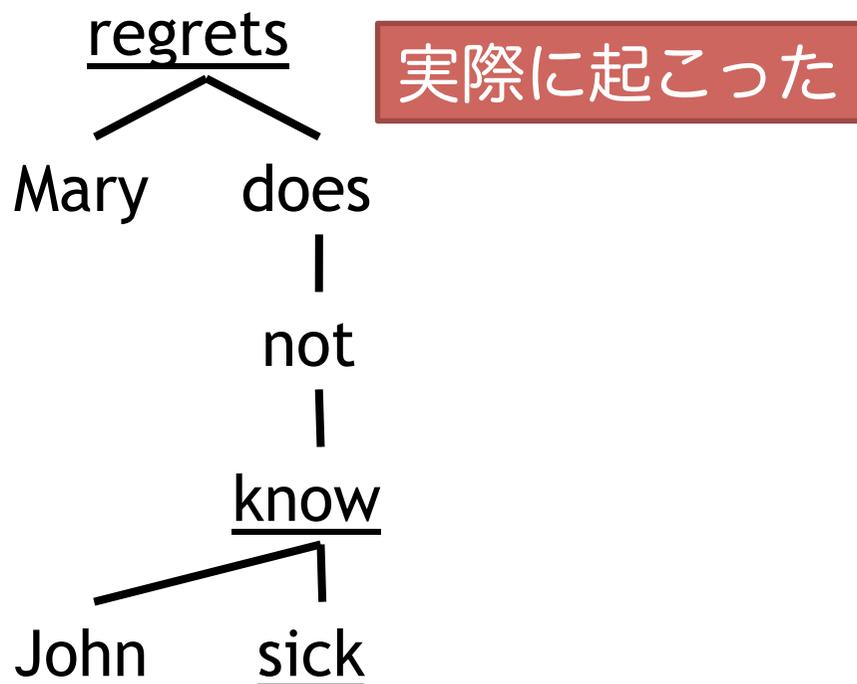
Mary regrets that John does not know he is sick.



## 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

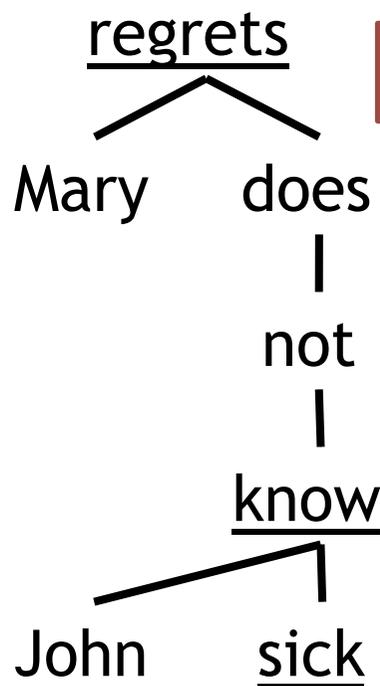
Mary regrets that John does not know he is sick.



# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.



実際に起こった

出力

出力

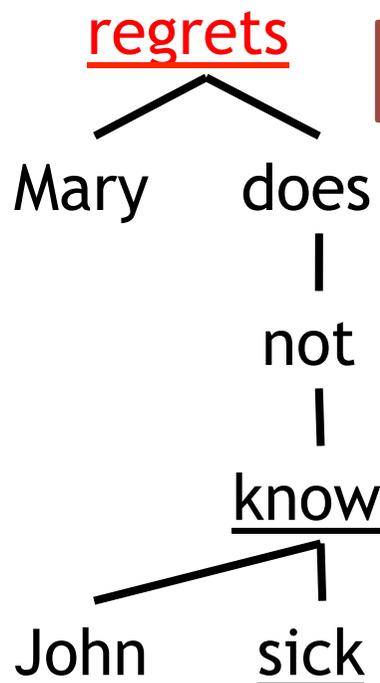
regrets:  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



実際に起こった

伝搬

更新なし

regrets → 起こった

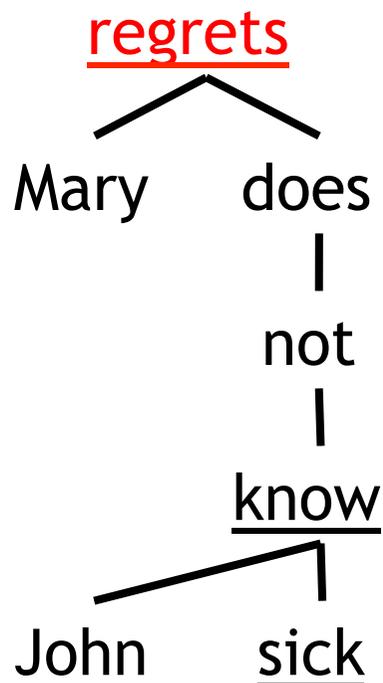
regrets:  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



実際に起こった

regrets → 起こった

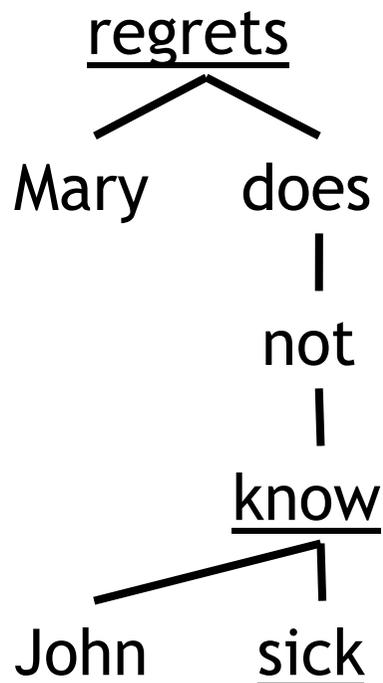
regrets:  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



実際に起こった

出力・更新なし

regrets:  
実際に起こった

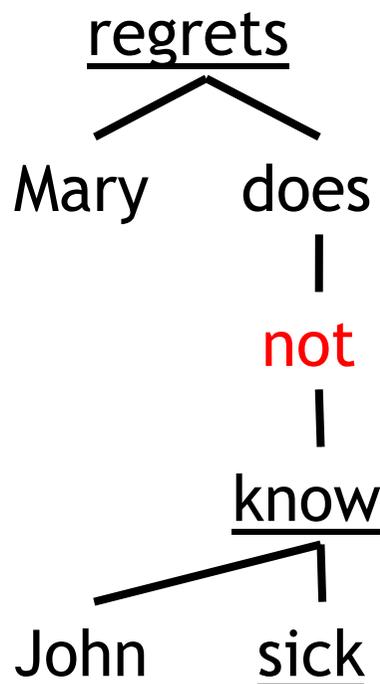
# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力

regrets:  
実際に起こった



実際に起こった

not → 起こっていない

更新

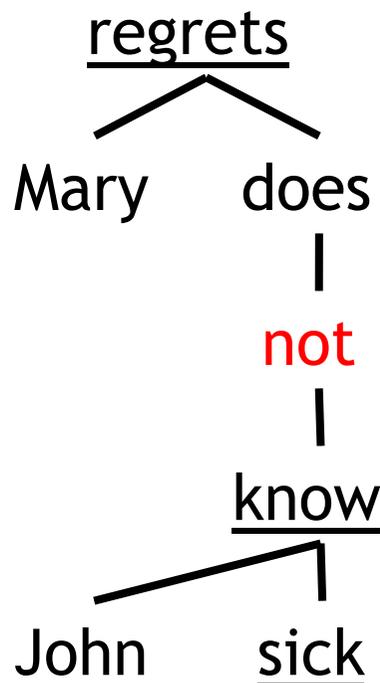
# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力

regrets:  
実際に起こった



実際には起こっていない

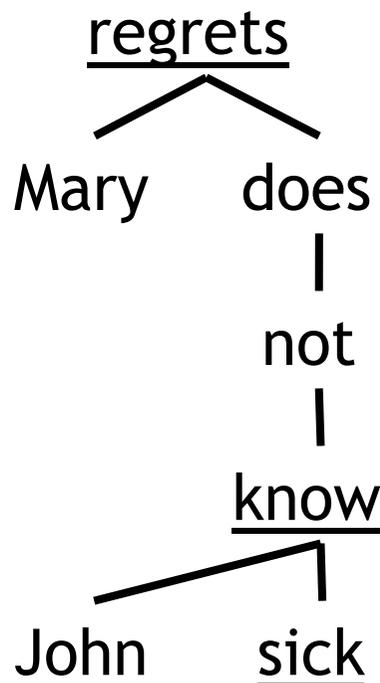
not → 起こっていない

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



regrets:  
実際に起こった

実際には起こっていない

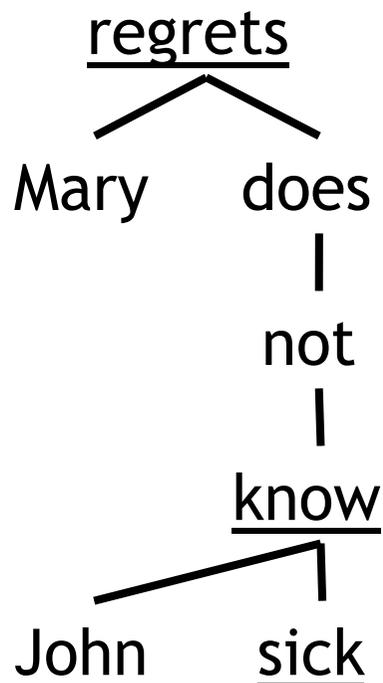
# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力

regrets:  
実際に起こった



出力

know:  
実際には起こっていない

実際には起こっていない

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力

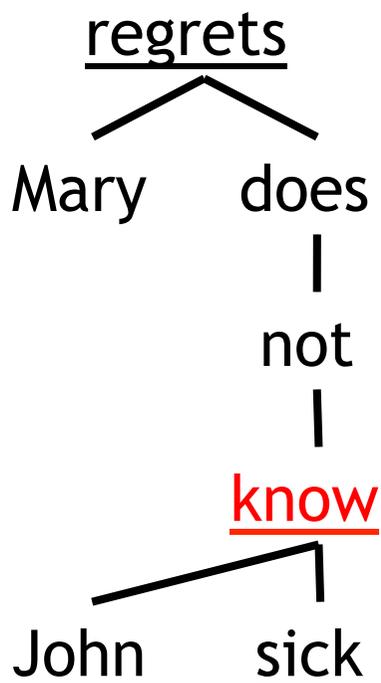
regrets:  
実際に起こった

know:  
実際には起こっていない

更新

実際には起こっていない

know →  
実際に起こった



# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

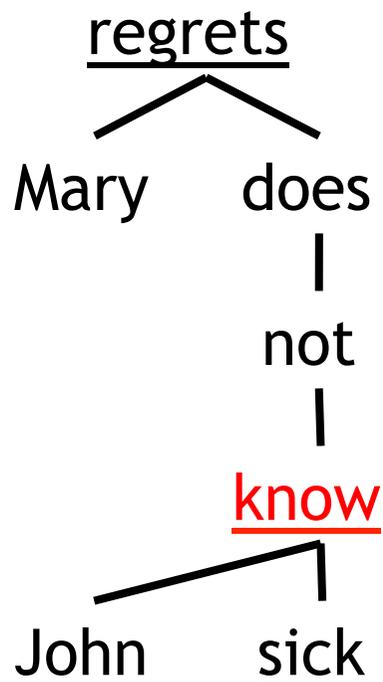
- 事実性に影響を与える手がかり表現を利用した、決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力

regrets:  
実際に起こった

know:  
実際には起こっていない



実際に起こった

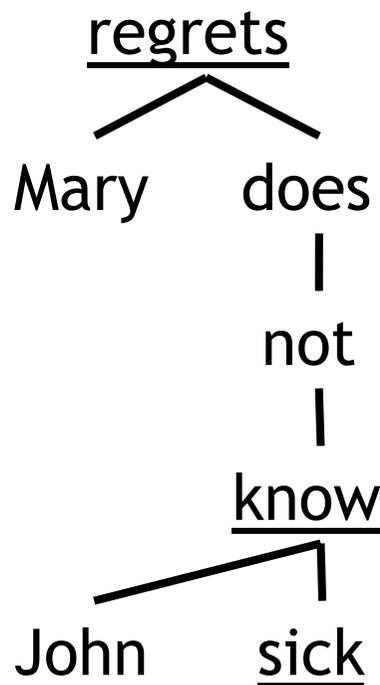
know →  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



regrets:  
実際に起こった

know:  
実際には起こっていない

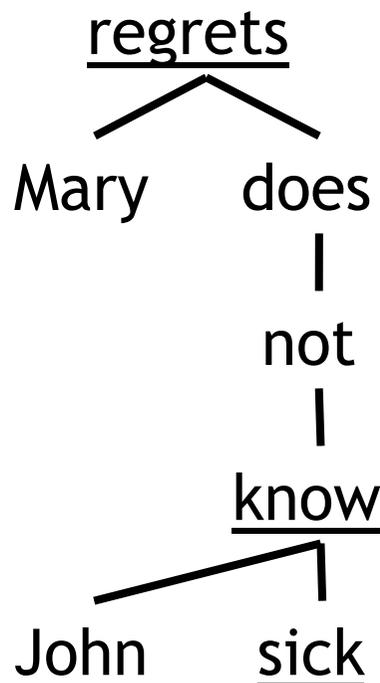
**実際に起こった**

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.

出力



regrets:  
実際に起こった

know:  
実際には起こっていない

出力

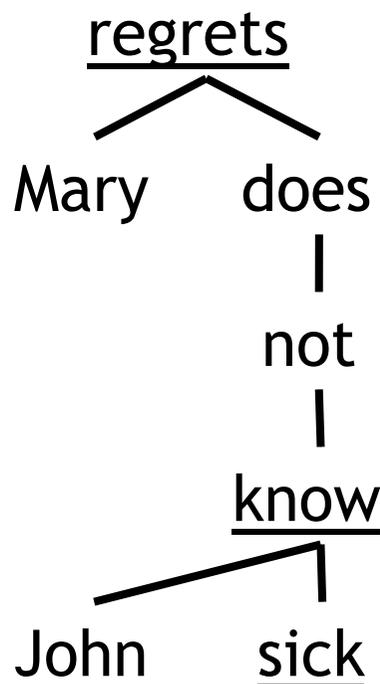
実際に起こった

sick:  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.



出力

regrets:  
実際に起こった

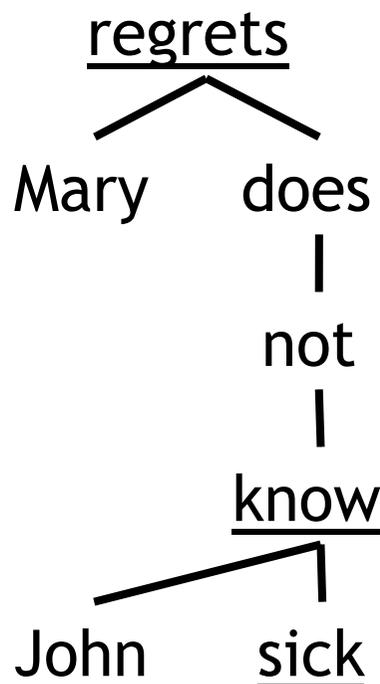
know:  
実際には起こっていない

sick:  
実際に起こった

# 語彙知識の組み合わせに基づく解析モデル[Saurí+ 07]

- 事実性に影響を与える**手がかり表現**を利用した、  
決定的解析モデル[Saurí+ 07]

Mary regrets that John does not know he is sick.



構成性を持ち、分析しやすい  
→日本語に適用・分析

出力

regrets:  
実際に起こった

know:  
実際には起こっていない

sick:  
実際に起こった

# 目次

- はじめに
  - 事実性とは
  - 本研究の目的
- 関連研究
- **日本語事実性解析器**
  - 事実性の定義
  - 使用する語彙知識
  - 解析モデル
- 実験
- 誤り分析
- おわりに

# 日本語事実性解析器：事実性の定義

- 事実性を確信度と肯否極性の組として定義
  - [Saurí+ 07]の定義を，日本語に合わせて一部変更

## 肯否極性

確信度

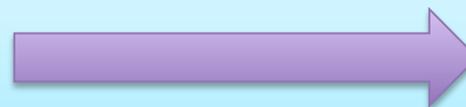
	<i>Positive (+)</i>	<i>Negative (-)</i>
<i>Certain (CT)</i>	実際に起こっている (CT+)	実際には起こっていない (CT-)
<i>Probable (PR)</i>	起こっている可能性が高い (PR+)	起こっていない可能性が高い (PR-)
<i>Underspecified (U)</i>	不明 (U)	

- 2軸にわかれているため，分析がしやすい

# 日本語事実性解析器：事実性の定義（例）

彼はさきほど部屋を出た。

「出る」：実際に起こったこと



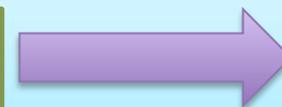
CT+

	<i>Positive (+)</i>	<i>Negative (-)</i>
<i>Certain (CT)</i>	実際に起こっている (CT+)	実際には起こっていない (CT-)
<i>Probable (PR)</i>	起こっている可能性が高い (PR+)	起こっていない可能性が高い (PR-)
<i>Underspecified (U)</i>	不明 (U)	

# 日本語事実性解析器：事実性の定義（例）

もう遅いから、彼は先に帰ったんだろう。

「帰る」：起こった可能性が高いこと



PR+

	<i>Positive (+)</i>	<i>Negative (-)</i>
<i>Certain (CT)</i>	実際に起こっている (CT+)	実際には起こっていない (CT-)
<i>Probable (PR)</i>	起こっている可能性が高い (PR+)	起こっていない可能性が高い (PR-)
<i>Underspecified (U)</i>	不明 (U)	

# 日本語事実性解析器：事実性の定義（例）

問題が発生するのを防いだ。

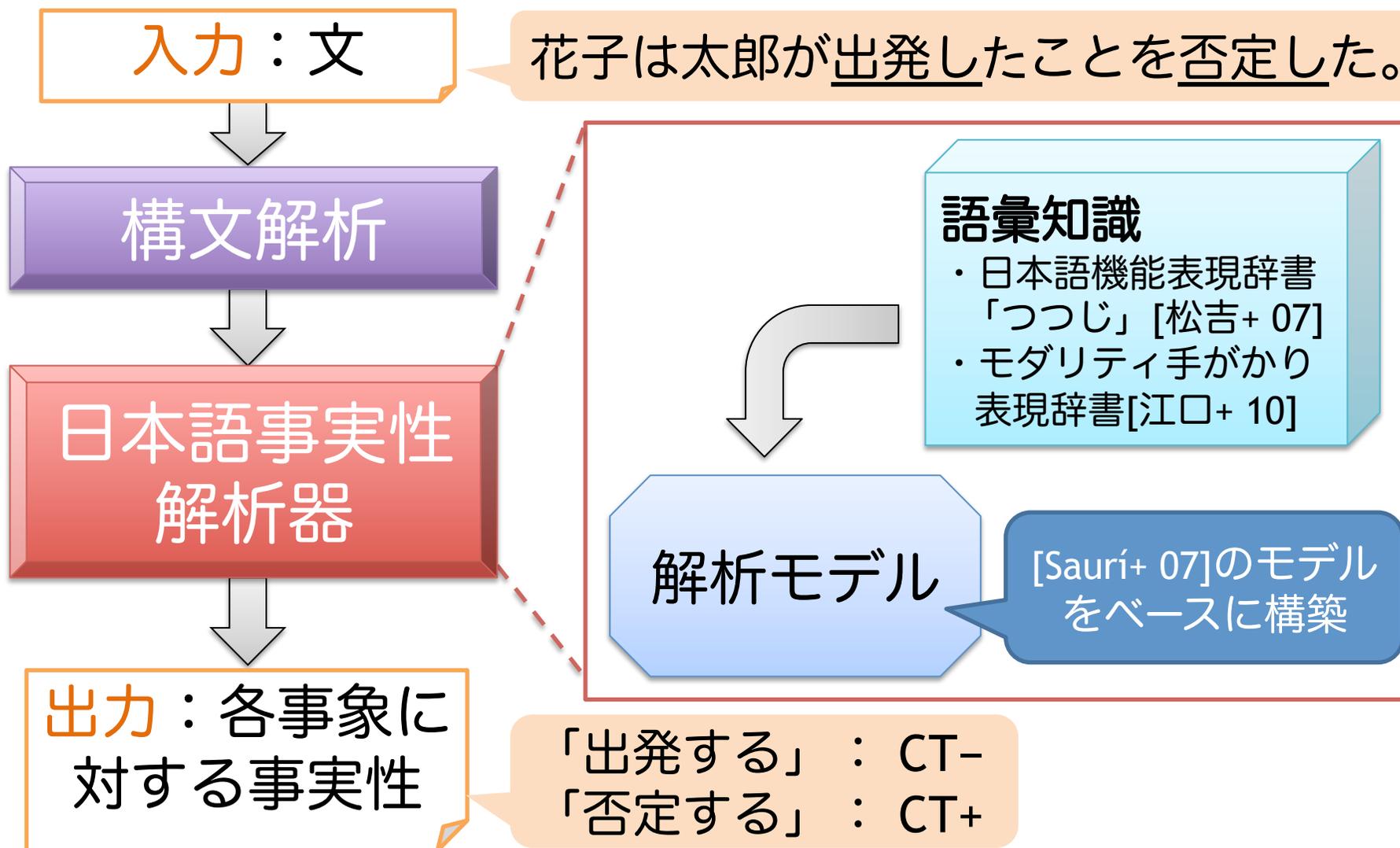
「発生する」：実際には起こっていないこと



CT-

	<i>Positive (+)</i>	<i>Negative (-)</i>
<i>Certain (CT)</i>	実際に起こっている (CT+)	実際には起こっていない (CT-)
<i>Probable (PR)</i>	起こっている可能性が高い (PR+)	起こっていない可能性が高い (PR-)
<i>Underspecified (U)</i>	不明 (U)	

# 日本語事実性解析器



# 解析モデル

- 語彙知識に基づいて、事実性の更新を行いながら伝搬を行っていくアルゴリズムを構築
  - [Saurí+ 07]のモデルを日本語に適用

事象間には依存関係があり、事実性を伝搬させることでそれを反映させる

花子は太郎が出発したことを否定した。

# 解析モデル

- 語彙知識に基づいて、事実性の更新を行いながら伝搬を行っていくアルゴリズムを構築
  - [Saurí+ 07]のモデルを日本語に適用

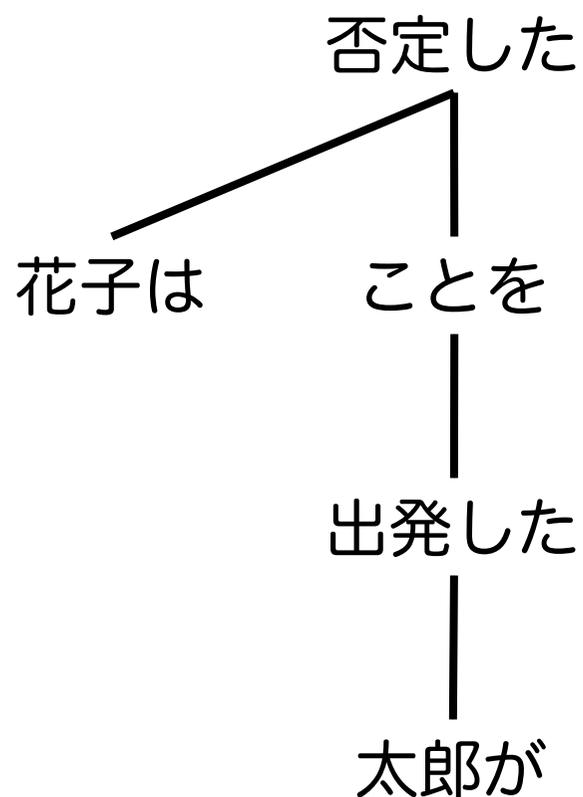
事象間には依存関係があり、事実性を伝搬させることでそれを反映させる

花子は太郎が出発したことを否定した。

「出発する」の事実性は「否定する」の影響を受ける

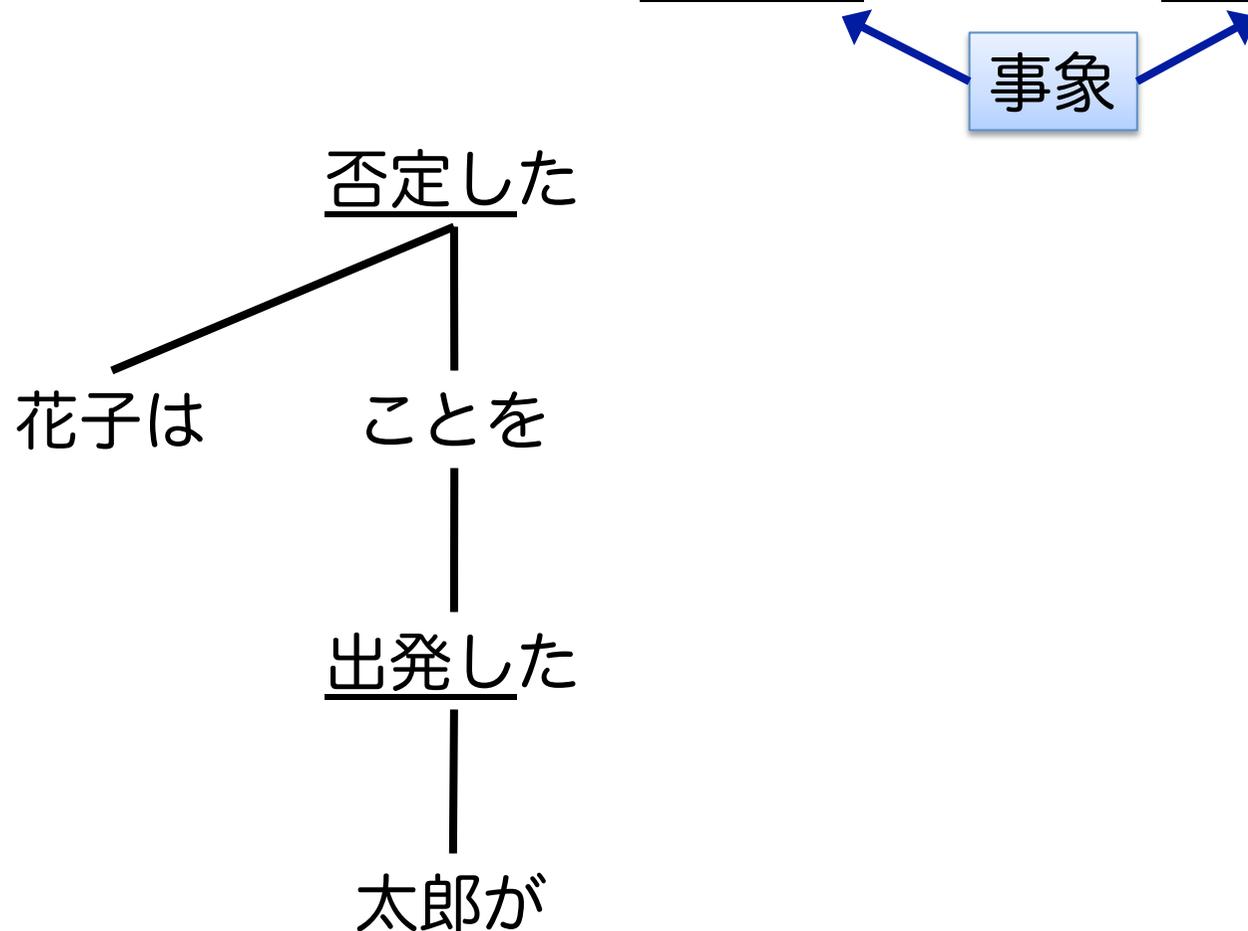
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



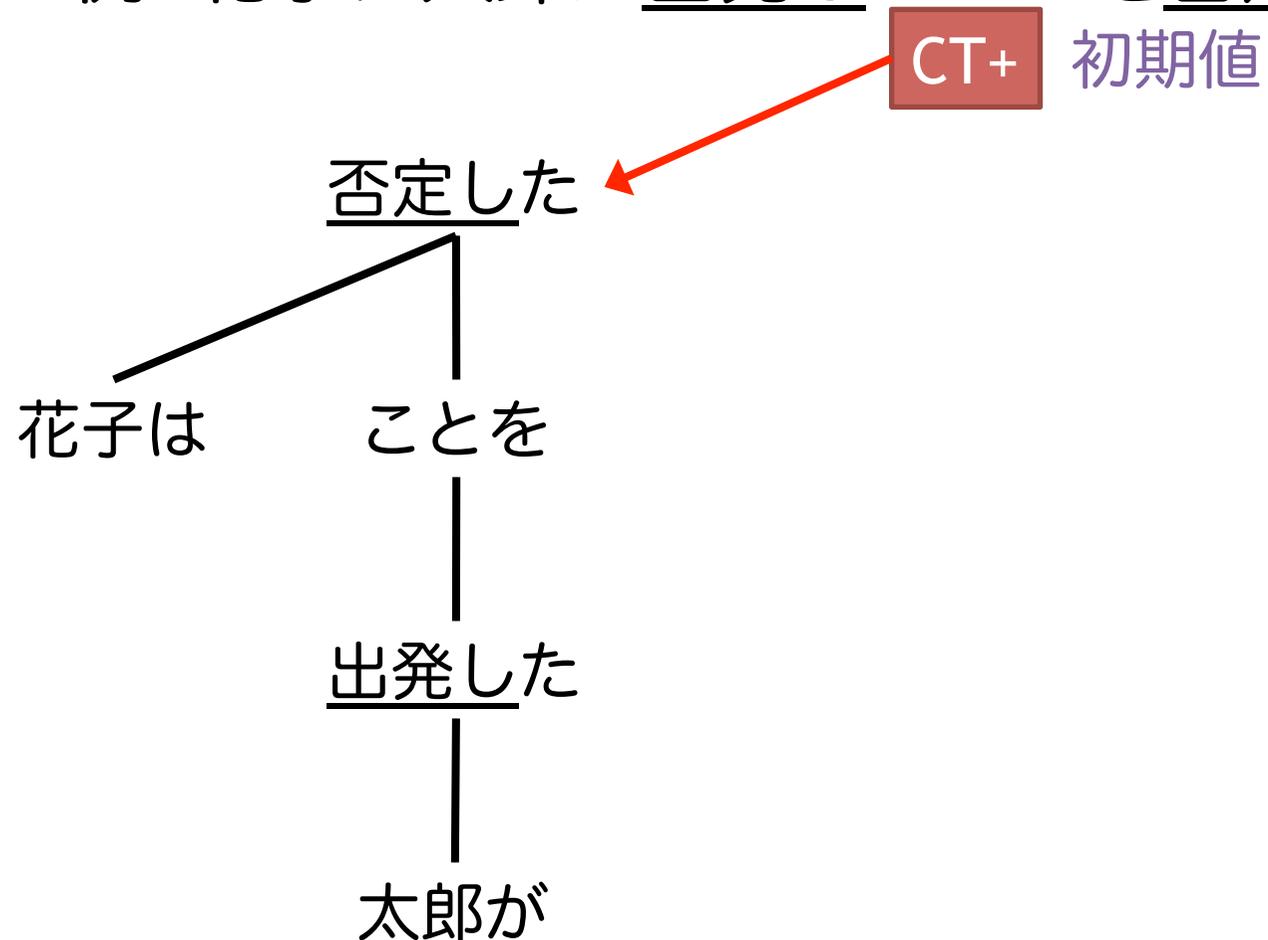
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



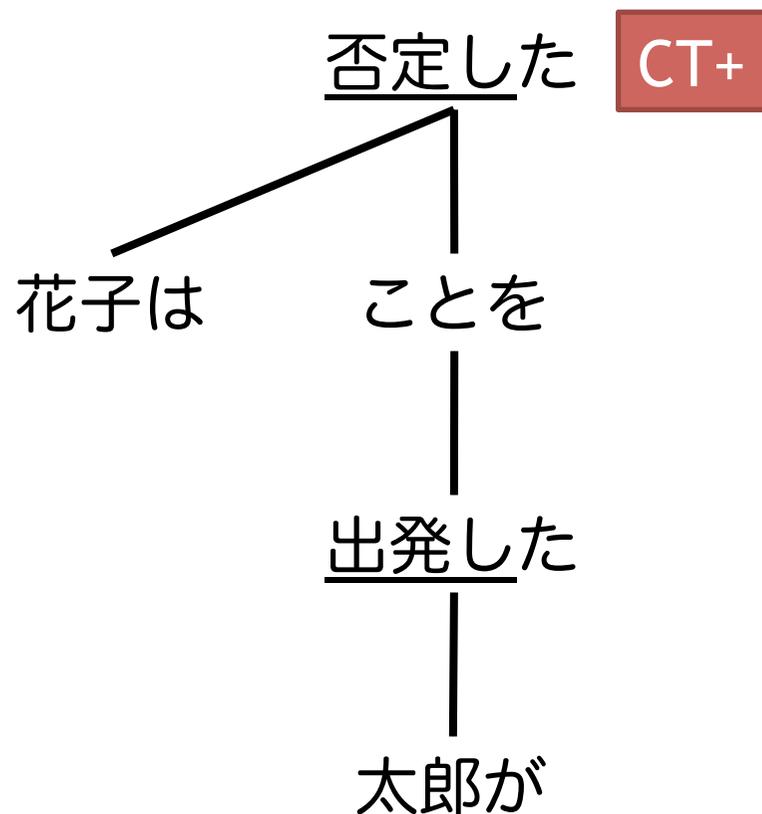
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



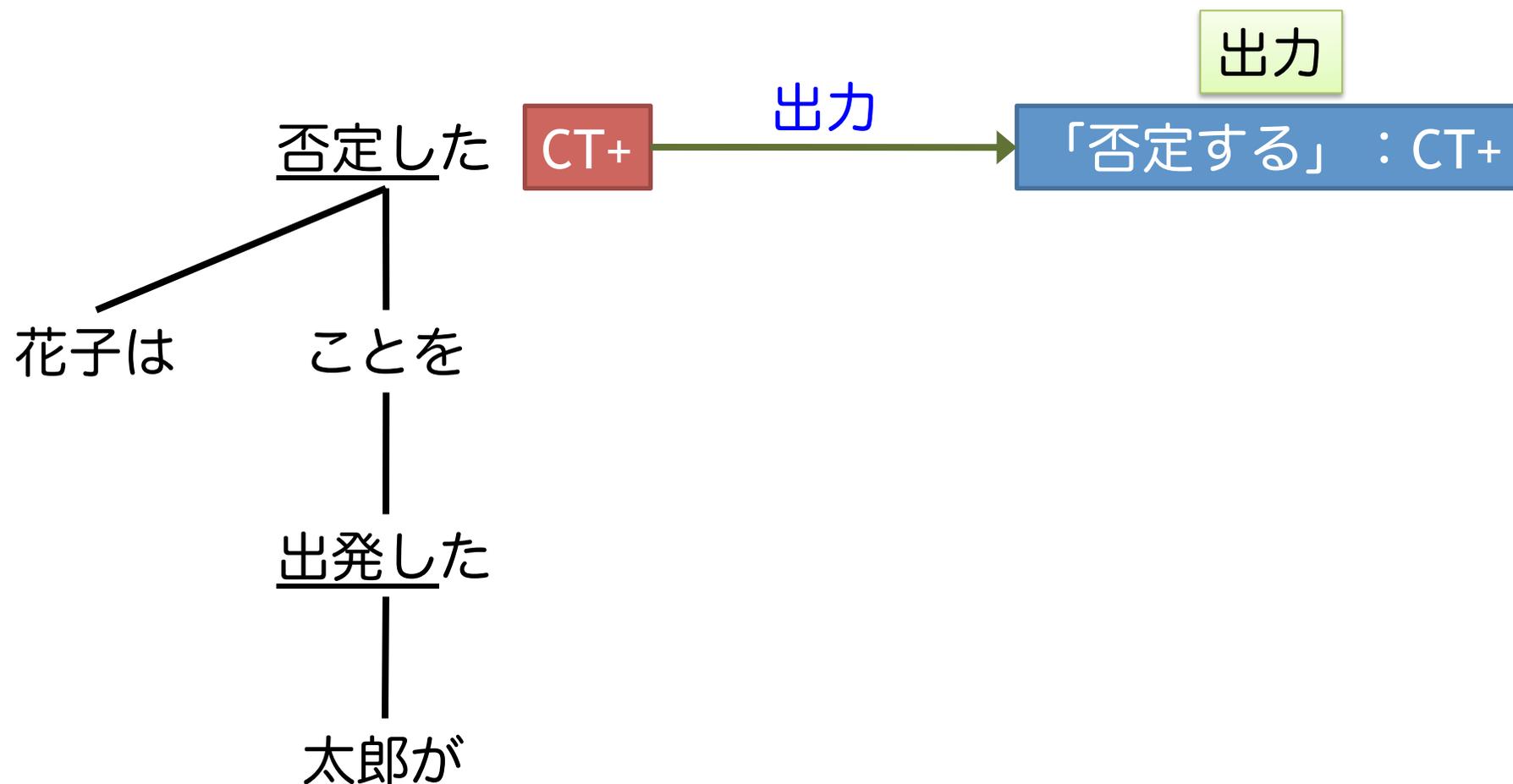
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



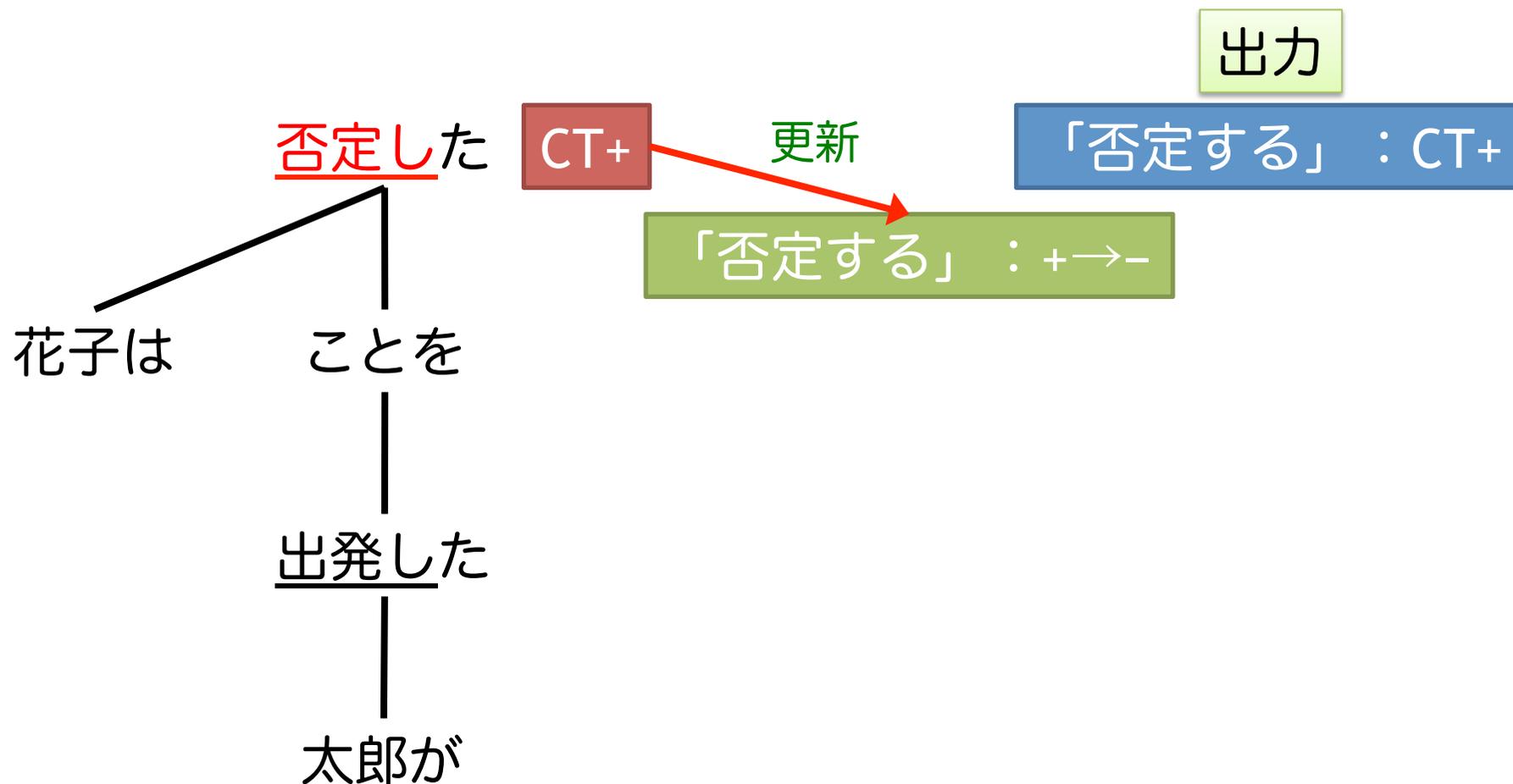
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



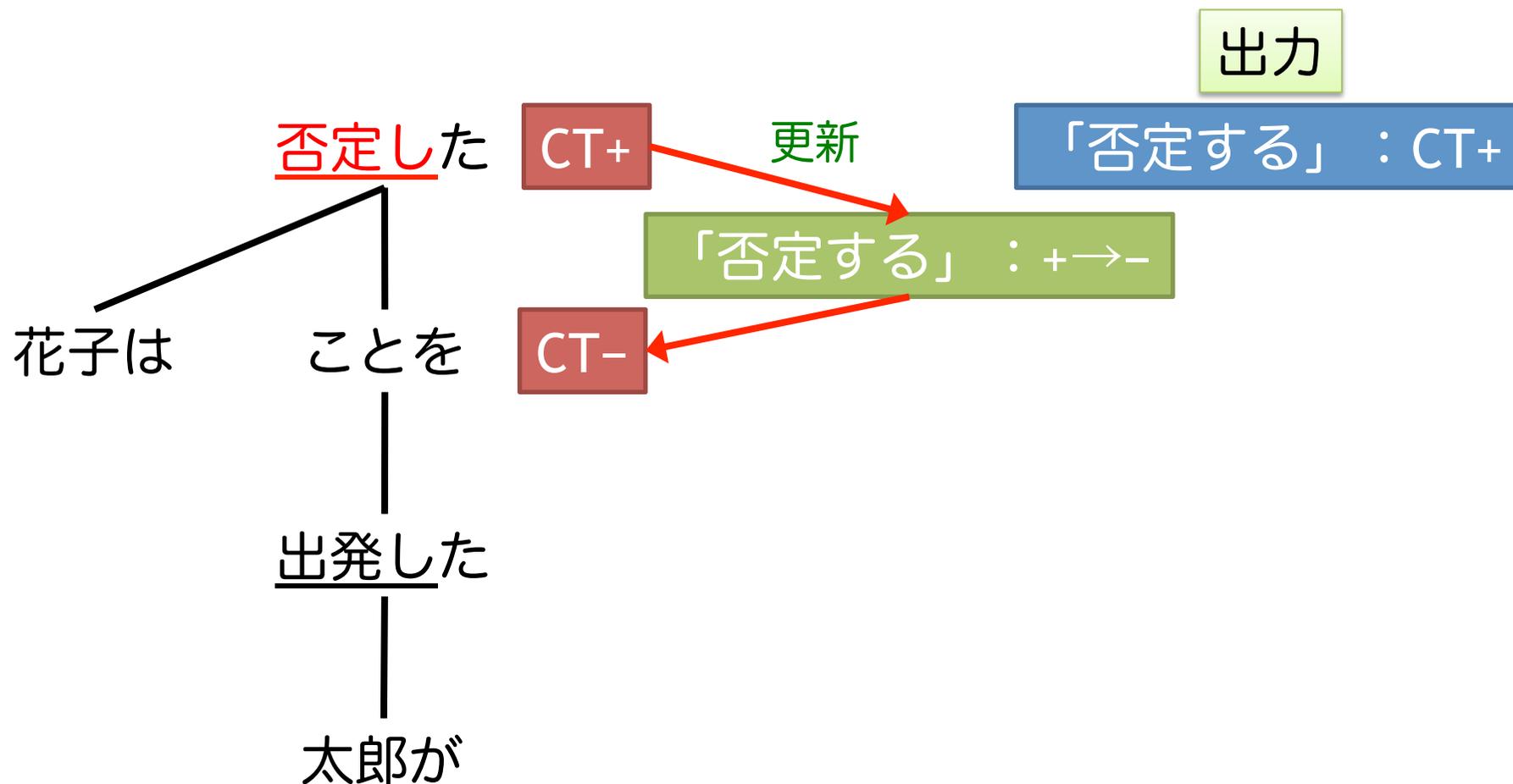
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



# 解析モデル

- 例. 花子は太郎が出発したことを否定した。

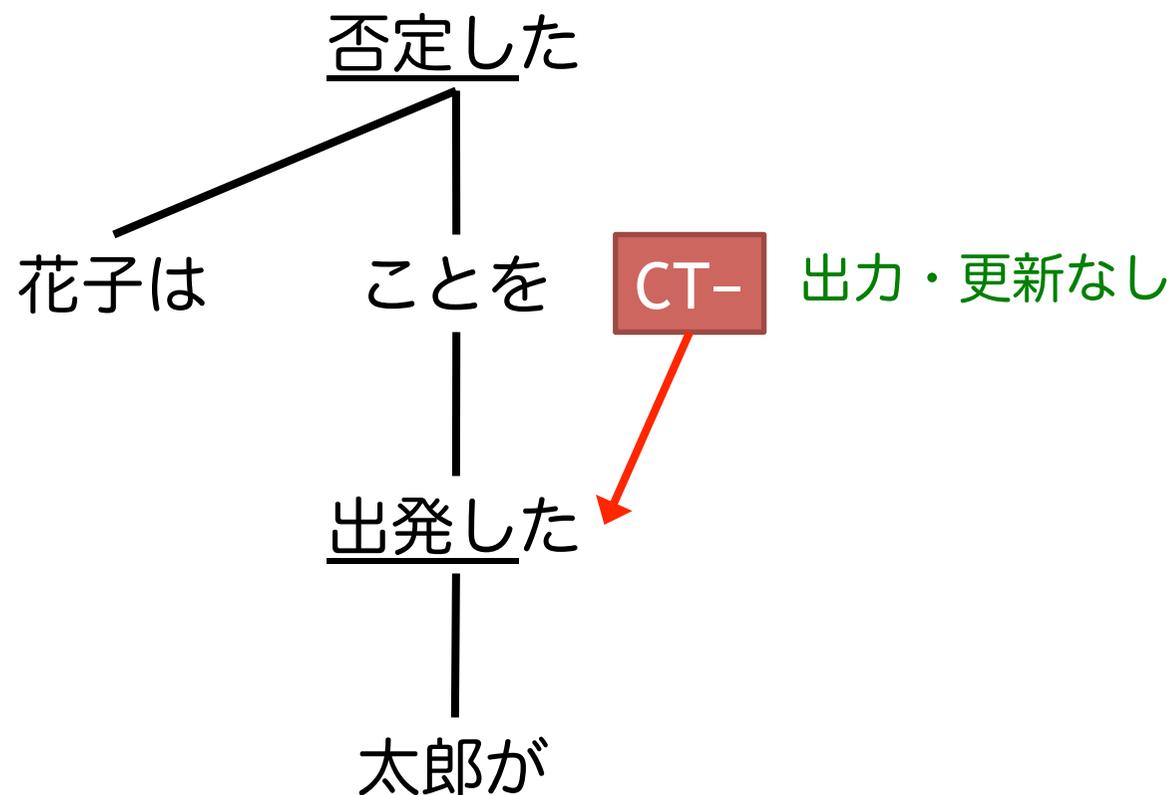


# 解析モデル

- 例. 花子は太郎が出発したことを否定した。

出力

「否定する」 : CT+

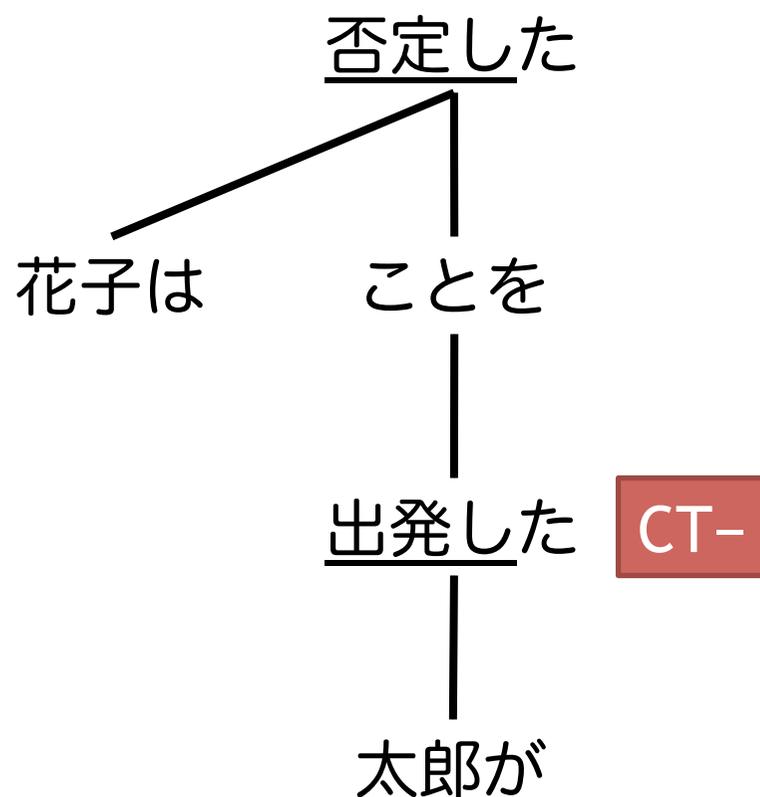


# 解析モデル

- 例. 花子は太郎が出発したことを否定した。

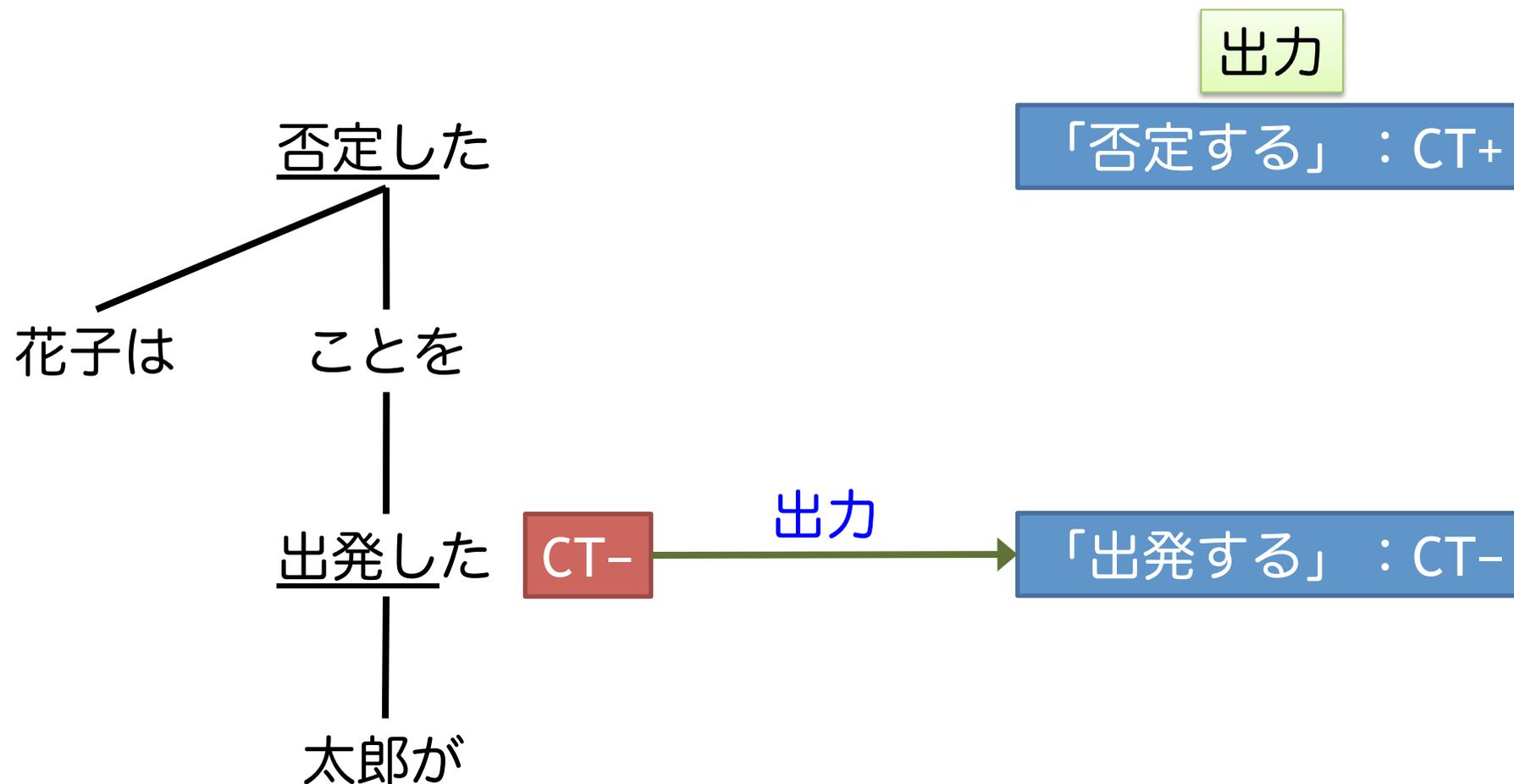
出力

「否定する」 : CT+



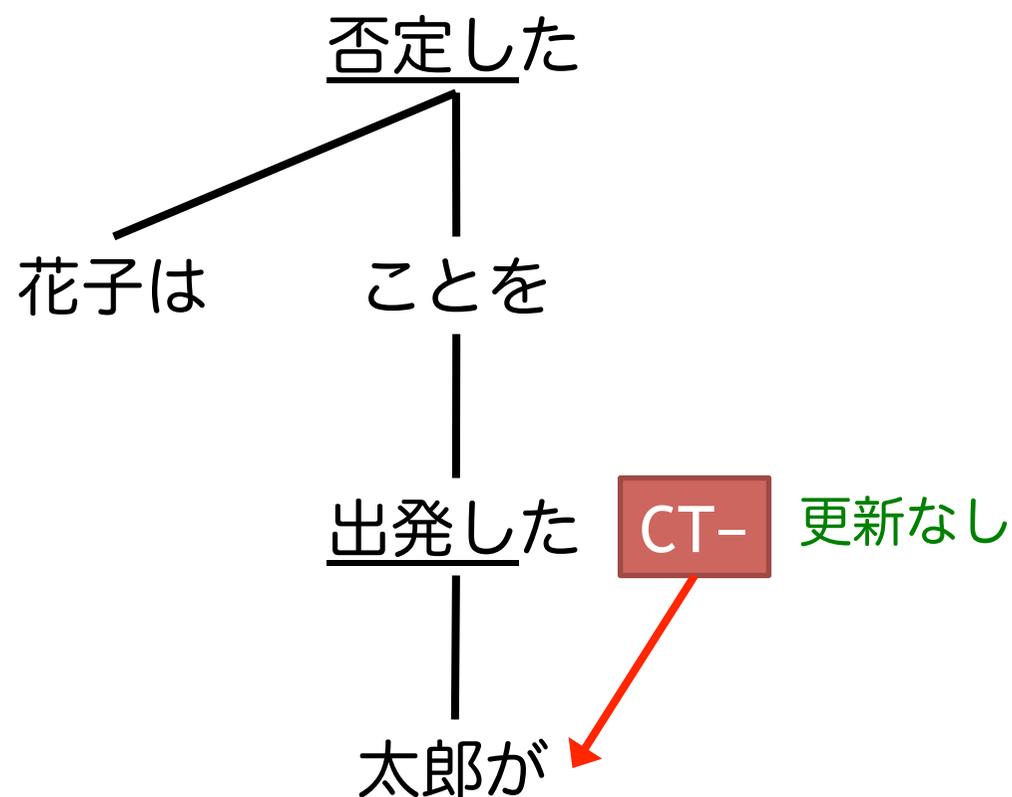
# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



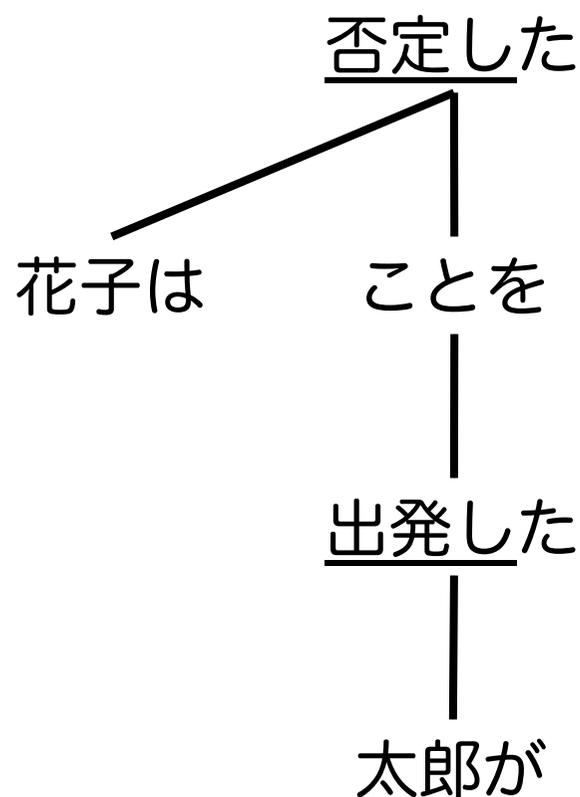
出力

「否定する」 : CT+

「出発する」 : CT-

# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



出力

「否定する」 : CT+

「出発する」 : CT-

CT-

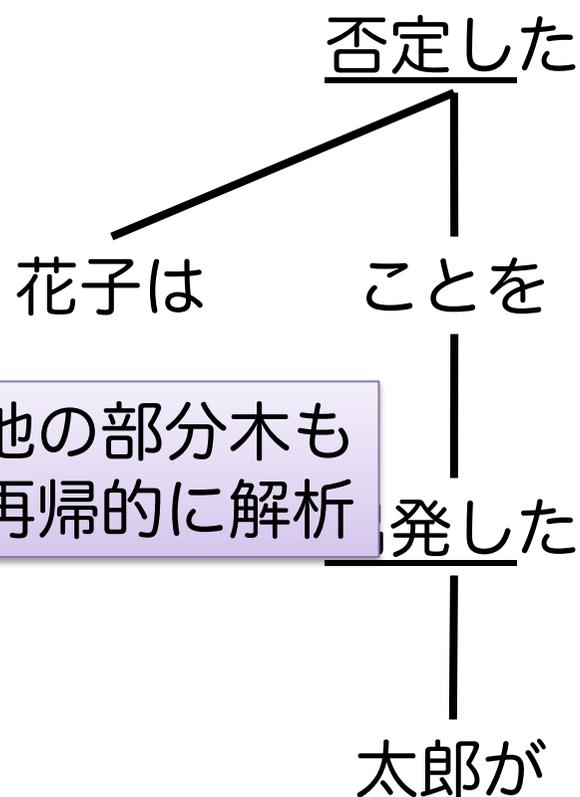
出力・更新なし

# 解析モデル

- 例. 花子は太郎が出発したことを否定した。

出力

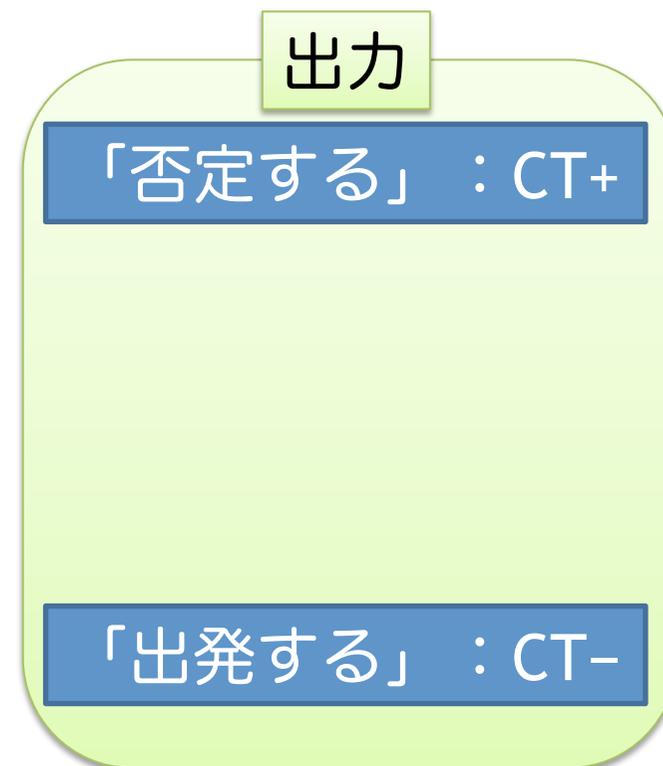
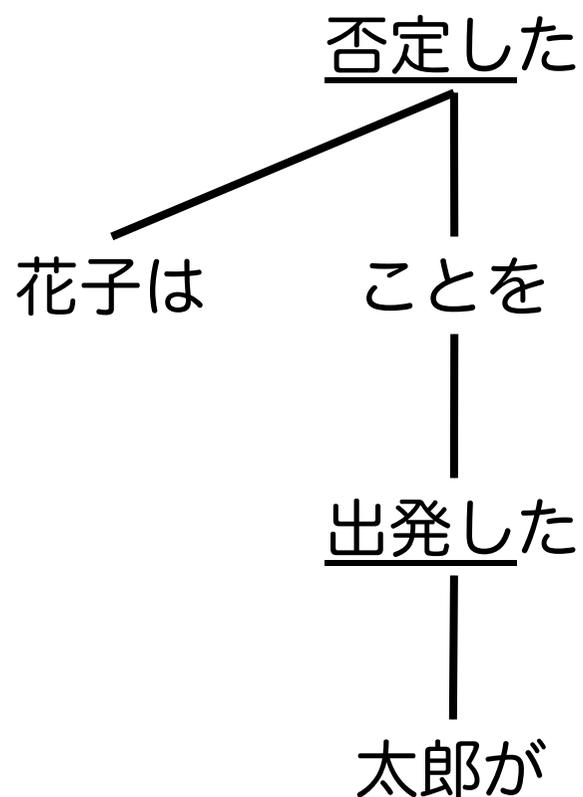
「否定する」 : CT+



「出発する」 : CT-

# 解析モデル

- 例. 花子は太郎が出発したことを否定した。



# 使用する語彙知識

- 事実性に影響を与える手がかり表現の語彙知識
  - 機能語：日本語機能表現辞書「つつじ」[松吉+ 07]
  - 内容語：モダリティ解析手がかり表現辞書[江口+ 10]

事実性に影響を与える表現が網羅的に収録されている

# 使用する語彙知識

- 事実性に影響を与える手がかり表現の語彙知識
  - 機能語：日本語機能表現辞書「つつじ」[松吉+ 07]
  - 内容語：モダリティ解析手がかり表現辞書[江口+ 10]

もう遅いから、彼は先に帰ったんだらう。

「帰る」：起こった可能性が高いこと



PR+

問題が発生するのを防いだ。

「発生する」：実際には起こっていないこと



CT-

# 使用する語彙知識

- 事実性に影響を与える手がかり表現の語彙知識

- 機能語：日本語機能表現辞書「つつじ」[松吉+ 07]

- 内容語：モダリティ

「だろう」：推量の機能語＝確信度→PR

もう遅いから、彼は先に帰ったんだだろう。

「帰る」：起こった可能性が高いこと



PR+

問題が発生するのを防いだ。

「発生する」：実際には起こっていないこと



CT-

# 使用する語彙知識

- 事実性に影響を与える手がかり表現の語彙知識

- 機能語：日本語機能表現辞書「つつじ」[松吉+ 07]

- 内容語：モダリティ

「だろう」：推量の機能語＝確信度→PR

もう遅いから、彼は先に帰ったんだだろう。

「帰る」：起こった可能性が高いこと



PR+

問題が発生するのを防いだ。

「発生する」：実際には起こっていないこと



CT-

「防いだ」：肯否極性を反転させる内容語＝肯否極性→-

# 目次

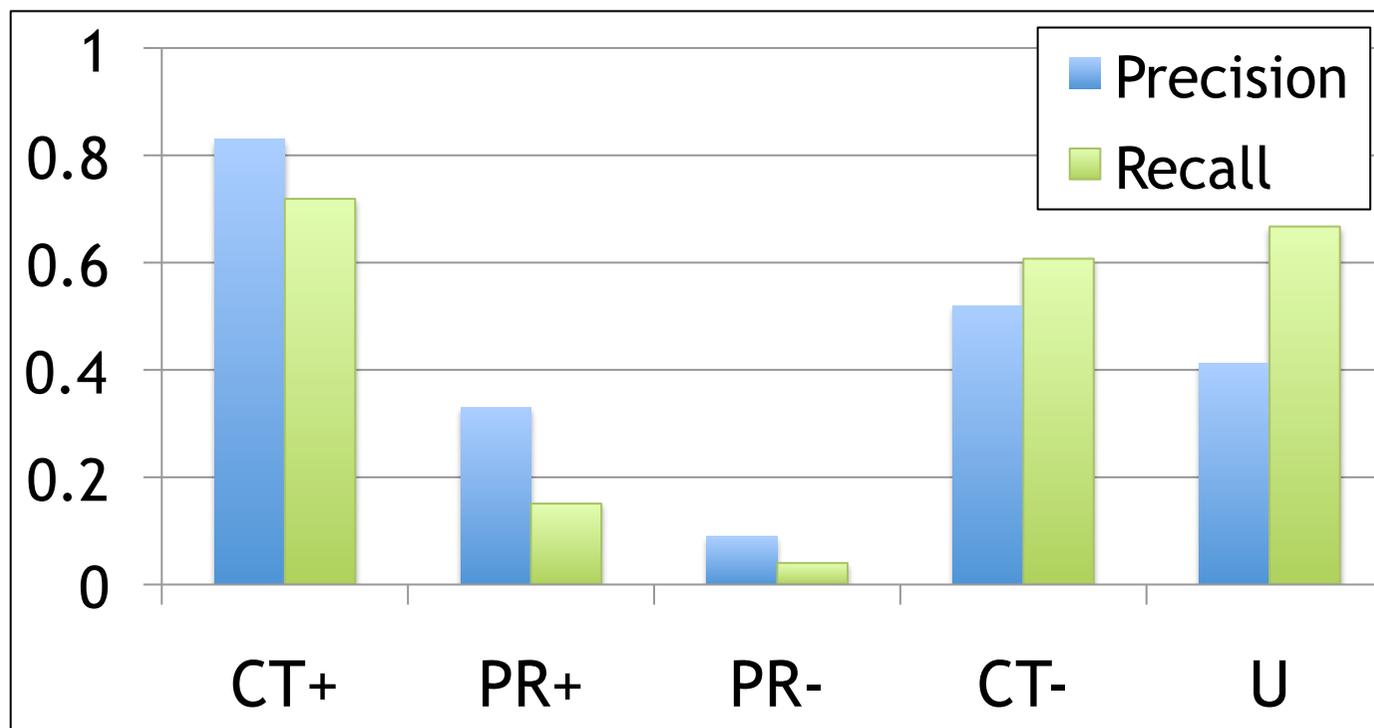
- はじめに
  - 事実性とは
  - 本研究の目的
- 関連研究
- 日本語事実性解析器
  - 事実性の定義
  - 解析モデル
  - 使用する語彙知識
- **実験**
- 誤り分析
- おわりに

# 実験

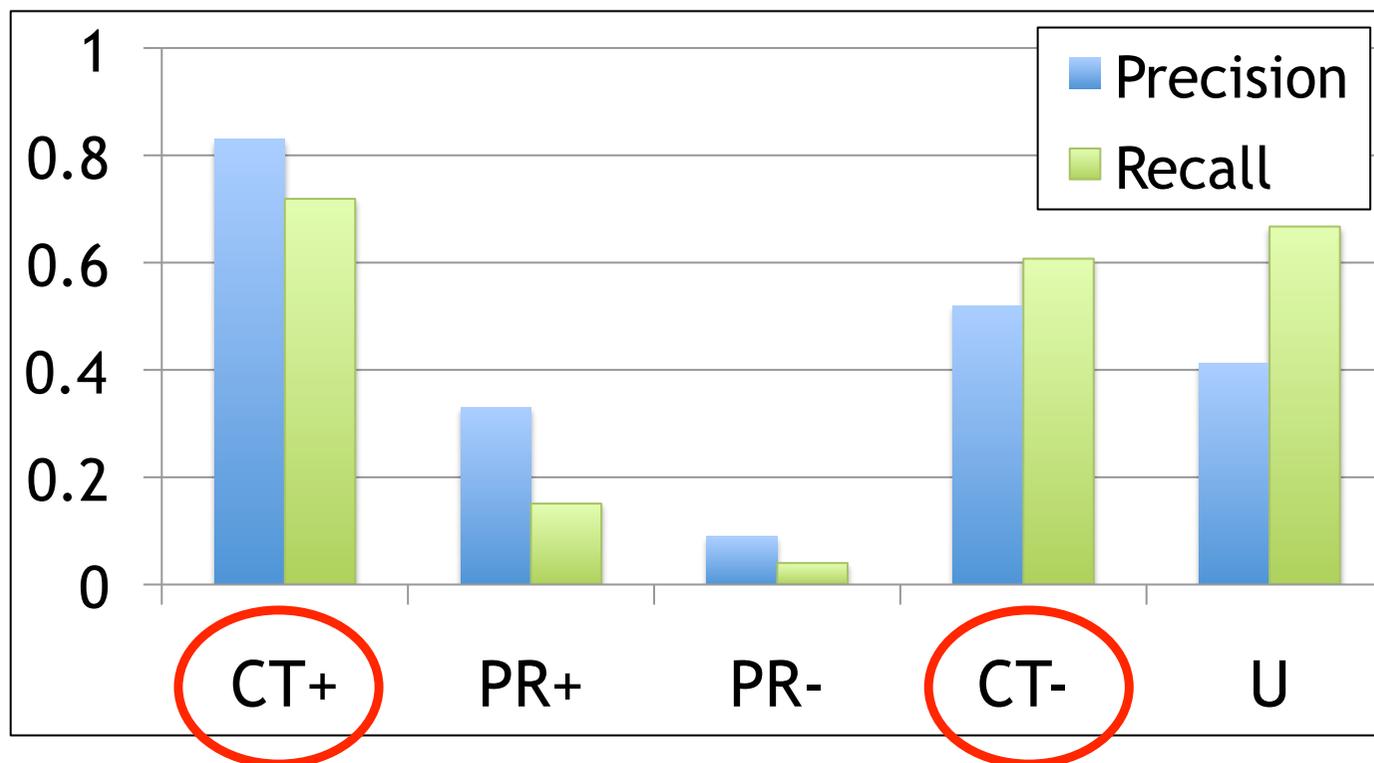
- 拡張モダリティタグ付与コーパス[松吉+ 10] のOC (Yahoo!知恵袋) 6,404文に対して適用
  - 入力：
    - 正解の形態素情報をもとに，構文解析を行った結果
    - 事実性を付与すべき事象の情報
  - 出力：
    - 各事象に対する事実性
  - 評価：
    - それぞれのラベル（CT+, CT-, PR+, PR-, U）に対して評価

事実性に影響を与える語彙知識に焦点を当て、  
分析を行う

# 実験：評価

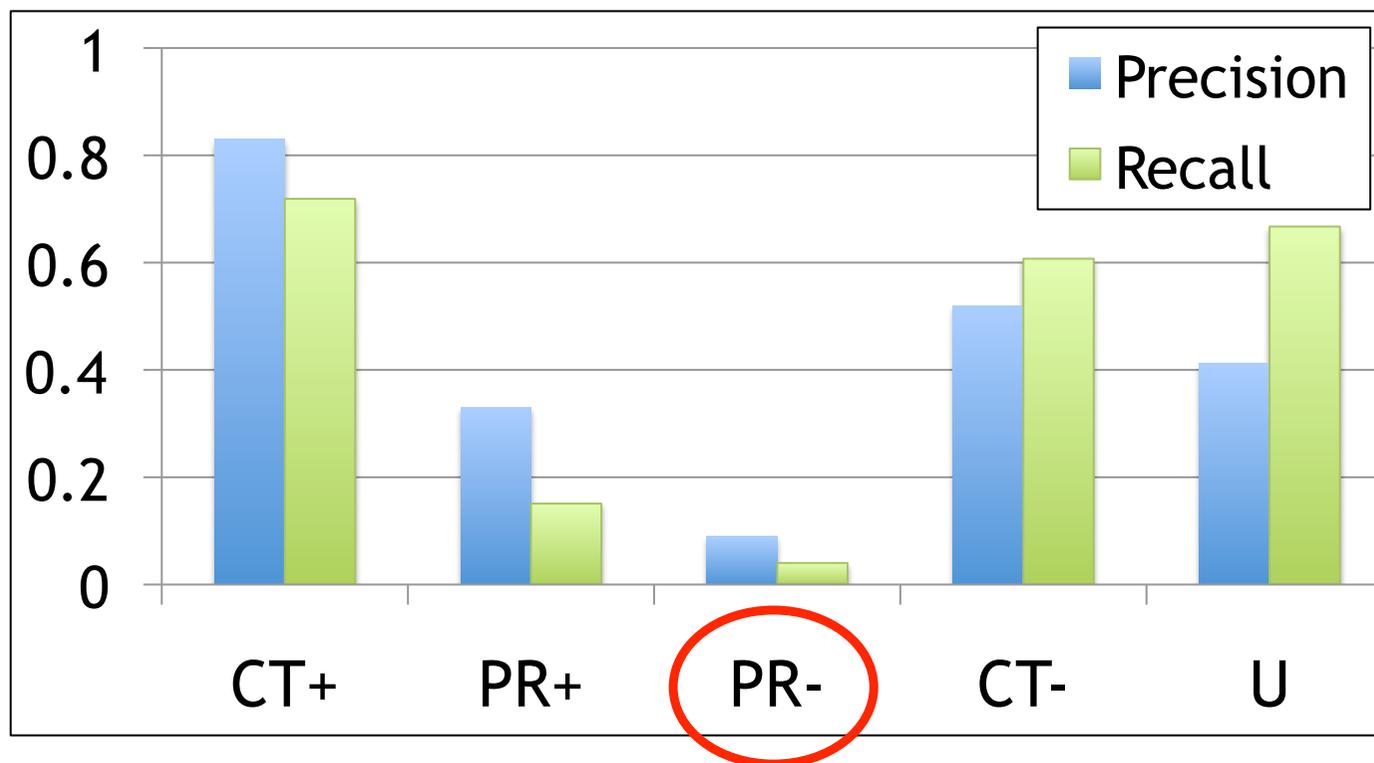


# 実験：評価



ナイーブな手法だが、CT+, CT-は比較的解けている

# 実験：評価



確信度，肯否極性両方の更新が必要であるため，PR-はあまり解けていない

# 目次

- はじめに
  - 事実性とは
  - 本研究の目的
- 関連研究
- 日本語事実性解析器
  - 事実性の定義
  - 解析モデル
  - 使用する語彙知識
- 実験
- 誤り分析
- おわりに

# 誤り分析

事実性に影響を与える語彙知識に  
焦点を当て、分析を行う

- 論点：
  - 現在利用可能な語彙知識のカバレッジ
  - 語彙知識の曖昧性の有無
  - 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無
  - その他

# 誤り分析

- 論点： 肯否極性に関する200の誤り事例を分類
  - 現在利用可能な語彙知識のカバレッジ (26%)
  - 語彙知識の曖昧性の有無 (9%)
  - 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無 (37%)
  - その他 (28%)
    - 構文解析誤り, アノテーション誤りなど

# 誤り分析

- 論点： 肯否極性に関する200の誤り事例を分類
  - 現在利用可能な語彙知識のカバレッジ (26%)
    - 内容語, 機能語
  - 語彙知識の曖昧性の有無 (9%)
  - 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無 (37%)
    - 手がかりとなる表現が影響を与える範囲の問題
  - その他 (28%)
    - 構文解析誤り, アノテーション誤りなど

# 誤り分析

- 語彙知識のカバレッジの問題（内容語）

サーバーは接続を解除しました。

- 「接続」正解：CT-, 出力：CT+

内容語「解除する」に関する知識不足

語彙知識の拡充が必要

# 誤り分析

- 語彙知識のカバレッジの問題（機能語）

いまいち説明しにくいです。

- 「説明する」正解：PR-, 出力：CT+

機能語「にくい」に関する知識不足

他の研究でも追加の必要性は  
言及されている[今村+ 11]

# 誤り分析

- 手がかりとなる表現が影響を与える範囲の問題

合理化して便利になることを追求できなかった。

- 「合理化する」正解：CT+, 出力：CT-

# 誤り分析

- 手がかりとなる表現が影響を与える範囲の問題

係り受け

合理化して / 便利になる / ことを / 追求できなかった。

- 「合理化する」 正解：CT+, 出力：CT-

「追求できなかった」の事実性が  
「合理化する」まで伝搬してしまう

# 誤り分析

- 手がかりとなる表現が影響を与える範囲の問題

係り受け

合理化して / 便利になる / ことを / 追求できなかった。

- 「合理化する」 正解：CT+, 出力：CT-

「追求できなかった」の事実性が  
「合理化する」まで伝搬してしまう

影響を与える範囲（スコープ）の特定が必要

# 事象の位置による違い

- 事象の位置に着目した正解率
  - 合理化して便利であることを追求できなかった。

	Accuracy
外側（文末）の事象	79.7% (4691/5885)
内側（文末以外）の事象	56.7% (5121/9032)

# 事象の位置による違い

- 事象の位置に着目した正解率
  - 合理化して便利であることを追求できなかった。

	Accuracy
外側（文末）の事象	79.7% (4691/5885)
内側（文末以外）の事象	56.7% (5121/9032)

「たぶん」などの副詞を考慮できていないため、十分ではないが、正解率は高い

# 事象の位置による違い

- 事象の位置に着目した正解率
  - 合理化して便利であることを追求できなかった。

	Accuracy
外側（文末）の事象	79.7% (4691/5885)
内側（文末以外）の事象	56.7% (5121/9032)

「たぶん」などの副詞を考慮できていないため、十分ではないが、正解率は高い

機能語により  
多くが正解できている

# 事象の位置による違い

- 事象の位置に着目した正解率
  - 合理化して便利であることを追求できなかった。

	Accuracy
外側（文末）の事象	79.7% (4691/5885)
内側（文末以外）の事象	56.7% (5121/9032)

「たぶん」などの副詞を考慮できていないため、十分ではないが、正解率は高い

スコープの問題が多く存在する

機能語により多くが正解できている

# 目次

- はじめに
  - 事実性とは
  - 本研究の目的
- 関連研究
- 日本語事実性解析器
  - 事実性の定義
  - 解析モデル
  - 使用する語彙知識
- 実験
- 誤り分析
- おわりに

## おわりに

- 語彙知識の組み合わせに基づいた事実性解析
  - 現在利用可能な語彙知識のカバレッジ
  - 語彙知識の曖昧性の有無
  - 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無

## おわりに

- 語彙知識の組み合わせに基づいた事実性解析
  - 現在利用可能な語彙知識のカバレッジ

特に内容語に対する語彙知識の不足

- 語彙知識の曖昧性の有無
- 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無

## おわりに

- 語彙知識の組み合わせに基づいた事実性解析
  - 現在利用可能な語彙知識のカバレッジ

特に内容語に対する語彙知識の不足

- 語彙知識の曖昧性の有無

存在しているが、既存研究によりカバーされている部分も多い

- 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無

## おわりに

- 語彙知識の組み合わせに基づいた事実性解析
  - 現在利用可能な語彙知識のカバレッジ

特に内容語に対する語彙知識の不足

- 語彙知識の曖昧性の有無

存在しているが、既存研究によりカバーされている部分も多い

- 局所的な語彙知識の組み合わせだけでは解析不可能な問題の有無

スコープが大きな問題