

知識に基づく構造的言語処理の確立と 知識インフラの構築

(2013.10－2019.3)



Research Director Sadao KUROHASHI (Kyoto University)

Main Research Collaborator Kentaro INUI
(Tohoku U)



Daisuke BEKKI
(Ochanomizu U)



Knowledge
Infrastructure

Language Understanding Knowledge Acquisition

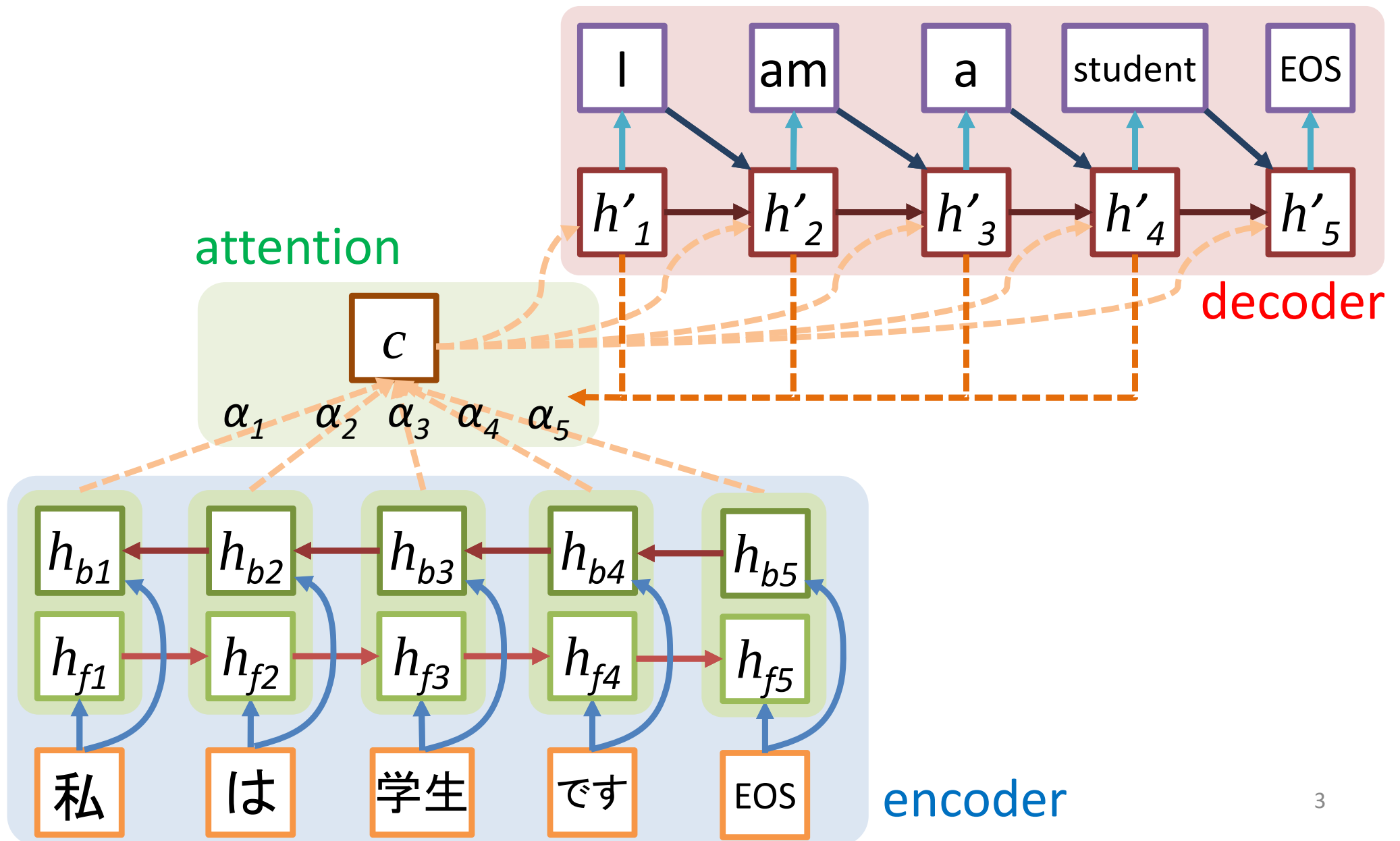
Deep Learning

Deep Learning

*Big Data
(Huge Text Corpora)*

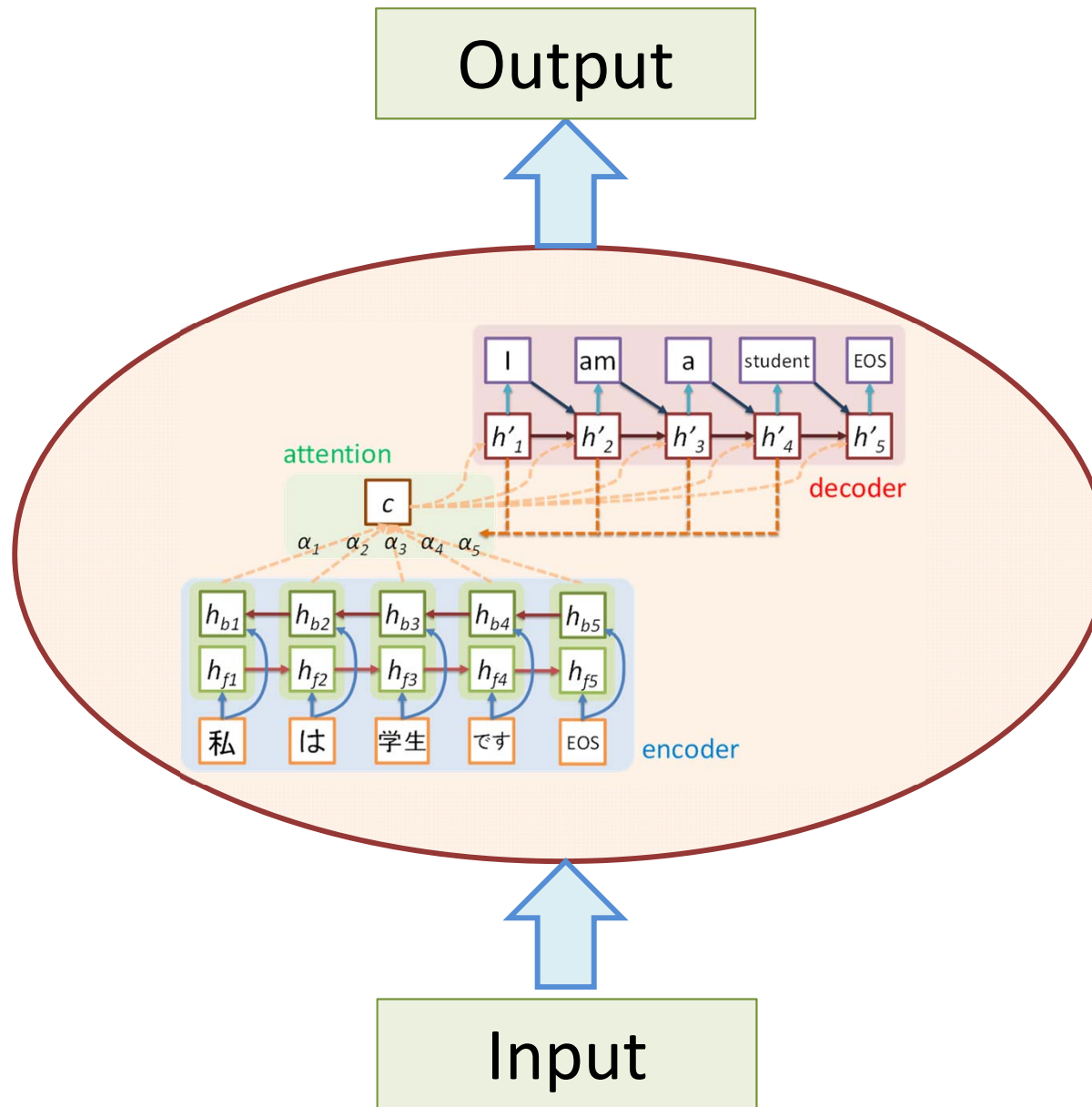
ニューラル機械翻訳モデル

Neural Machine Translation Model [Bahdanau+ 2015]



End-to-End 学習

End-to-End Learning



Deep Learning

Pros

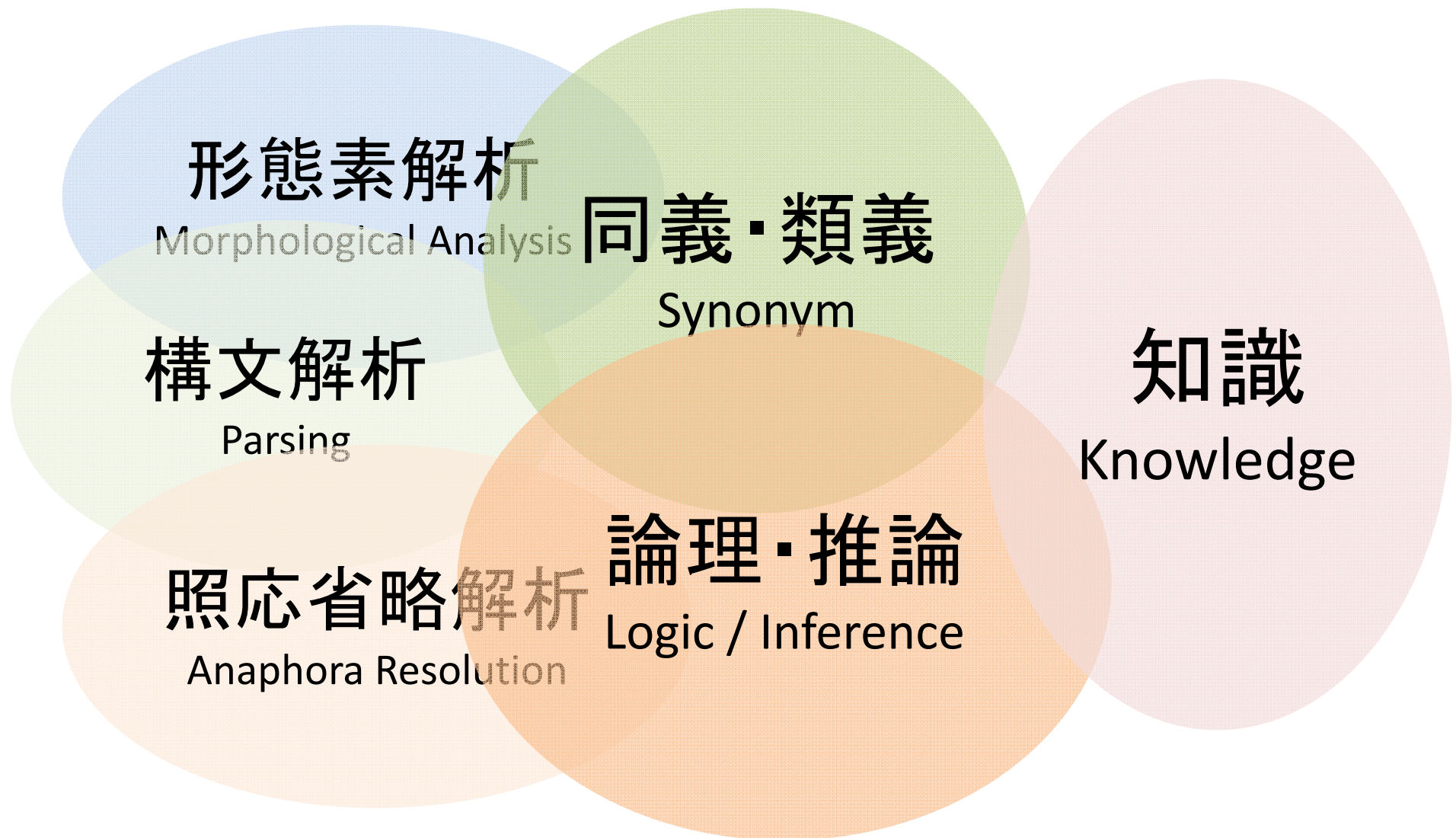
- 意味表現

- End-to-end learning

Cons

- Data hungry

NLPのアプローチ



実課題への応用

Application of NLP to Real World Tasks

Complaint Purchase Center

(不満買取センター)

Complaints about **Hamburger Shops**

ハンバーガーはおいしいけど、

hamburger tasty

ソースが多すぎて、口の周りが必ず汚れる。

sauce much mouth mouth always dirty

ソース多すぎて食べにくい。

sauce too much hard to eat

テリヤキソースとマヨネーズが多すぎて

teriyaki sauce mayo too much

食べにくいのが超不満!

hard to eat very discontent

ソースが垂れて最後に残り過ぎてもったいない。

sauce drool last left too much wasteful

日曜日の昼間は喫煙席は、いらないと思います。

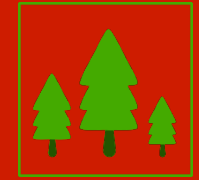
Sunday daytime smoking area unnecessary

店内タバコ臭い。喫煙席デカすぎ。

shop tobacco stink smoking area too big

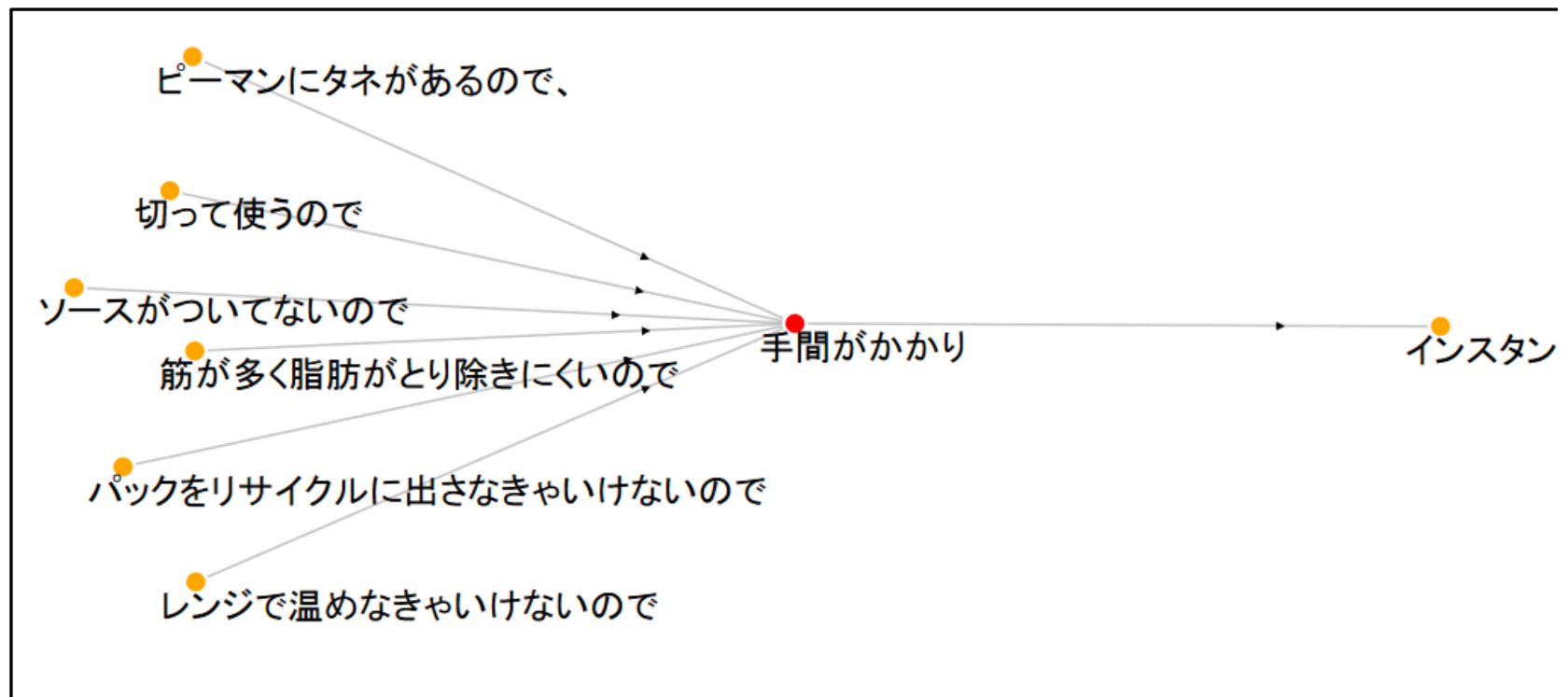


構造的言語処理の 実世界課題への適用



黒橋CREST・京都大学・柴田知秀

- 意見集約：因果関係グラフ構築
 - 対象：不満買取センター
 - 手法：1. 形態素・構文・省略解析・因果関係抽出を適用
2. 同義・反義判定や類似度に基づきクラスタリング



構造的言語処理の 実世界課題への適用



黒橋CREST・京都大学・柴田知秀

行政サービスの対話ボットシステム

行政ホームページ

粗大ごみの出し方について知りたい。
No.13341

回答

粗大ごみの収集は有料です。
金属を含むものは30cm以上、金属を含まないものは50cm以上が粗大ごみです。

粗大ごみの出し方は、次の通りです。

- 1 収集申込みをする
- 2 粗大ごみ処理券を購入する
- 3 収集日に粗大ごみ処理シールを貼って出す

1 収集申込みをする

出す粗大ごみの一番長いところを測ってからお申し込みください。
お申し込みの際に、処理手数料、収集日、受付番号を案内いたします。また、収集場所の確認を行います。

お住まいの地域別収集日の3日前（12月31日～1月3日、土、日曜日を除く）までにお申し込みください。申込みが多い場合は、次回の収集日になる場合がありますので、お早めにお申し込みください。

2 粗大ごみ処理券を購入する

粗大ごみ処理手数料は長さなどにより、200円、500円、1,000円の3区分があります。
申込みの際に確認した金額分の「川崎市粗大ごみ処理券」を購入してください。
「粗大ごみ処理手数料の減免の申請（美化活動や生活保護世帯など）」は生活環境事業所に直接お問い合わせください。
（管轄の生活環境事業所は下記の関連ホームページをご参照ください。）

購入先
市内のコンビニエンスストア、市内の郵便局（ゆうちょ銀行）

対話フローチャート生成

粗大ごみの出し方について知りたい。

「粗大ごみの出し方について知りたい。」についてお答えします

粗大ごみの収集は有料です。
金属を含むものは30cm以上、金属を含まないものは50cm以上が粗大ごみです。

どれが聞きたいですか

- 1 収集申込みをする
- 2 粗大ごみ処理券を購入する
- 3 収集日に粗大ごみ処理シールを貼って出す

編集 追加 削除

1 収集申込みをする

1 収集申込みをするについてお答えします
出す粗大ごみの一番長いところを測ってからお申し込みください。
お住まいの地域別収集日の3日前（12月31日～1月3日、土、日曜日を除く）までにお申し込みください。
申込みが多い場合は、次回の収集日になる場合がありますので、お早めにお申し込みください。

編集 追加 削除

行政対話ボット

既読 午後 9:35 粗大ゴミを出したい

「粗大ごみの出し方について知りたい。」についてお答えします
粗大ごみの収集は有料です。
金属を含むものは30cm以上、金属を含まないものは50cm以上が粗大ごみです。
粗大ごみの出し方は、次の通りです。
1 収集申込みをする
2 粗大ごみ処理券を購入する
3 収集日に粗大ごみ処理シールを貼って出す
どれが聞きたいですか

午後 9:35

既読 午後 9:36 収集申込み

「1 収集申込みをする」についてお答えします
出す粗大ごみの一番長いところを測ってからお申し込みください。
お申し込みの際に、処理手数料、収集日、受付番号を案内いたします。
また、収集場所の確認を行います。
お住まいの地域別収集日の3日前までにお申し込みください。
申込みが多い場合は、次回の収集日になる場合がありますので、お早めにお申し込みください。

午後 9:36

粗大ごみ処理券を購入してください。
「粗大ごみ処理手数料の減免の申請（美化活動や生活保護世帯など）」は生活環境事業所に直接お問い合わせください。

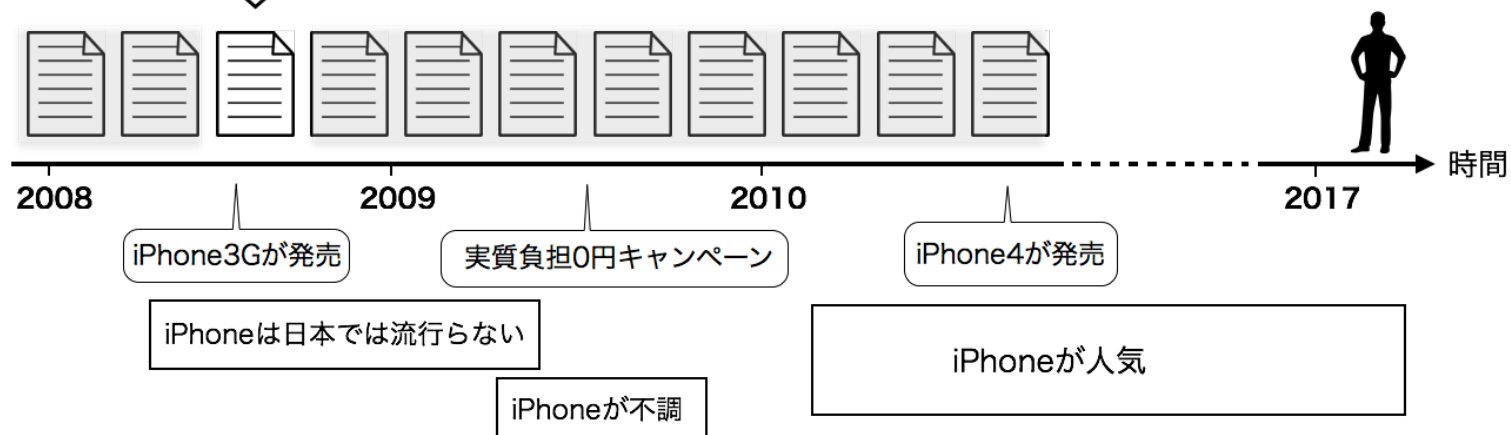
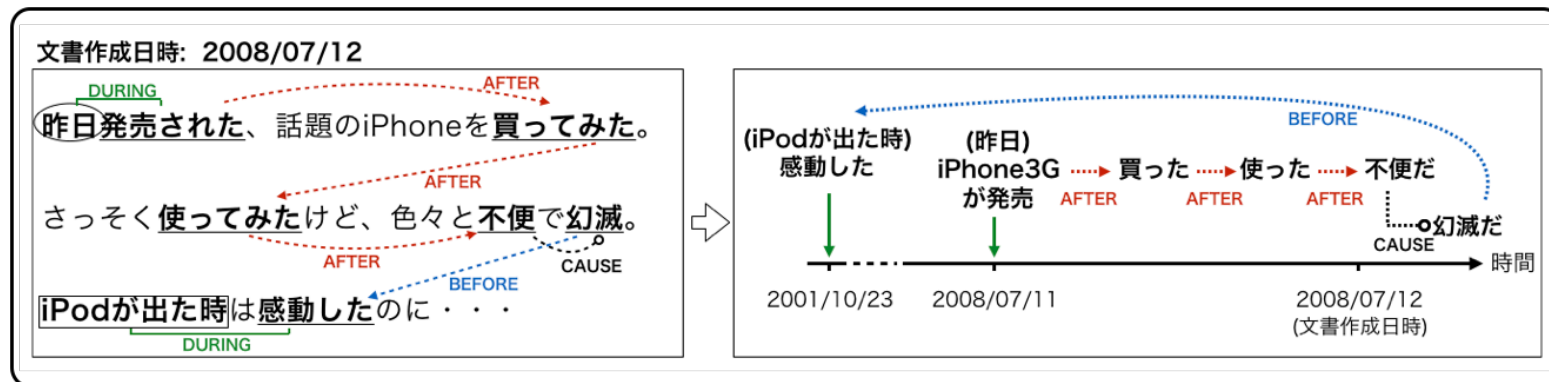
編集 追加 削除

時間・事象表現の解釈に基づく テキストの時間的構造化



黒橋CREST・京都大学・坂口智洋

Web上には大量のテキスト情報があるが、現在の視点からは古い情報も多いため、時間軸を考慮した分析が必要



時間・事象表現の解釈に基づく テキストの時間的構造化



黒橋CREST・京都大学・坂口智洋

機械がテキストを理解するためには、テキストが「いつ」の「何」に言及しているかを正しく理解する必要がある

事象間の時間関係解析

テキスト中の事象表現間の時間的前後関係を
検出するモデルを提案 (F値14)
[Shibata+, TACKBP17]

タイムライン構築

テキスト中の事象を時間軸上に配置する、2
段階機械学習モデルを提案 (F値32)
[Sakaguchi+, CICLING17]

事象-時間コーパス

タイムラインの学習や評価のため、述語に
時間情報を付与した日本語コーパスを構築
(113文書, 856文) [Sakaguchi+, LREC18]

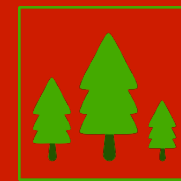
DocID: 39896, Document Creation Date: 2010-06-16

1. **Pre-orders** of the recently **announced** iPhone 4 began **Tuesday**.
2010-06-15
2. The newest iPhone, iPhone 4, was **introduced** by Apple CEO Steve Jobs at the company's 2010 Worldwide Developer 's Conference on **7 June**.
2010-06-07
3. He **praised** it as "the biggest leap we've taken since the original iPhone."

Timeline of iPhone 4

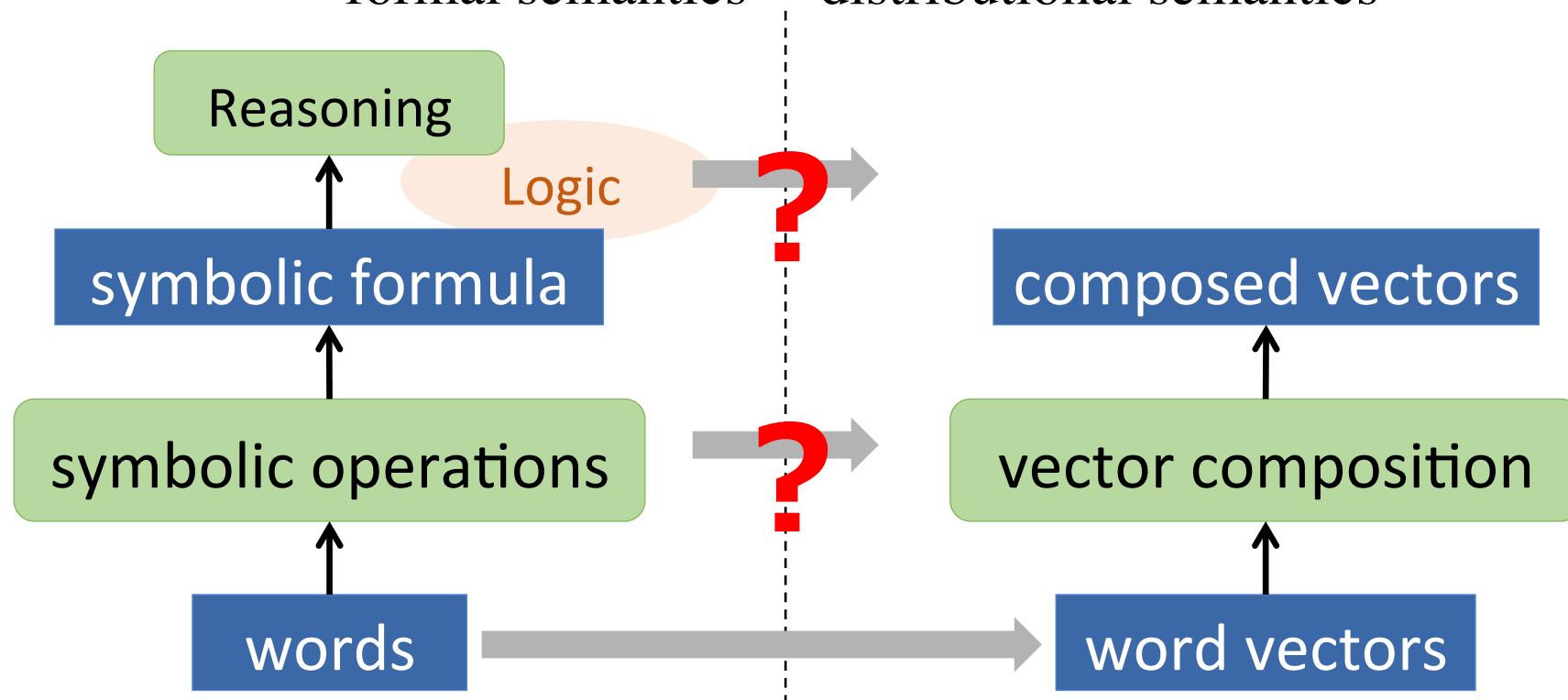
<time value>	<event>
2010-06-07	39896-1-announced
2010-06-07	39896-2-introduced
2010-06-07	39896-3-praised
2010-06-15	39896-1-pre-orders

意味役割を区別する加法的分散表現



黒橋CREST・東北大学・田然

- 自然言語の形式意味論（記号による推論）と
分布意味論（ベクトル表現）の融合を目指す
formal semantics distributional semantics



意味役割を区別する加法的分散表現



黒橋CREST・東北大学・田然

- なぜ単語ベクトルを足し算してフレーズの意味を表現できるか？ [Tian+, Machine Learning Journal 2017]

“dance hall”: dance + hall $\approx$ ballroom

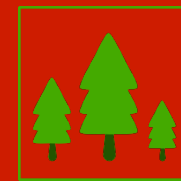
- 語順→意味役割を区別するにはどうすれば？
[Tian+, ACL 2016]

trans(arrest, _SBJ) $\approx$ policeman, gestapo, ...

trans(arrest, _OBJ) $\approx$ guilty, accomplice, ...

- この分散表現を使った自然言語推論を目指す！
(Demoあり)

賛否表明パターンと行列分解に基づく賛否モデリング



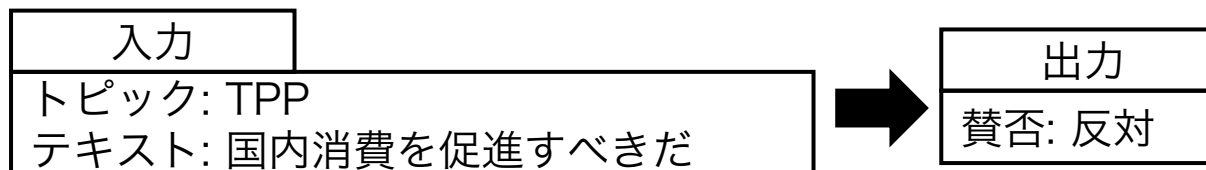
黒橋CREST・東北大学・佐々木彬

・ 目的

- 「Aに賛成する人はBにも賛成する」あるいは「Aに反対する人はBにも反対する」といった**人々の賛否の傾向をモデリング**
- このような知識は賛否分類などに有用

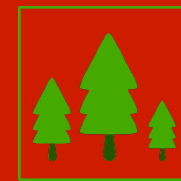
賛否分類

- テキストと特定の賛否対象 (トピック) を入力として賛否 (賛成/反対) を出力するタスク



- Web上の大規模テキストに蓄積された膨大な意見の俯瞰などに繋がる

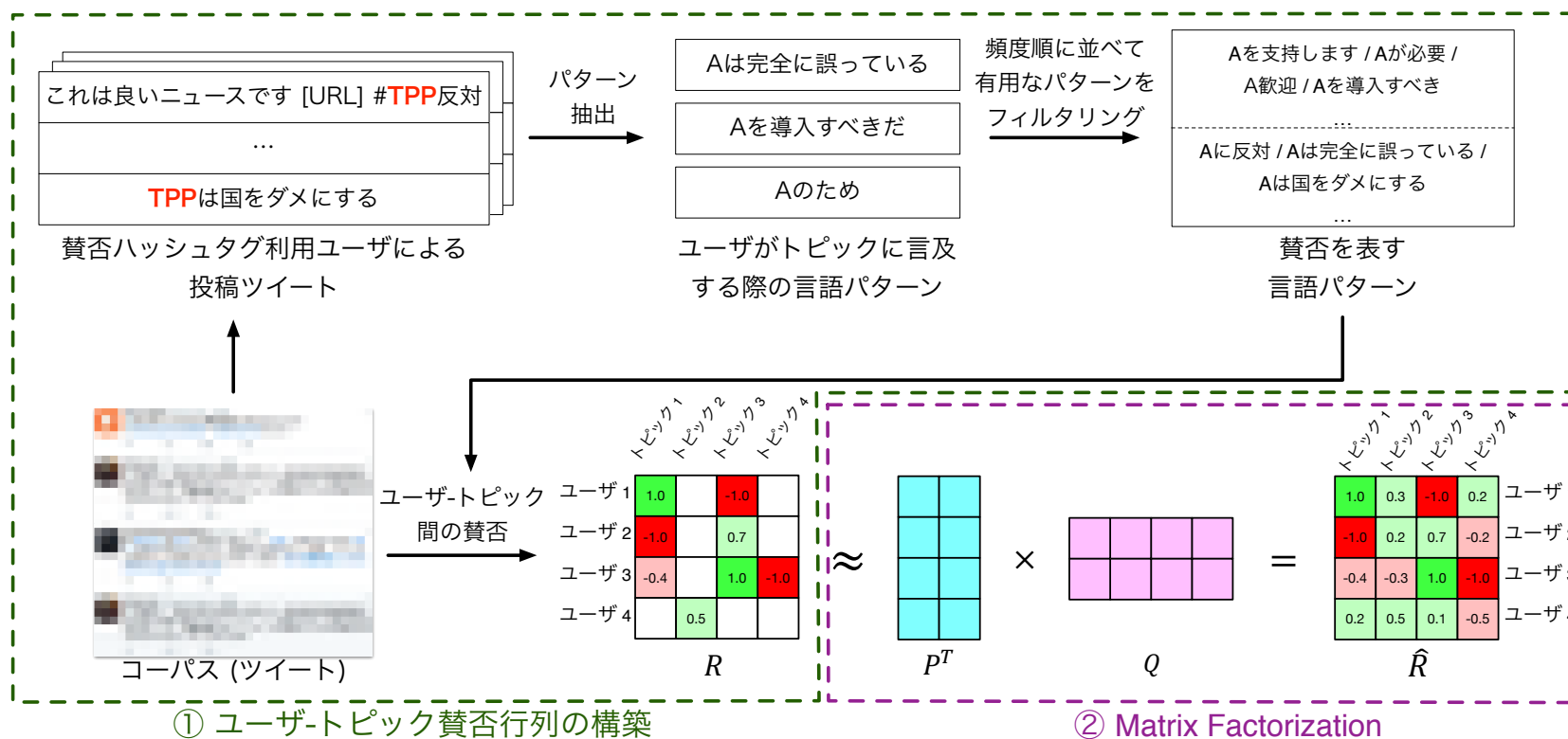
賛否表明パターンと行列分解に基づく賛否モデリング



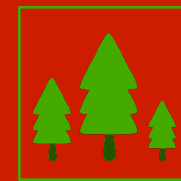
黒橋CREST・東北大学・佐々木彬

・アイデア

- ・ Matrix Factorizationを利用
- ・ これにより、**ユーザの一部の賛否表明から他のトピックへの賛否も予測**



自然演繹に基づく証明を用いた 文間の意味的関連性の学習



黒橋CREST・東京大学・谷中瞳

Semantic Textual Similarity (STS) (Agirre et al, 2012) :
文の意味を正しく表現し，文間類似度を予測するタスク

論理推論：否定や数量表現の解析
など，文の頑健な意味解析を実現

機械学習：特徴量から
ソフトな推論を実現

A: *There is no child sawing logs in the park*

B: *All boys are not cutting wood outside*

文A→文Bの含意関係の証明：

$\neg \exists x(\text{child}(x) \wedge \text{saw}(x, \text{logs}) \wedge \text{in}(x, \text{park}))$

$\rightarrow \forall x(\text{boy}(x) \rightarrow \neg(\text{cut}(x, \text{woods}) \wedge \text{outside}(x)))$



予測：
文間類似度/含意関係

論理推論の実行過程から文間の意味的関連性に関する
特徴を抽出し、文間の意味的関連性を予測

自然演繹に基づく証明を用いた 文間の意味的関連性の学習



黒橋CREST・東京大学・谷中瞳

