レシピの言語処理の現状

森 信介　笹田 鉄郎　前田 浩邦

京都大学

2013 年 8 月 18 日

Table of Contents

レシピテキストの自然言語処理

- レシピ ＝ 材料リスト ＋ 手順テキスト
  - 手順テキスト

> 1. 豆腐は水気を切り、えびは背わたを取って粗く刻む。
> 2. ボウルに豆腐と海老をいれる。
> 3. 玉ねぎ、卵、パン粉、小麦粉、塩、こしょうを入れてよくかき混ぜる。
> 4. 小判型にして中火で焼く。
> 5. 醤油をかけて食べる。

- 手順テキストの理解
  - レシピ検索
  - 調理補助システム
  - etc.

自然言語処理

1. 形態素解析 (単語分割 + 品詞推定 + (読み推定))
   - 文中の単語の認定
2. 固有表現認識
   - 実世界の物体や行動に対応する単語列
     例: 組織名, 人名, 地名, 日付, 時間, 金額, 割合 (MUC[1])
3. 係り受け解析
   - 単語や固有表現間の統語的関係
4. 述語項構造解析
   - 単語や固有表現の動作に対する意味的役割

就任 (subj: ゴーン$_{\textbf{person}}$ 氏, i-obj: 日産$_{\textbf{org.}}$ の 社長)

[1] Message Understanding Conference

レシピテキスト

- 文が比較的単純
  - 主観や時制などの問題がほとんどない
  - 言語理解の中間目標
  - 著作性がない $\Rightarrow$ 再配布可能 (判例なし)
- 一般的自然言語処理ツールでは困難
  - 独特の単語・表現
  - 多くは UGC[2] (推敲不足・誤記)

    例: クリーム コーン 、 牛乳 、 ナツメグ を 振り入れ
    例: タルト 生地 を 綿棒 で 4 mm の 薄 さ に の す

[2] User-Generated Content

分野適応の必要性 [森 12]

- BCCWJ[前川 09] のコアデータで単語分割器を学習
  - 代表性のある約 5 万文 (この質では過去最大)
- 各分野の学習コーパスを追加
  - 部分的アノテーション (後述)
- 各分野でテスト (F 値)

分野	一般	医薬品情報	特許文	レシピ	twitter
テスト文	3,680	1,250	500	728	50
作業時間	–	11 時間	12 時間	10 時間	90 分
適応前	99.32	96.75	97.25	96.70	96.52
適応後	–	98.98	97.70	97.05	97.17

分野適応の必要性 (つづき)

- 固有表現認識
  - 固有表現の定義が違う

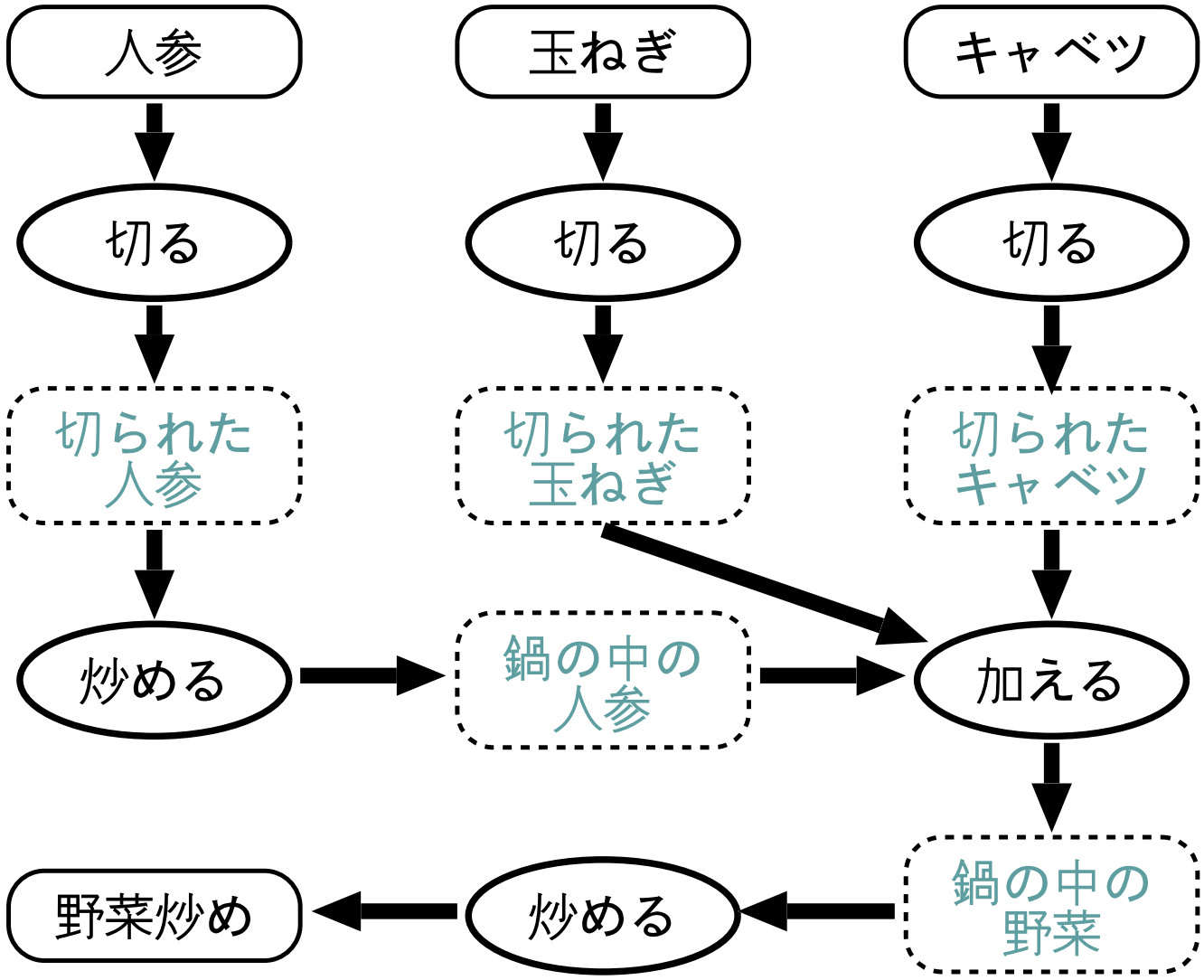
一般: 組織名, 人名, 地名, 日付, 時間, 金額, 割合

料理: 食材, 量, 道具, 継続時間, 食材の状態, 道具の状態, 調理者の動作, 食材の動作

- 係り受け解析
  - 単語分割と同様に精度低下
  - 学習コーパスの追加で解決

フローグラフ

- 抽象表現 [Momouchi 80] [Hamada 00] [山肩 07]



テキスト解析

最先端の言語処理 + 分野適応

1. 単語分割 [Neubig, Mori, et al. 11]
   - 文中の単語の認定
   - 活用語の原形推定
   - KyTea (きゅーてぃー) (Cf. 茶筌, MeCab, JUMAN, ...)
2. 固有表現認識
   - 実世界の物体や行動に対応する単語列
   - 種類は独自設定
     食材 (F), 量 (Q), 道具 (T), 継続時間 (D),
     食材の状態 (Sf), 道具の状態 (St),
     調理者の動作 (Ac), 食材の動作 (Af)

テキスト解析 (つづき)

3. 係り受け解析 [Flannery, Mori, et al.]
   - 単語や固有表現間の統語的関係
   - EDA (Cf. CaboCha, KNP, ...)

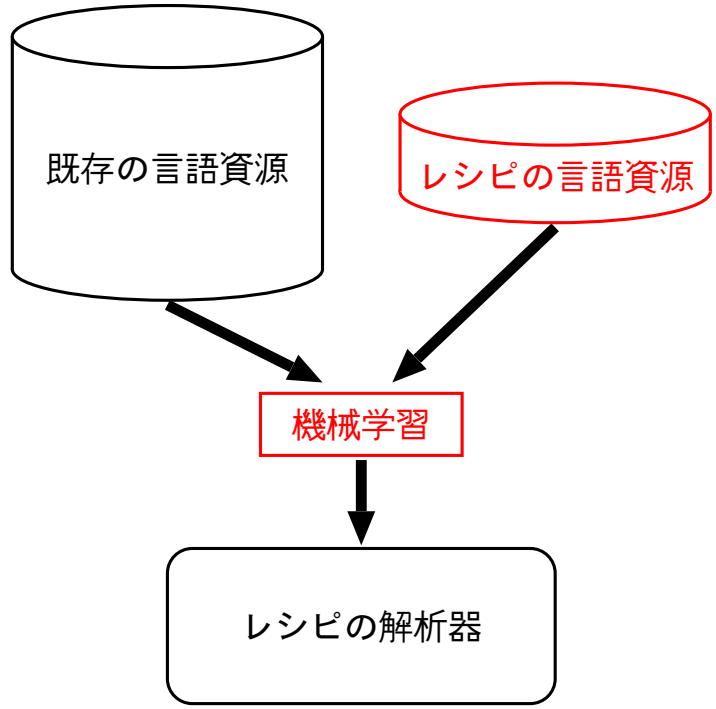
4. 述語項構造解析 [Yoshino, Mori, et al.]
   - 単語や固有表現の動作に対する意味的役割
   - ツール未公開

出力

煮立て$_{\mathbf{Ac}}$(ヲ:水-４００-ｃｃ$_{\mathbf{F}}$, デ:鍋$_{\mathbf{T}}$)

レシピテキストへの適応

- レシピテキストは一例
  - 一般的な分野適応の方法を追求 [森 12]



- 機械学習部分と適応対象の言語資源を総合設計

言語資源

- 既存: 一般分野のフルアノテーション

出典	文 数	文 字 数	固有表現数	係り受け数
BCCWJ	53,899	1,834,784	–	–
辞書の例文	11,700	197,941	–	136,109
新聞記事	9,023	398,569	–	254,402

BCCWJ: 現代日本語書き言葉均衡コーパス [前川 09]

- レシピテキスト : フルアノテーション

出典	文 数	文 字 数	固有表現数	係り受け数
固有表現 認識の学習	242	7,023	1,523	–
テスト	724	19,966	3,797	12,426

Step1. 単語分割 (単語の同定)

- 入力: 文
  水４００ｃｃを鍋で煮立て、沸騰したら中華スープの素を加えてよく溶かす 。
- 出力: 単語列

  水|４-０-０|ｃ-ｃ|を|鍋|で|煮-立-て|、|
  沸-騰|し|た-ら|中-華|ス-ー-プ|の|素|を|
  加-え|て|よ-く|溶-か|す|。

  - |: 単語境界あり
  - -: 単語境界なし

※ 活用語尾の分割 ⇒ 活用語の正規化

単語，品詞，形態素

1. 単語
   - 意味や職能を有する最小の言語単位
   - 文字列 (平均 1.4〜2.0 文字程度)
2. 品詞
   - 10〜15 程度の文法範疇 (例：名詞，動詞)
   - 細分類と呼ばれる下位分類 (例：固有名詞，上一段活用)
3. 形態素
   - 形式的・文法的機能を担う単語またはその一部

※自然言語処理における定義は便宜的

単語 ≈ 形態素

品詞体系

- 品詞大分類 (基準により多少異なる)

単語 (短単位)	自立語	活用しない	主語になる	名詞
			修飾語になる	副詞 連体詞
			独立語になる	接続詞 感動詞
		活用する		動詞 形容詞 形容動詞
	付属語	活用しない 活用する		助詞 助動詞
	その他			記号

- 後処理で利用
  - 機械学習 (係り受け解析, etc.)
  - パターンマッチ

単語分割基準

- 文法家の助けを借りて決定
  - 本来は言語処理の目的に応じて設計すべき
  - 現実にはツールやコーパスにより規定

1. 基準書

   例: 『現代日本語書き言葉均衡コーパス』形態論情報規程集改定版 [小椋 09]

2. 実例 (単語分割済みコーパス)

   水|４－０－０|ｃ－ｃ|を|鍋|で|煮-立-て|、|
   沸-騰|し|た-ら|中-華|ス-ー-プ|の|素|を|
   加-え|て|よ-く|溶-か|す|。

単語定義の粒度

- 短い単位は被覆率が高い (⇔ 未知語率が低い)
  - | 活用語 | + | 語尾 | ≪ | 活用語 | × | 語尾 |
- 語幹で用言 (動詞, 形容詞, 形容動詞) を表現
  - 原形に戻す処理が不要
  - 一部語義曖昧性の増加 (例: 行-く v.s. 行-なう)
- 応用 (後処理) ではしばしば長い単位が望まれる
  - 意味 (翻訳)
  - 読み (連濁など)
  - 係り受け (複合動詞)

⇒ 固有表現として対応

例) | 中–華 | ス–ー–プ | の | 素 | (全体で食材)

単語の定義

- 短単位: できる限り分割
  - Cf. 『現代日本語書き言葉均衡コーパス』形態論情報規程集改定版 [小椋 09]

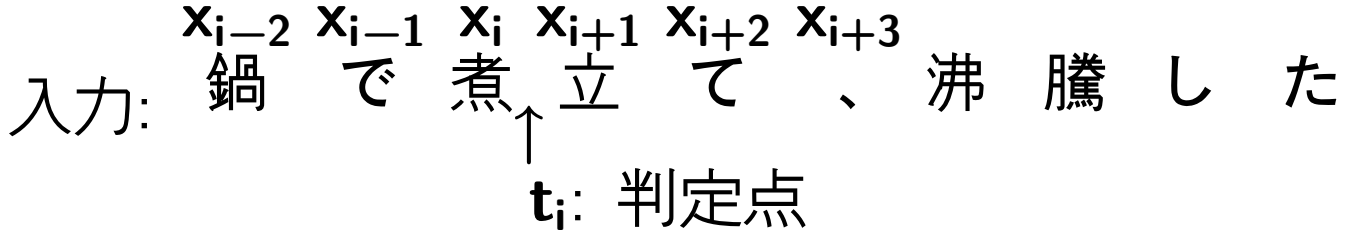
※ 活用語尾の分割 ⇒ 活用語の正規化

- 語幹を標準形として同一判定
  - 例) 焼-く ＝ 焼-い-た (同じ動作)
- 形態素解析 (MeCab, JUMAN など) では原形を推定
  - 未知語の場合に活用型を指定して登録

- 文字列のみ指定
  - 語幹の特定のみ必要
  - 品詞は一般分野のコーパスから推定

点予測による単語分割 (KyTea [Neubig 11])

- 文字間の 2 値分類問題 ⇒ 線形 SVM による実装



- 素性

  Char (type) 1-gram feature:

  -3/鍋 (K), -2/で (H), -1/煮 (K), 1/立 (K), 2/て (H), 3/、(S)

  Char (type) 2-gram feature:

  -3/鍋で (KH), -2/で煮 (HK), -1/煮立 (KK), 1/立て (KH), 2/て、(HS)

  Char (type) 3-gram feature:

  -3/鍋で煮 (KHK), -2/で煮立 (HKK), -1/煮立て (KKH), 1/立て、(KHS)

部分的アノテーションコーパスから学習可能
⇒ 柔軟なコーパス作成!
⇒ 迅速・安価な分野適応!

部分的アノテーションコーパス

- 文は複数の判定箇所を含む
- 一部の判定箇所のみにラベル付与

1. 未知語候補の抽出 [Mori 96] (あるいは解析誤りの文)
2. 単語境界の修正作業

玉ねぎ (頻度=1362)

… | 玉 – ね – ぎ | は 薄 切 り 、 ピ ー マ ン は 薄 い 輪 …
… マ リ ネ 液 を 作 り 、 （ 1 ） の | 玉 – ね – ぎ | ・ …
… 約 6 分 加 熱 す る 。 | 玉 – ね – ぎ | は 粗 み じ ん …

こん (頻度=1338)

… 移 し 、 「 | こ – ん – ぶ | だ し 」 、 半 ず り 白 ご …
… 入 れ 、 両 面 を | こ – ん – が – り – と | 色 づ く ま で …
… 2 つ 切 り 、 | れ – ん – こ – ん | は 皮 を む い て 8 …

文脈情報の重要性

- 一般分野から Web(Yahoo!知恵袋) への分野適応
  `http://www.phontron.com/kytea/dictionary-addition.html`
  (2011 年 11 月 25 日)

- 単語分割の精度

モデル	精度 (F 値)
適応なし	95.54%
辞書追加 (文脈なし)	96.75%
コーパス追加 (文脈あり)	97.15%

- 約 75〜80%の精度向上は辞書追加により実現可能
  - 多くの言語処理応用ではここまで

- 残りの 20〜25%の精度向上には文脈情報が必要

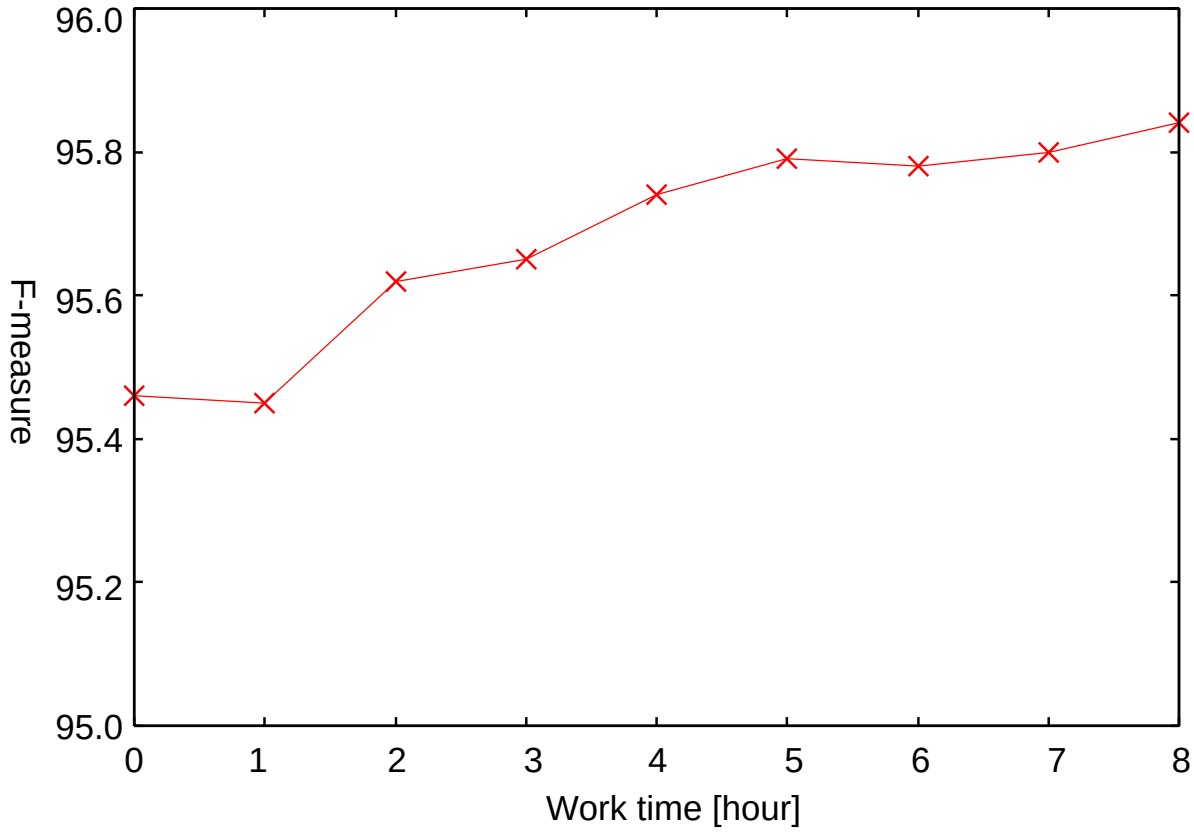
一般モデルとその分野適応

- 一般モデル: BCCWJ, UniDic, など
- 適応モデル: 未知語候補への部分的アノテーション
  - 8 時間
- 評価基準: F 値 (再現率と適合率の調和平均)

再現率 = **LCS**/出力
適合率 = **LCS**/正解

※**LCS**：**longest common subsequence** 最長共通部分系列

学習曲線



- 一般モデルでは不十分 (一般分野: 99%程度)
- さらなる作業が必要
- 作業時間にしたがって精度向上

Step 2. 固有表現認識

- 固有表現 (Named Entity)
  - 実世界の物体や動作に対応する単語列

    例: 組織名, 人名, 地名, 日付, 時間, 金額, 割合 (MUC)

> ９９ 年 ３ 月$_{\textbf{date}}$ カルロス ゴーン$_{\textbf{person}}$ 氏
> が 日産$_{\textbf{org.}}$ の 社長 に 就任

- BIO2 記法 (Begin, Intermediate, Other)

> ９９/B-Dat 年/I-Dat ３/I-Dat 月/I-Dat
> カルロス/B-Per ゴーン/I-Per 氏/O が
> 日産/B-Org の/O 社長/O に/O 就任/O

- 系列ラベリング問題 (HMM, CRF)
  - タグセット $= \{\mathbf{B}, \mathbf{I}\} \times \text{NE-Type} \cup \{\mathbf{O}\}$

- 精度: 80% ～ 90% (1 万文程度の学習コーパス)

レシピの固有表現認識

- 固有表現
  - 実世界の物体や動作に対応する単語列
  - 一般的には、人名、組織名、時間、 ...
  - 定義はタスク依存 $\Rightarrow$ 一般分野コーパスがない
- レシピの固有表現を独自に設定:
  食材 (**F**), 量 (**Q**), 道具 (**T**), 継続時間 (**D**),
  食材の状態 (**Sf**), 道具の状態 (**St**),
  調理者の動作 (**Ac**), 食材の動作 (**Af**)

水$_{\mathbf{F}}$ ４００ ｃｃ$_{\mathbf{Q}}$ を 鍋$_{\mathbf{T}}$ で 煮立て$_{\mathbf{Ac}}$ 、沸騰し$_{\mathbf{Af}}$ たら
中華 スープ の 素$_{\mathbf{F}}$ を 加え$_{\mathbf{Ac}}$ て よく 溶か$_{\mathbf{Ac}}$ す 。

点予測による固有表現認識

部分的アノテーションコーパスから学習可能
⇒ 柔軟なコーパス作成!
⇒ 迅速・安価な分野適応!

1. BIO2 表現 (1 単語に 1 つの固有表現タグ)
   水/B-F ４００/B-Q ｃｃ/I-Q を/O 鍋/BT で/O
   煮立て/B-Ac 、/O 沸騰/B-Af し/I-Af たら/O
   中華/B-F スープ/I-F の/I-F 素/I-F を/O 加え/B-Ac
   て/O よく/O 溶か/B-Ac す/O 。/O

2. 部分的アノテーションコーパスから単語のタグを推定するロジスティック回帰を構築 (KyTea "`-solver 6`")
   - Cf. CRF の学習にはフルアノテーションが必要

点予測による固有表現認識 (つづき)

3. 各単語に対して可能なタグと確率を出力

P(y|w)	w 水	４００	ｃｃ	を	・・・
B-F	0.62	0.00	0.00	0.00	・・・
I-F	0.37	0.00	0.00	0.00	・・・
B-Q	0.00	0.82	0.01	0.00	・・・
y I-Q	0.00	0.17	0.99	0.00	・・・
B-T	0.00	0.00	0.00	0.00	・・・
⋮	⋮	⋮	⋮	⋮	⋱
O	0.01	0.01	0.00	1.00	・・・

点予測による固有表現認識 (つづき)

3. 各単語に対して可能なタグと確率を出力

P(y|w)	w 水	４００	ｃｃ	を	···
B-F	0.62	0.00	0.00	0.00	···
I-F	0.37	0.00	0.00	0.00	···
B-Q	0.00	0.82	0.01	0.00	···
y I-Q	0.00	0.17	0.99	0.00	···
B-T	0.00	0.00	0.00	0.00	···
⋮	⋮	⋮	⋮	⋮	⋱
O	0.01	0.01	0.00	1.00	···

4. 解釈可能な最適タグ列を探索

   例: “**F-I Q-I**” は解釈不可能

初期モデルと分野適応

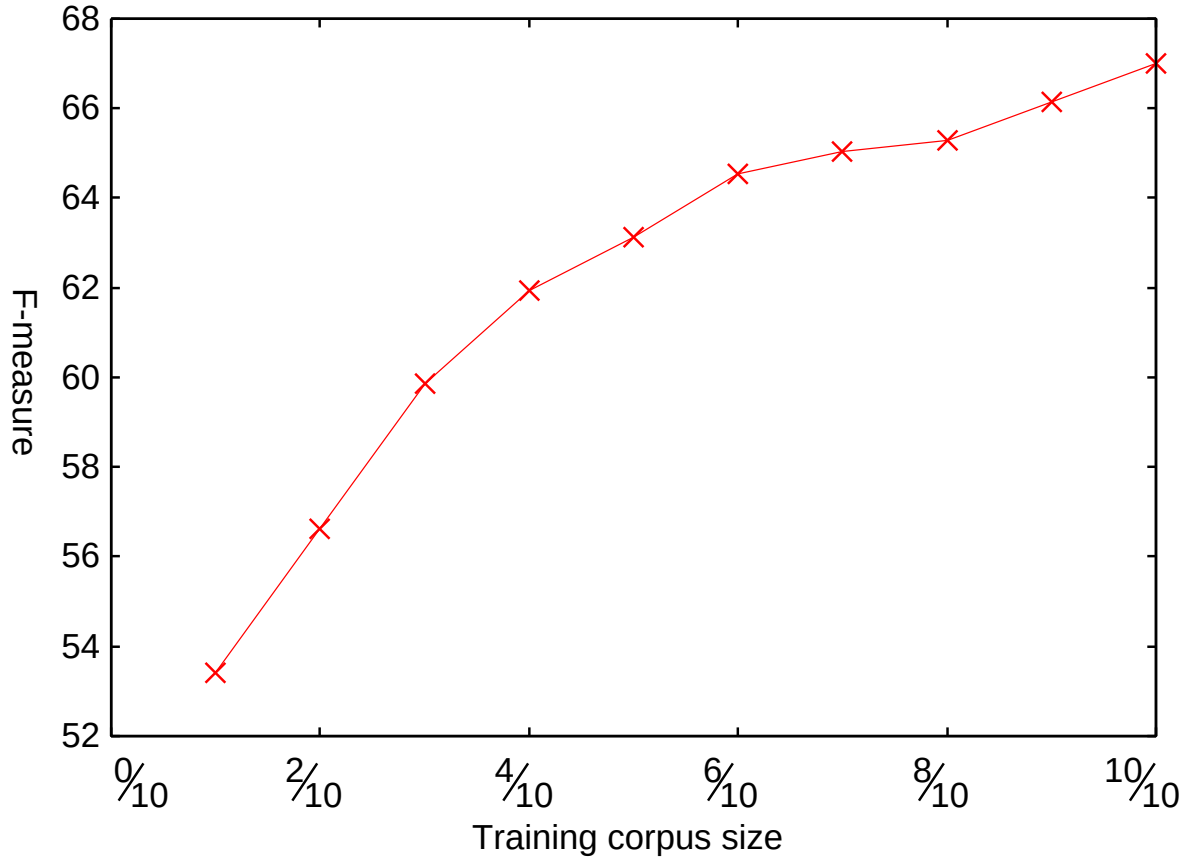
- 肉じゃがのレシピ (242 文) にタグ付与 (5 時間)

  ↑ 良くない設定 ⇒ 無作為抽出に変更中

- 初期モデル: 1/10 を利用
- 適応モデル: 2/10 から 10/10 を利用

学習曲線

- F 値

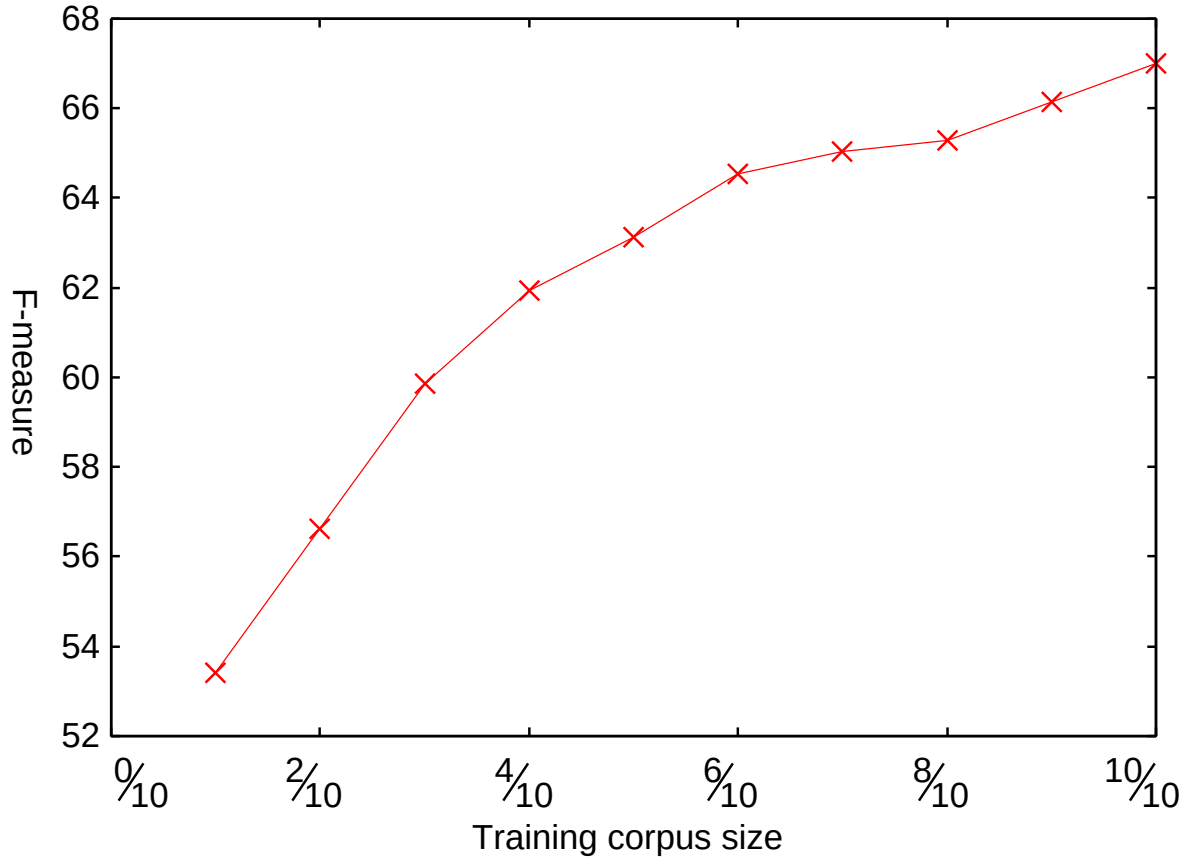


- 一般的な固有表現認識タスクよりかなり低い

  ex. 学習 = 11,000 文で 83.1%, 1,038,986 語で 90.0%)

学習曲線

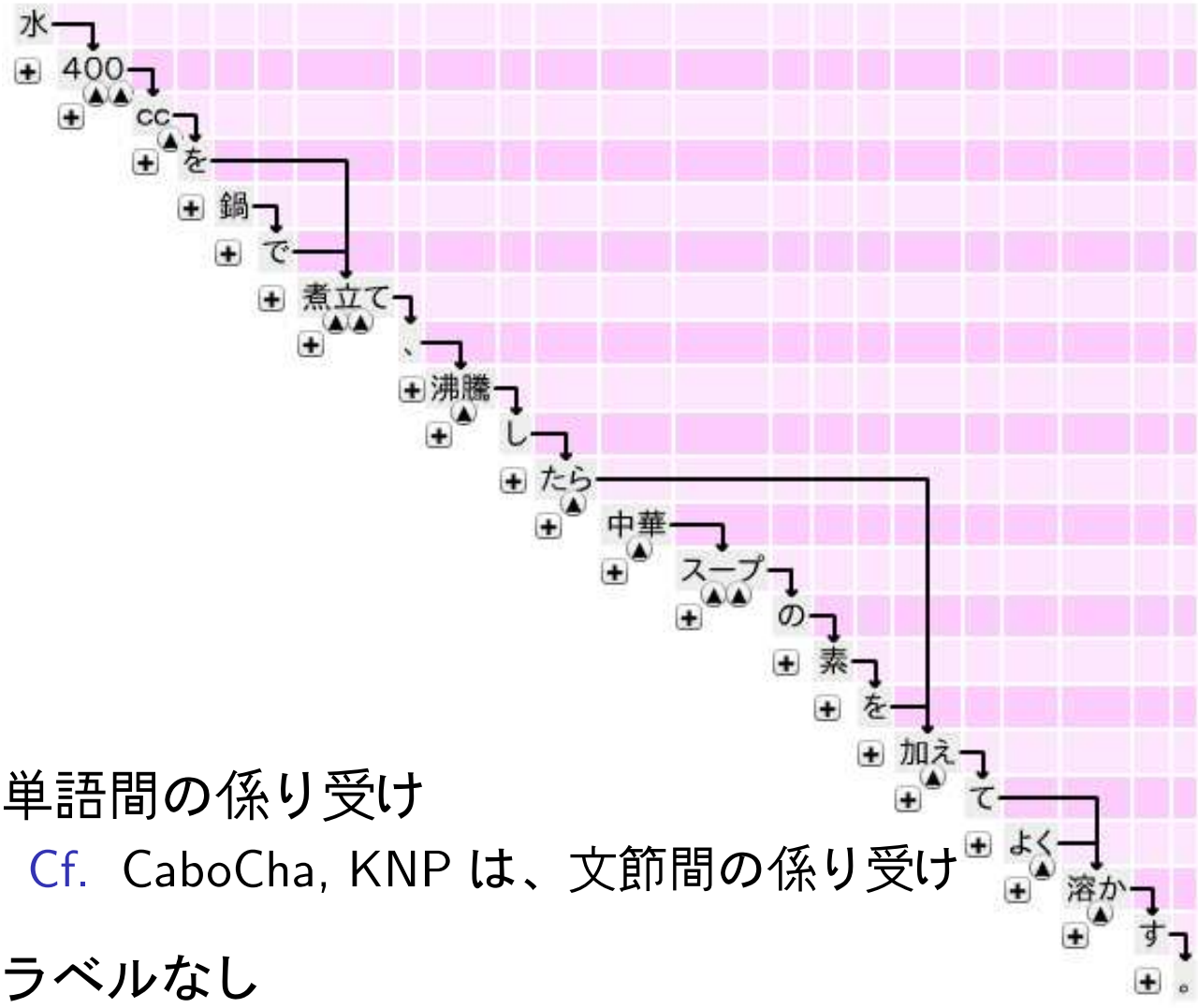
- F 値



- アノテーション作業にしたがって急激に上昇
  - ex. 5 時間 (243 文) ⇒ 250 時間 (12,150 文)

Step 3. 係り受け解析

- 文の統語構造



  - 単語間の係り受け

    Cf. CaboCha, KNP は、文節間の係り受け

  - ラベルなし

点予測による係り受け解析 (EDA) [Flannery 11]

- 点予測による最大全域木 (MST)
  1. 全ての単語間の係り受けスコアを計算

$$\sigma(\langle \mathbf{i}, \mathbf{d_i} \rangle, \vec{\mathbf{w}}), \qquad \text{ここで } \mathbf{w_i} \text{ は } \mathbf{w_{d_i}} \text{ に係る}$$

  2. エッジスコアの合計が最大になる全域木 (MST) を選択

$$\hat{\vec{\mathbf{d}}} = \underset{\vec{\mathbf{d}} \in \mathbf{D}}{\mathbf{argmax}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \sigma(\langle \mathbf{i}, \mathbf{d_i} \rangle, \vec{\mathbf{w}})$$

部分的アノテーションコーパスから学習可能
⇒ 柔軟なコーパス作成!
⇒ 迅速・安価な分野適応!

点予測による係り受け解析 (つづき)

- スコア計算の素性

牡蠣 を 広島 に 食べ に 行 く

w_{i-3} w_{i-2} w_{i-1} w_i w_{i+1} w_{i+2} w_{i+3}

w_{d_i-3} w_{d_i-2} w_{d_i-1} w_{d_i} w_{d_i+1} w_{d_i+2} w_{d_i+3}

F1 係り元 w_i と係り先 w_{d_i} の距離
F2 w_i と w_{d_i} の表記
F3 w_i と w_{d_i} の品詞
F4 w_i と w_{d_i} の前後 3 単語の表記
F5 w_i と w_{d_i} の前後 3 単語の品詞

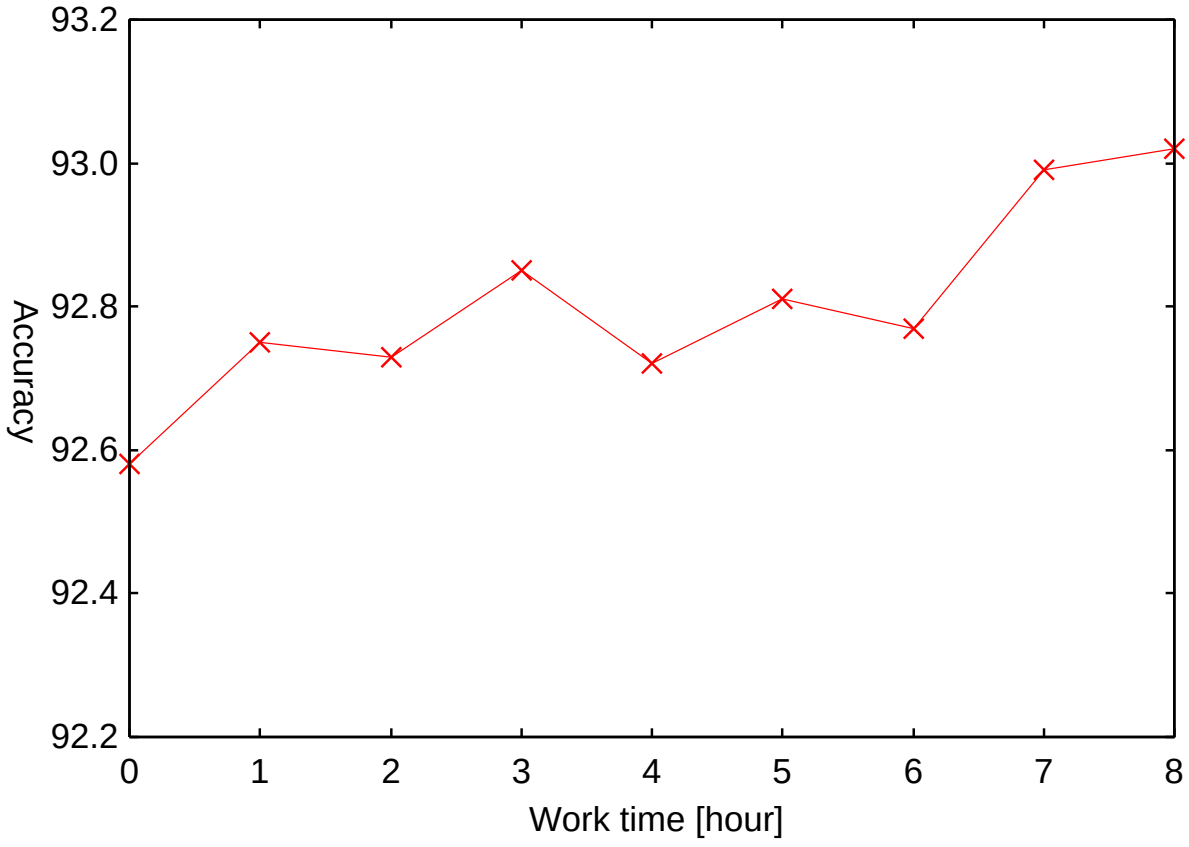
一般モデルとその分野適応

- 一般モデル: 約 2 万文から学習
  - 英語表現辞典の例文: 11,700 文, 145,925 語
  - 日経新聞の記事: 9,023 文, 263,425 語
- 分野適応: 新出の名詞と助詞の組に係り先を付与
  1. 既存のアノテーションに含まれない名詞と助詞の列を見つける
  2. 名詞から用言までの係り受けを付与する

     ｃｃ → を → ... 煮立て
  3. 8 時間の作業

結果

- 学習曲線



- 一般分野に対する精度 (96.83%) と比べて低い
- 作業時間にしたがって精度向上

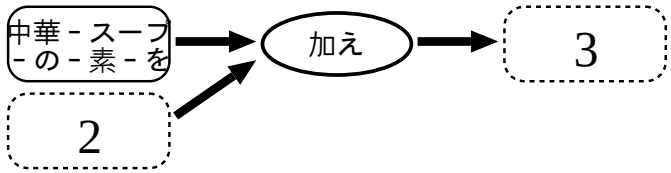
Step 4. 述語項構造解析

- 現状は規則に基づく方法
  - 点予測による機械学習 [Yoshino, Mori, et al.]
- 有向グラフの最小の単位に対応

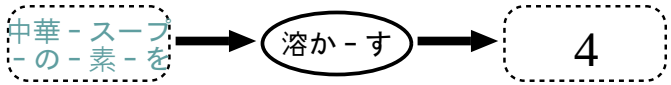
1. 煮立て$_{\mathbf{Ac}}$(Chef, 水$_{\mathbf{F}}$ ４００ ｃｃ$_{\mathbf{Q}}$ を, 鍋$_{\mathbf{T}}$ で)

2. 沸騰-し$_{\mathbf{Af}}$(Food), たら

3. 加え$_{\mathbf{Ac}}$(Chef, 中華 スープ の 素$_{\mathbf{F}}$ を, 水$_{\mathbf{F}}$ に)



4. 溶か-す$_{\mathbf{Ac}}$(Chef, 中華 スープ の 素$_{\mathbf{F}}$ を)



機械学習による述語項構造抽出

- 言語処理として確立していない
  - 大規模なコーパスがない
    - 現象の「密度」が低い
    - アノテーションの基準策定が困難
- 動的素性を使わない設計
- 点予測による機械学習 [Yoshino, Mori, et al.]

部分的アノテーションコーパスから学習可能
⇒ 柔軟なコーパス作成!
⇒ 迅速・安価な分野適応!

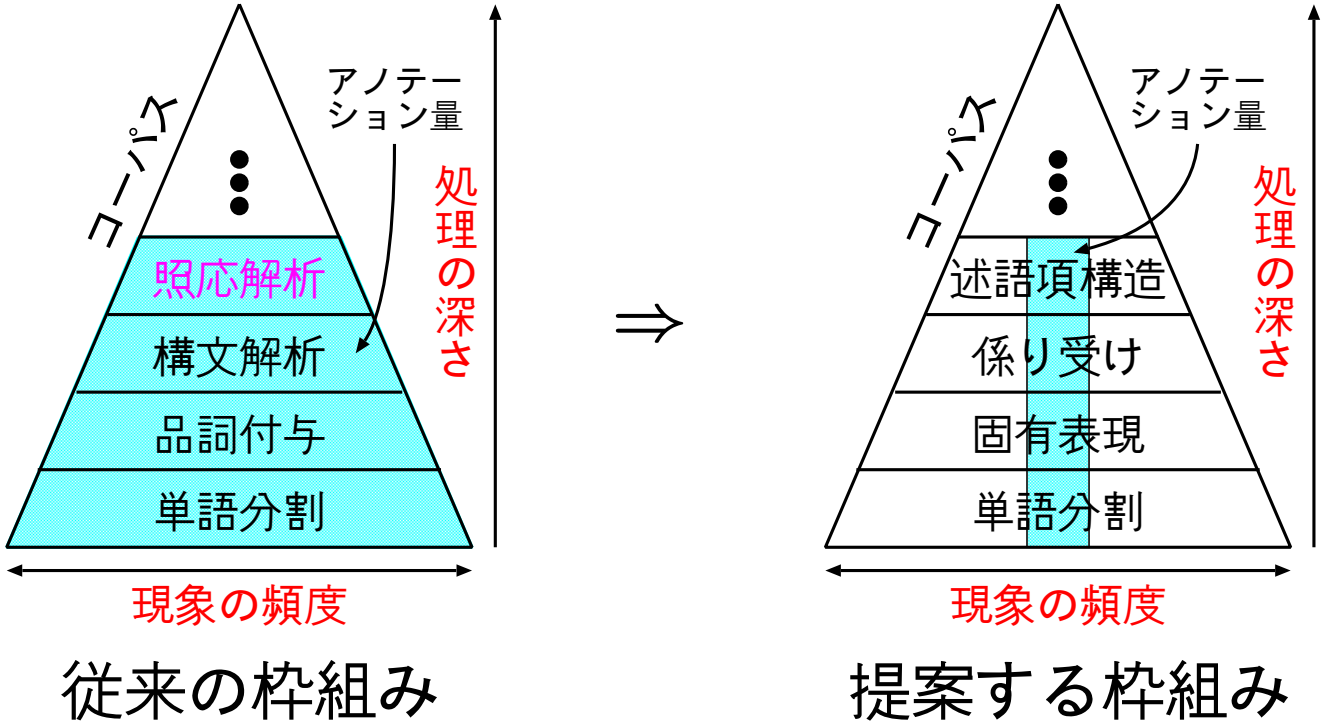
部分グラフ抽出の評価

1. テスト コーパス: 無作為抽出の 100 レシピ

出典	文 数	文 字 数	固有表現数	係り受け数
テスト	724	19,966	3,797	12,426

2. 学習コーパス
   - 単語分割:
     (BCCWJ + etc.) + 部分的アノテーション
   - 固有表現認識:
     肉じゃが 1/10 + 9/10 (設定が良くない)
   - 係り受け解析:
     (辞書の例文 + 新聞記事) + 部分的アノテーション
   - 述語項構造解析: 規則による方法 ⇒ 機械学習

各段階の言語資源を独立となるように設計



従来の枠組み　　提案する枠組み

- 点予測で容易に実現
- (統一的の) 系列予測でも実現可能のはず
  - 異なる処理段階の統一は昔から課題

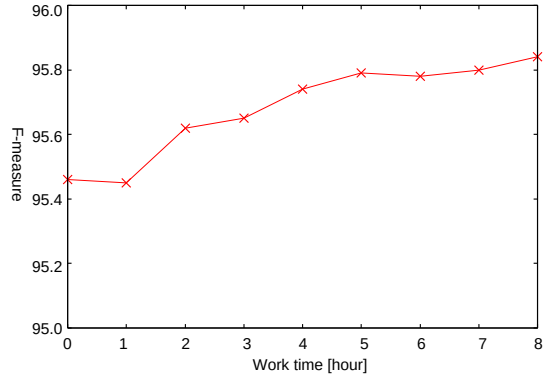
各処理の結果のまとめ

Step 1. 単語分割

一般モデル: 95.46%

⇓ (8 時間)

分野適応後: 95.84%

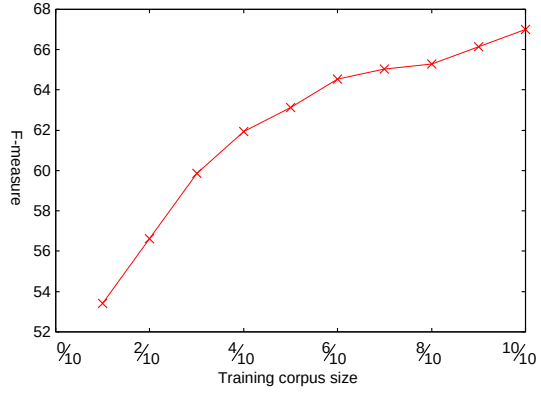


Step 2. 固有表現抽出

初期モデル: 53.42%

⇓ (5 時間)

資源追加後: 67.02%

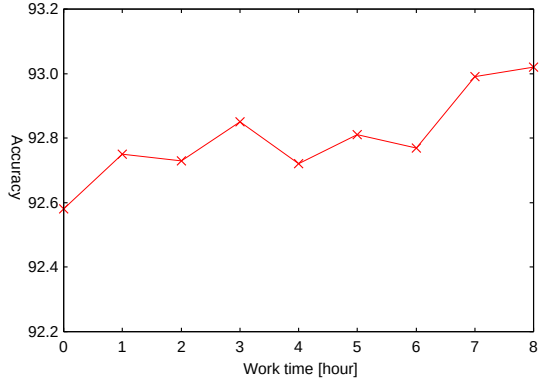


Step 3. 係り受け解析

一般モデル: 92.58%

⇓ (8 時間)

分野適応後: 93.02%



部分グラフ抽出の評価

1. 述語項構造 (有向グラフの部分グラフ)
   - 述語と項の組

     例: ⟨ 煮立て, を:水-４００-ｃｃ ⟩, ⟨ 煮立て, で:鍋 ⟩

   - F 値

     初期モデル: 42.01%　多くの研究では辞書追加程度

     ⇓ (8 + 5 + 8 時間)　28.0%のエラーを削減!

     資源追加後: 58.27%

   - 依然として低い F 値
     - さらなるアノテーション (21 時間 $\ll \infty$)
     - 固有表現認識が問題 (67.02% $\ll$ 90%)
     - それぞれの処理のみを適応した結果を定量的に比較!!

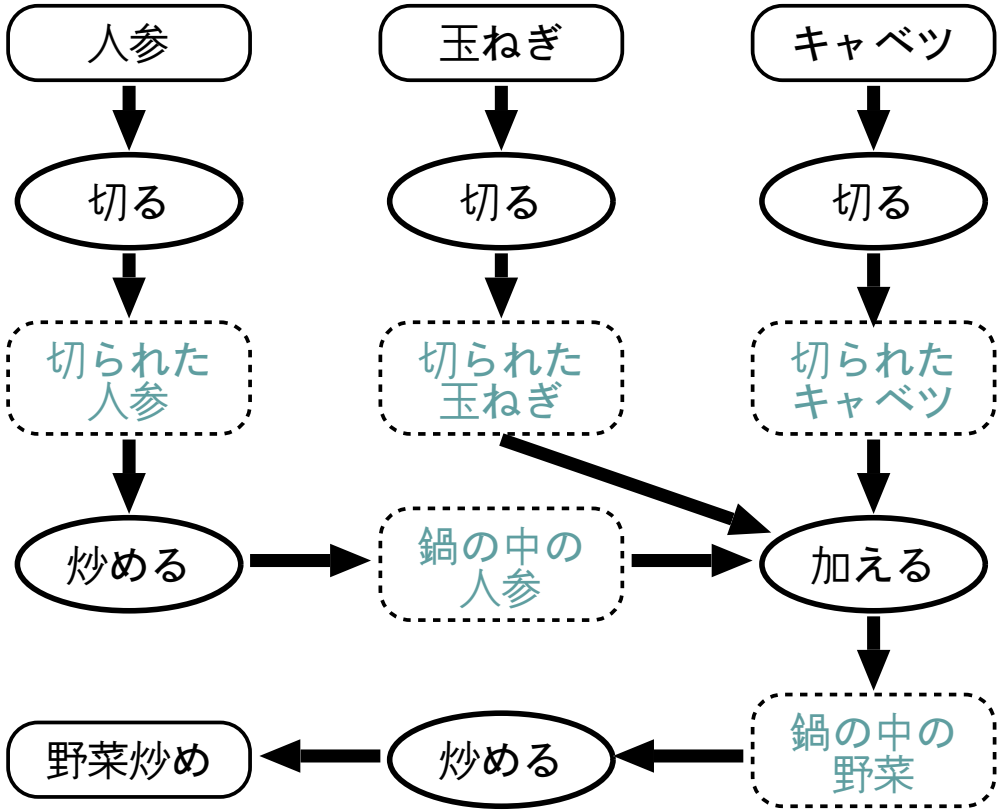
未解決事項 (or 研究段階)

- 単語の同一性
  例: たまねぎ = タマネギ = 玉葱 = 玉ねぎ = ...
  - 読み推定である程度解決可能
- 固有表現 (物体) の包含関係
  例: 新-玉ねぎ $\subset$ 玉ねぎ
  例: にんじん $\subset$ 野菜
  - 主辞 (最後の単語) の同一性である程度解決可能
- 動作の包含関係 (あるいは含意)
  例: Mix = {加える, 混ぜる, ...}
  例: 炒めた $\Rightarrow$ 温かいはず
- 物理実体
  例: 少々 = ??g

ここから
現在取り組み中

レシピテキストからフローグラフへの変換

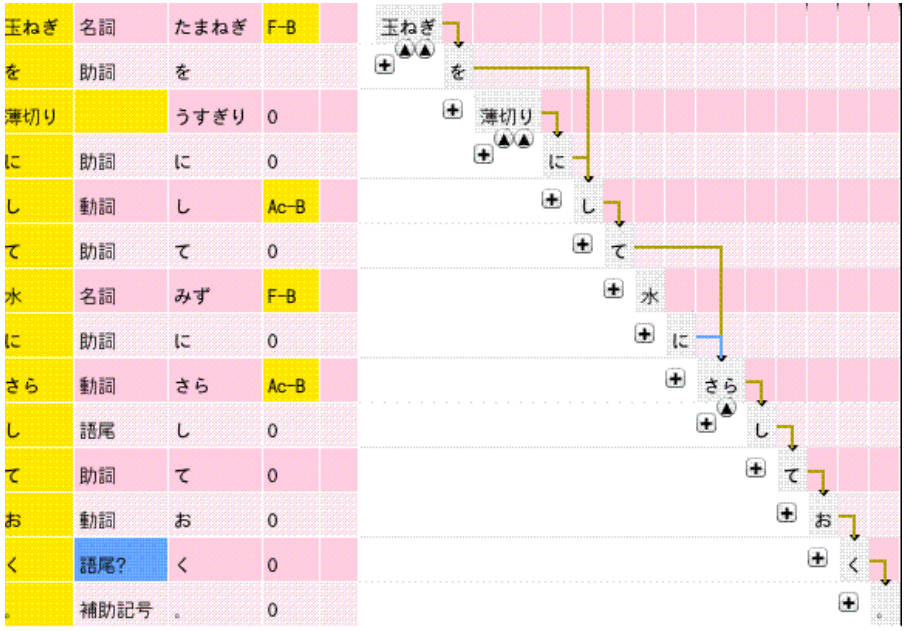
1. 固有表現認識
2. 固有表現をノードとして最大全域木
   - 動作ノードは動作による生成物でもある



※ 木にならない場合もある (例: 食材の分離)

各処理の学習コーパスの充実

1. 単語分割
2. 固有表現認識
3. 係り受け解析
4. 述語項構造解析



玉ねぎ	名詞	たまねぎ	F-B
を	助詞	を	
薄切り		うすぎり	0
に	助詞	に	0
し	動詞	し	Ac-B
て	助詞	て	0
水	名詞	みず	F-B
に	助詞	に	0
さら	動詞	さら	Ac-B
し	語尾	し	0
て	助詞	て	0
お	動詞	お	0
く	語尾?	く	0
。	補助記号	。	0

アノテーションツール PNAT (現在 1〜3 に対応)

- 各処理の部分的アノテーション大幅増量
  - 部分的アノテーションからの系列予測学習 ($\neq$ 点予測)
- 各処理の改善による全体の精度の定量的評価
  - どの処理のアノテーションに注力?
  - アノテーション or 手法の改善?

レシピテキストの言語処理

- 進捗状況

処理	設計	論文	十分な精度
単語分割	☑	☑	☐
固有表現認識	☑	☐	☐
係り受け解析	☑	☑	☐
述語項構造解析	☑	☑	☐
フローグラフ推定	☐	☐	☐

- 応用
  - レシピ検索
  - 調理シーンの映像処理とのマッチング
  - 対話システムによる教示

References

- Flannery, D., Miyao, Y., Neubig, G., and Mori, S.: Training Dependency Parsers from Partially Annotated Corpora, in *Proceedings of the Fifth International Joint Conference on Natural Language Processing* (2011)
- Hamada, R., Ide, I., Sakai, S., and Tanaka, H.: Structural Analysis of Cooking Preparation Steps in Japanese, in *Proceedings of the fifth international workshop on Information retrieval with Asian languages*, No. 8 in IRAL '00, pp. 157–164 (2000)
- Momouchi, Y.: Control Structures for Actions in Procedural Texts and PT-Chart, in *Proceedings of the Eighth International Conference on Computational Linguistics*, pp. 108–114 (1980)

- Mori, S. and Nagao, M.: Word Extraction from Corpora and Its Part-of-Speech Estimation Using Distributional Analysis, in *Proceedings of the 16th International Conference on Computational Linguistics* (1996)
- Neubig, G., Nakata, Y., and Mori, S.: Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (2011)
- Yoshino, K., Mori, S., and Kawahara, T.: Predicate Argument Structure Analysis using Partially Annotated Corpora, in *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (2013)

- 山肩 洋子, 角所 考, 美濃 導彦■調理コンテンツの自動作成のためのレシピテキストと調理観測映像の対応付け, 電子情報通信学会論文誌, Vol. J90-DII, No. 10, pp. 2817–2829 (2007)
- 小椋 秀樹, 小磯 花絵, 冨士池 優美, 原 裕■『現代日本語書き言葉均衡コーパス』形態論情報規程集改定版, 国立国語研究所内部報告書 (2009)
- 森 信介■自然言語処理における分野適応, 人工知能学会誌, Vol. 27, No. 4 (2012)
- 前川 喜久雄■代表性を有する大規模日本語書き言葉コーパスの構築, 人工知能学会誌, Vol. 24, No. 5, pp. 616–622 (2009)