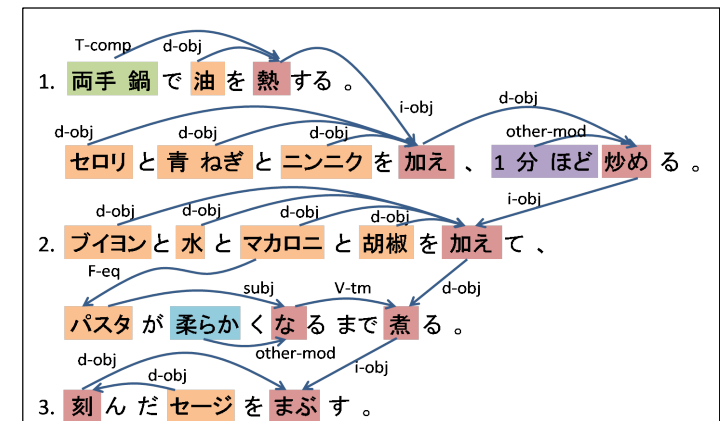
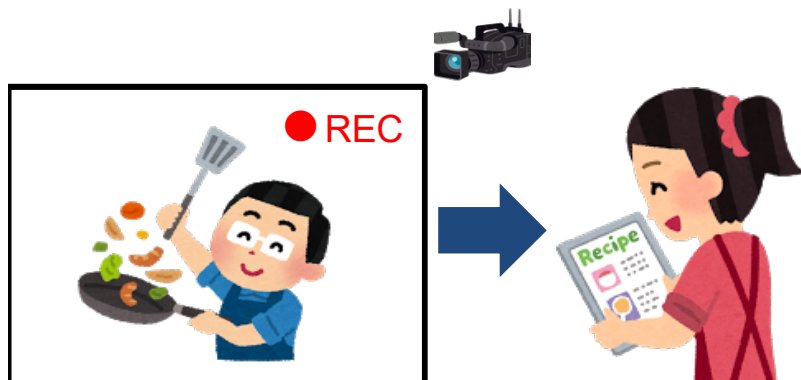


手順実施動画からの手順書生成

京都大学 学術情報メディアセンター

森 信介



講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

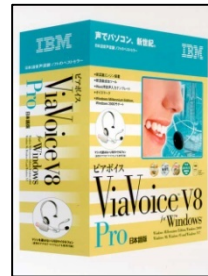
自己紹介

- 使用言語 (今でも学習中)
 - 日本語, Englis., França..., Itali..., Esp...., Po.....
- 研究経歴
 - 京都大学 学士 → 修士 → 博士

- 自然言語処理

- 日本アイ・ビー・エム

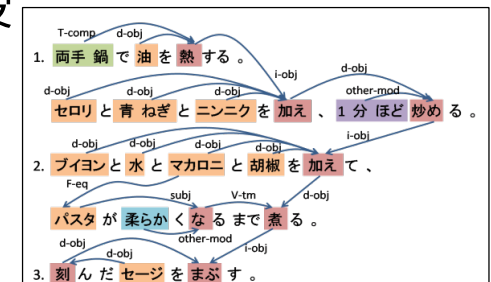
- 音声認識・合成



- 京都大学学術情報メディアセンター 准教授 → 教授

- 自然言語処理

- メディア融合

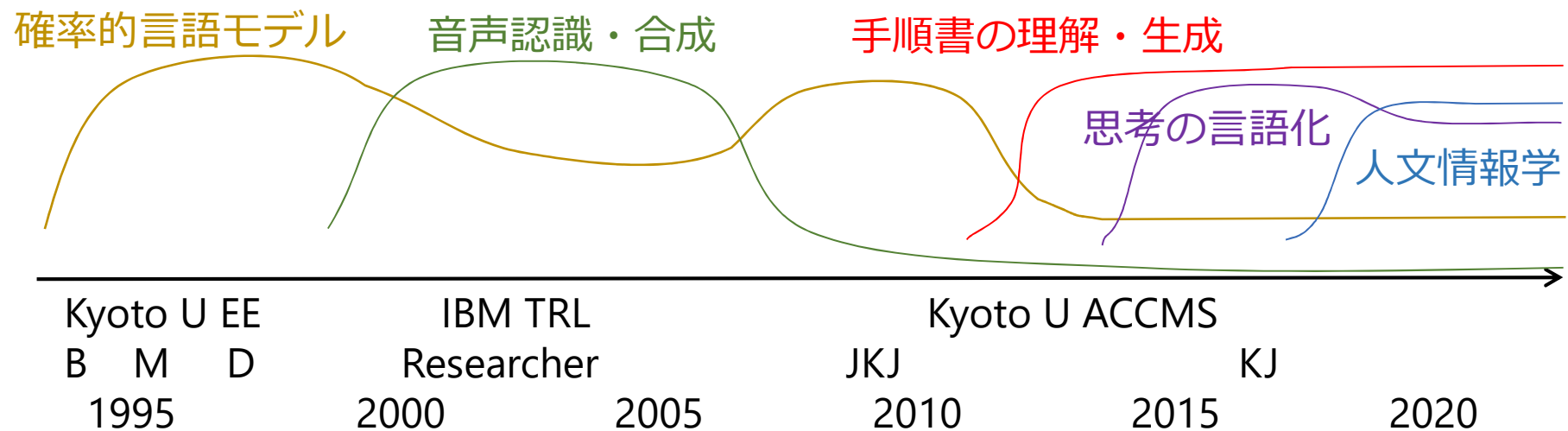


NL/CVIM/PRMU

- NL研発表
 - デビュー戦 NL-96
 - 合計27件, 筆頭17件
 - 幹事 (2012-2015)
 - 屋久島開催 (2013年11月)
 - 自然言語処理シンポジウム (2014年~, NL, SLP, NLC, SP)
- CVIM/PRMU研発表
 - デビュー戦なう

研究テーマ

活動実績ベースで



自然言語処理研究者として

- 統計的手法
 - 卒論の1994年では先駆的, 現在は当然
- 基礎的な言語解析
 - 形態素解析・固有表現認識・係り受け解析
 - 照応解析・意味解析・文法理論
- 応用課題の提案・解決 (データ作成 → モデル構築)
- 対象言語は主に日本語 (引用数が稼ぎにくいだが…)
 - 論文発表
 - 言語資源・ツールの公開

立ち位置

- 統計的手法による言語処理
 - 生成的 → 識別的 → ニューラル ネット →
 - 「既存課題をモデルを変えて改善」に食傷気味
 - 統合的処理が容易に
- エンジニア志向 **dogfooding** (drinking my own champagne)
 - 言語 (学) に興味があるが…
 - 起業したい人が来て会社を残していくも…
 - 言語資源 (データ) の蓄積
 - 社会実装に活用



講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

基礎的言語処理

1. 単語分割・品詞推定・読み推定 [LREC10, ACL11, IS11]

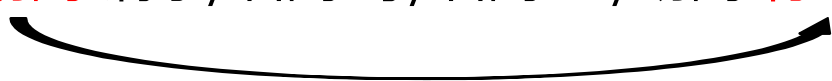
↓ 彼/名詞 と/助詞 清水/名詞 寺/名詞 に/助詞 行く/動詞

2. 用語認識 (固有表現認識) [PACLING15, ACL16]

↓ 彼/名詞 と/助詞 清水/名詞 寺/名詞 に/助詞 行く/動詞

3. 係り受け解析 [IJCNLP11, NLP12, LREC14, IWPT15]

↓ 彼/名詞 と/助詞 清水/名詞 寺/名詞 に/助詞 行く/動詞



4. 共参照解析 [IJCNLP13, LREC16]

彼/名詞 と/助詞 清水/名詞 寺/名詞 に/助詞 行く/動詞

= バラク・オバマ = 音羽山清水寺

多様な実テキストに強い手法

- 柔軟な言語資源活用

- 部分的アノテーションからの学習

例) 玉-ね-ぎ|と|に-ん-じ-ん|を?み?じ?ん?切?り?に?し?ま?す

- 単語分割・品詞推定・読み推定

- KyTea [LREC10, ACL11, IS11]

- 言語資源の追加は加法的な効果 [LRE16]

- 用語認識 (固有表現認識)

- POWNER [PACLING15, ACL16] → N3ER

- 係り受け解析

- EDA [IJCNLP11, NLP12, LREC14, IWPT15]

形態素解析

1. 単語分割
2. 品詞推定
3. その他タグ (読み, 原型, 正規表記, ...)

INPUT: 吾輩は猫である。

OUTPUT:

我輩	ワガハイ	ワガハイ	我が輩	代名詞		
は	ワ	ハ	は	助詞-系助詞		
猫	ネコ	ネコ	猫	名詞-普通名詞-一般		
で	デ	ダ	助動詞	助動詞-ダ	連用形-一般	
ある	アル	アル	有る	動詞-非自立可能	五段-ラ行	終止形-一般

。

。

補助記号-句点

EOS

隠れセミマルコフモデルに基づく教師なし完全形態素解析
内海 慶, 塚原 裕史, 持橋 大地, NLP2015

- KyTea [Neubig+ ACL11][InterSpeech11]

固有表現認識

- 固有表現: **単語列**とその種類

種類: 人名, 地名, 組織名, 金額, 日付, 時間, 割合, 人工物

- 実世界に対応するのは単語列 (単語の定義の問題)
- 拡張固有表現: 約200種

彼/名詞	Other
と/助詞	Other
清水/名詞	Loc-Begin
寺/名詞	Loc-Intermediate
に/助詞	Other
行く/動詞	Other

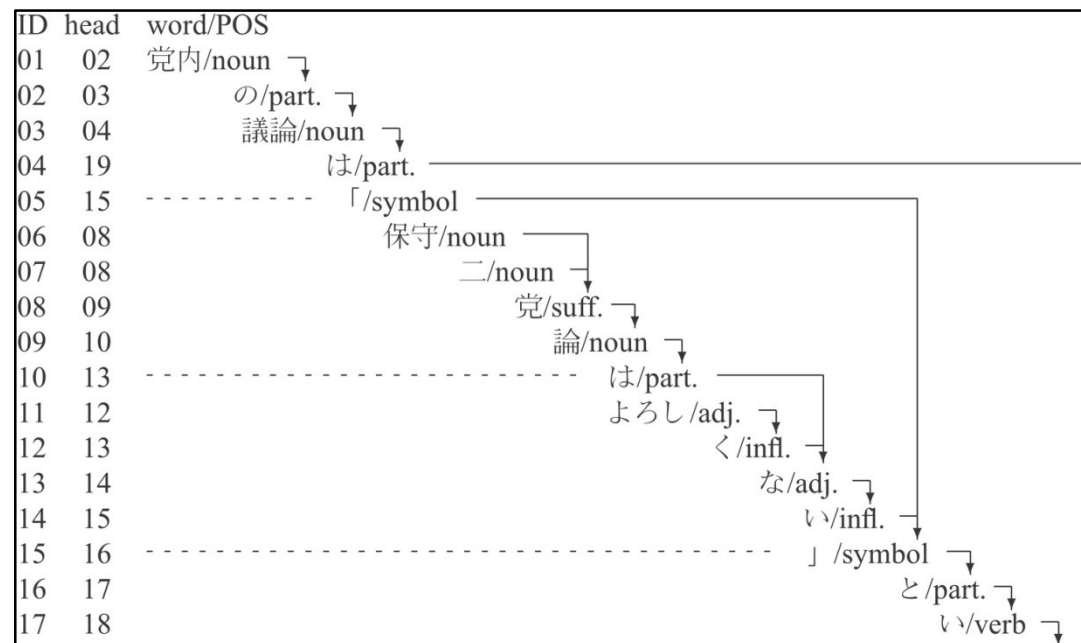
- 種類の定義は課題依存
- PWNER [Sasada+ PACLING15] → N3ER

係り受け解析

- 単語単位

- 係り先の推定

- 係り受けの種別の推定



- EDA [Flannery+ IJCNLP11] [Flannery+ IWPT15]

統計的仮名漢字変換

世界初の
統計的仮名
漢字変換

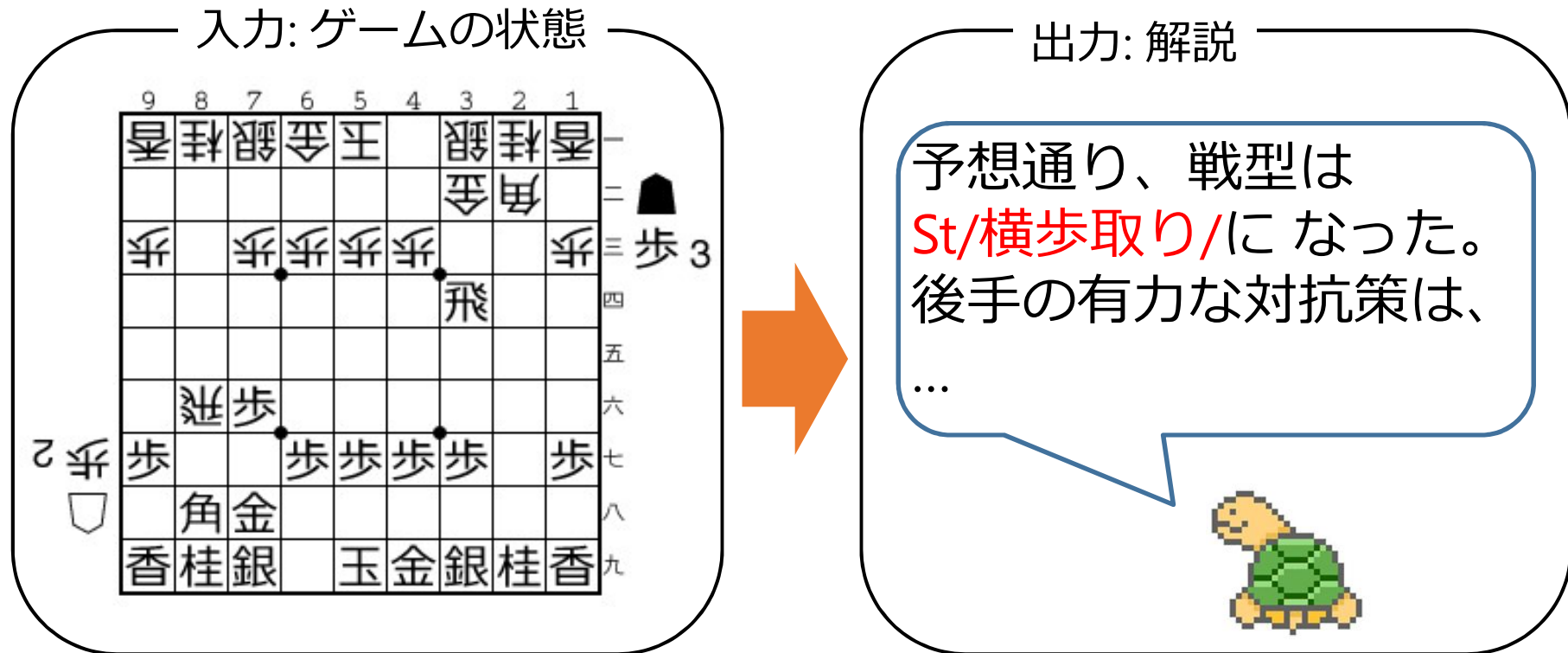
- [NL125][情処論99] (第58回電気科学技術奨励賞)
 - Google, Microsoft, Apple と自分が使っている (特許をとってれば…)
- [Coling-ACL06] Drinking my own champagne
 - 大規模テキストの部分文字列の列挙



```
File Edit Options Buffers Tools Help
補題 5.6
次元 1 の ネーター 整域 においては、
鏡:** *scratch* All L2 (Lisp Interaction)
WMA:000268 少し一般化して、単項イデアル整域上の有限生成加群に対して
WMA:001132 ゼンシュタイン、ユークリッド整域、三角形、二つの平方数の
WMA:001144 数体C の部分環であるから、整域である。
WMA:001198 ンシュタイン整数環は素元分解整域である。
WMA:001202 ュタイン整数環はユークリッド整域であり、よって一般の環論
-E:** *kwic* All L6 (Fundamental)
a: [整域] (S) s: 聖域 d: 西域 f: 声域 g: せいいき
```

- [Takahashi+ EMNLP15]
 - ユーザーの変換ログからの言語知識の獲得

データ分析と未来予測の言語化



- 科研 挑戰的萌芽 (2014-2017) → 亀甲助教
 - [Kameko+ CIG15] [LREC16] [Ushiku+ SIGIR17] [Matsuyoshi+ LREC18]

利用可能な資源

2014年2月8日 第7回朝日杯将棋オープン戦決勝
羽生善治三冠 対 渡辺明二冠

2014/02/08 14:30
第7回朝日杯将棋オープン戦
東京・有楽町朝日ホール
持ち時間: 各40分

残り時間 終局
▲0時間00分
△0時間00分

▲渡辺明

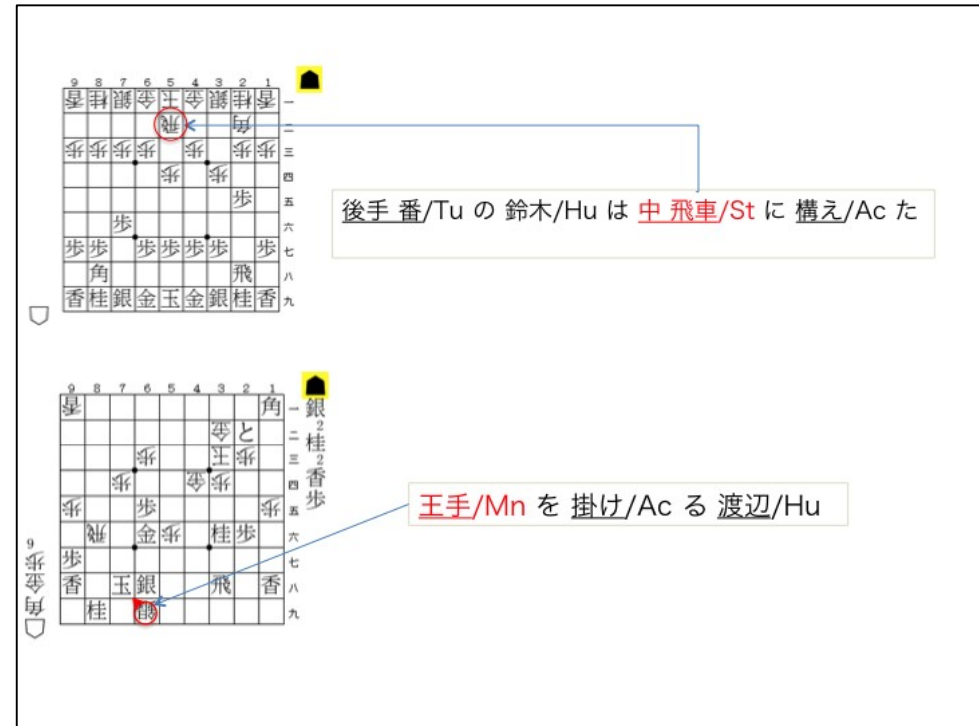
角行 金将
歩兵 歩兵 歩兵

この手を指したあと、渡辺が額に手をやって天をあおいだ。
「△9九角に駒を使うわけは、いいないので▲8八銀で...ん、これは、かなり詰みとということですか。△同角成▲同玉
△7九銀と追って、これは詰んでるように見えます」(佐藤康光九段)

- Position and Commentary
- Junisen C1, C2
 - 2003~2013
 - 3,786 match
 - 212,066 sent.

将棋解説コーパス [Mori+ LREC16]

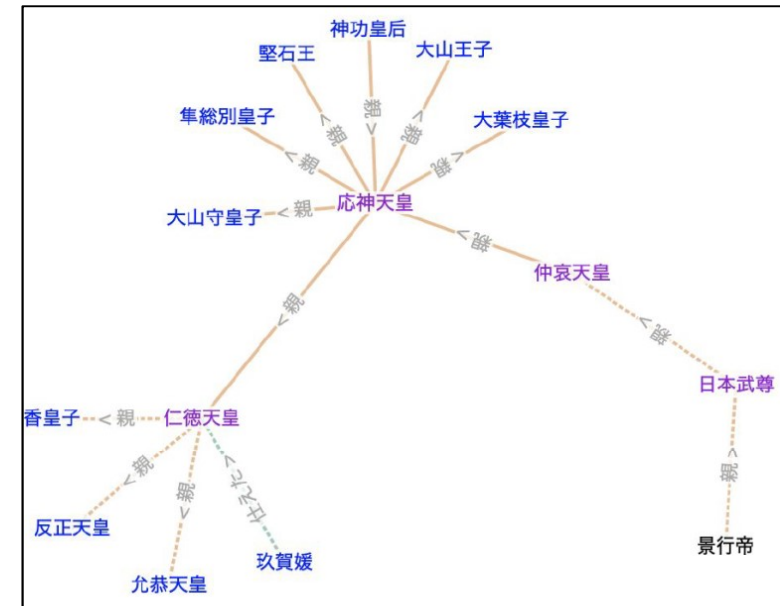
- 2,508文と対応する局面
- 人手アノテーション
 - 単語分割
 - 用語 (将棋固有表現)
 - 21種
 - BIO2 タグ体系



- <http://res.lsta.media.kyoto-u.ac.jp/data/game/>

人文情報学

- 辞典の知識グラフ化
 - 芳賀人名 [白井+ じんもんこん20]
 - 平安人物志 [本澤+ CH125]
- 時空間表現の絶対値推定
 - 前近代日本語
- 歴史資料画像と言語表現の対応学習



– 誰か…

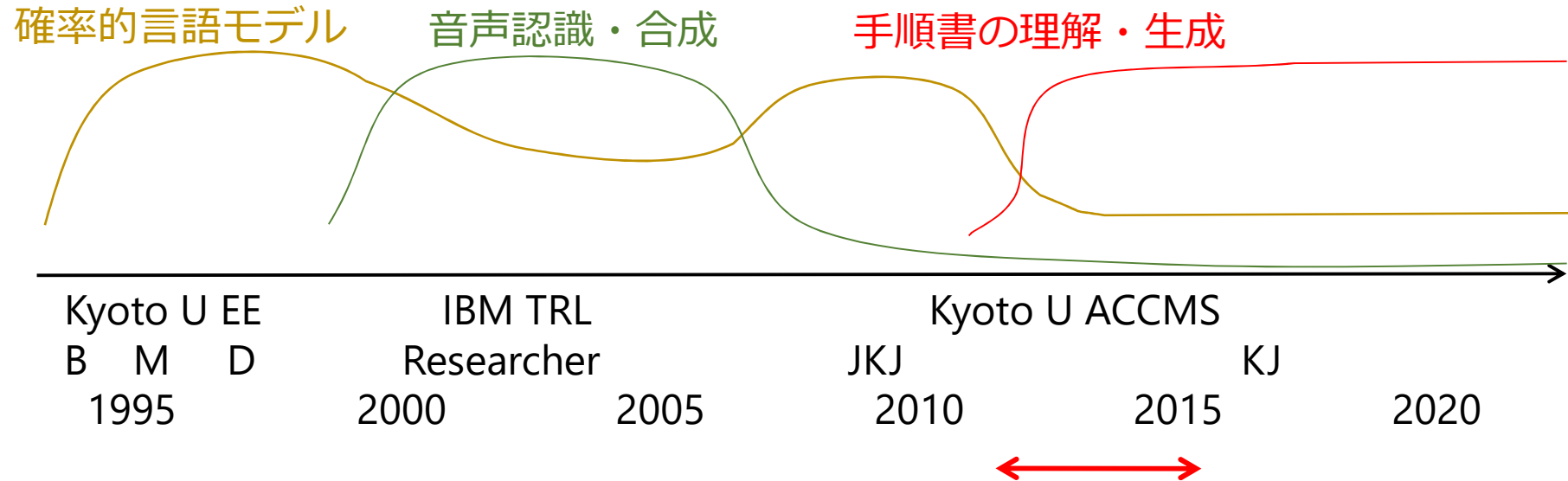


【資料名称】	高廻り2号古墳出土 <u>舟形埴輪</u> 複製
【備考】	原品: 大阪市教育局委員会蔵 幅25.5cm 長さ126.5cm 破片5点あり

講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

端緒 (あるいは動機)



- 2011年 東日本大震災

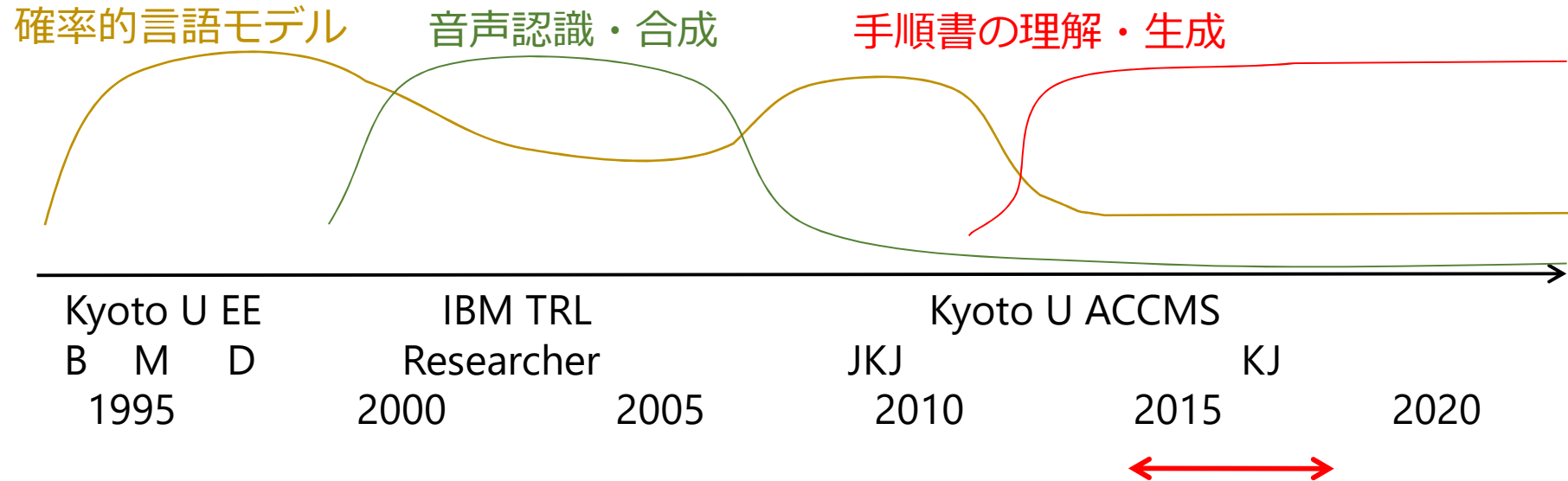
- 実施記録の重要性 (2014年 STAP細胞騒動)

- 2012年 美濃先生 の基盤 (A) の分担

- 食材, 道具, 動作の認識を連携させた調理行動の認識

私的 Vision&Lang. 元年

端緒 (あるいは動機)



- 2014年-2017年 基盤 (B)

- 作業実施映像からの手順文書の自動生成

- 学術的作業実施映像 (学術情報メディアセンター構成員なので)



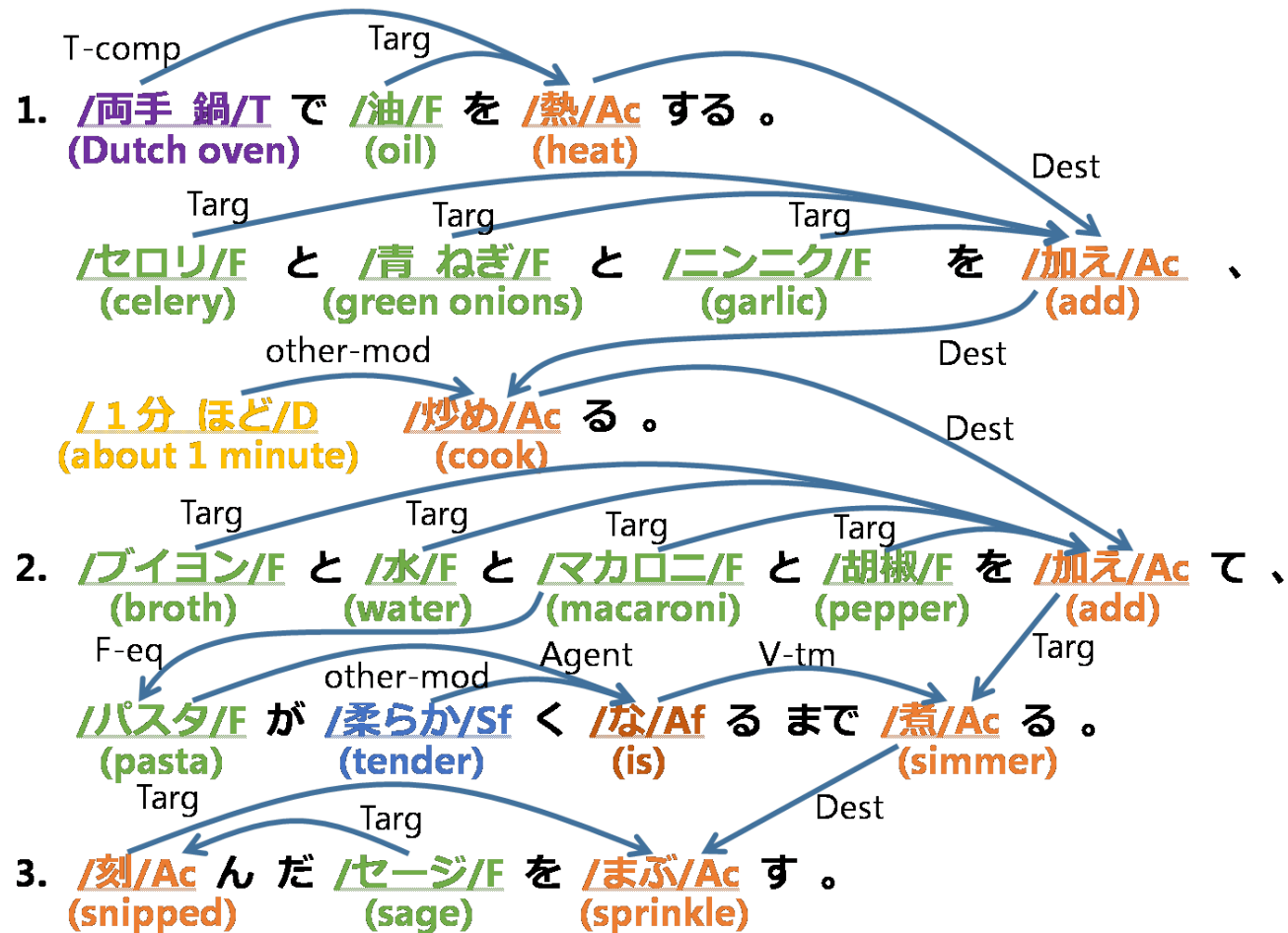
と書店で
手順書を
物色

- 言語生成, 入力: 音声 → 動画

講演の概要

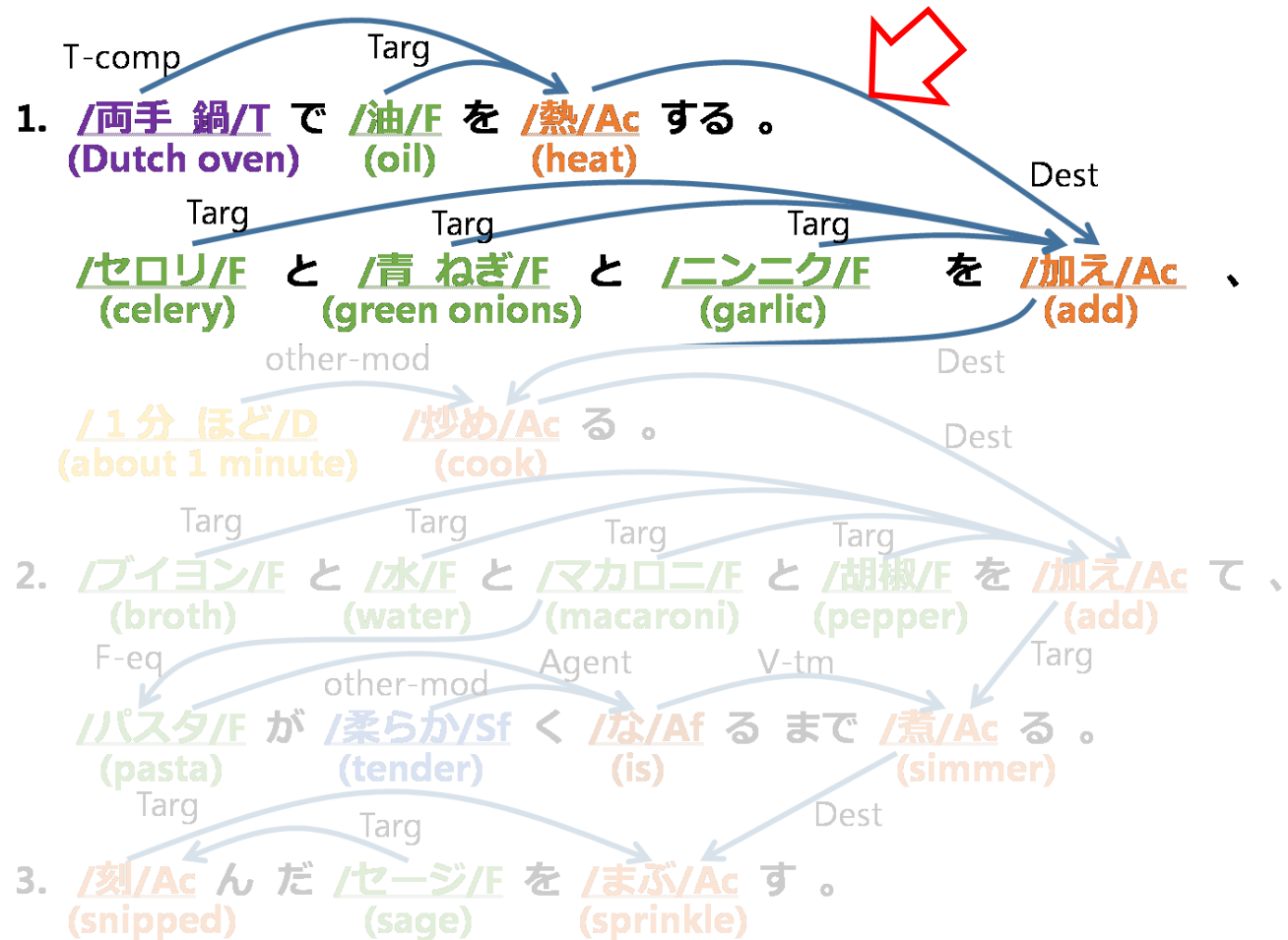
- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

Recipe Flow Graph [LREC14]



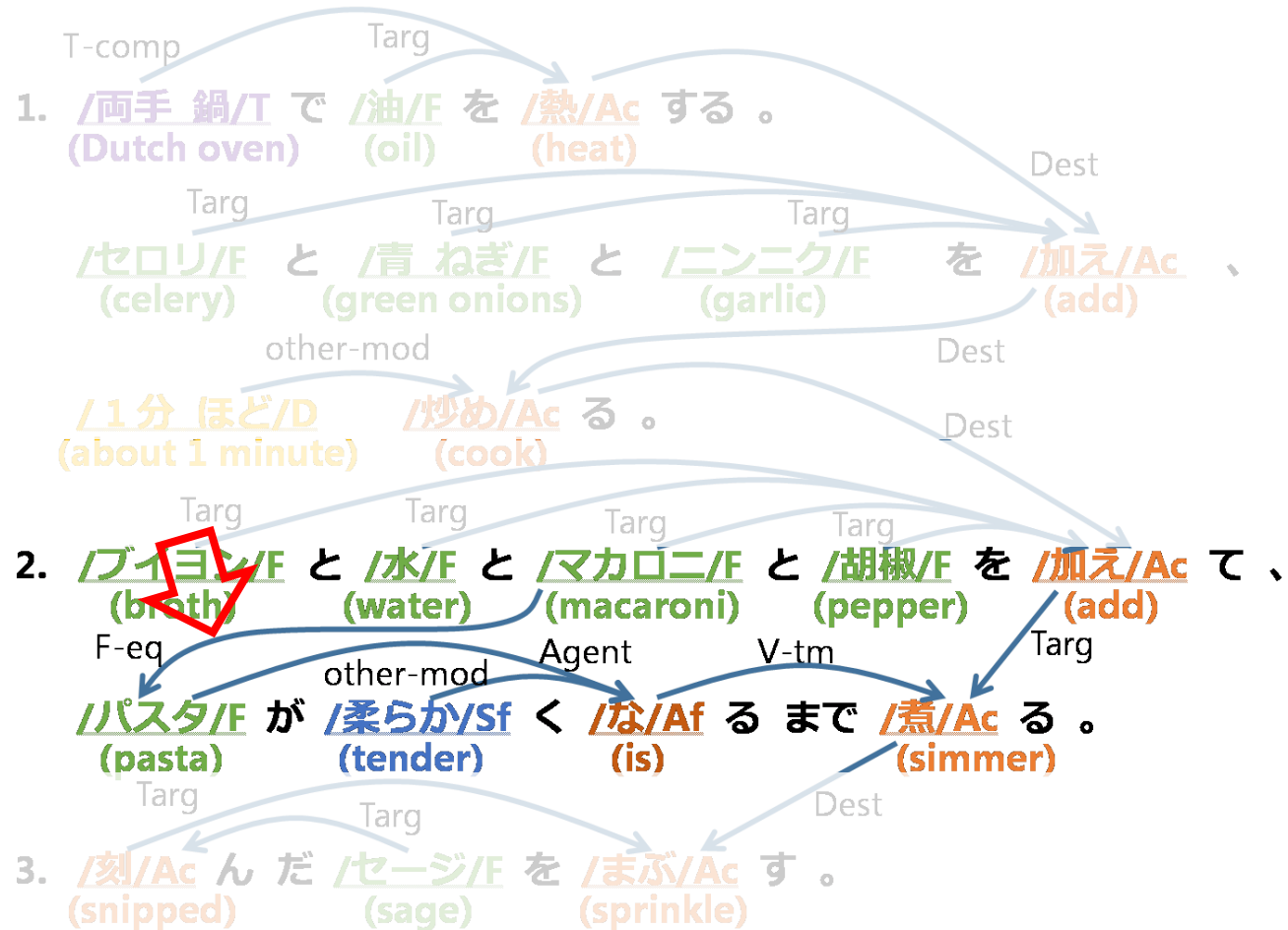
- 最終生成物を根とするDAG (Directed Acyclic Graph)

注目点 1



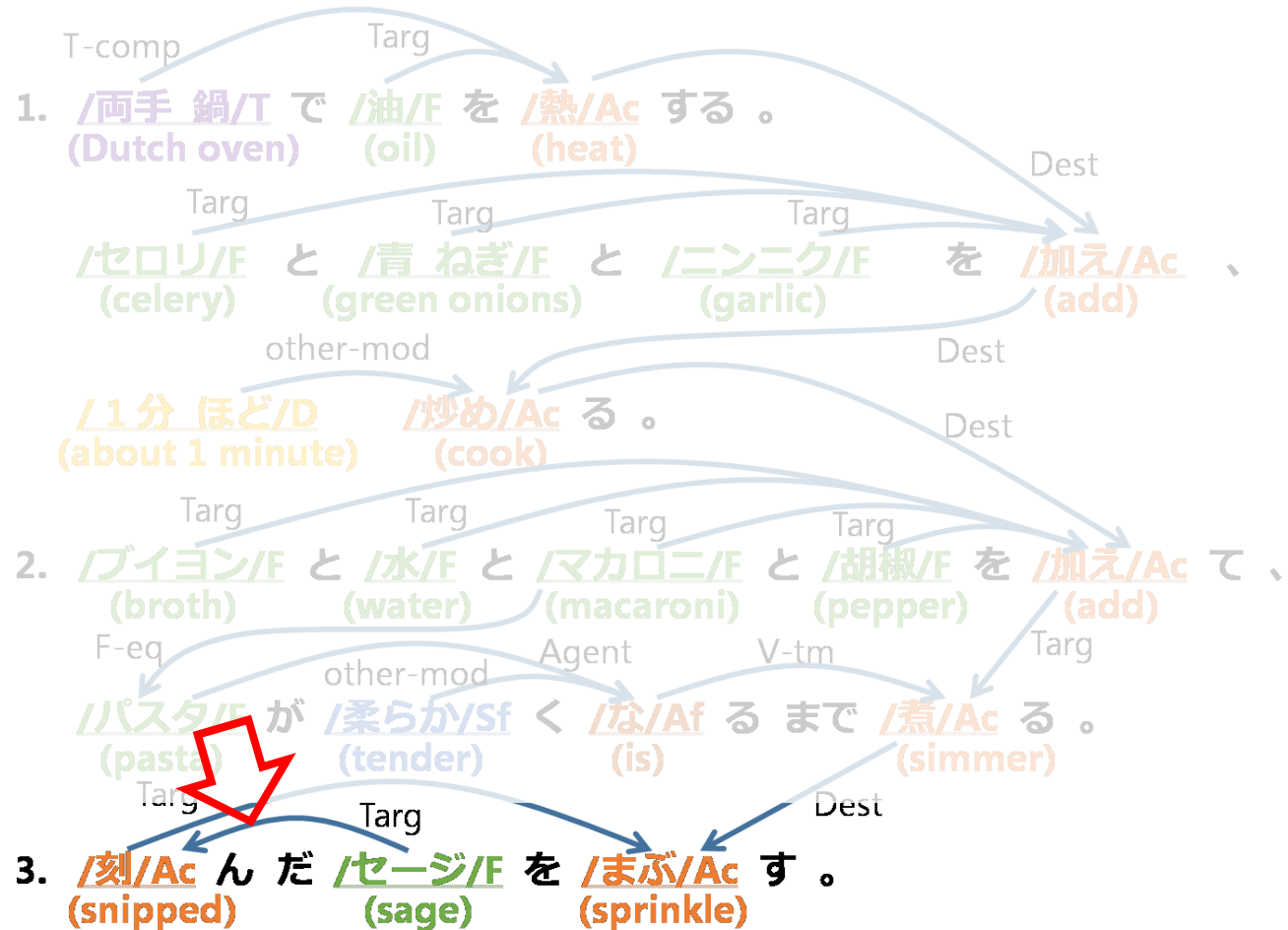
- ある処理の出力が別の処理の入力

注目点 2



- 同一物体の別表現による参照（上位概念・比喻）

注目点 3



- 過去分詞 (名詞修飾) による動作指示

r-NE Vertex Labels

Tag	Meaning	Remarks
F	Food	Eatables including intermediate products,
T	Tool	Knife, container, etc.
D	Duration	Duration for cooking
Q	Quantity	Quantity of food
Ac	Action by chef	Verbs representing chef's actions (stem only)
Af	Action by food	Verbs representing food's actions (stem only)
Sf	Food state	Food's initial or intermediate states
St	Tool state	Tool's initial or intermediate states

Arc Labels

Label	Meaning	Remark
Agent	Action agent	Relationship with actions (Ac or Af)
Targ	Action target	
Dest	Action destination	
T-comp	Tool complement	Tool used in an action
F-comp	Food complement	Food used as a tool
F-eq	Food equality	Identical food
F-part-of	Food part-of	Refer to a part of a food
F-set	Food set	Refer to a set of foods
T-eq	Tool equality	Identical tool
T-part-of	Tool part-of	Refer to a part of a tool
A-eq	Action equality	Identical action (Ac, Af)
V-tm	Head verb for timing, etc.	
other-mod	Other relationships	

r-FG Corpus [LREC14] [Yamakata+ LREC20]

- Open Resource
 - Solution = Source (Program)
+ Resource (Data)
- r-FG Corpus (recipe Flow Graph)
 - 日本語: 266レシピ
 - 英語: 300レシピ



手順書の「理解」 [Maeta+ IWPT15]

1. 両手鍋で油を熱する。

(In a Dutch oven, heat oil.)

セロリと青ねぎとニンニクを加え、

(Add celery, green onions, and garlic.)

1分ほど炒める。

(Cook for about 1 minute.)

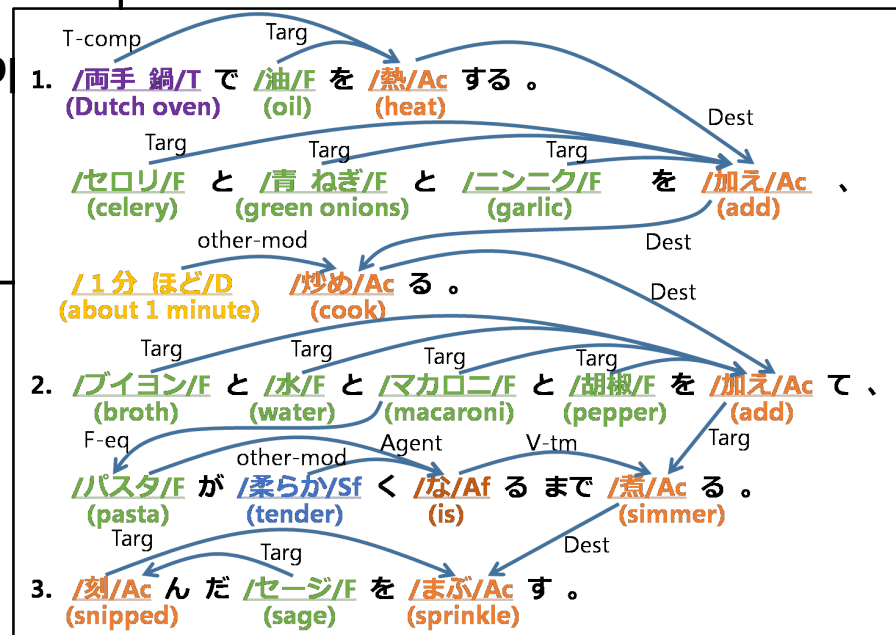
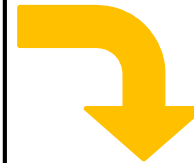
2. ブイヨンと水とマカロニと胡椒を加えて、 パスタが柔らかくなるまで煮る。

(Add broth, water, macaroni, and pepper and simmer until the pasta is tender.)

3. 刻んだセージをまぶす。

(Sprinkle the snapped sage.)

1. 単語分割 (WS)
2. 用語認識 (NER)
3. フローグラフ推定



手順書

1. 両手鍋で油を熱する。

(In a Dutch oven, heat oil.)

セロリと青ねぎとニンニクを加え、1分ほど炒める。

(Add celery, green onions, and garlic. Cook for about 1 minute.)

2. ブイヨンと水とマカロニと胡椒を加えて、

(Add broth, water, macaroni, and pepper,

パスタが柔らかくなるまで煮る。

and simmer until the pasta is tender.)

3. 刻んだセージをまぶす。

(Sprinkle the snipped sage.)

Step 1. 単語分割 (WS)

Result:

1. 両手 鍋 で 油 を 熱 する 。
(In a Dutch oven, heat oil.)

セロリ と 青 ねぎ と ニンニク を 加え 、 1 分 ほど 炒め る 。
(Add celery, green onions, and garlic. Cook for about 1 minute.)

2. ブイヨン と 水 と マカロニ と 胡椒 を 加え て 、
(Add broth, water, macaroni, and pepper,

パスタ が 柔らか く な る ま で 煮 る 。
and simmer until the pasta is tender.)

3. 刻 ん だ セージ を まぶ す 。
(Sprinkle the snipped sage.)

Step 1. 単語分割 (WS)

- 各文字間の単語境界の有無の2値分類
 - 最小の問題に特化 (品詞は不要)
- KyTea [Neubig+ ACL11] <http://www.phontron.com/kytea>
 - 一般コーパス (BCCWJ/UniDic) で学習
 - 約95%
- 言語資源追加
 - レシピ用語に集中して部分的アノテーション
 - 例) 玉-ね-ぎ|と|に-ん-じ-ん|を?み?じ?ん?切?り?に?し?ま?す
 - 約99%

Step 2. 用語認識 (NER)

Result:

1. /両手 鍋/T で /油/F を /熱/Ac する。
(Dutch oven) (oil) (heat)

/セロリ/F と /青 ねぎ/F と /ニンニク/F を /加え/Ac、
(celery) (green onions) (garlic) (add)

/1 分 ほど/D /炒め/Ac る。
(about 1 minute) (cook)

2. /ブイヨン/F と /水/F と /マカロニ/F と /胡椒/F を /加え/Ac て、
(broth) (water) (macaroni) (pepper) (add)

/パスタ/F が /柔らか/Sf く /な/Af る まで /煮/Ac る。
(pasta) (tender) (is) (simmer)

3. /刻/Ac ん だ /セージ/F を /まぶ/Ac す。
(snipped) (sage) (sprinkle)

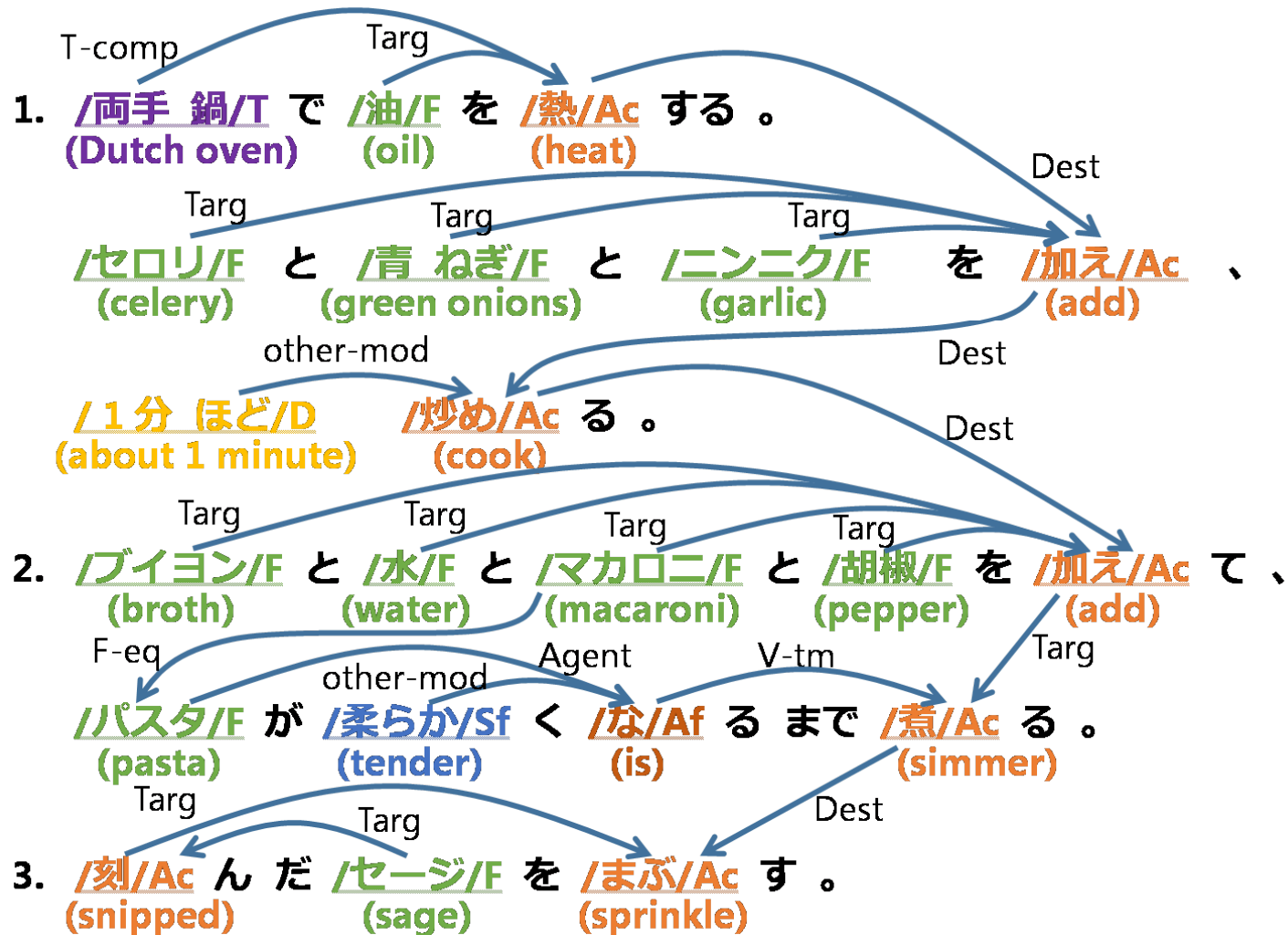
Step 2. 用語認識 (NER)

- 固有表現認識と同じ課題
 - PWNER [Sasada+ PACILNG15] <http://www.lsta.media.kyoto-u.ac.jp/resource/tool/PWNER/>
- 独自の用語 (固有表現) の定義
- コーパス作成
 - 200フローグラフ + 200 NE 付与レシピ
- 約90% (F1)
 - 活用可能 (画像との対応, 食材や道具の頻度計算, etc.)

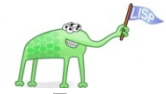
Step 3. フローグラフ推定 [Maeta+ IWPT15]

Parsing of a document, not a sentence

Result:



Step 3. フローグラフ推定 [Maeta+ IWPT15]



- MST Parsing [McDonald & Nivre CL11] + α
 1. Arc score の推定 (Logistic regression)
 2. 最大全域木 (MST; Maximum Spanning Tree) 探索
 - 1st order
 3. 枝の追加
 - DAG 制約
 - LR score $\leq$ penalty(n)
- Arc 単位で 約72%

解析精度

Task	Input	F
単語分割 (WS)	Raw Recipe Texts	98.6
用語認識 (NER)	Gold WS results	90.7
フローグラフ推定 (FGE)	Gold WS/NER results	72.1
WS + NER + FGE	Raw Recipe Texts	51.6

- Vision民, Search民, et al.
 - 高精度を実現せよ! 手段は問わん!
 - トップ会議に通ったとか言っても無駄

Cf. 音声認識 = 音響モデル + 言語モデル (音響屋 v.s. 言語屋)

解析精度

Task	Input	F
単語分割 (WS)	Raw Recipe Texts	98.6
用語認識 (NER)	Gold WS results	90.7
フローグラフ推定 (FGE)	Gold WS/NER results	72.1
WS + NER + FGE	Raw Recipe Texts	51.6

1. 単語分割

- 十分高精度

2. 用語認識

- 一般の固有表現認識と同程度
- 主要クラス (F, T, Ac) は約95%

- 検索などの応用



- 用語と画像の対応

- レシピ生成

解析精度

Task	Input	F
単語分割 (WS)	Raw Recipe Texts	98.6
用語認識 (NER)	Gold WS results	90.7
フローグラフ推定 (FGE)	Gold WS/NER results	72.1
WS + NER + FGE	Raw Recipe Texts	51.6

3. フローグラフ推定

- 実用的とはいえない精度 → ロボ制御
 - 動作の単位での処理 (例: 動画クリップとの対応)
- データの追加・手法の改良

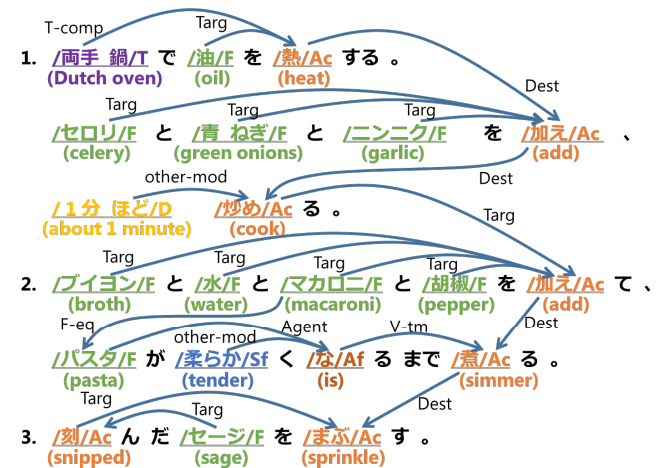
応用

Procedural text

1. 両手鍋で油を熱する。
セロリと青ねぎとニンニクを加え、
1分ほど炒める。
2. ブイヨンと水とマカロニと胡椒を加えて、
パスタが柔らかくなるまで煮る。
3. 刻んだセージをまぶす。

Flow graph

[Mori+ LREC14]



Cooking robot

[Bollini+, ISER13]



www.denso.co.jp

Smart kitchen

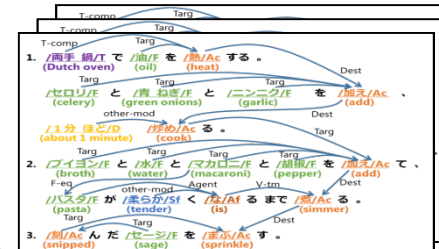
[Hashimoto+, IPMU08]



Cut garlic
Then add it

Intelligent search

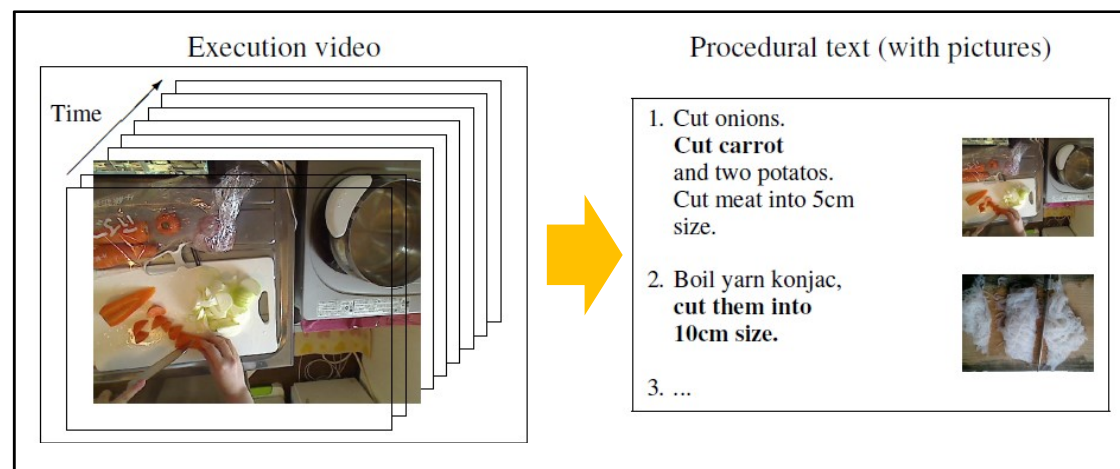
[Yamakata+, SocInfo13]



講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

手順実施動画からの手順書生成



- 音声認識（仮名漢字変換）の 音声（平仮名） → 動画
- 多岐にわたる応用
 - 作業実施映像の高機能アーカイブ
 - 教示システム（スマートキッチン）
- 一般化を意識しつつ調理・レシピで研究

実験動画を残す時代に

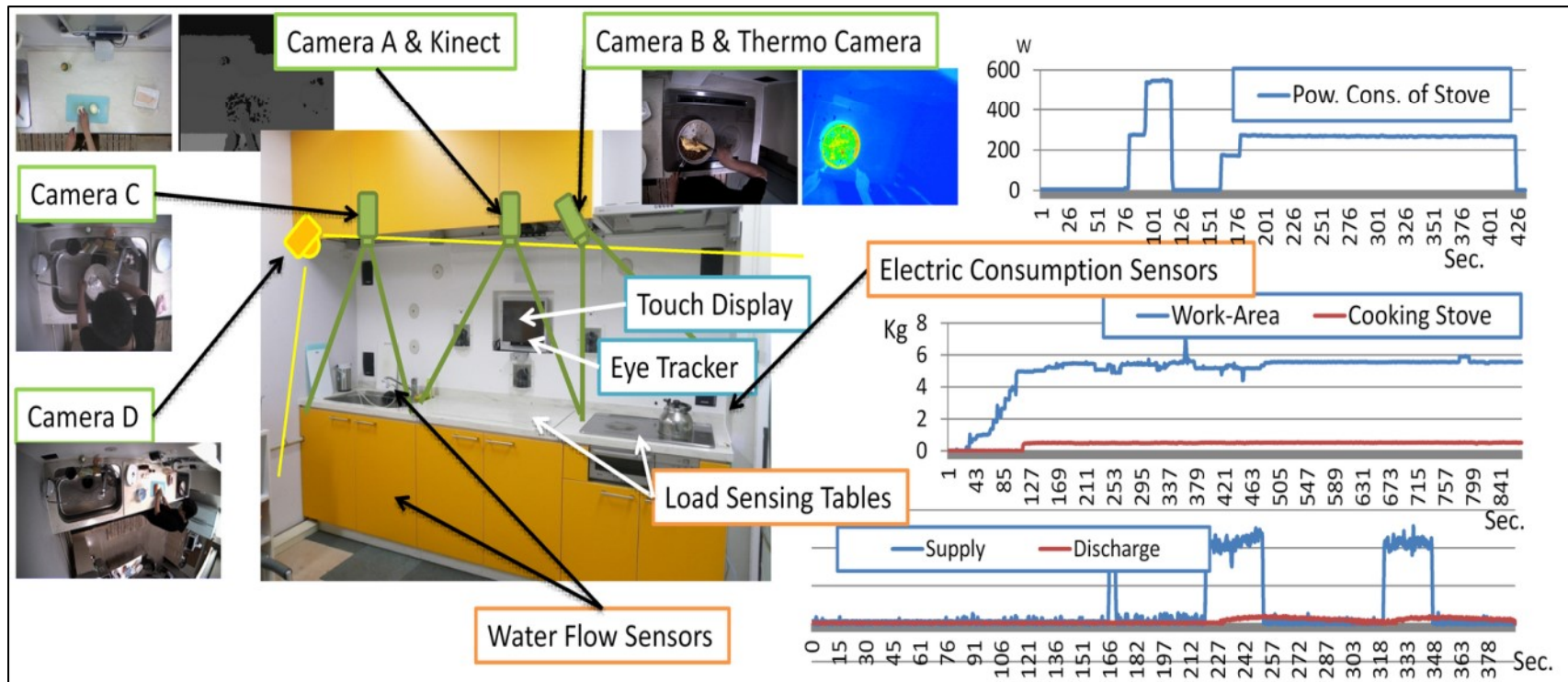
KUSK Dataset [Hashimoto+ CEA15]



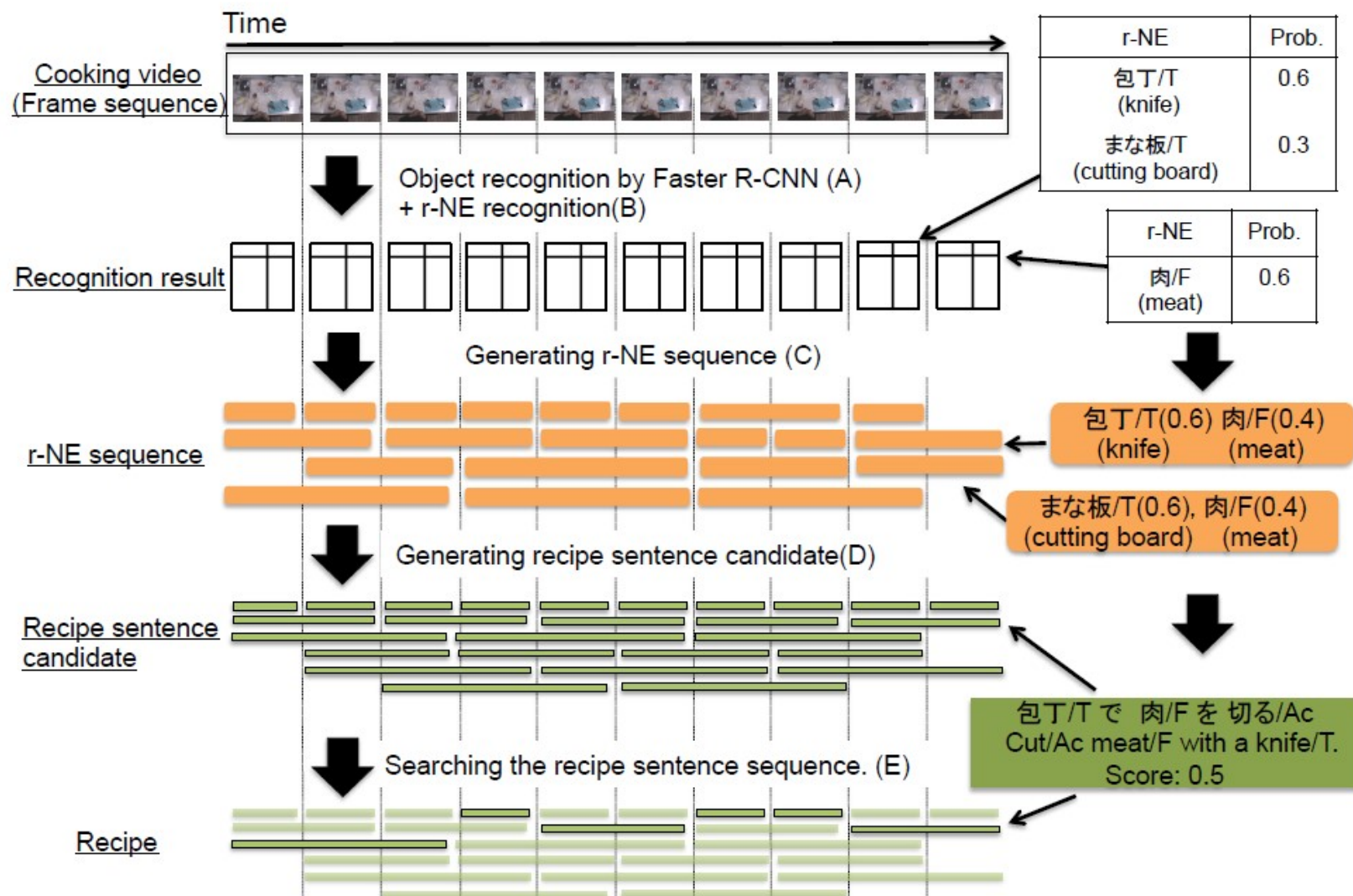
- 20 recipes
- 40 cooking videos

EXAMPLE

Recipe says; *cut chicken into bite-size pieces*
Then human will do ...



物体認識+言語モデル [Ushiku+ IJCNLP17]



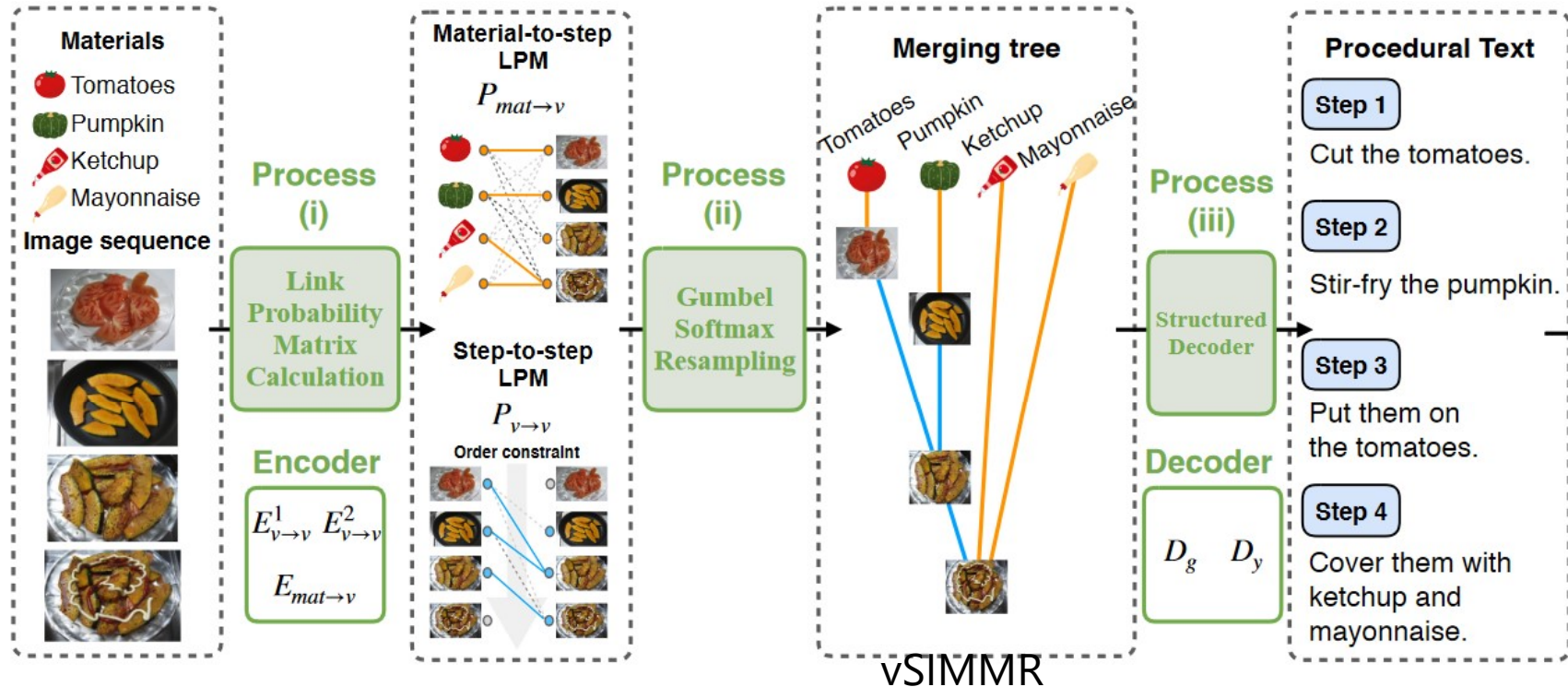
vSIMMR [Nishimura+ LREC20]



- 構造があることが面白い
- 構造と写真のあるレシピコーパス
 - SIMMR [Jermurawong & Habash EMNLP15] の拡張

	視覚情報	構造情報	レシピ数
Cookpad Image Dataset	写真列	なし	1,715,595
レシピフローグラフ	なし	あり	266
vSIMMR	写真列	あり	2,103

End-to-End [Nishimura+ IEEE-A20]



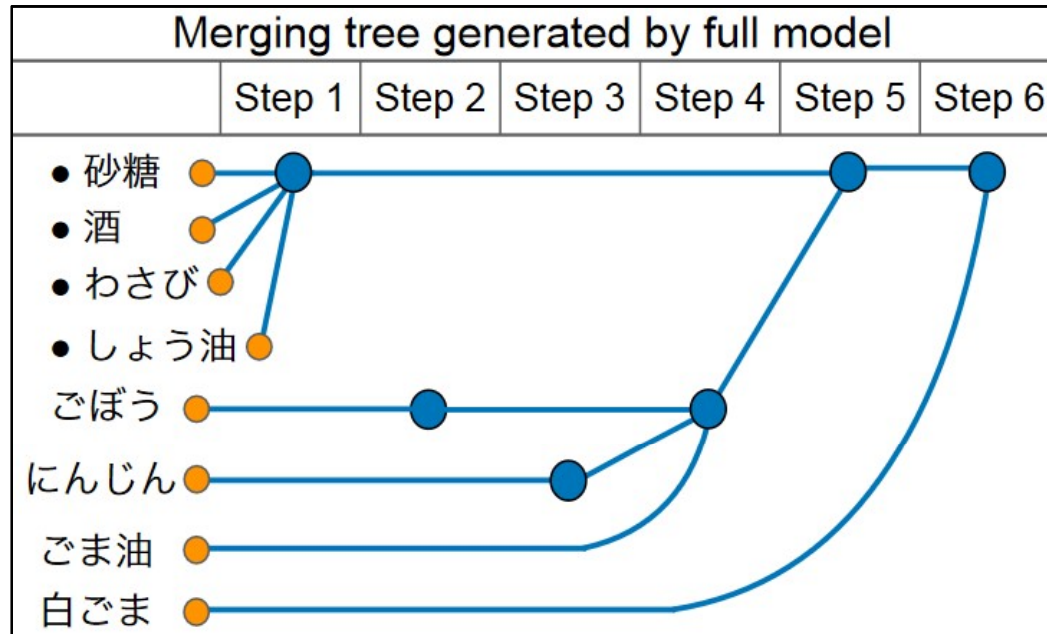
- 構造 (材料の併合) を考慮した文生成
 - グラフも出力可能

文生成例

Ingredients	ごぼう, にんじん, ● 砂糖, ● 酒, ● しょう油, ● わさび, 白ごま, ごま油					
						
Models	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
RetAttn (full model)	材料を用意します。●の調味料を合わせておきます。	ごぼうはさがきにし、水にさらしておきます。	にんじんは千切りにします。	フライパンにごま油を熱し、人参を炒めます。	人参がしんなりしたら、ごぼうを入れて、全体がしんなりするまで炒めます。	全体に火が通ったら、調味料を入れて、よく混ぜます。できあがり！
Ground Truth	●の調味料を混ぜておきます。	ごぼうは皮を剥いて5cmぐらいの長さで細切りにし、水にさらしておきます。	にんじんを細切りにします。	熱したフライパンにごま油をひいてゴボウを炒めます。ごぼうに火がとおったらにんじんを入れて炒めます。	にんじんに火がとおったら①の調味料を入れて汁気がなくなるまで炒めます。	火を止めて、白ごまをふりかけて混ぜればできあがりです。

- 使えるレベルの例もある
 - 「ごぼう v.s. にんじん」の世界知識が必要

副産物



• 食材併合のグラフ

→ 人の作業の**模倣**ロボ

用語は必要?, **文は不要**



S-expression
 (混ぜる
 (ふりかける 白ごま
 (入れる (合わせる ●)
 (炒める
 (炒める
 (さらす
 (ささがき ごぼう)
 水))
 (細切り にんじん))))))

講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道筋

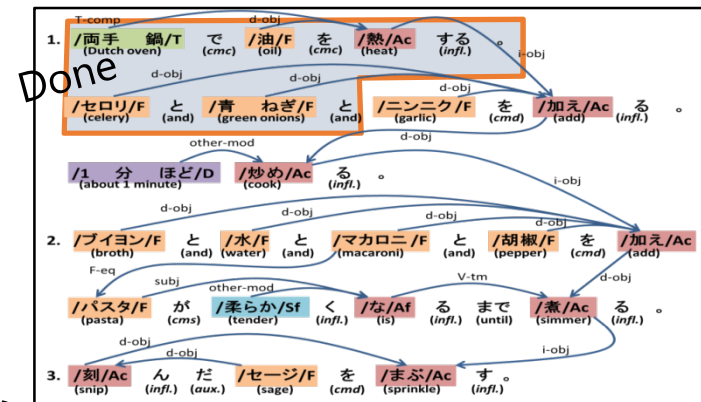
スマートキッチン

- 調理者に次の調理行動を教示

- レシピをフローグラフで保持
- 動画認識で進捗を把握
- 次の調理動作を発話

- 調理者の負荷が低い時には雑談

- レシピ推薦 (スキルの観点で)
- 食材等の蘊蓄



診察/治療

- 医療現場の一人称動画・音声
 - 診察/治療の記録
 - マルチメディア教材



Tocomii 社長 (MD) 自ら収録

生命科学実験



1. /Transfer/Ac /1 - 10 ml/Q of /culture/F overnight to /Eppen/T .
2. /Centrifuge/Ac for /5 minutes/D .
3. Thoroughly /resuspend/Ac /pellet/T with /250 μ l/Q of /Cell Resuspension Solution/F .
4. /Add/Ac /250 μ l/Q of /Cell Lysis Solution/F to each /sample/F .

- 実験動画とプロトコル

- 作業実施映像の高機能アーカイブ
- 教示システム
- 論文執筆補助・教材作成

講演の概要

- 自己紹介
 - 本講演以外の研究
- 手順実施動画からの手順書生成
 - 端緒（あるいは動機）
 - 手順書の構造化
 - 手順書生成
- 実用に向けて
 - 応用例
 - 社会実装への道

社会実装への道

- 動機は何だったか?
 - エンジニア志向
 - 論文発表は生き残り戦略
- 調理・レシピで一定の精度を実証
 - 「現場」にお邪魔してデータ収集
- 現場に有用性を体感してもらう
 - 高機能アーカイブを速やかに試作すべし!!
- 会社は作れる!
 - *Linfer*, Tocomii, LegalForce

