

レシピテキストのためのフローグラフの定義

森 信介¹ 山肩 洋子² 笹田 鉄郎³ 前田 浩邦⁴

^{1,3} 京都大学 学術情報メディアセンター ^{2,4} 京都大学 情報学研究科

〒606-8501 京都市左京区吉田本町

¹forest@i.kyoto-u.ac.jp

1 はじめに

自然言語処理の目標の1つは理解であるが、文章一般に対して理解の良い定義を与えるのは非常に難しい。しかしながら、手続きを記述した文章に対しては、概ね理解を定義できる。Momouchi [1] は、手続き文章をフローグラフとして表すことを提案し、文章からの自動変換を試みている。この取り組みは、コーパスに基づく手法が主流になる以前のもので、今は利用可能な言語資源がない。

このような背景の下、理解の表現形式としてフローグラフを用い、手続き文章にフローグラフを付与したコーパスを作成している。本論文では、このコーパス作成の取り組みについて述べる。典型的な利用方法は、文章からの自動変換であろう。語彙や表現を限定することでこれが現実的となるように、レシピに注目することとした。レシピに限定することの他の利点として、実施環境が狭く制御下におくことができ、実施映像を収録することが容易であることが挙げられる。実際、映像処理の研究者は、複数のカメラやセンサーを備えたキッチンを構築し、調理映像を収録している [2]。我々のコーパスに含まれるレシピの一部は、この取り組みと共有している。それを用いることで、記号接地の研究 [3] [4] や画像や映像からの文生成 [5] などに対して、より現実的な状況に対して取り組める。

実際、調理に関しては、索引生成を目的として、レシピを「木」に変換し、調理映像と照合する研究がある [6]。この研究におけるレシピ木の定義は、本論文で我々が提案するフローグラフの定義ほどの完成度ではない。実際、多くはないが食材の分離などは木で表すことができない。レシピには、省略、名詞のみならず動詞の共参照、部分全体関係など様々な言語現象が現れ、これをすべて表現できる必要がある。

本論文では、これらの現象を扱うことができるフローグラフの定義とそれに基づくコーパス作成について述べる。我々のコーパスを用いることで、固有表現認識、述語項構造解析 [7]、共参照解析 [8] をはじめとして、レシピテキスト全体の「理解」の研究が可能になる。

2 レシピテキスト

通常、レシピは以下の3つの部分から構成される。

1. タイトル
2. 食材リスト (材料名および分量)
3. 調理手順 (手順番号および手順説明文)

以上に加えて、調理手順の説明のための写真や完成した料理の写真、あるいは、レシピ作成者や利用者のコメントなどの情報が付与されている場合もあるが、本研究では扱わない。

本稿で紹介するフローグラフは、調理手順の部分の意味内容を表現する。レシピテキストの例を図1に、この調理手順に対応するフローグラフを図2にそれぞれ示す。このレシピは、調理作業が複雑で、かつ説明文の解釈における難度も高い事例である。

2.1 レシピフローグラフ

レシピで記述される手順には、依存関係があるものもあれば、ないものもある。例えば、調理者は、ニンジン茹でる前にニンジン切る必要があるが、ニンジン茹でる前にジャガイモ切る必要はない。調理の手順を表現するために、フローグラフを用いる。ほとんどのレシピは木で表現できるが、木で表現できないレシピも存在する。典型的な例は、食材の分割である。したがって、フローグラフは、無閉路有向グラフ (DAG; Directed Acyclic Graph) $G = (V, E)$ であるとする。ここで、 V は頂点の集合であり、 E は辺の集合である。図2にその例を示す。フローグラフの頂点は、食材や道具、あるいは動作などに対応する。辺は、それらの関係を表す。一般のDAGとは異なり、レシピフローグラフは根と呼ばれる特別の頂点がある。これは、料理の完成品に対応する。以下の節では、頂点と辺についての説明を行う。

料理名：なんなんなん♪

材料 (二人分)
 ☆トマト約2個
 ☆たまねぎ(小)約1個
 ☆七味唐辛子適宜
 ☆オリーブオイル適宜
 ☆バジル適宜
 ☆酢適宜
 ☆レモン汁適宜
 ☆イタリアン
 エッセンス適宜
 ☆豆板醤適宜
 ☆塩適宜
 キャベツ少量
 ビーマン約1個
 とろけるチーズ4枚
 ウィナー4本
 ケチャップ適宜

作り方

- 1 サルサソースを作る。トマトは角切り、たまねぎはみじん切りにし、☆の材料を全てボールに入れて混ぜる。味を見ながら調味料を加えてください。
- 2 キャベツは千切り、ビーマンは輪切りにする。
- 3 ナンにとろけるチーズを巻いたウィンナーをのせ、かる〜くオーブントースターでチンする。
- 4 チーズのとけた上から、サルサソース、切った野菜(キャベツ、ビーマン)をのせる。
- 5 ケチャップを好きなようにかけてめしがれ♪

図 1: 「なんなんなん」のレシピ

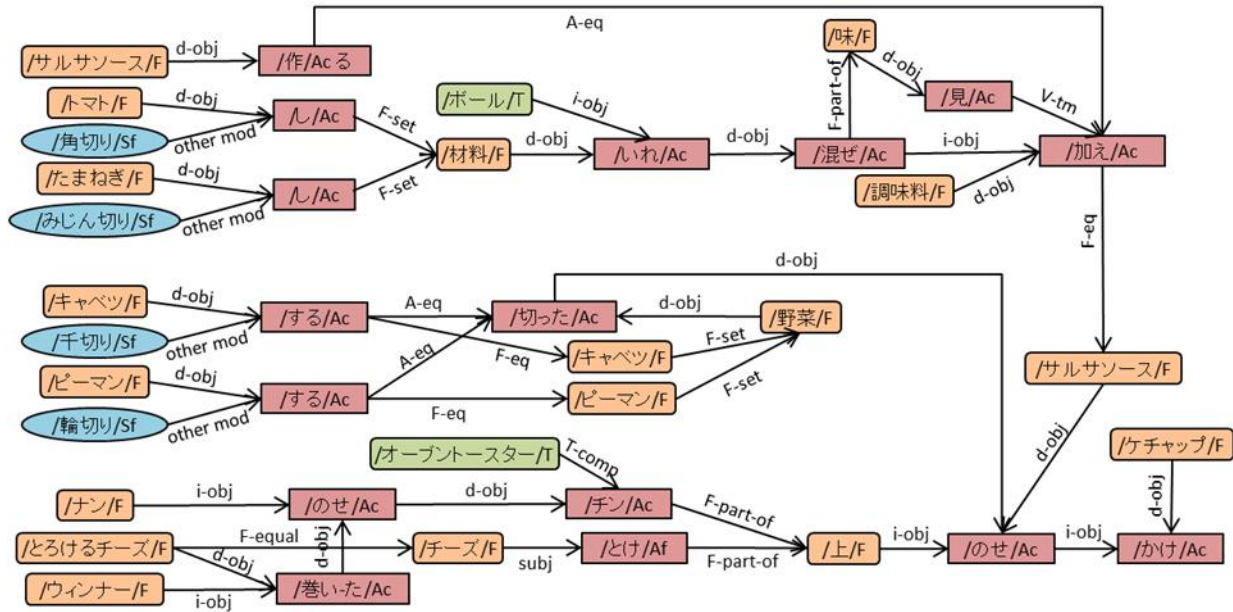


図 2: 「なんなんなん」のレシピから生成されたフローグラフ

3 フローグラフの頂点

フローグラフの頂点は、食材や動作を表す単語列からなる。通常、ある実態に対応する単語列を固有表現と呼ぶので、本コーパスにおいてもこれを踏襲する。表 1 に、固有表現の種類(クラス)を掲げる。この表からわかるように、通常の固有表現と異なり、調理者や食材の動作を表す用言も含む。以下では、各固有表現クラスについて順に説明する。例の表記にあたっては、単語境界を空白で示し、固有表現となる単語列をスラッシュ(/)で囲み、最後にクラスを示す記号を付加している。

3.1 F: 食材

食材は、材料リストに記載されているような材料名に加えて、「タネ」「生地」といった中間食材名や、料理の完成品を含む。野菜のなどの捨てる部分など、食材とすべきか判断に迷う表現もあるが、食せる、捨てる、量が変化するならば F と判定する。

表 1: 固有表現の種類

タグ	意味	備考
F	食材	中間・最終生成物を含む
T	道具	調理道具や器など
D	継続時間	概数表現を含む
Q	分量	概数表現を含む
Ac	調理者の動作	語幹のみ
Af	食材の動作	語幹のみ
Sf	食材の状態	
St	道具の状態	

典型的な例は、以下の通りである。ただし修飾句は F に含めない。

例) 新しい /卵/F を /割/Ac って

商品名などの固有名詞は、全体を食材とする。

例) /めんつゆの素/F

以下の「水分」は捨てるものだが F とする。

例) /春雨/F の /水分/F を /切/Ac り

単独で食材を指す番号なども食材とする。

例) /1/F を

ただし、次の例のように、形容詞的に使われる番号は、より具体性の高い単語列の方のみを F とする。

例) 1 の /材料/F を

3.2 T: 道具

鍋、蓋、包丁、コンロなど、調理道具や器を道具とする。手や指などの調理者の身体の一部も道具のような振る舞いをする場合があり、T とする。食せない、量が変化しない点以外は、概ね「食材」の判定と同じである。

典型的な例は、以下の通りである。

例) /3 分/D /レンジ/T を /し/Ac てから

ただし、直後にサ変動詞語幹が続き、動作を表現している場合は、後述する調理者の動作とする。

例) /3 分/D /レンジ/Ac する

火加減や道具の一部を指す表現も道具とする。

例) /弱火/T で /煮/Ac る

「水」や「手」など、それを使って、食材等に動作を行う場合は、道具とする。

例) /水/T で /洗/Ac って

例) /手/T で /洗/Ac って

3.3 D: 継続時間

加熱時間や冷却時間など、加工の継続時間を示す。数字と単位のほか、それらに対する概数表現等も含める。典型的な例は以下の通りである。

例) /12 ~ 15 分/D 間 /煮込/Ac みます

例) /5 分くらい/D

例) /2 日後くらい/D が /食べ時/Af です!

3.4 Q: 分量

食材の量を指定する表現を分量とする。数字部分や単位に加えて、それらに対する概数表現等も含める。典型的な例は以下の通りである。

例) /人参/F /3 ~ 4 cm くらい/Q

3.5 Ac: 調理者の動作

調理者が主語となる用言で、主に他動詞である。表層表現の比較により同一性を判定できるようにするために、活用語尾を含めない。また、補助的動詞は含めない。典型的な例は以下の通りである。

例) /切/Af った /ニンジン/F を

使役や否定の助動詞を伴う場合のみ、これらの助動詞語幹までを含む。

例) /沸騰させ/Ac たら

目的語など格助詞で示される「項」を含めない。

例) /皮/F を /む/Ac いて

複合動詞は全体を調理動作とする。

例) /薄力粉/F を /振るいいれ/Ac

開始や完了などをあらかず補助的な動詞は含まない。

例) /煮込/Ac んでいく

例) /煮た/Af ってくる

動詞派生名詞やサ変名詞などの事態性名詞も動作とする。

例) /ねぎ/F を /みじん切り/Ac。

例) /ねぎ/F を /みじん切り/Ac する。

例) /ねぎ/F を /みじん切り/Sf に。

例) /ねぎ/F を /みじん切り/Sf に /する/Ac。

例) /みじん切り/Sf の /ねぎ/F を 入れ/Ac

3.6 Af: 食材の動作

食材が主語となる用言もある。その場合には、Af とする。範囲や判定の基準は Ac と同じである。

3.7 Sf: 食材の様態

用意された食材様態の初期状態を表す表現、並びに Ac や Af の影響で遷移する(した)状態を表す表現を Sf とする。ただし、量の表現は Q である。

例) /柔らか/Sf く /な/Af る

例) /色/Sf が /変わ/Af る

例) /濃いめ/Sf の /調味料/F

3.8 St: 道具の様態

用意された道具様態の初期状態を表す表現、並びに Ac や Af の影響で遷移する (した) 状態を表す表現を St とする。

例) /強火/St の /フライパン/T で /炒め/Ac する

例) /オープン/T を /150度/St に /予熱/Ac する

「よく」「ざっくり」などの副詞表現は、固有表現としない。

4 フローグラフの辺

前節で述べた固有表現について、文章中の固有表現間に以下であげる 3 つのうちのいずれかの関係が認められる場合、それらの固有表現に対応する頂点の間に枝を張り、ラベルとしてその関係の種類を付与する。

- 述語項関係がある場合 (ゼロ照応も含む)
- 同種の 2 つの固有表現の間に何らかの参照関係がある場合
- 2 つの動作の間に何らかの関係がある場合

ラベルの一覧を表 2 に掲げた。以下では、これらを順に説明する。

4.1 主語 (ガ格, subj) ・ 対象 (ヲ格, d-obj) ・ 方向 (ニ格, i-obj)

レシピにおいて最も基礎的な情報は、調理者がどの食材にどのような加工を行ったかということである。主語 (ガ格) ・ 対象 (ヲ格) ・ 方向 (ニ格) は、調理者の動作とその対象食材を関連付けることから、最も重要な関係であると言える。

主語 (ガ格) は多くの場合、『食材 (F)』と『食材の動作 (Af)』を関係づける。たとえば、「/水/F が /沸騰/Af したら」という文は次のような枝を生成する。

例) $\boxed{/水/F} \xrightarrow{\text{subj}} \boxed{/沸騰/Af}$

対象 (ヲ格) は、『調理者の動作 (Af)』とその対象となる『食材 (F)』とを関係づける。たとえば、「/チーズ/F を /おろ/Ac して」という文は次のような枝を生成する。

例) $\boxed{/チーズ/F} \xrightarrow{\text{d-obj}} \boxed{/おろ/Ac}$

それ以前の調理動作によって生成された食材を対象として調理動作を行う場合は、その対象食材が文中に表れてい

ない場合がある。たとえば、「玉葱を炒めたあと取り出す」とある場合、「/取り出/Ac」した対象の食材は、「/玉葱/F を /炒め/Ac」した結果として生成された食材である。このように、ある動作の前の食材と後の食材を別のものとして区別し、後の参照においては、どちらであるかを明示する。

例) $\boxed{/取り出/Ac} \xrightarrow{\text{d-obj}} \boxed{/炒め/Ac}$

方向 (ニ格) は、食材を『食材 (F)』に混ぜる場合や、食材を『容器 (T)』に入れる場合など、主に調理動作と対象食材の行き先を関係づける。たとえば「/材料/F を /ボール/T に /入れ/Ac て」という場合は以下のようになる。

例) $\boxed{/材料/F} \xrightarrow{\text{d-obj}} \boxed{/入れ/Ac}$

例) $\boxed{/ボール/T} \xrightarrow{\text{i-obj}} \boxed{/入れ/Ac}$

また、対象 (ヲ格) と同様、方向 (ニ格) においても、「フライパンを熱します。玉葱を入れて...」のように、行き先の食材や道具が明記されていない場合がある。この場合は、その行き先の道具や食材に最後に加えた調理動作を指定する。

例) $\boxed{/過熱/Ac} \xrightarrow{\text{i-obj}} \boxed{/入れ/Ac}$

ここで、主語・対象・方向の関係を表す文において、必ずしもそれぞれガ・ヲ・ニの助詞が使われているとは限らない。そこで、次のような基準を導入した。

- 文に格助詞「ガヲニ」が使われている場合
意味ではなく文法に従って「ガ格」「ヲ格」「ニ格」を指定する。たとえば「鍋にバターを熱し」という文では、意味的には「鍋ヲ熱し」なので「道具ヲ」格が正しいが、文では「方向ニ」が使われているので、「方向ニ」格とする。
- 文で「ガヲニ」が使われていない場合
意味に従って格を決定する。たとえば「鍋もバターを熱し」という文では、「鍋」に格助詞「ガヲニ」が使われていないので、意味を考える。意味的には「鍋ヲ」なので、「道具ヲ」格とする。

4.2 道具ヲ格 (T-comp) ・ 食材ヲ格 (F-comp)

道具ヲ格は、『調理者の動作 (Ac)』と、それを実行するのに使用される『道具 (T)』を関連づける。「皮をむく」という動作は、包丁を使った場合とピーラーを使った場合で異なるので、このレシピフローグラフを映像における調理作業認識のためのシナリオとして利用する際には、この道具は非常に重要な要素となる。たとえば「アルミホイルで覆い」という文は以下のようになる。

表 2: 辺のラベル

ラベル	意味	概説
subj	主語 (ガ格)	主に「が」や「は」で表される動作と主語の関係
d-obj	対象 (ヲ格)	主に「を」で表される動作と対象の関係
i-obj	方向 (ニ格)	主に「に」で表される動作と方向や場所の関係
T-comp	道具デ	主に「で」で表される動作とその手段の関係
F-comp	食材デ	主に「で」で表される動作とその手段の関係
F-eq	同一の食材	既出の食材とそれに対する参照表現
F-part-of	食材の一部	既出の食材とその一部に対する参照表現
F-set	食材の集合	既出の複数種の食材とその全体に対する参照表現
T-eq	同一の道具	既出の道具とそれに対する参照表現
T-part-of	道具の一部	既出の道具とその一部に対する参照表現
A-eq	同一の動作	既出の動作とそれに対する参照表現
V-tm	動作を行う	別の動作を行う条件やタイミングを示す句の動詞
other mod	その他の修飾語句	

例) $\boxed{/アルミオイル/T} \xrightarrow{T-comp} \boxed{/覆/Ac}$

また、『調理者の動作 (Ac)』を実行する道具が食材である場合もある。これは食材デ格で指定する。たとえば、「/オリーブオイル/F で /焼/Ac きます」という文は以下のようなになる。

例) $\boxed{/オリーブオイル/F} \xrightarrow{F-comp} \boxed{/焼/Ac}$

4.3 同一の食材 (F-eq)・食材の部分 (F-part-of)・食材の集合 (F-set)

すでに登場した食材と同一の物体を指している固有表現を関連付けるためにいくつかのラベルを用意した。たとえば「1) 玉ねぎを切ります、2) 1) を炒めます」の場合、2 個目の「1)」は「手順 1)」の結果 (すなわち炒めた結果) と同一食材なので、手順 1) の最後の動作を「同一の食材 (F-eq)」とする。

例) $\boxed{/1)/F} \xrightarrow{F-eq} \boxed{/炒め/Ac}$

以前に登場した食材の一部を指しているとは判断できる場合、たとえば「1) 玉ねぎの皮を剥いて」という文においては、「皮」は「玉ねぎ」の一部であると考えられるため、「皮」を「玉ねぎ」の『食材の一部 (F-part-of)』に指定する。

例) $\boxed{/皮/F} \xrightarrow{F-part-of} \boxed{/玉ねぎ/F}$

以前に登場した食材の集合を指しているとは判断できる場合、たとえば「1) 玉ねぎと人参を切ります。2) 野菜を炒めます」という文章では、「野菜」が「玉ねぎと人参」

を指していると考えられるため、「野菜」の「食材の集合 (F-set)」に「玉ねぎ」と「人参」を指定する。

例) $\boxed{/玉ねぎ/F} \xrightarrow{F-set} \boxed{/野菜/F}$
 $\boxed{/人参/F} \xrightarrow{F-set} \boxed{/野菜/F}$

4.4 同一の道具 (T-eq)・道具の一部 (T-part-of)

すでに登場した道具と同一の物体を指している単語同志を関連付ける。たとえば「蓋をして 10 分煮込みます。蓋を外してさらに煮詰めます」の場合、蓋はどちらも同一物体と考えられるので、以下のようにする。

例) $\boxed{/蓋/T} \xrightarrow{T-eq} \boxed{/蓋/T}$

以前に登場した道具の一部を指しているとは判断できる場合、たとえば「圧力鍋で煮ます。10 分たったら蓋を外し」の場合、「蓋」は「圧力鍋」の蓋であると考えられるため、「蓋」の『道具の一部 (T-part-of)』に「圧力鍋」を指定する。

例) $\boxed{/圧力鍋/T} \xrightarrow{T-part-of} \boxed{/蓋/T}$

4.5 同一の動作 (A-eq)

実際には一度しか行わない動作が、文中では複数回記述される場合がある。たとえば「玉葱₁をみじん切り₂します。先ほど切₃った玉葱₄を」の場合、後者の「切₃」は前者の「みじん切り₂」を指示しているにすぎず、実際には行わなくてよい。このような場合に、両固有表現を『同一の動作 (A-eq)』によって関連づけ、実際に行う動作と行わ

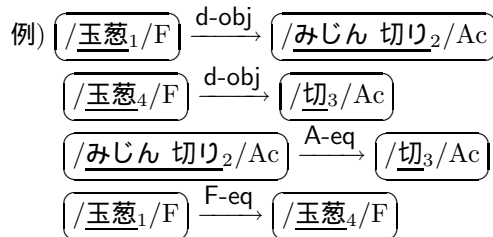
表 3: 1 レシピあたりの頂点のラベル数

固有表現タグ	数
F	11.65
T	4.12
D	0.57
Q	0.88
Ac	13.72
Af	1.72
Sf	3.20
St	0.25
合計	36.11

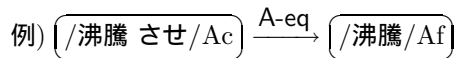
表 4: 1 レシピあたりの辺のラベル数

辺のラベル	数
subj	1.89
d-obj	15.23
i-obj	7.05
T-comp	1.58
F-comp	0.59
T-eq	0.51
F-eq	3.19
T-part-of	0.27
F-part-of	2.46
F-set	0.11
V-tm	0.81
A-eq	0.56
other-mod	5.12
合計	39.37

ない動作を区別する。この例の場合には、共参照も含めて以下のような部分グラフになる。

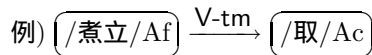


上述の関係は『調理者の動作 (Ac)』だけでなく、『食材の動作 (Af)』にも発生する。たとえば『水を沸騰させる。沸騰したら...』の場合、以下ようになる。



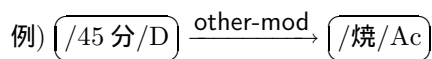
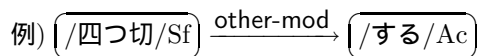
4.6 動作を行うタイミング (V-tm)

動作を行うタイミングは調理を行う上で重要な指示情報であり、「したら」「するまで」「しないように」といった記述で表現される。たとえば「煮立ってきたら灰汁を取り」の場合は以下のようにする。



4.7 その他の修飾語句 (other-mod)

その他のすべての修飾関係はこの関係によって記述される。たとえば次のようなものがある。



5 フローグラフコーパス

前述までの基準にしたがって、無作為に抽出した 91 レシピを手でフローグラフに変換した。固有表現タグ付与や必要個所の単語分割を含めて、1 レシピあたり平均で約 40 分を要した。本節では、我々のフローグラフコーパスの統計情報を提示する。

5.1 統計情報

頂点のラベルごとの 1 レシピあたりの数を表 3 に掲げる。この表からもわかるように、食材 (F) と調理者の動作 (Ac) が頻度という観点でもっとも重要ということがわかる。表 4 は、辺のラベルごとの 1 レシピあたりの数である。主に食材と動作を関係づける d-obj が最多で、その方向や場所を指定する i-obj が続く。我々のコーパスにより、同一の料理のレシピ間でのあるいは異なる料理の間での様々な比較が可能となるであろう。

表 3 の合計と表 4 の合計を比べると、辺の数が頂点の数よりも少し多いことがわかる。もしすべてのフローグラフが木であれば、辺の数は頂点の数よりちょうど 1 少なくなるはずである。実際の差は 3 を少し超える程度であるから、我々のフローグラフは、ほぼ木だといってよい。

6 おわりに

本論文では、日本語の調理手順文章の意味表現の詳細について述べた。データを観察した結果、表現形式としては、

木ではなく無閉路有向グラフ (DAG) が適切であることが分かった。実際に、約 100 のレシピをフローグラフに変換し統計情報を提示した。

我々のフローグラフコーパスは、ラベルにテキストの単語列を出現位置情報とともに保持しているのので、部分的単語分割コーパスとして単語分割器 [9] [10] の学習に用いることができる。同様に、頂点の固有表現タグを用いて、固有表現認識の研究にも活用することが可能である。

我々のコーパスは、調理手順テキストの自然言語理解の結果とみることもでき、述語項構造解析や共参照解析からレシピ文章全体の理解に取り組むことも可能となる。

謝辞

本研究は科研費 (課題番号: 23500177, 23700144, 24240030) の助成を受けたものである。ここに謝意を表する。

参考文献

- [1] Yoshio Momouchi. Control structures for actions in procedural texts and pt-chart. In *Proceedings of the Eighth International Conference on Computational Linguistics*, pp. 108–114, 1980.
- [2] Atsushi Hashimoto, Naoyuki Mori, Takuya Funatomi, Masayuki Mukunoki, Koh Kakusho, and Michihiko Minoh. Tracking food materials with changing their appearance in food preparing. In *Proceedings of the Second Workshop on Multimedia for Cooking and Eating Activities*, 2010.
- [3] Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. Grounding action descriptions in videos. *Transactions of the Association for Computational Linguistics*, Vol. 1, No. Mar, pp. 25–36, 2013.
- [4] Haonan Yu and Jeffrey Mark Siskind. Grounded language learning from video described with sentences. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2013.
- [5] Yezhou Yang, Ching Lik Teo, Hal Daumé III, and Yiannis Aloimonos. Corpus-guided sentence generation of natural images. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 2011.
- [6] Reiko Hamada, Ichiro Ide, Shuichi Sakai, and Hidehiko Tanaka. 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.
- [7] Mitchell Marcus, Grace Kim, and Mary Ann Marcinkiewicz. The penn treebank: A revised corpus design for extracting predicate argument structure. In *Proceedings of the International Symposium on Spoken Dialogue*, pp. 77–81, 1993.
- [8] Xiaofeng Yang, Jian Su, GuoDong Zhou, and Chew Lim Tan. An np-cluster based approach to coreference resolution. In *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 226–232, 2004.
- [9] Graham Neubig and Shinsuke Mori. Word-based partial annotation for efficient corpus construction. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, 2010.
- [10] Graham Neubig, Yosuke Nakata, and Shinsuke Mori. Pointwise prediction for robust, adaptable japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011.