

# 言明間の意味的関係の 体系化とコーパス構築

村上浩司† 増田祥子‡ 松吉俊†  
乾健太郎† 松本裕治†

奈良先端科学技術大学院大学†  
大阪府立大学‡

# Web情報の信憑性評価

- 例) バナナダイエットでダイエットできる？
  - 科学的に根拠があるのかどうか知りたい
  - 検索結果の上位のページのほとんどは、方法やその効果、耳にした噂等の記載が殆ど



Google 検索結果上位は...

朝ごはんをバナナにするだけでやせる！  
そんな朝バナナダイエットが人気です

バナナには酵素以外にも、ダイエットの天敵である  
脂質や糖質の代謝を促してくれるビタミンB群が豊  
富に含まれているため、ダイエットに効果がある

バナナを朝食に食べるだけで体質が改善し、楽々  
痩せ体質に変わっていく方法を詳しくご紹介します

私の場合、6ヶ月でマイナス13キロでした。

# Web情報の信憑性評価

– 例) バナナダイエットでダイエットできる？

- 科学的に根拠があるのかどうか知りたい
- 検索結果の上位のページのほとんどは、方法やその効果、耳にした噂等の記載が殆ど



よくよく探してみると...

実際に減っているのは水分であり、体脂肪が減っている訳ではありません

確かにバナナは、低カロリーで、食物繊維も摂取できる食材なんだけど、逆に言えばそれだけ。決して「魔法のダイエット食」ではありません

『バナナに含まれる酵素・アミラーゼが体の代謝を良くする』これは明らかな間違いである

# Web情報の信憑性評価

朝ごはんをバナナにするだけでやせる！  
そんな朝バナナダイエットが人気です

バナナには酵素以外にも、ダイエットの天敵である  
脂質や糖質の代謝を促してくれるビタミンB群が豊  
富に含まれているため、ダイエットに効果がある

バナナを朝食に食べるだけで体質が改善し、楽々  
痩せ体質に変わっていく方法を詳しくご紹介します

私の場合、6ヶ月でマイナス13キロでした。

実際に減っているのは水分であり、体脂肪が減ってい  
る訳ではありません

確かにバナナは、低カロリーで、食物繊維も摂取できる  
食材なんだけど、逆に言えばそれだけ。決して「魔法の  
ダイエット食」ではありません

『バナナに含まれる酵素・アミラーゼが体の代謝を良く  
する』これは明らかな間違いである

良いこと、悪いこと、  
意見がバラバラ



俯瞰的に見たい！

# 言論マップ生成に必要な意味的關係とは？

## 言論マップ出力例

着目言明 **バナナダイエットでダイエットできる？**

**着目** [朝、バナナを食べるだけでダイエットできる]

- ・朝、バナナを食べるだけで、気軽に痩せられる…
- ・朝食にバナナを食べるだけで痩せられる「朝バナナダイエット」
- …

対立

**対立** [バナナを食べれば痩せるなんてことはない]

- ・バナナ食べて痩せるなんて、有り得ない訳です。
- ・朝バナナと水を飲んだだけで、痩せるわけがありません！
- …

**根拠** [バナナの持つ酵素の働きはダイエット効果がある]

- ・バナナに含まれる酵素を使ってダイエットできる
- ・バナナの持つ酵素の働きにより、ダイエット効果がある
- ・バナナは糖質分解酵素が脂肪燃焼を活発にする
- …

対立

**根拠** [バナナの酵素が代謝を高めることはない]

- ・果物の中に含まれている…酵素が、体の中で活性を持つ、なんて話はない。
- ・バナナでは、含まれる酵素が代謝を高めるなんてことは絶対にありえない。
- …

**根拠** [バナナ食物繊維で便秘解消してダイエットできる]

- ・食物繊維も豊富だからお通じもよくなりダイエットにも…
- ・繊維も多く「便秘」にも効果があり…
- …

**根拠** [私はバナナダイエットで1カ月に3キロ痩せました]

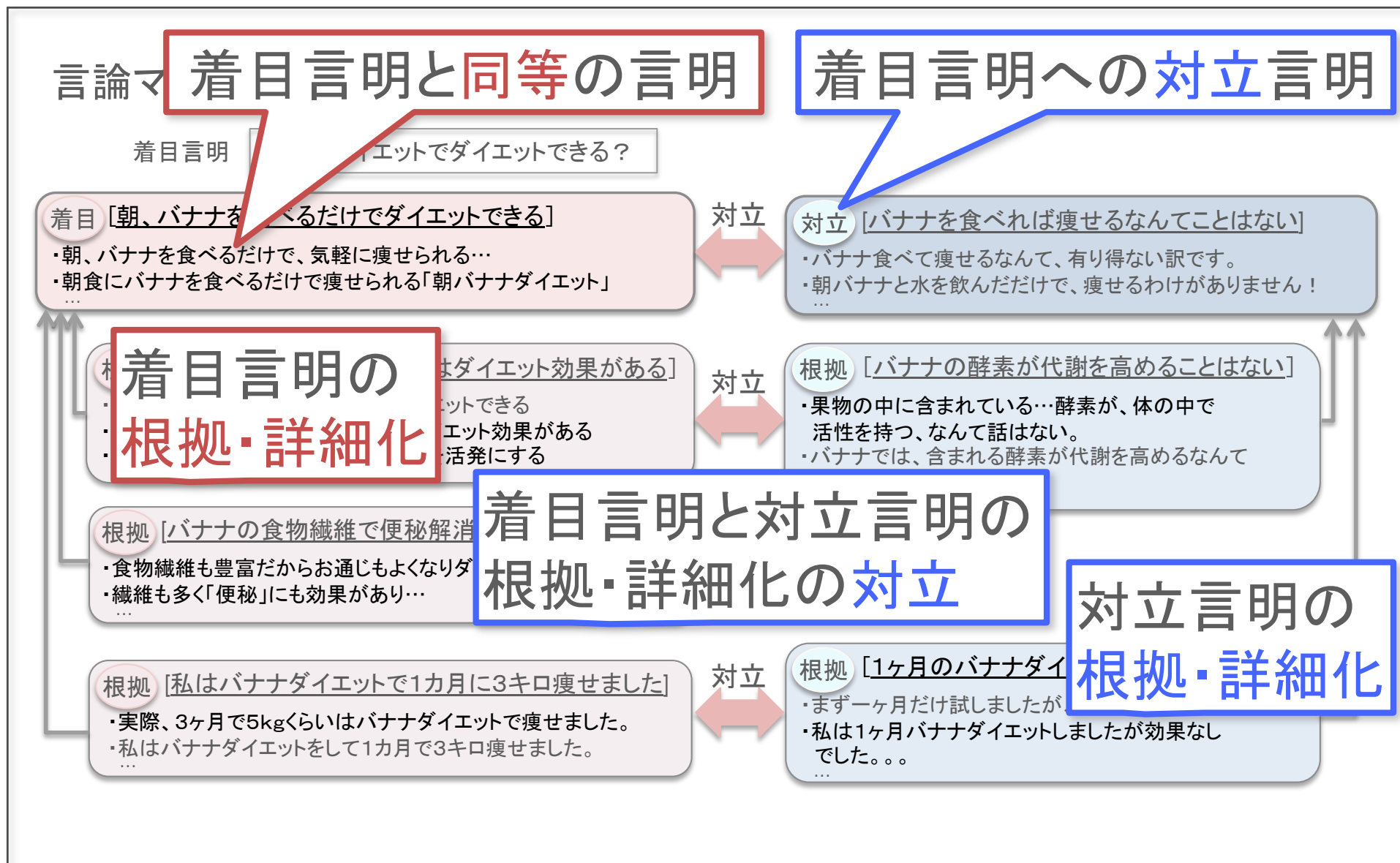
- ・実際、3ヶ月で5kgくらいはバナナダイエットで痩せました。
- ・私はバナナダイエットをして1カ月に3キロ痩せました。
- …

対立

**根拠** [1ヶ月のバナナダイエットで効果なしでした]

- ・まず一ヶ月だけ試しましたが、やせませんでした。
- ・私は1ヶ月バナナダイエットしましたが効果なしでした。。。
- …

# 言論マップ生成に必要な意味的關係とは？



# 言論マップ生成に必要な意味的關係とは？

## ・言論マップ

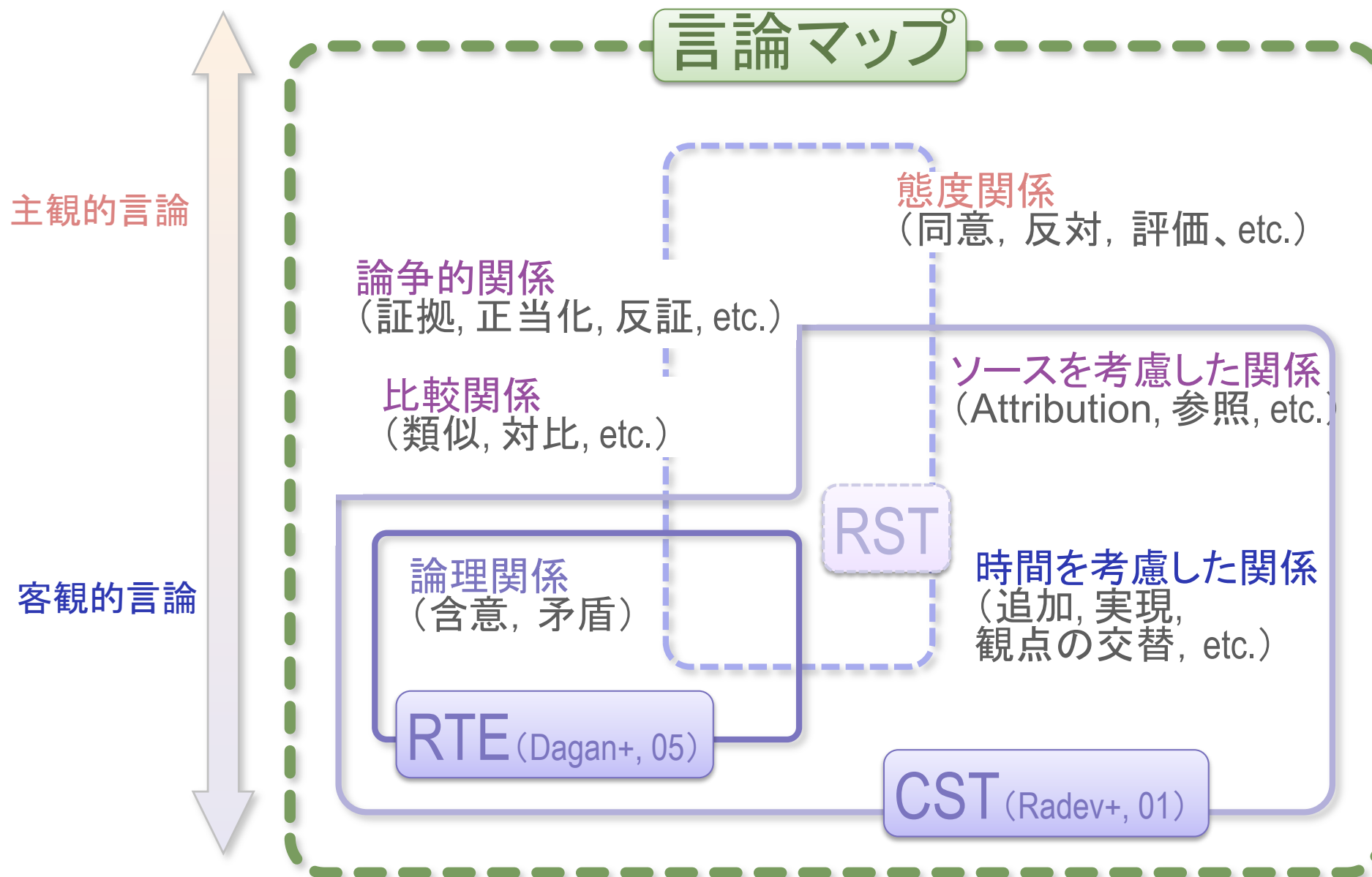
- ユーザが自分の着目する言明の信憑性を判断するための支援を行う
- ユーザに提示する言明間の関係は  
〈同等〉、〈対立〉、〈根拠、詳細化〉

## ・関係認識のための訓練、評価

→ コーパスが必要

・現状、関係認識コーパスがない → 新規構築

# 言明間関係の対象領域





# 事例収集の戦略

どんな関係の事例をどうやって収集するのか？

言論マップ生成に必要な関係

- 同等 : 2事例の対に付与できる関係
- 対立 : 2事例の対に付与できる関係
- 根拠 : 1事例中に存在する関係
- ソースを考慮 : 1事例中に手がかり  
(Attribution)

# 事例収集の戦略

どんな関係の事例をどうやって収集するのか？

言論マップ生成に必要な関係

- ・同等 : 2事例の対に付与できる関係  
事例対を作成して関係付与
- ・対立 : 2事例の対に付与できる関係  
事例対を作成して関係付与
- ・根拠 : 1事例中に存在する関係  
修辞構造解析により関係を付与
- ・ソースを考慮 : 1事例中に手がかり  
(Attribution) 事例に対してアノテーション

# 事例収集の戦略

どんな関係の事例をどうやって収集するのか？

言論マップ生成に必要な関係

- ・**同等** : 2事例の対に付与できる関係  
事例対を作成して関係付与
- ・**対立** : 2事例の対に付与できる関係  
事例対を作成して関係付与
- ・**根拠** : 1事例中に存在する関係  
修辞構造解析により関係を付与
- ・**ソースを考慮** : 1事例中に手がかり  
(Attribution) 事例に対してアノテーション

# 事例収集の戦略

本研究で扱う関係と扱わない関係

関係	対応	事例収集
ソースを考慮した関係	事実性解析用のコーパスにて行う	事例収集を行わない
根拠関係	RSTなどにより関係を特定し、コーパスを作成	事例収集を行わない
同等、対立などの関係	2事例間の関係を付与するコーパスを作成	事例収集を行う

# 関係認識：関連研究

1. RTE (Recognizing Textual Entailment Challenge)
  - ・テキスト(t)と仮説(h)間の関係を認識 (Dagan+ 2005)
  - ・対象：〈含意〉、〈矛盾〉、〈関係無し〉の3値分類
2. CST (Cross-document Structure Theory)(Radev+ 2001)
  - ・文書内の修辞構造解析を文書横断構造解析に拡張
  - ・あるイベントに関連する複数新聞社の記事群を対象
  - ・記事中の2文間の関係を認識
  - ・対象：〈同等〉、〈詳細〉など18種類の関係を定義
3. 文書横断文間関係コーパス (衛藤+ 2005)
  - ・CSTを元に日本語に適応したコーパス構築
  - ・14種類の関係を再定義

・RTE,CST共に対象は事実関係、意見などは扱わない

# コーパスで扱う意味的關係

關係カテゴリー	意味的關係	対応	
論理關係	同義	AとBが同時に成り立つ	A:死刑は犯罪を抑止する効果がある B:死刑制度には犯罪抑止力がある
	矛盾	AとBが同時に成り立たない	A:水産庁は、調査捕鯨を中止すると発表した B:日本は調査捕鯨を強行すると発表した
	含意	Aが成り立つときBが成り立つ	A:バラクオバマはアメリカ合衆国の大統領である B:バラクオバマは1月に大統領に就任した
態度關係	意見一致	異なる判断主体が同義、もしくは含意の言明を表す	A:私は捕鯨が正しいとは思えない B:俺の彼女は捕鯨に反対している
	意見対立	異なる判断主体が矛盾する言明を表す	A:私はマイナスイオンに効果があるとは思えない B:両親はマイナスイオンは体に良いと考えている
	評価極性一致	ある側面に対する何らかの評価をA,Bが行い、それらの極性が一致	A:私はiPodは持ち運びに便利だと思う B:自分ではiPodの値段は安いと思う
	評価極性対立	ある側面に対する何らかの評価をA,Bが行い、それらの極性が対立	A:私はiPodは持ち運びに便利だと思う B:僕はiPodは使い勝手が悪いと感じる
	評価(P/N/O)	Aに対しBが(P)ポジティブ、(N)ネガティブ、(O)その他の評価、を行う	A:日本では少子化が進行している B:少子化は非常に問題であると思う(N評価)
負例	負例	AとBの間に關係がない、上記の意味的關係に当てはまらない	A:遺族の気持ちを無視した死刑制度廃止には反対だ B:死刑制度廃止を目指す弁護士の活動は怪しい

# 言明間意味的関係コーパス構築(1)

Web文書中の文

- 複数の言明から構成されることが多い
- 2文間に関係を付与することは困難

(例)

実文A:

イソフラボンを摂取すると  
女性ホルモンと同じ成分があり  
ガン予防に  
効果がある。

実文B:

イソフラボンを食べれば  
体内の代謝に作用するため  
ガンの予防に  
効く。

AとBの間の意味的関係は??

# 言明間意味的関係コーパス構築(1)

Web文書中の文

- 複数の言明から構成されることが多い
- 2文間に関係を付与することは困難

(例)

実文A:

イソフラボンを摂取すると  
女性ホルモンと同じ成分があり  
ガン予防に  
効果がある。

A,Bに別情報

実文B:

イソフラボンを食べれば  
体内の代謝に作用するため  
ガンの予防に  
効く。

AとBの間の意味的関係を付与するのは困難  
なので。。



# 言明間意味的関係コーパス構築(1)

Web文書中の文

- 複数の言明から構成されることが多い
- 2文間に関係を付与することは困難

(例)

実文A:

イソフラボンを摂取すると  
女性ホルモンと同じ成分があり  
ガン予防に  
効果がある。

削除



実文B:

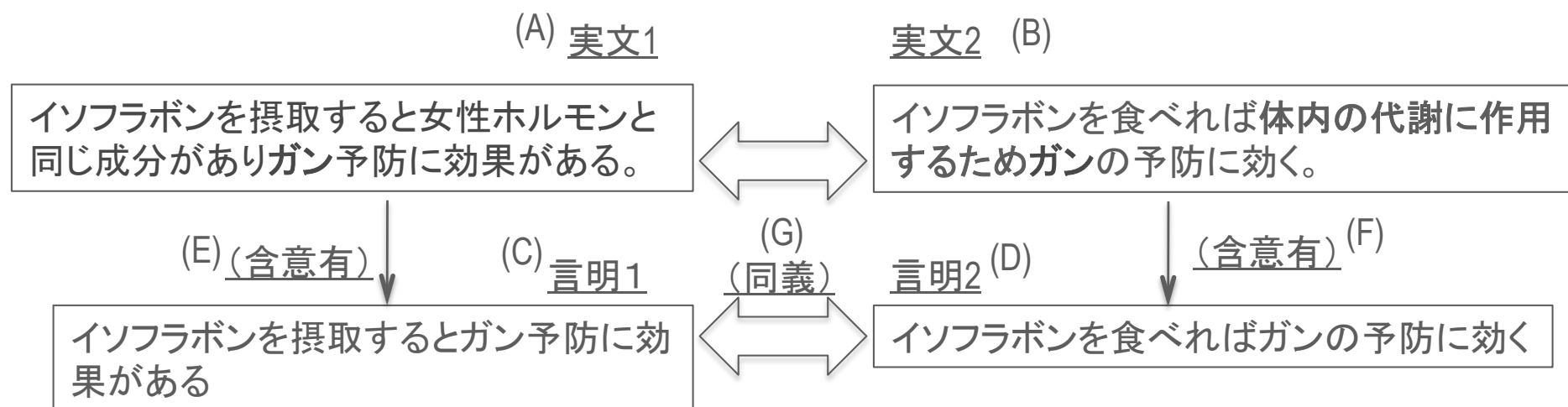
イソフラボンを食べれば  
体内の代謝に作用するため  
ガンの予防に  
効く。

AとBの間の意味的関係 : 同義

実文から適当な長さで取り出した言明間に関係を付与する

# 言明間意味的關係コーパス構築(2)

## 1エントリを構成する情報



実文 (A,B) : Web文書中の文

言明 (C,D) : 実文から作成した言明

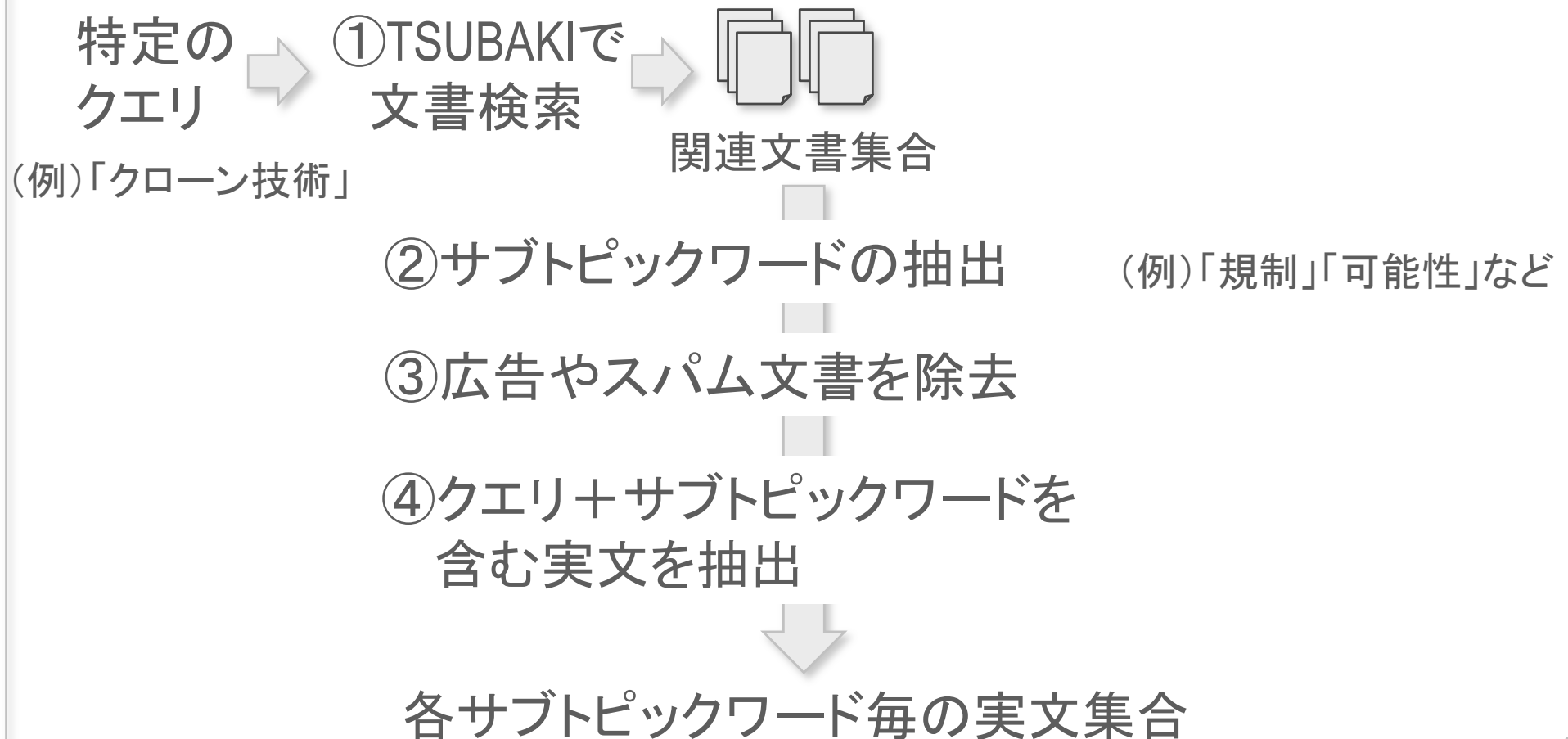
含意フラグ (E,F) : 実文と言明との間の含意関係

言明間関係 (G) : 言明 (C) と (D) との意味的關係

# コーパス構築の手順(1)

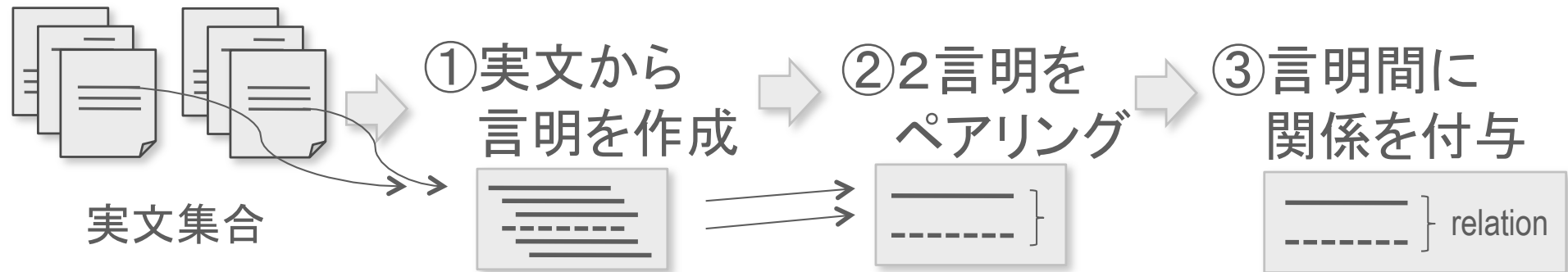
- A. Web上の文書群から関連する実文を取得
- B. 言明ペアを作成し、意味的關係を付与

## A(実文取得)の手順



# コーパス構築の手順(2-1)

## B(言明ペアへ意味的關係付与)の手順



### 課題:

- ・ペア作成の作業量は、最大で  $O(n^2)$   
→ 人手による、すべてのペアへの作業は不可能
- ・実文から”意味のある”言明を作成
- ・負例を作らないようにペアを作成

効率よく“關係付与できそうな”言明ペアを作成するには？

## コーパス構築の手順(2-2)

### 言明ペア作成方法

- A. 一定数の言明を実文から作成し、言明をペアにする
- B. 関係が付与できそうな実文をペアにし、それぞれ言明を作成

#### 考慮点

- ・言明作成の人的コストを抑える
- ・関係の付かない言明を少なく抑えて言明ペアを作成
- ・機械でできることは機械をつかう  
→ Bを採用

## コーパス構築の手順(2-3)

### 関係が付与できそうな実文ペアの作成

- ・正例となり得る言明ペアを含む実文ペア
  - 同一単語、言い換え可能な語、表現を含む
- 類義表現は知識により吸収
  - ・名詞の上位下位関係(約10万)、類義語の辞書(約180万)
  - ・動詞項構造辞書(約6万エントリ)
- 文字Unigram(平仮名、漢字)、単名詞、  
複合名詞、形容詞、形容動詞を素性
- 実文間類似度をコサイン距離を用いて計算、候補作成

A. インディカーの燃料はおなじみエタノールだが、今年からバイオエタノールになって、より環境に配慮されている

B. 燃料には、バイオエタノールとして、環境にも配慮し、ターボはブレーキや廃熱を利用したものになるようだ

A. 代替燃料バイオエタノールによる地球環境対策が世界の流れとなり、政府も増奨に本腰を入れ始めた

B. バイオエタノールは高い可能性を秘めた高品質の燃料であり、現在、我々が直面する環境問題に対処し得る潜在能力を持っている

# コーパス構築の現状

- ・2/26現在  
（正例:368、負例:312）
- ・3月末を目処に3,000エントリのコーパス構築
- ・コーパスの公開へむけて準備中  
<http://cl.naist.jp/stmap/corpus>

## コーパス中のエントリ例

ただ、バイオエタノールの混合方式（日本では上限3%）を巡って、ETBE方式を主張する石油業界と直接混合方式を主張する環境省が対立し、普及の障害となる可能性がある

↓ 含意あり

バイオエタノールの混合方式（日本では上限3%）を巡って、石油業界と環境省が対立している

環境省は将来的にバイオエタノールの直接混合型への変換を主張するが、一方の石油業界はETBE化して混入する方式を訴え、現在対立が続いている

↓ 含意あり

[バイオエタノールの混合方式に関して]、環境省と石油業界は現在対立が続いている

同義

# まとめ

- ・ 言論マップ生成課題で取り扱う言明
- ・ 言明間意味的関係の体系化と事例収集の戦略
  - 言論マップ生成に必要な言明間の関係
  - コーパス構築の対象を態度関係、論理関係に限定
- ・ 体系を元にしたコーパス構築
  - コーパスを構成する情報
  - 言明ペアの作成手順



# 今後の課題

## 1. 条件節等の取り扱い(限定関係の定義)

A: キシリトールは虫歯予防に役立ちます。

B: キシリトールは、基本的には規則正しい食生活(間食を含む)ときちんとしたブラッシング、フッ素の使用、プラスとしてキシリトールを毎食後とることで虫歯の予防に役立つのです。

A: キシリトール入りのチューインガムなども虫歯予防に役立ちます。

B: ガムにキシリトール以外の甘味料が含まれている場合、それが砂糖や水飴といった虫歯の原因になる甘味料だと、キシリトールも虫歯予防にはなりません。

## 2. 修辞構造解析を用いた根拠関係コーパスの開発

## 3. コーパスの規模の拡大(3,000 → 6,000)