

# 自然言語処理による 情報編集

乾健太郎

東北大学 情報科学研究科

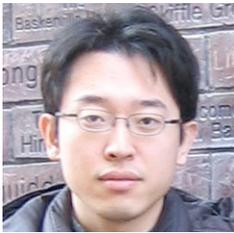
# 東北大学 自然言語処理研究室

Since 2010

## 研究スタッフ



乾健太郎  
教授



岡崎直観  
准教授



井之上直也  
助教



松林優一郎  
特任助教



田 然  
特任助教



折田奈甫  
特任助教



水本智哉  
特任助教



水野淳太  
NICT研究員



### • メンバー 42名

- 教員 7名、研究員 3名、事務 1名
- プログラマ 2名、言語分析者 2名
- 博士課程 9名、修士課程 12名

### • メンバーの受賞 37件 (2010～)

- 国際会議の賞 3件
- 言語処理学会年次大会賞 6件
- 言語処理学会論文賞 3件
- 情報処理学会山下賞 2件
- 国内若手奨励賞 12件、他



高良倉吉  
○  
**琉球の時代**  
大いなる歴史像を求めて

---

ちくまぶっくす 1980／ひるぎ社 1981／ちくま学芸文庫 2012  
ISBN:4480094431

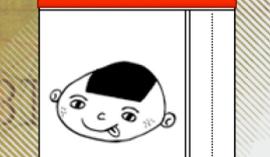
---

編集：加藤雄弥・富川益郎・平野洋子 装幀：(ちくま) 神田昇和

---

真夏がくると、今年も熱いオキナワが滲みってくる。

ISIS EditTour エディットツアー  
6万冊の本の森で  
「編集」を体験  
8.17 開催 | 詳細 →  
ISIS 編集学校



<http://1000ya.isis.ne.jp/>

# 人の知の活動はすべて編集である

## 64編集技法（松岡正剛，知の編集工学，2001）

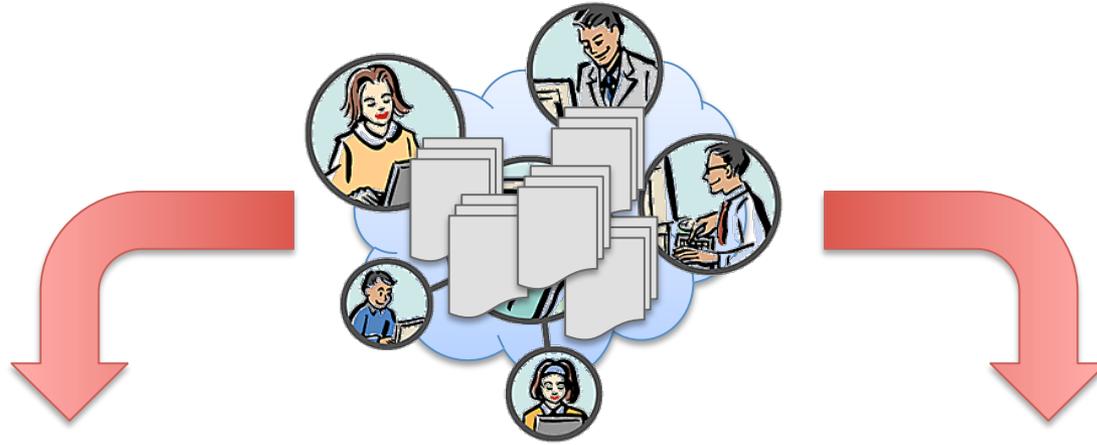
収集，選択，分類，流派，系統，編定，要約，凝縮，  
原型，模型，列举，順番，規則，配置，交換，比較，  
適合，競合，共鳴，結合，比較，推理，境界，地図，  
図解，注釈，引用，例示，暗示，相似，擬態，象徴，  
輪郭，強調，変容，歪曲，不調，諧謔，意匠，装飾，  
模擬，補加，削除，保留，代行，速度，構造，形態，  
生態，焦点，報道，統御，筋道，脚本，場面，劇化，  
遊戯，競技，翻訳，通訳，周期，曲節，総合，創造



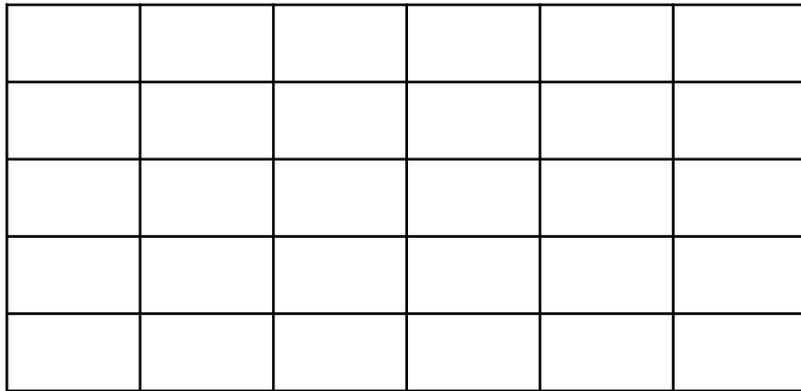




# 情報の編集＝組織化

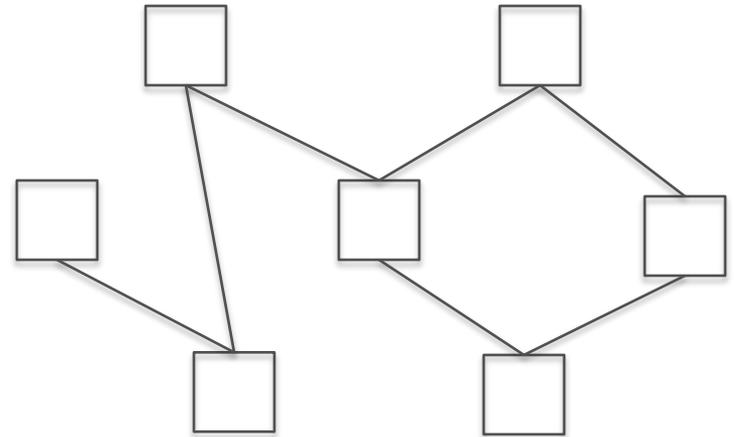


## 分類による組織化



属性

## 関係による組織化



同意, 対立, 根拠

# 分類による組織化

賛否の意見から世の中の関心を掘り起こす

(渡邊ら, 2013)

靖国神社への首相参拝  
に中国が反対している  
ことに配慮した。



地元民は基地に反対し  
ていない



誰が	何に	賛否
中国	首相参拝	反対
地元民	基地に	— (賛成?)

# 世の中の関心を自動的に掘り起こす

朝日新聞社とライセンス契約(2013)

Twitter 賛成・反対意見ランキング

all

ツイート数	意見	キーワード1	キーワード2	キーワード3
10139	原発	住民	福島	東電
9859	TPP	安倍	自民党	左翼
9025	児童ポルノ禁止法改正案	声明	協会	日本
4917	憲法改正	調査	世論	前回
4093	Jリーグの2ステージ制	浦和	横断幕	サポ
3515	憲法96条改正	古賀	国会	議員
3362	消費税の増税	原発	TPP	民主党
2617	原発廃炉	演説	福島	首相
2487	戦争	同じ	日本	命
2368	財源としての国債増発	インフラ	公共	化
2323	東京五輪招致	署名	ネット	分
1986	アベノミクス	浜	矩子	理由
1977	安倍政権	投票	民主党	反日
1765	外国人参政権	投票	鳩山	民主
1549	銃規制	アメリカ	理由	TPP



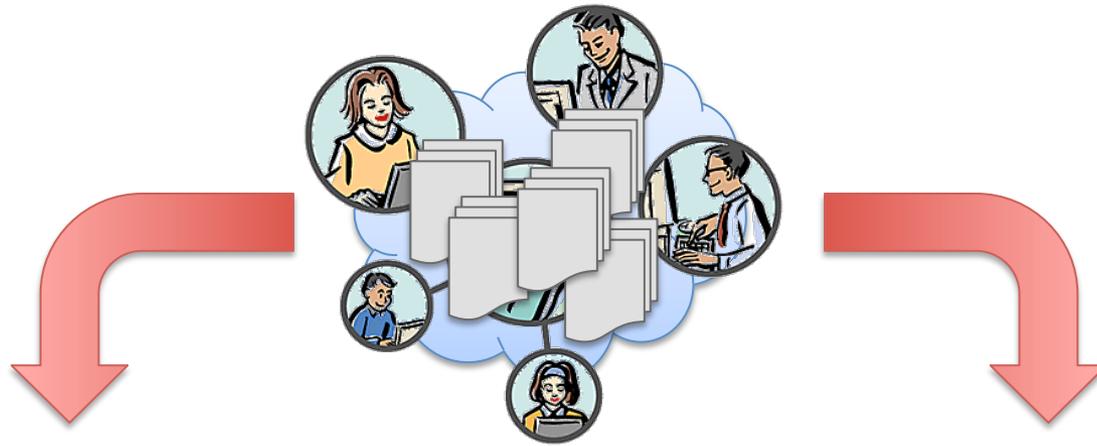
# 「報道に耐えうる」とのギャップ

多くの人が 東京五輪に 反対しているけど、  
意見者                      対象                      反対

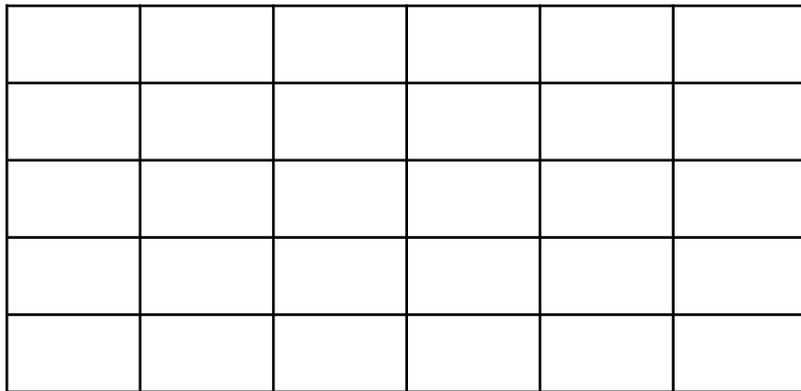
俺はやったらよいと思うよ

**省略、婉曲、皮肉、、、**  
**一方で、手法の透明性も**

# 情報の編集＝組織化

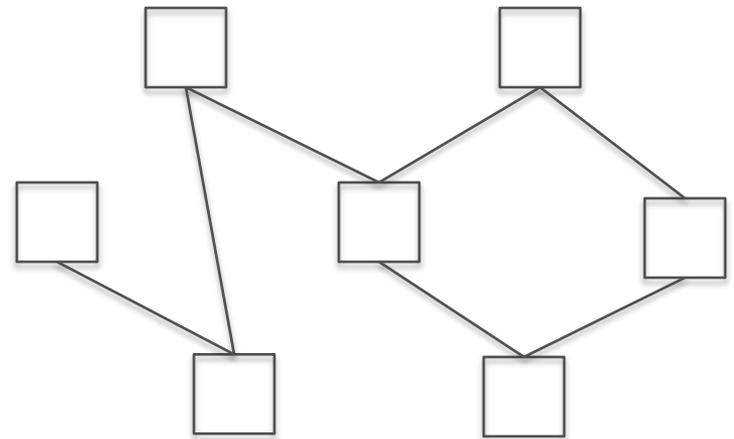


## 分類による組織化



属性

## 関係による組織化



同意, 対立, 根拠

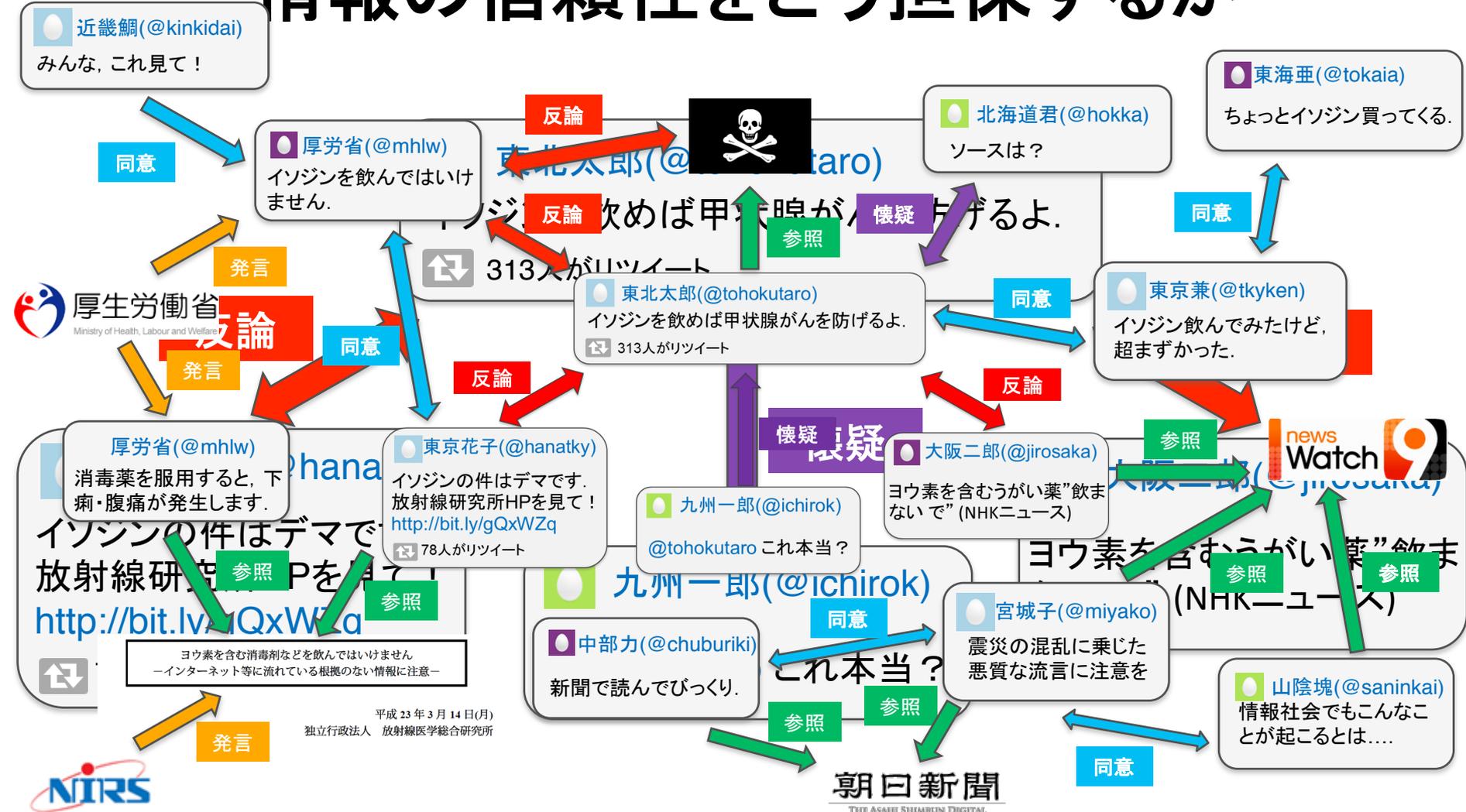
# 関係による組織化

## 情報の信頼性をどう担保するか



# 関係による組織化

## 情報の信頼性をどう担保するか



# 言論マップ



質問

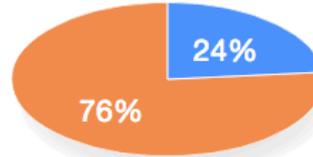
イソジンで被曝を防げる？

肯定

否定

39

★根拠 21



12

★根拠 1

肯定

被曝を軽減するにはイソジンがいい

否定

イソジンは被曝対策には使えません

# 平時の一般のトピックでも

コラーゲンは肌に良い

質問：

**同意** (459文) フェミダンプラスはコラーゲンとヒアルロン酸の働きで、  
噛むだけでお肌つやつや！若返りと美肌を提供します log1 log2

- コラーゲンは健康な肌ならきれいな網目構造になっており、まるでベッドのスプリングのように表皮を支えて弾力のある肌を保っています log1 log2

フェミダンプラスはコラーゲンとヒアルロン酸の働きで、  
噛むだけでお肌つやつや！若返りと美肌を提供します log1

**対立** (76文) ただ、研究職の私から言わせてもらおうと、コラーゲンは体  
に入ったからといって、肌に潤いをもたらすわけではありません log1 log2

- お肌にハリがなくなってきたり、たるんできているのかな、と思うときにコラーゲンやヒ log1 log2

more

ゲンですが、肌か  
たしわに対しての効

が良く、他の哺乳類・魚類とは違って、人間の肌に対して類をみないなじみがあるので  
す log1 log2 more

- ただしコラーゲンは肌以外にも骨や血管・内臓にも使われるので不規則な生活で  
体内がぼろぼろだと先に体内の改善に使われるので不規則な生活だといくら一生懸

**根拠** (20文)

- 生まれたばかり  
みたいに肌が柔
- また、アミノ酸合  
とされています
- コラーゲンはお  
ら補給するのが効

ニッピコラーゲンといったコラーゲンですが、肌か  
ら浸透しないので、出来てしまったしわに対しての効  
果は期待できません log1 log2

う事なので一緒にビタミンCやミネラルをとりましょう log1 log2 more

**根拠** (10文) 美肌づくりに欠かせないコラーゲンですが、実はコラ  
ーゲンは分子が大きいので、お肌に入りにくい！と  
いう難点があります log1 log2

# 同意・対立・根拠の認識

事実性(モダリティ)解析 + 構造的アライメント + 関係パターン認識

*Collagen has beautiful skin effects.*  
コラーゲンには美肌効果がある

+ 存在  
+ existence

- 存在  
- existence

コラーゲンをサプリメントで飲んでも、  
*Even when taken as a supplement,*

途中でアミノ酸に分解されてしまうので、  
*(collagen) ends up broken down into amino acids so,*

(Φが)肌をきれいにする効果は期待できません。  
*we cannot expect (collagen) to be effective at making skin pretty.*

# 否定文の解析

そのイベントは起こったのか起こっていないのか

流出しなかった（流出しなかった）

流出したではないか（流出した）

流出しただけでない（流出した）

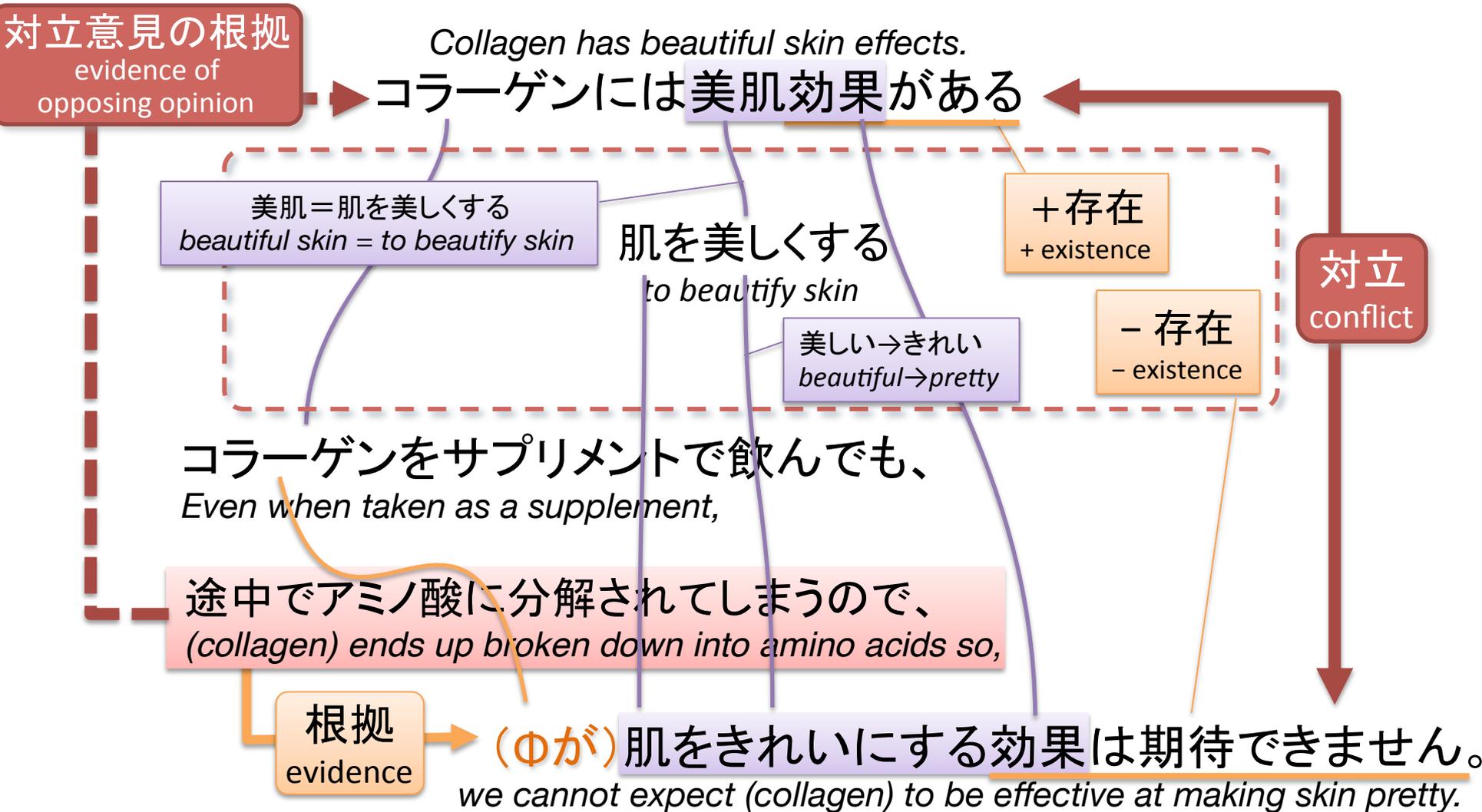
流出を防いだ（流出しなかった）

流出を抑えた（流出した，でも少し）

流出を止めたい（流出した，今も流出）

# 同意・対立・根拠の認識

事実性(モダリティ)解析 + 構造的アライメント + 関係パターン認識

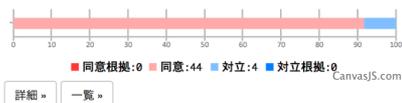


# 言論マップで利用している語彙知識の例

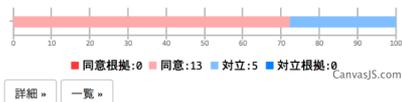
- 日本語WordNet (Bond+ 2009) : 56,741 synsets, 92,241 語
  - 過剰: 冗多, 過多, 冗長, 余った, 余計, 余分
  - 効果: 効目, 実効, 効, 利き目, 効き目, 利き, 効力, 効き, 機能, 効能
  - 病気 ⇔ 具合い, 丈夫, 健康, フィットネス, 健全, 多幸
- 事象間関係知識 (Matsuyoshi+ 2008) : 約140万対
  - <誰か>が<何か>を摂取する – <誰か>が<何か>を取る (上位語)
  - <何か>が普通だ – <何か>が異常だ (反義)
- Wikipediaから抽出した同義語: 113,401 synsets
  - HIDランプ–高圧放電灯–高輝度放電ランプ
  - バイオ燃料–エコ燃料–バイオエネルギー
- Wikipediaから抽出した上位下位関係 (Sumida+ 2008) : 約350万対
  - イソフラボン–女性ホルモン様成分
  - キシリトール–甘味添加物
- 文脈類似度に基づく動詞間含意関係 (ALAGIN) : 約170万対
  - 引き起こす–もたらす, 予防する–防げる, ケアする–防止する

# あらゆる命題で

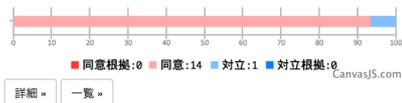
ヨーグルトが花粉症に効く



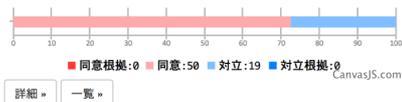
トマトが花粉症に効く



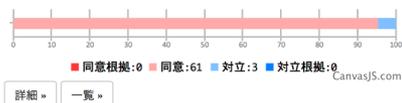
お灸で逆子は治る



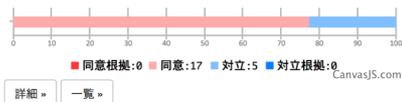
食物繊維が便秘を予防する



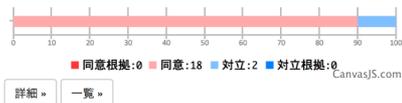
ヨーグルトがダイエットに効く



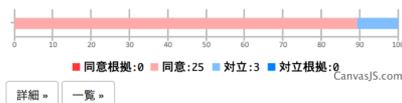
乳酸菌で便秘を解消する



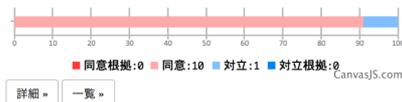
食事療法でがんは治る



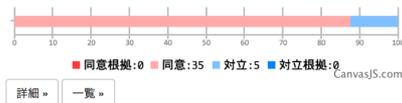
腰痛体操で腰痛を予防する



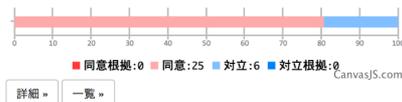
マイナスイオンが疲労回復に効く



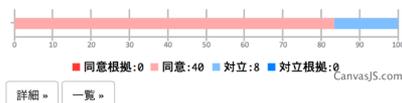
キシリトールが虫歯を予防する



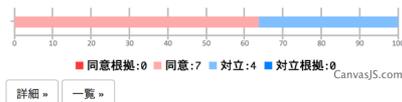
脂肪吸引でセルライトを解消する



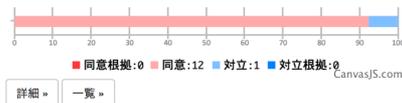
食物繊維が大腸がんを予防する



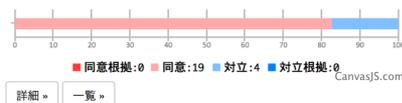
タマネギが糖尿病に効く



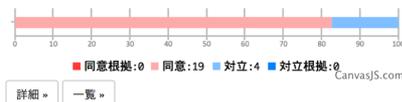
漢方で口内炎を治す



システインがそばかすに効く



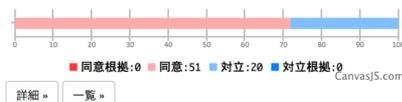
漢方で子宮筋腫は治る



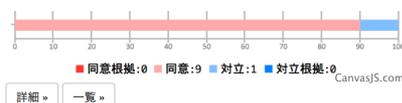
カテキンがダイエットに効く



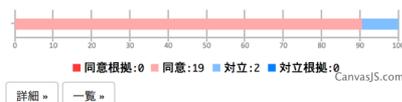
ダイエットでリバウンドを防ぐ



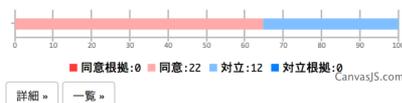
マッサージでセルライトを解消する



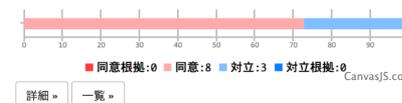
乳酸菌で腸内環境を改善する



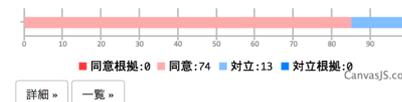
タミフルでインフルエンザは治る



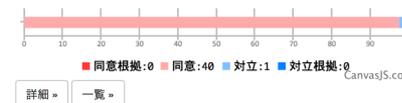
糖質制限食で血糖は改善する



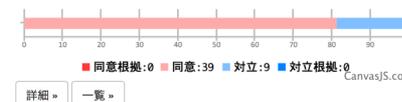
コーヒーが便秘に効く



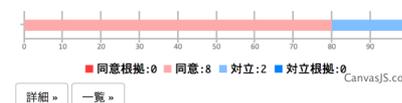
活性水素が活性酸素を抑える



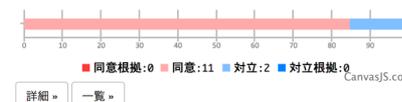
化粧水がニキビに効く



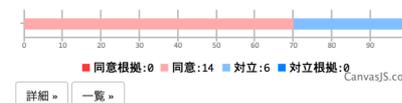
正しいブラッシングで歯周病を防ぐ



うがいで歯周病は治る



ダイエットで脂肪肝を克服する



# 含意・矛盾を超える関係

「うがいは風邪を予防する」

意味論的含意・矛盾

語用論的効果による同意・対立

含意

「うがいをすれば風邪を引かない」

同義

「うがいは風邪を防ぐ」

矛盾

「うがいでは風邪予防に効果がない」

条件付き同意・対立

「外出後に必ずうがいをすれば風邪予防の効果は見込める」

例示(同意)

「うがいのおかげで風邪を引いたことがない」

逆向き含意

「うがいは風邪に良い」

裏

「うがいをしないと風邪を引きやすい」

比較

「手洗いよりうがいの方が風邪予防に効果的だ」

程度制限

「うがいをすると少しは風邪予防になる」

例示(対立)

「うがいをしていたのに風邪を引いてしまった」

# 「うがいは風邪を予防する」

## 意味論的含意・矛盾

### 含意

「うがいをすれば風邪を引かない」

### 同義

「うがいは風邪を防ぐ」

### 矛盾

「うがいでは風邪予防に効果がない」

13%

### 矛盾(著者)+含意(非著者)

「うがいでは風邪を予防できると聞いたが、違うと思う」

9%

## 語用論的効果による同意・対立

### 条件付き同意・対立

「外出後に必ずうがいをすれば風邪予防の効果は見込める」

### 例示

「うがいのおかげで風邪を引いたことがない」

### 逆向き含意

「うがいは風邪に良い」

### 裏

「うがいをしないと風邪を引いてしまう」

### 比較

「手洗いよりうがいの方が風邪予防に効果的だ」

### 程度制限

「うがいをすると少しは風邪予防になる」

### 対立例示

「うがいをしていたのに風邪を引いてしまった」

68%

### 同意例示(著者)+矛盾(非著者)

「うがいの風邪予防効果は疑問視されているが、少なくとも私には効果があった」

10%

# Time to collaborate

分類と関係に基づく情報編集

言語処理技術は着実に進歩

「報道に耐えうる」とのギャップ

今の技術と3年5年後の技術

相談してください。具体例で。