

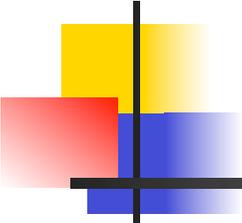
事象間関係知識の整備と 類似・対立認識への応用

†¹奈良先端科学技術大学院大学、†²大阪府立大学

†³独立行政法人 情報通信研究機構

松吉 俊†¹、村上 浩司†¹、増田 祥子†^{1,†2}、

松本 裕治†¹、乾 健太郎†³

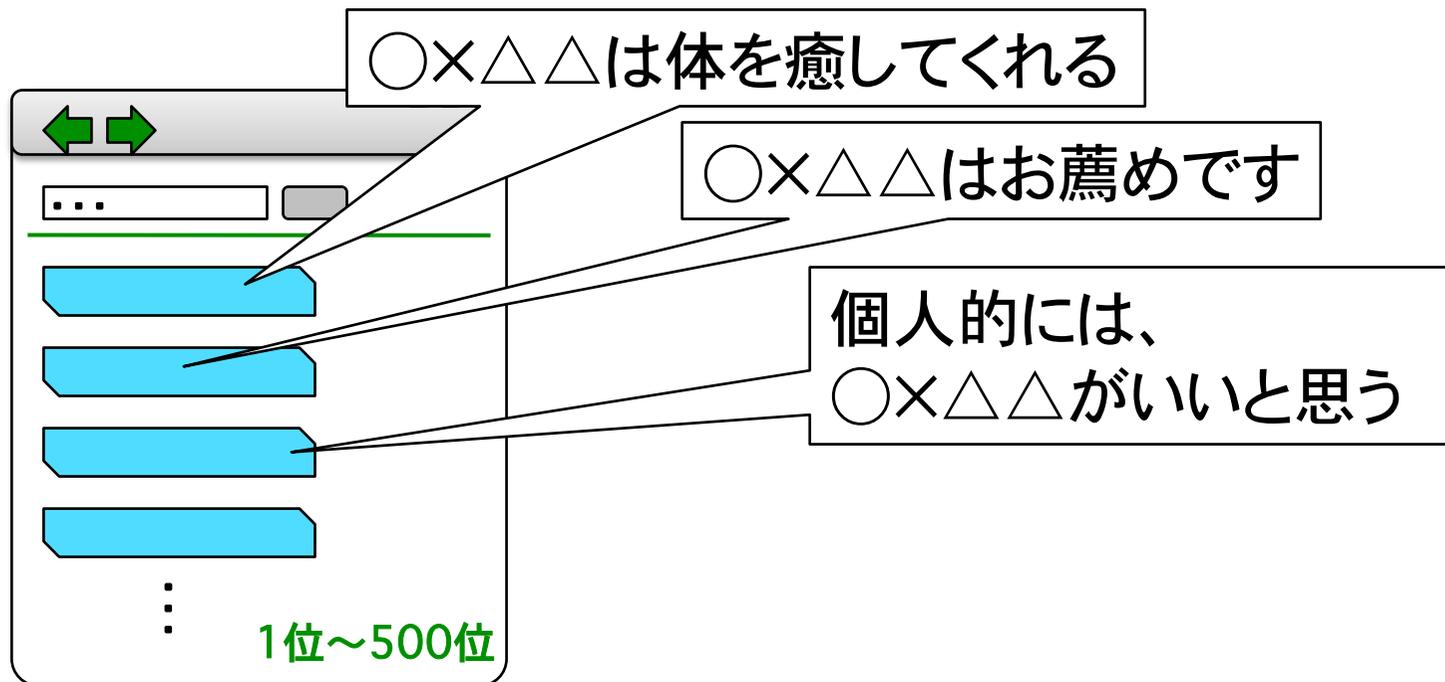


発表概要

1. 背景と目的
2. 言論マップと言論マップ生成システム
3. 事象間関係知識の整備
4. 整備した関係知識の小規模な評価実験
5. まとめ

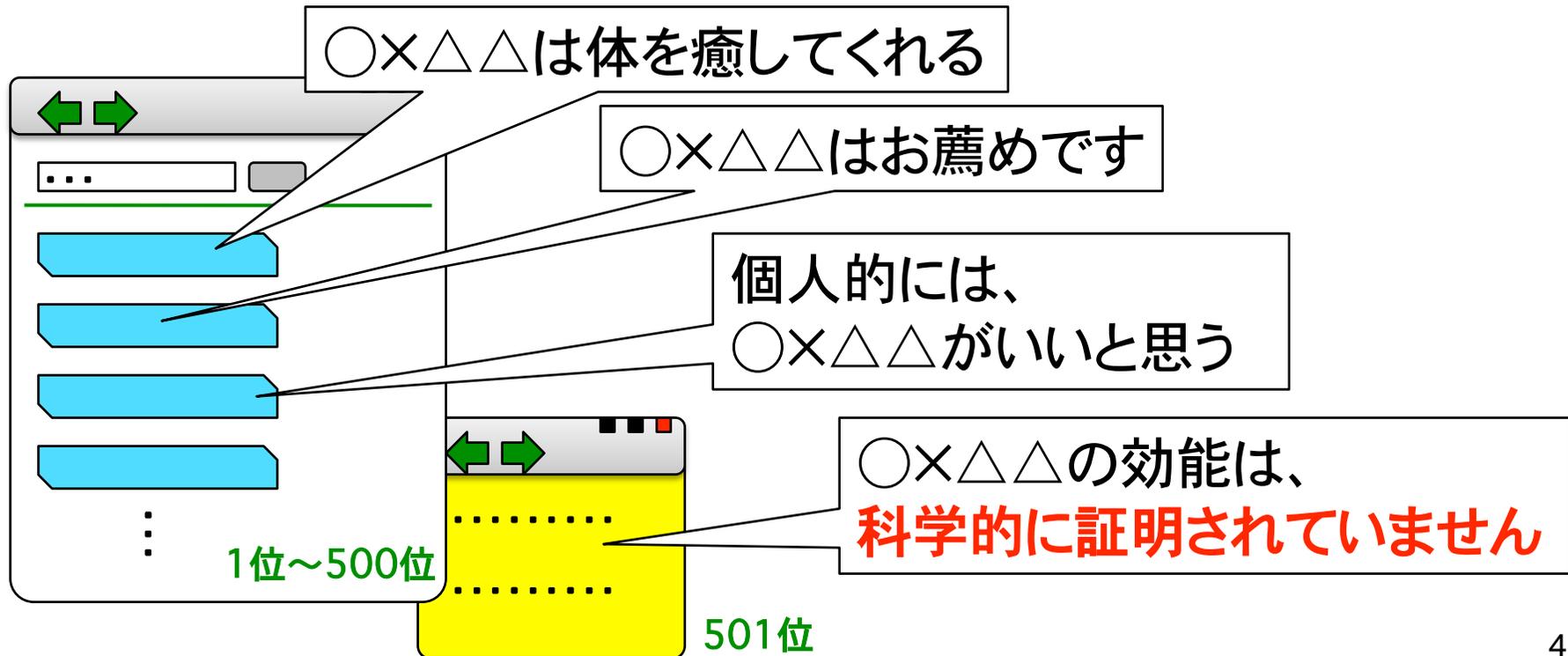
背景(1/3)

- ウェブ上には大量のテキスト情報が存在し、そこでは様々なトピックに関して多角的な意見が述べられている



背景(2/3)

- ウェブ上には大量のテキスト情報が存在し、そこでは様々なトピックに関して多角的な意見が述べられている



背景(3/3)

- ウェブ上には大量のテキスト情報が存在し、

そのすべてが真実であるとは限らない
(根拠が全く記述されていないことがある)

○×△△は体を癒してくれる

○×△△はお薦めです

対立する意見へのリンクが
必ず存在するとは限らない
(隠れた関係に気づけない
可能性がある)

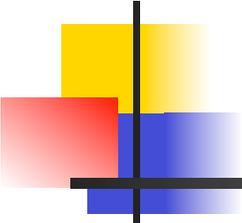
個人的には、
○×△△がいいと思う

○×△△の効能は、
科学的に証明されていません

隠れた対立

⋮
1位~500位

501位



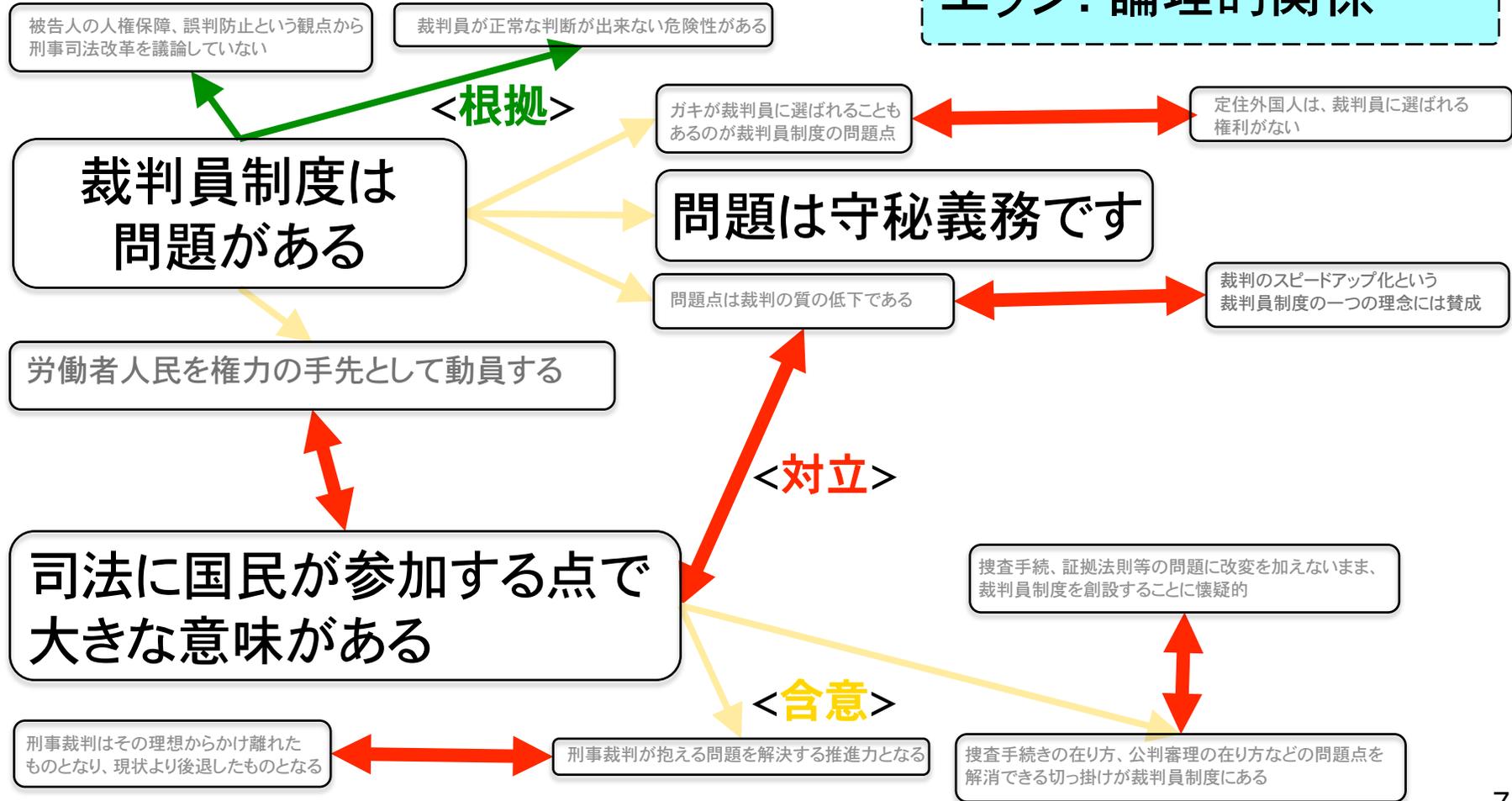
目的

- テキスト情報間に存在する論理的関係を解析し、**言論集合の俯瞰図**を提供する
 - 各言論の信憑性を判断する作業を支援し、情報の偏りや思い込みによる誤信の可能性を抑えることを目指す

言論マップ

■ 例) トピック「裁判員制度」

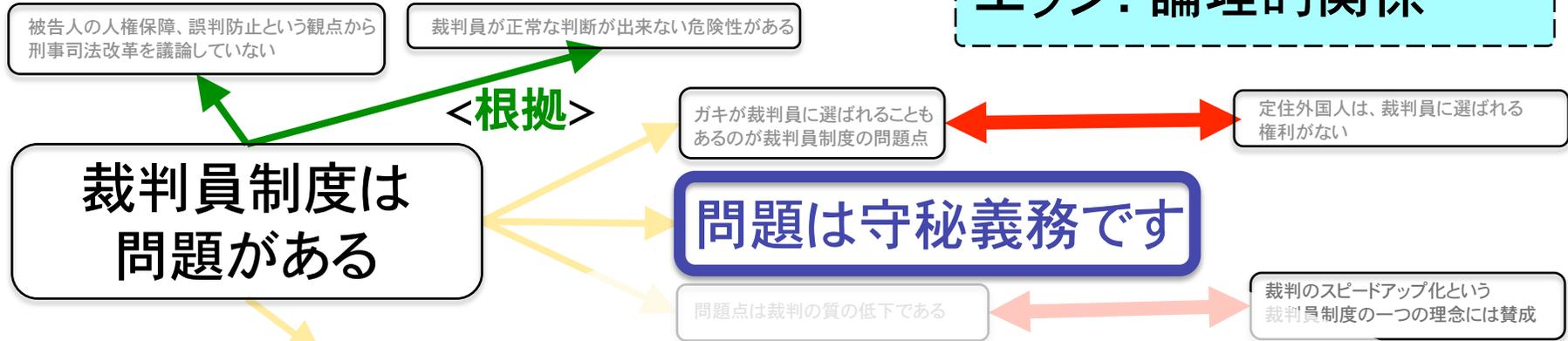
ノード: 類義言論の集合
エッジ: 論理的関係



言論マップ

■ 例) トピック「裁判員制度」

ノード: 類義言論の集合
エッジ: 論理的関係



労働者人民を権力の

司法に国民
大きな意味

刑事裁判はその理想からかけ離れたものとなり、現状より後退したものの

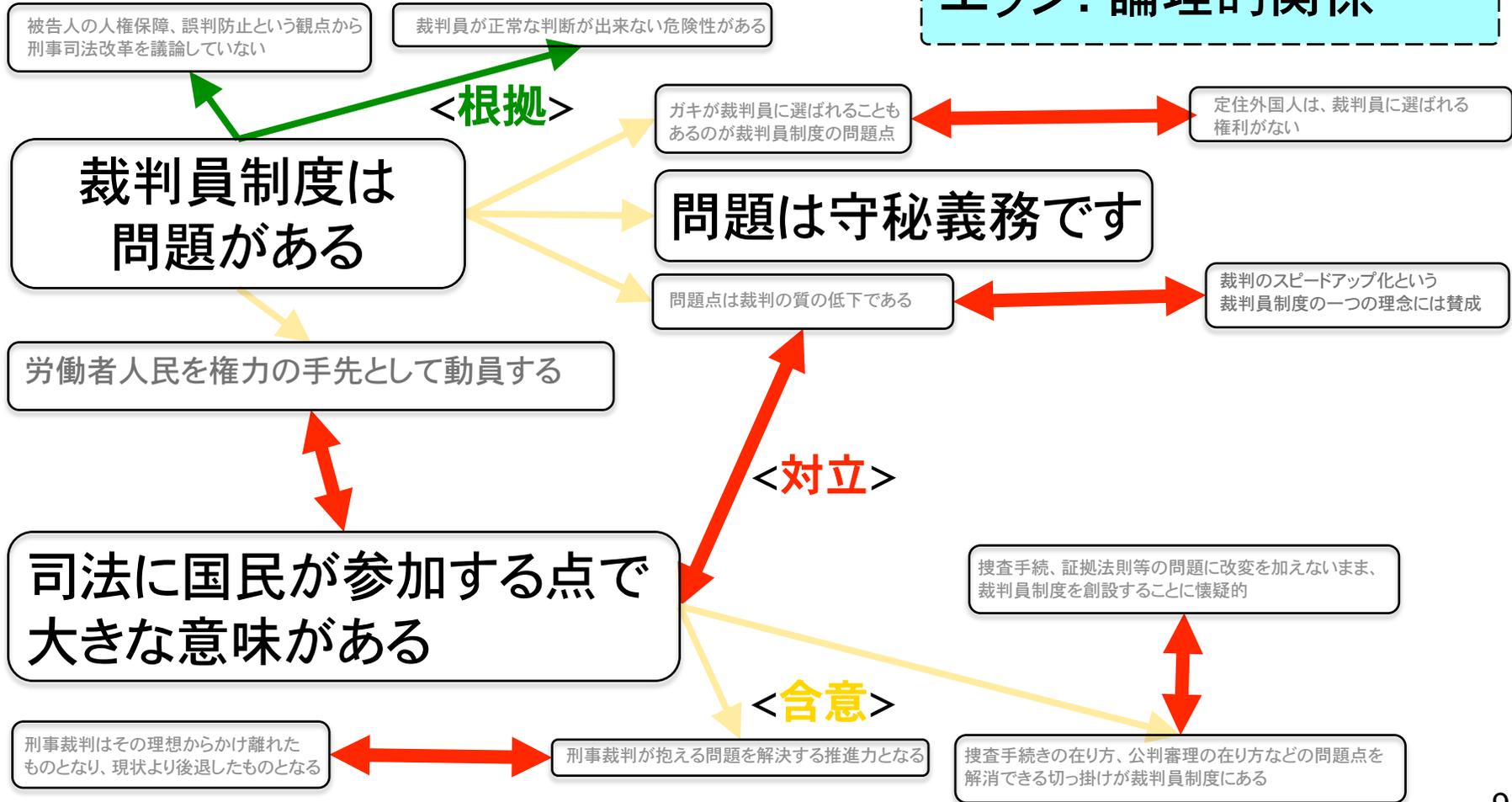
＜問題は守秘義務です＞

- そして、今、この裁判員制度の中で最も問題だと言われているのが守秘義務のことだと思います。
- この制度で、最も問題なのは、守秘義務だと考えています。
- 裁判員法案は、裁判員への広範な守秘義務、裁判員への情報提供を処罰するなど問題点はあります。
- 裁判上知りえた事実を守秘義務によって終生口外してはならない規定が裁判員の負担になる。

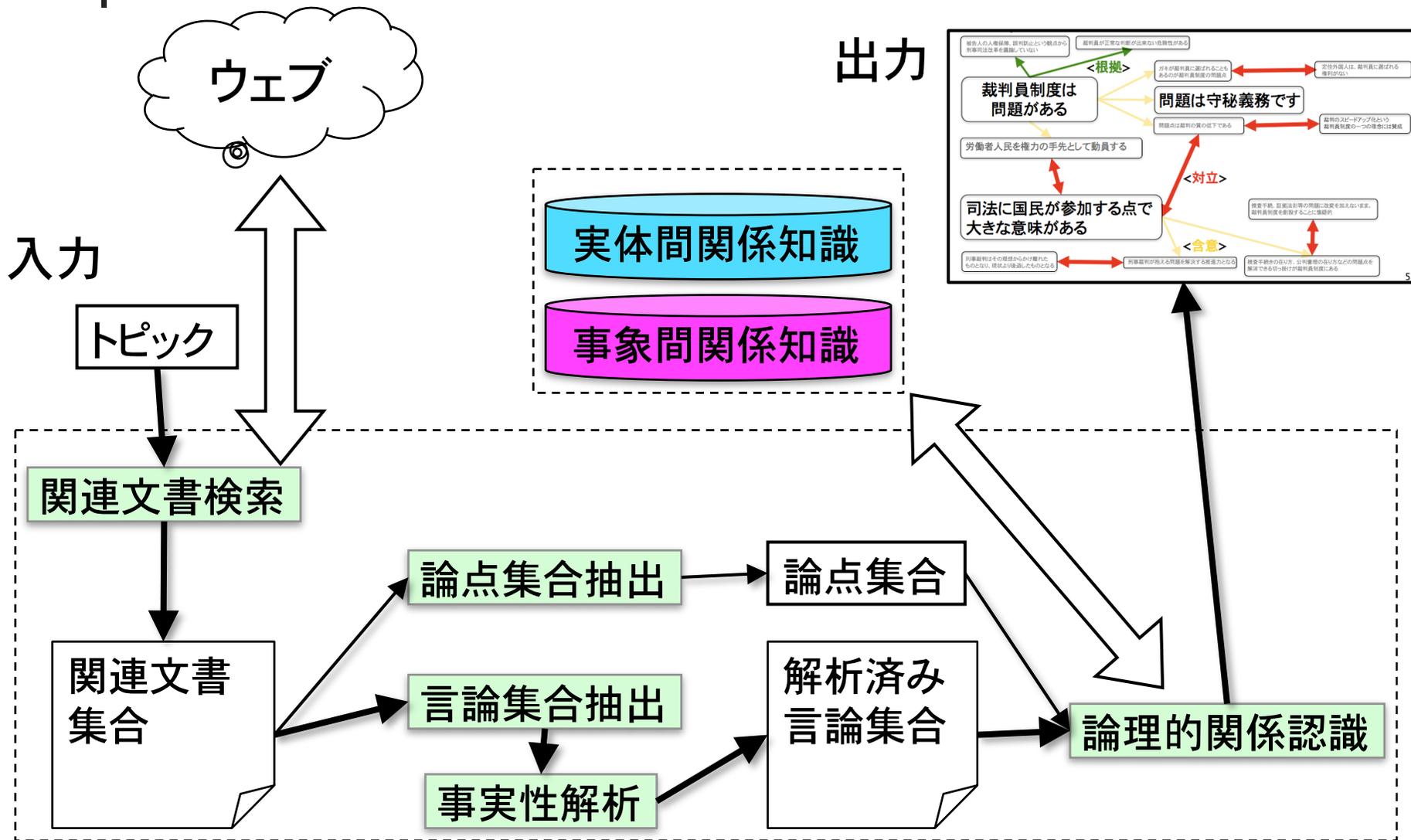
言論マップ

■ 例) トピック「裁判員制度」

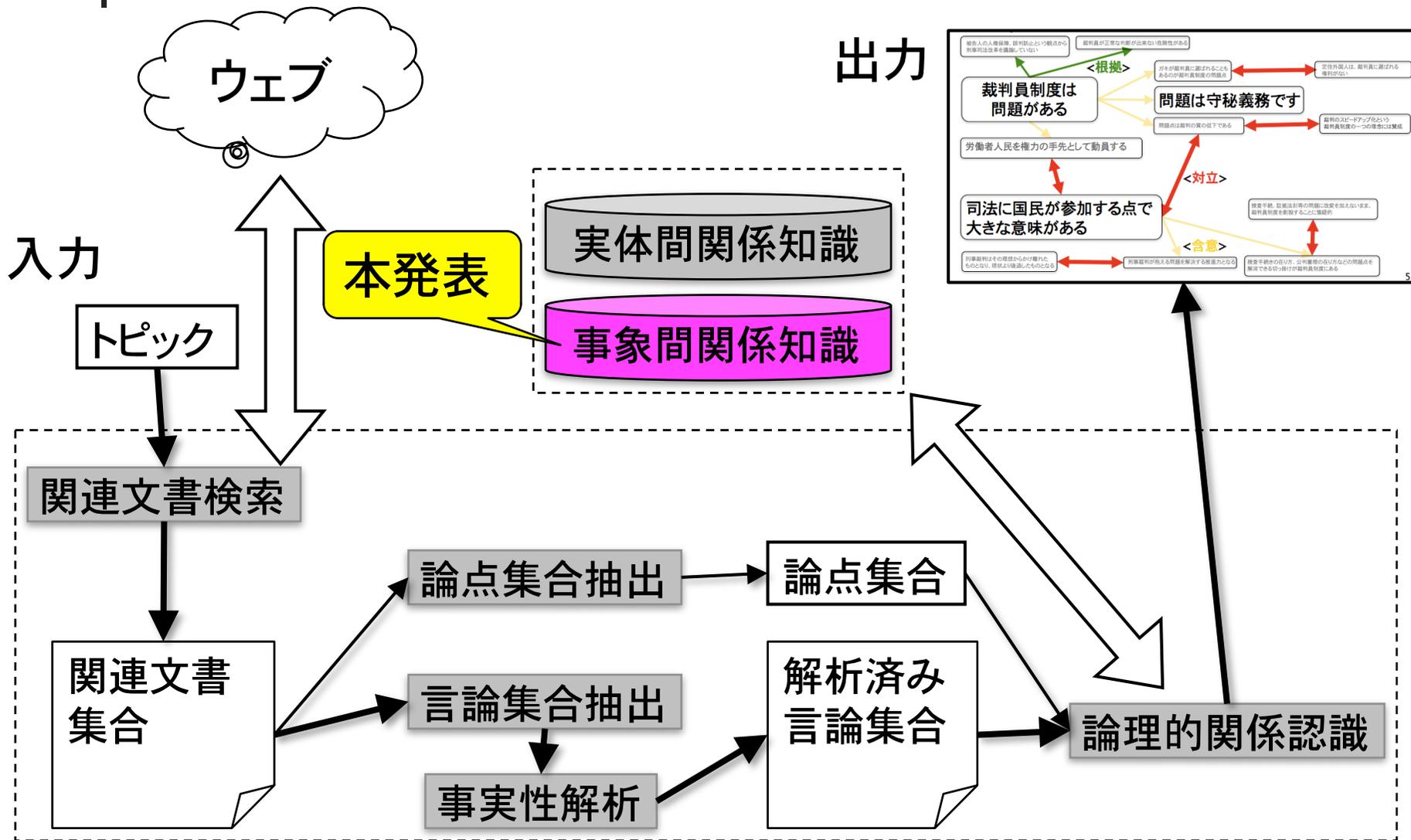
ノード: 類義言論の集合
エッジ: 論理的関係

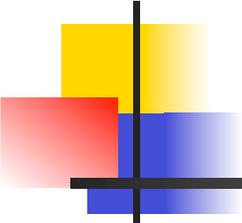


言論マップ生成システム



言論マップ生成システム





発表概要

1. 背景と目的
2. 言論マップと言論マップ生成システム
- 3. 事象間関係知識の整備**
4. 整備した関係知識の小規模な評価実験
5. まとめ

事象間関係知識

- 事象 = 述語項構造の形式で表される用言
 - <X>が<Y>に賛成する
 - <X>が<Y>に同意する
- 事象間関係知識

事象間関係知識

類義: <X>が<Y>に賛成する ⇔ <X>が<Y>に同意する
反義: <X>が<Y>に賛成する ⇔ <X>が<Y>に反対する

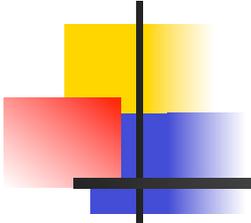
国民が新しい制度に賛成する

国民が新しい制度に反対する

<類義>

<反義>

国民が新しい制度に同意する



関連研究

■ 英語

■ 大規模なデータベース

- WordNet (Fellbaum 1998)、
FrameNet (Baker et al. 1998)、
VerbOcean (Chklovski et al. 2004)

■ 日本語

■ 大規模なデータベース

- (Bond et al. 2008; 大西ら 2008; 竹内ら 2008)

■ コーパスからの知識獲得手法

- (Inui et al. 2005; Torisawa 2006; 阿部ら 2008)

事象間関係知識データベース

	項構造		関係知識		
	基本	下位事象を含む	シソーラス (類義・反義)	二項関係	その他
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約600語 [約700 語義]	約4千語 [約7千語義] のシソーラス (竹内ら 2008)	約4万6千件 (大西ら 2008)	
形容詞	約700語 (進行中)			約700語 (進行中)	
形容動詞	約2千語 (進行中)			約2千語 (進行中)	
複合用言 (慣用句等)					
名詞述語					

基本項構造

	項構造		関係知識		
	基本	下位事象 を介して	シソーラス (類義語)	二項関係	その他
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約 [約 語	<ul style="list-style-type: none"> • <X>が<Y>を<Z>に<u>使う</u> • <X>が<Y>に<u>賛成する</u> • <X>が<Y>に<u>詳しい</u> • <X>が<Y>が<u>好きだ</u> • <X>が<Y>に<u>役に立つ</u> • <X>が<u>二進も三進も行かない</u> • <X>からの<u>電話だ</u> 		
形容詞	約700語 (進行中)				
形容動詞	約2千語 (進行中)				
複合用言 (慣用句等)					
名詞述語					

下位事象を含む項構造

	項構造	
	基本	下位事象を含む
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約600語 [約700 語義]
形容詞	約700語 (進行中)	
形容動詞	約2千語 (進行中)	
複合用言 (慣用句等)		
名詞述語		

- 部長が部下にノルマを達成することを命令した
- 部長が部下にノルマの達成を命令した
- 部長が部下にノルマを達成しろと命令した
- 部長が部下にノルマを達成するように命令した
- 部長が部下にノルマを達成するようにと命令した

上位事象: 部長が部下に何かを命令する
下位事象: **部下が**ノルマを達成する

下位事象を内包する構造を持つ用言の分類

分類型	語義数	例
決意型	25	決意する、思い立つ
約束型	28	約束する、請け負う
命令型	93	命令する、せがむ
言う型	1	言う
許可型	16	許可する、禁止する
提案型	3	提案する、発案する
発言型	170	発言する、申し上げる
感覚型	124	思う、認める
認識型	54	見る、聞こえる
感情型	143	驚く、感心する
態度型	71	歓迎する、叱る
計	728	

下位事象の主格

ことを、のを、を、(基本形)と、
(意向形)と: 上位のガ格

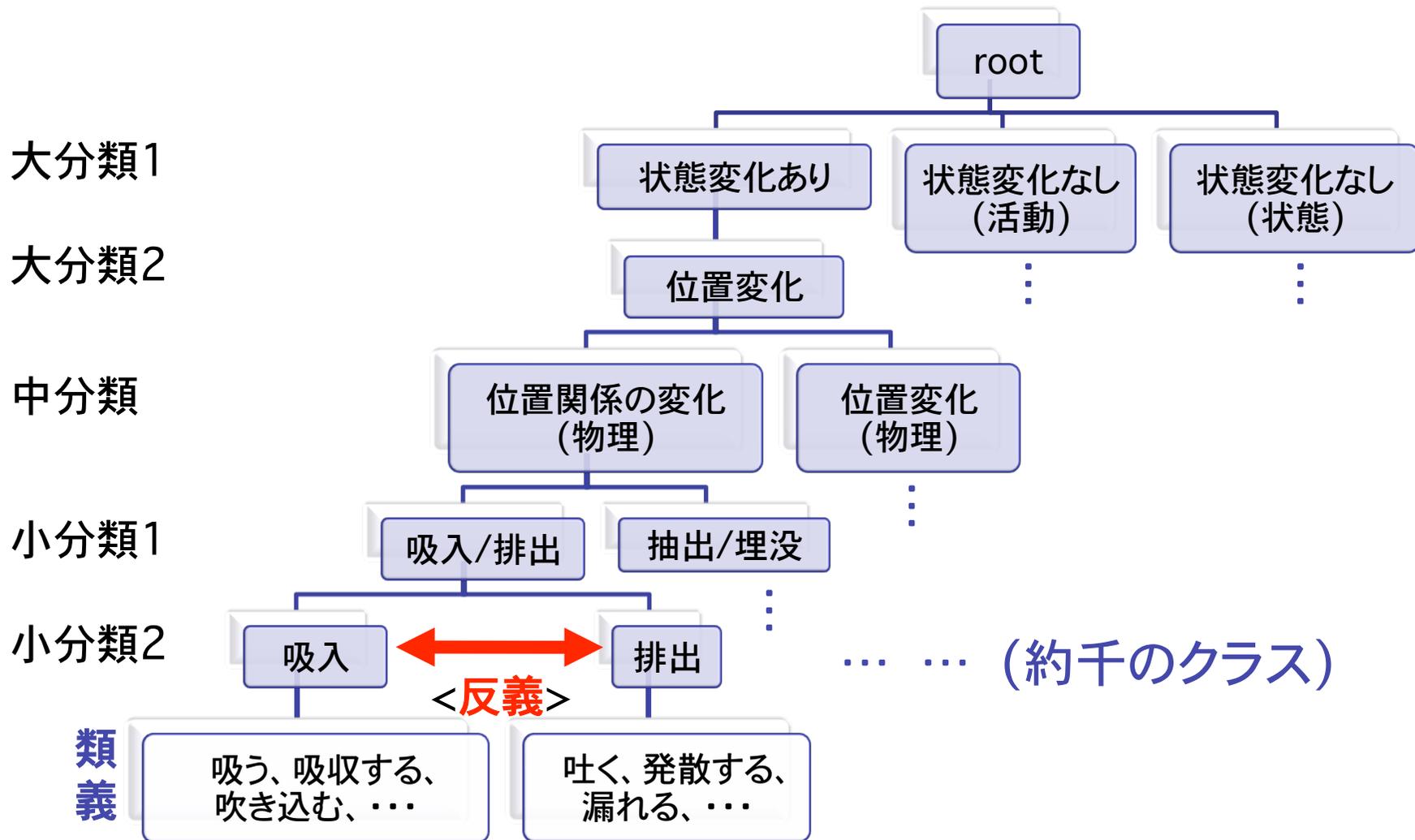
ことを、のを、を、(命令形)と、
ように、ようにと: 二格

ことを、のを: 下位のガ格、
を: ノ格、
(意向形)と: 上位のガ格、
ように、ようにと: 二格

用言のシソーラス

	項構造		関係知識		
	基本	下位事象を含む	シソーラス (類義・反義)	二項関係	その他
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約600語 [約700 語義]	約4千語 [約7千語義] のシソーラス (竹内ら 2008)	約4万6千件 (大西ら 2008)	
形容詞	約700語 (進行中)			約700語 (進行中)	
形容動詞	約2千語 (進行中)			約2千語 (進行中)	
複合用言 (慣用句等)					
名詞述語					

動詞項構造シソーラス(竹内ら 2008)



二項関係

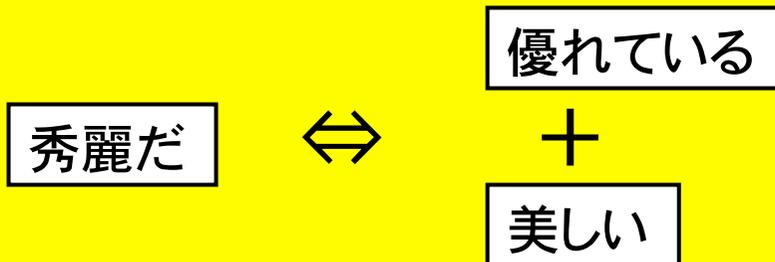
	頂構造	関係知識	
		二項関係	その他
動	<p>国語辞典における語釈文から人手で収集</p> <ul style="list-style-type: none"> • 同義(言い換え): <X>が<Y>を取り押さえる ⇔ <X>が<Y>をつかまえる • 同義・上位: <X>が<Y>を使いこなす ⇒ <X>が<Y>を使う • 結果(状態): <X>が<Y>を起こす ⇒ <Y>が直立する • 手段: <X>が<Y>を沸かす ⇒ <X>が<Y>に熱を加える 	約4万6千件 (大西ら 2008)	
形		約700語 (進行中)	
形		約2千語 (進行中)	
複 (性)			
名			

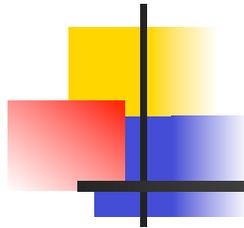
9種類の二項関係

二項を超えた関係

	頂構造	関係知識
	<p>順序(Ruppenhofer et al. 2006):</p>	
動詞	約1万 [約3万] (大西 竹内ら)	その他
形容詞	約70 (進行)	
形容動詞	約2千 (進行)	
複合用言 (慣用句等)		
名詞述語		

必要十分:





現状

	項構造		関係知識		
	基本	下位事象を含む	シソーラス (類義・反義)	二項関係	その他
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約600語 [約700 語義]	約4千語 [約7千語義] のシソーラス (竹内ら 2008)	約4万6千件 (大西ら 2008)	
形容詞	約700語 (進行中)			約700語 (進行中)	
形容動詞	約2千語 (進行中)			約2千語 (進行中)	
複合用言 (慣用句等)					
名詞述語					

エントリー例

言論マッププロジェクト - 事象オントロジー version0.03

http://gooseberry.local/~matuyosi/outside/a202stma| Google

Yahoo! JAPAN 最新ニュース

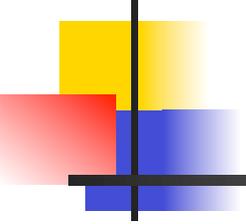
事象オントロジー

大竹辞書 version0.07

「吸収する」

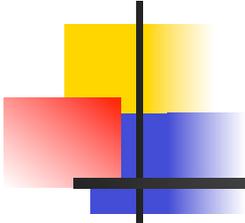
ID	04992 / 0011538-0-0-0-0-x0
表記	吸収する (きゅうしゅうする)
意味クラス	状態変化あり.位置変化.位置関係の変化 (物理).吸入/排出.吸入
が格	<何か>.causer
を格	<何か B>.対象

1. <何か>が<何か B>を吸収するということは、<何か>が<何か B>を<何か>に吸い込むということで、逆も成り立つ
2. <何か>が<何か B>を吸収するということは、<何か>が<何か B>を吸い取るということで、逆も成り立つ
3. <何か>が<何か B>を吸収するということは、同時に、<何か>が<何か B>を取り入れることも起こる
4. <何か>が<何か B>を吸収するということは、<何か>が<何か B>を自分のものとするということであるが、逆は成り立つかどうか分からない



発表概要

1. 背景と目的
2. 言論マップと言論マップ生成システム
3. 事象間関係知識の整備
- 4. 整備した関係知識の小規模な評価実験**
5. まとめ



類義・対立認識への応用

- 言論マップ生成システム(村上ら 2008)を利用
- 評価コーパス(述語項構造レベル、2トピック):
 - 「ステロイド」: 4,347 言論
 - 「喫煙」: 5,989 言論
- 評価観点:
 1. 評価コーパス内に、類義関係にある言論を少なくとも1つ持っている言論を抽出できるか
 2. 人手によるクラスタリング結果を与えたときに、人間と同じような対立関係認識ができるか

結果

「ステロイド」	精度	再現率
類義認識	80%(822/1026)	63%(822/1300)
対立認識	45%(14/31)	26%(14/54)

「喫煙」	精度	再現率
類義認識	80%(501/630)	51%(501/980)
対立認識	59%(20/34)	13%(20/158)

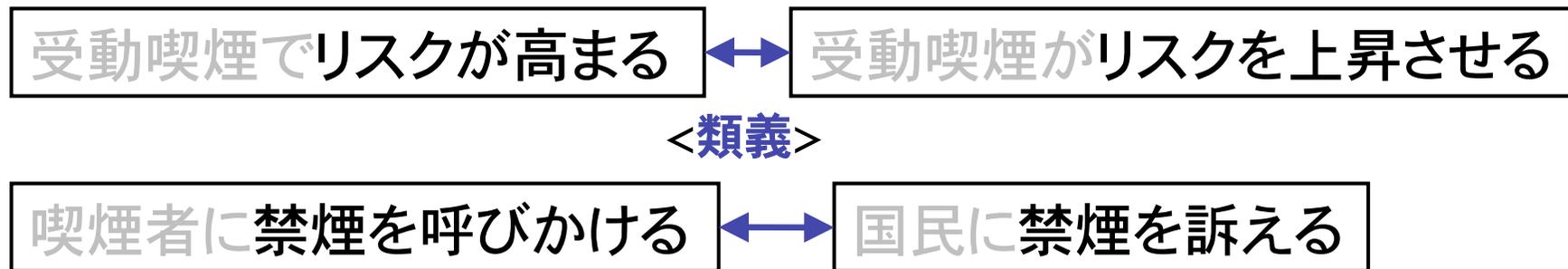
再現率が低い理由:

ニコチン中毒になる ↔ ニコチン依存症になる

(現在の言論マップ生成システムは、
言論内の名詞間の関係を全く考慮していないため、
<類義>と認識できない)

考察

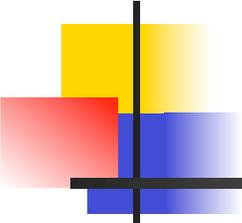
■ 成功例:



■ 失敗例:

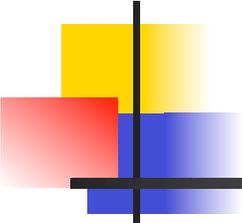


“動詞⇔「形容詞の連用形」+なる、する”の形式の
関係知識を収集する必要がある



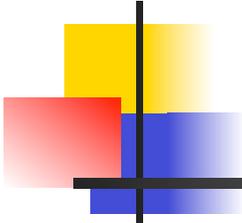
発表概要

1. 背景と目的
2. 言論マップと言論マップ生成システム
3. 事象間関係知識の整備
4. 整備した関係知識の小規模な評価実験
5. まとめ



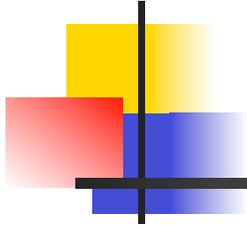
まとめ

- 言論マップ生成に必要なとなる
事象間関係知識データベースと
その現状について報告した
 - 収録エントリー: 語義単位で動詞約3万語
 - 関係知識:
 - 約7千語義の動詞項構造シソーラス
 - 約4万6千件の二項関係



今後の予定

	項構造		関係知識		
	基本	下位事象を含む	シソーラス (類義・反義)	二項関係	その他
動詞	約1万4千語 [約3万語義] (大西ら 2008; 竹内ら 2008)	約600語 [約700 語義]	約4千語 [約7千語義] のシソーラス (竹内ら 2008)	約4万6千件 (大西ら 2008)	
形容詞	約700語 (進行中)			約700語 (進行中)	
形容動詞	約2千語 (進行中)			約2千語 (進行中)	
複合用言 (慣用句等)					
名詞述語					



ご静聴ありがとうございました