

日本語文を想定した 知識処理ツール

拓殖大学工学部情報工学科
石川 勉

講演内容

- ◎ ツール開発の経緯
 - ・ 言葉で考えるコンピュータの研究
- ◎ 想定する知識表現法
- ◎ 知識変換ツール(CONV)
- ◎ 推論エンジン(ARSK)
 - ・ 概略推論法ARSK
 - ・ 拡張型順序ソート論理の導入
- ◎ 類似度算出ツールと共起情報DB
- ◎ 開発ツールと応用例(雑談型対話システム)のデモ

2

ツール開発の経緯

本講演での諸ツールは以下で公開予定(一部公開中)
<http://www.cs.takushoku-u.ac.jp/ai/html/demo.html>

3

研究のねらい

- ・ 質問応答システムの高度化 (IBM Watson)
- ・ 自然に対話できるロボットの実現 (Apple Siri)
- ・ ヒューマンインターフェースの柔軟化

究極的なねらいはロボットの脳 !!

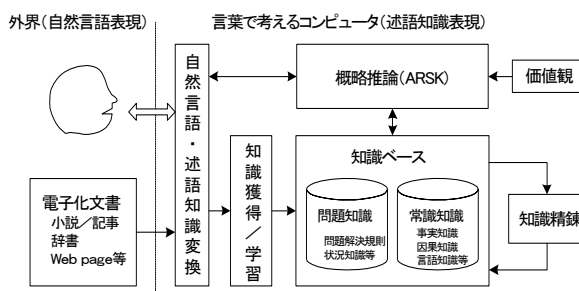
“意味処理の高度化”が必須

“言葉を理解し、言葉で考え、言葉で答える”コンピュータ
(言葉で考えるコンピュータ)

前提条件: 全処理の自動化

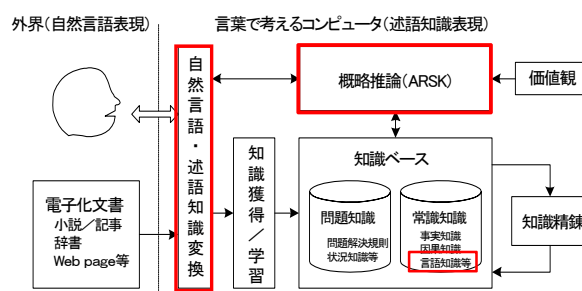
4

言葉で考えるコンピュータ



5

ツール化部分



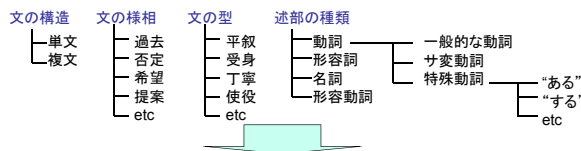
6

知識表現法

対象とした自然言語文

文のレベル(概念レベル)の分類*

- 一般的な事態を表す文 「春は雨が降る」
 - 個別的な現象を表す文 「昨日は雨が降った」
 - 主体の判断を表す文 「今日は雨が降るだろう」
 - 事態あるいは判断を表現・伝達する文 .. 「昨日は雨が降ったようだよ」
- } 論理では 扱えない



このような文の全てを対象とし、これを述語論理に変換する

基本的な知識表現形式

述語式 $L = s_1, \dots, s_n P (r_1:t_1, \dots, r_n:t_n)$

P: 述語, r_i : ラベル, t_i : 引数, s_i : 識別子

様態	記号	表現例	ラベル	格(意味)	説明
受身	u	～れる	agt	主体(agent)	意図的に動作を行う主体
希望	w	～したい	obj	対象(object)	動作・思考の対象
否定	!	～しない	gol	目標(goal)	動作の終点
過去	*	～した	sou	源泉(source)	動作の起点
⋮	⋮	⋮	⋮	⋮	⋮

太郎は自宅に帰った ⇒ *帰る(agt:太郎, gol:自宅)

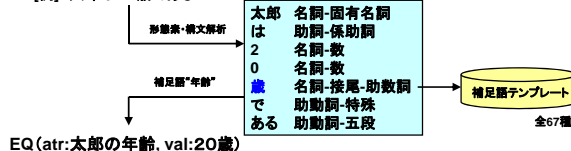
◎ 等値関係の表現形式

$EQ(atr:t_1, val:t_2)$

EQ: 等値関係, atr: 属性, val: 値, t_i : 引数

主部と述部の関係を示す ⇒ 主部に対して補足語を補う

【例】 太郎は20歳である



EQ(atr: 太郎の年齢, val: 20歳)

複文の基本的な表現法

太郎は花子が読んだ本を買った

花子は太郎が買った本を読んだ

変換

従来の一階述語論理

*買う(agt:太郎,obj:本) ∧ *読む(agt:花子,obj:本)

本知識表現

*買う(agt:太郎,obj:[本]{*読む(agt:花子,obj:#)})

*読む(agt:花子,obj:[本]{*買う(agt:太郎,obj:#)})

◎ 各種従属節の表現

○ 名詞節, 連体節, 連用節 (条件文を除く)

主節の述語式に埋め込む繰り返し構造 ⇒ [a]{L}

L': 節述語(述語式の形で従属節を表現), a: L'に修飾される語

刑事は犯人が隠れている家を探している
⇒ 探す(agt:刑事,obj:[家]{隠れる(agt:犯人,gol:#)})

○ 並列節

主節と従属節をそれぞれ独立の文として扱う

並列節 ⇒ $L_1 \wedge L_2$

太郎は本屋に行って本を買った
⇒ *行く(agt:太郎,gol:本屋) ∧ *買う(agt:太郎,obj:本)

○ 連用節(条件文)

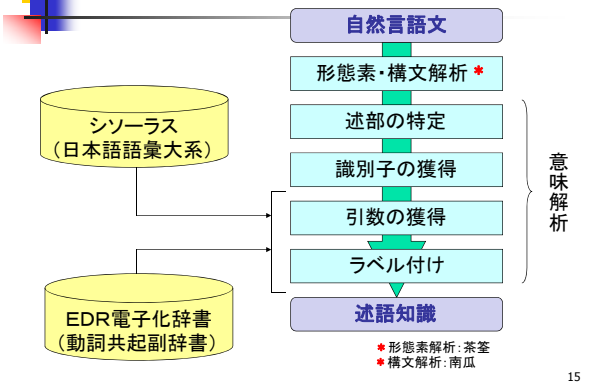
連用節(条件文) \rightarrow $L_1 \rightarrow L_2$ (普遍的な事実)
 \rightarrow $cnd:[a]\{L\}$ (特定の個体)

秋が来れば葉が落ちる
 \Rightarrow 来る(obj:秋) \rightarrow 落ちる(obj:葉)

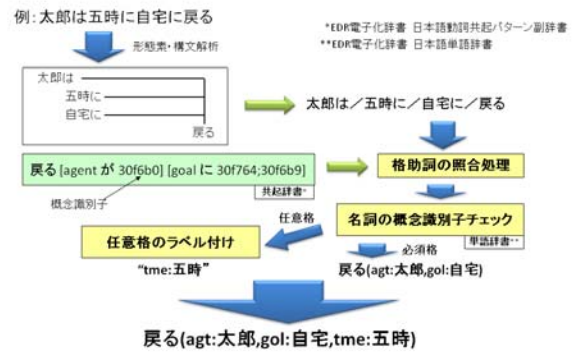
特定の個体
 太郎が来るならば次郎は来ない
 \Rightarrow !来る(agt:次郎, cnd:[条件]{来る(agt:太郎)})

知識変換ツール (CONV)

知識変換の流れ



◎ 述部が動詞の場合



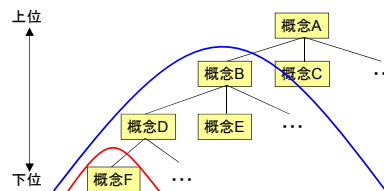
○ 任意格の決定

格助詞	第1候補	第2候補	第3候補	第4候補	第5候補	第6候補	第7候補	第8候補
で	plc*	imp	son	mat	cau*	tme	cnd	
に	gol	tme	plc*	son	cau*	mat	imp	mnr
を	obj	gol	son	plc*				
から	tfr	sou	cau*	mat				
まで	plc*	gol	tto					
へ	gol	plc*						
が	agt	tme	obj					
と	注)「と」に関しては別手法で処理する							
より	sou							
にて	plc*	imp	son	mat	cau*	tme	cnd	
として	as							
副詞	mnr							
によって	cau*	imp						

* 否定のチェックを行う

- 必須格との重複
- 格助詞の前方名詞に対するカテゴリ情報との照合*

[第1チェック]: カテゴリの照合範囲を厳格に指定したチェック
 [第2チェック]: 第1チェックでの照合範囲を拡大したチェック



* 照合範囲のチェックには日本語語彙体系を利用

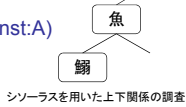
◎ 述部が名詞の場合

「AはPである」の文型が対象

- A(主部)とP(述部)に上位下位やクラス-インスタンスの関係がある

⇒ ラベル inst を用いて表現 ⇒ P(inst:A)

鰯は魚である ⇒ 魚(inst:鰯)



シソーラスを用いた上下関係の調査

- AとPにクラス-インスタンスの関係がない

⇒ EQ述語を用いて表現 ⇒ EQ(atr:A,val:P)

太郎は20歳である ⇒ EQ(atr:太郎の年齢,val:20歳)

◎ 述部が形容詞の場合

文型と格助詞ごとにラベルを決定(14種の変換パターン)

文型・格助詞	ラベル付き知識
AはP	P (obj: A)
AはBがP	P (obj: A, s-obj: B)
AはBに(は)P	
AはBと(は)P	
AはBでP	
AはBまでP	P (obj: A, s-gol: B)
AからBまでP	
AはBからP	P (obj: A, s-sou: B)
AまでBからP	
AとBはP	P (obj: A)∧P (obj: B)
AとBはCがP	P (obj: A, s-obj: C)∧P (obj: B, s-obj: C)
AはBよりP	P (obj: A, s-omp: B)
AはBよりCがP	P (obj: A, s-obj: C, s-omp: B)
AはBがCよりP	

○ 表現例

- 花子は美しい
美しいa(sbj:花子@)
- 私は歯が痛い
痛いa(sbj:私,s-obj:歯)
- 先生は女生徒に甘い
甘いa(sbj:先生,s-obj:女生徒)
- この靴は私には大きい
大きいa(sbj:この靴,s-obj:私)
- 私の考えは彼の考えに近い
近いa(sbj:私の考え,s-obj:彼の考え)
- 私は彼と親しい
親しいa(sbj:私,s-obj:彼)
- 私は彼とは親しい
親しいa(sbj:私,s-obj:彼)
- 私は試験勉強で忙しい
忙しいa(sbj:私,s-obj:試験勉強)
- 我が家は駅まで遠い
遠いa(sbj:我が家,s-gol:駅)
- 我が家から駅まで遠い
遠いa(sbj:我が家,s-gol:駅)
- 根室は札幌から遠い
遠いa(sbj:根室,s-sou:札幌)
- 花子は良子より美しい
美しいa(sbj:花子,s-cmp:良子)

特殊動詞を含む文の変換

特殊動詞「ある」や「する」等は使用する文によって、多くの意味を持つ

事故は空港であった ⇒ 事故は空港で発生した

国会議事堂は東京にある ⇒ 国会議事堂は東京に存在する

会社が違法行為をする ⇒ 会社が違法行為を行う

彼は首にマフラーをする ⇒ 彼は首にマフラーを身に着ける

する(agt:会社,obj:違法行為) する(agt:彼,gol:首,obj:マフラー)

- 文型と名詞のカテゴリ情報を利用して意味判別
- 文の書き換えを行った後、述語知識化

日本語基本動詞用法辞典と日本語語彙大系から

「する」の意味を20種に分類

- ① 自分の意思である動作・行為を行う ⇒ [例] 会社が違法行為をした
- ② ある職業・役割・係り等に従事する ⇒ [例] 本のセールスをする
- ③ ある物や人を別の物や人に変える ⇒ [例] 私は鰯を刺身にする
- ④ せき・くしゃみ等の生理現象を起こす ⇒ [例] 父は一年前に病気をした
- ⑤ 何かを身に着ける ⇒ [例] 彼は首にマフラーをした

1. 動詞の置き換え

2. EQ述語の利用

3. 識別子の利用

4. テンプレートの利用

1. 動詞の置き換え

「する」の意味を8種に分類

「成る」、「存在する」、「置く」、「有する」
「行う」、「変わる」、「決める」、「着る」

「行う」の規則

文型
A {が/は} (B {に/で}) C を (D で) する

A~Dの名詞のカテゴリ
A:人/組織/動物 , B:A/物/場所 , C:行為 , D:場所

[例] 会社が違法行為をする ⇒ 会社が違法行為を行う ⇒ 行う(agt:会社,obj:違法行為)

2. EQ述語の利用

文型
A {が/は} B {を/も} する

A, Bの名詞のカテゴリ
A:物 , B:形容詞+形/単位

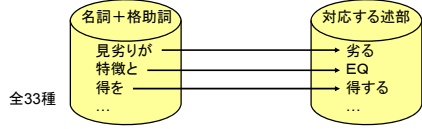
[例] あの車は300万円もする ⇒ EQ(atr:あの車の値段,val:300万円)

3. 識別子の利用

規則: 動詞+(よう)としている

[例] 猫が屋根から降りようとしている ⇒ g降りる(agt:猫,sou:屋根)

4. テンプレートの利用



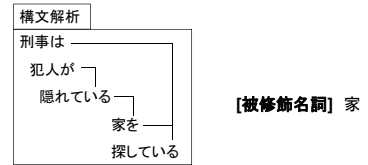
[例] 先生は指導力に見劣りがする ⇒ 先生は指導力に劣る
⇒ 劣る(agt:先生,gol:指導力)

25

複文の変換法

○名詞節、連体節、連用節(条件文以外)

刑事は犯人が隠れている家を探している

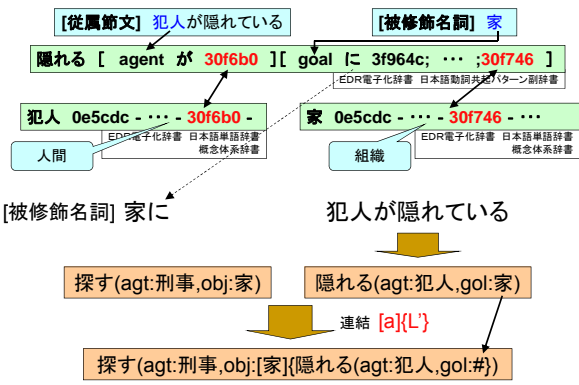


[主節文] 刑事は家を探している

[従属節文] 犯人が隠れている

探す(agt:刑事,obj:家)

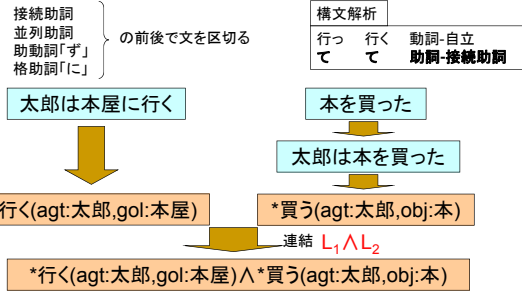
26



27

○並列節

太郎は本屋に行って本を買った

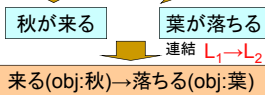


28

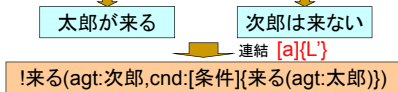
○連用節(条件文)

接続語: たら、ば、と、なら、でも、たって、ては、とすると
とすれば、としたり、となると、かぎり、ばあい

秋が来れば葉が落ちる 一般的な概念



太郎が来るならば次郎は来ない 特定の個体



29

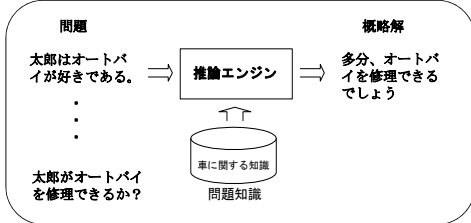
推論エンジン (ARSK)

30

概略推論法 (ARSK) の狙い

知識間の類似性を利用して問題知識の不完全さを補い概略的な解を導く

動作イメージ



31

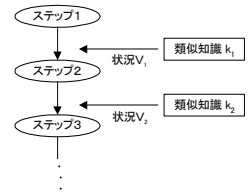
ARSKの定義

$$K \not\vdash G$$

$$K \cup K_S \vdash G$$

$$K_S = \{k_i \mid k \in K, k_i \approx_{v_i} k\}$$

ARSK : Approximate Reasoning with Similar Knowledge



知識の類似性判定には概念ベース等を利用

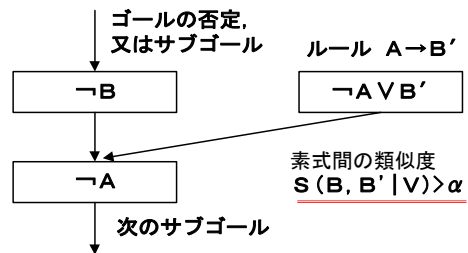
32

◎ 不完全知識に対する働き

	知識の欠落		誤った知識の存在	
	従来	ARSK	従来	ARSK
真偽問題 (Yes/No)	No	No	No	No
	Yes	Yes	Yes	Yes
穴埋め問題 (x = ?)	なし	$x \approx A$ (正)	$x = B$ (誤)	$x = B$ (誤) $x \approx A$ (正)

33

◎ 類似導出

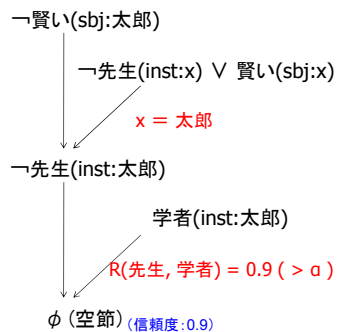
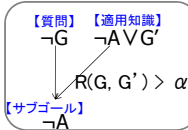


34

◎ ARSKの推論例

- xが先生ならばxは賢い
- 太郎は学者である
- 太郎は賢いですか?

\neg 先生(inst:x) \vee 賢い(sbj:x)
学者(inst:太郎)
賢い(sbj:太郎)



35

◎ サブゴールの信頼度

類似知識を適用して推論を進める

- サブゴールに信頼度の付与
- サブゴールの信頼度 $R < \eta$ のとき停止

n 番目のサブゴールの信頼度:

$$R = \prod_{i=1}^n s_i$$

s_i : i 番目の導出での類似度

36

ARSKの改良

- ARSKの課題
 - 基本的に知識構成要素は単語
 - ⇒ 名詞句等の複雑な概念を扱えない
 - 知識の類似性のみを利用
 - ⇒ 知識間の意味的な包含関係(⊃)を考慮していない

- 名詞句等の複雑な概念を処理可能 (名詞句内の変数も許容)
- 知識間の意味的な包含関係を利用可能

に改良

37

◎ 概念間の意味的な包含の判定

- 基本的な考え方
 - 一般的に $a \supset b, c \supset d$ のとき $ac \supset bd$
 - ただし、慣用的な表現等により必ずしもこれは成り立たない

例) 自動車 会社 大型 の 魚
 IU ~~II~~ II II ~~IU~~ IU
トラック 会社 大型 の 金魚

複合概念・名詞句の判定基準を設定

38

◎ 本推論法での処理例

① xが外国の企業で働くならば xは豊かである

② 太郎はヨーロッパの貿易会社で働く

③ 太郎は裕福ですか？

↓

一働く<(agt:x, plc:外国の企業) ∨ 豊か<(sbj:x)

働く<(agt:太郎, plc:ヨーロッパの貿易会社)

裕福<(sbj:太郎)

一裕福<(sbj:太郎)

一働く<(agt:x, plc:外国の企業) ∨ 豊か<(sbj:x)

x = 太郎
R(裕福, 豊か) = 0.8 > α

一働く<(agt:太郎, plc:外国の企業) (信頼度 0.8)

働く<(agt:太郎, plc:ヨーロッパの貿易会社)

外国の企業 ⊃ ヨーロッパの貿易会社

φ (空節) (信頼度 0.8)

39

拡張型順序ソート論理の導入

- 知識表現法
 - 順序ソート論理ベースの表現法 (基本は前述の表現法)
- 推論法
 - C.Beierle等の拡張型順序ソート論理の枠組みをさらに拡張
 - 一般概念や変数、定数を含んだ名詞句などの複合概念をソートとして扱えるように拡張

例) 変数を含む概念: “[x:顧客]の車”
 一般概念: “会社の車”

40

◎ 順序ソート論理

- 一階述語論理の変数や定数を、ソートと呼ばれる限定されたドメインの1つの要素とする
 - ソートsに属する個体tを “t:s” と表す
 - 例) c: 医者
- ソート間の上位下位の階層関係(ソート階層)
 - 下位のソートに含まれる要素は、必ず上位のソートに含まれる
 - 例) “医者” ⊆ “専門家” であるならば、“全ての専門家xが賢ければ医者cは賢い”

41

◎ 複合的な概念をソートとする利点

- C.Beierle等の拡張型順序ソート論理
 - ソートは1つの概念

例) 企業の建物の鍵
 UI 全てのソートを定義!
 会社のビルの鍵

- 本推論法
 - ソートは複合的な概念

例) 企業 の [x:建物] の 鍵
 UI UI II 単語のみの定義でOK
 会社 の [c:ビル] の 鍵

42

◎ 知識表現法WPL(述語式)

[定義1] Pを述語, t_i を引数(項), r_i を述語に対する引数の関係を表すラベルとしたとき, 以下のように表現された式Lを述語式という.

$$L = P(r_1 : t_1, r_2 : t_2, \dots, r_n : t_n)$$

ここで, 引数 t_i は, S_i をソート, x_i を変数, a_i を定数としたとき, 以下のように表現される.

$$t_i = x_i : S_i \text{ or } a_i : S_i \text{ or } x_i : T \text{ or } a_i : T$$

43

◎ ソート概念

[定義2] 以下の構成から成る概念をそれぞれ, 基本概念(W), 節付概念(Wp), 個体概念(gs), 複合概念(G), ソート概念(S)と呼ぶ.

- $W = (\text{副詞 or 形容詞 or 名詞})^0 + \text{名詞}$
- $Wp = [S]\{Lp\}$
- $gs = [x : S] \text{ or } [c : S] \text{ or } [x : T] \text{ or } [c : T]$
- $G = (\text{修飾概念} + \text{助詞})^{*1} + \text{主概念}$
 ここで, 修飾概念 = $W \text{ or } Wp \text{ or } gs$
 主概念 = W
- $S = W \text{ or } Wp \text{ or } G$

*0, *1 はそれぞれ, 0 回, 1 回以上の繰り返し 44

◎ WPLの表現例

- “太郎は医者である”
 医者(inst:太郎:T)
- “花子は美しい”
 美しい(sbj:花子:T)
- “刑事は犯人が隠れている家を探している”
 探す (agt: c_1 :刑事, obj: c_2 :[家]{隠れる(agt: c_3 :犯人, plc:#)})
- “花子は太郎が経営している会社の社員である”
 [c_1 :会社]{経営する(agt:太郎:T, obj:#)}の社員 (inst:花子:T)
- “都会の会社で働く労働者は裕福である”
 裕福 (sbj: x_1 :[労働者]{働く(agt:#, plc: x_2 :[x_3 :都会]の会社)})

45

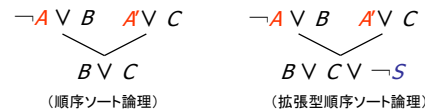
推論の基本的な考え方

■ 述語式間の照合による導出処理の繰り返し

[導出処理]

“ $\neg A \vee B$ ”と“ $A \vee C$ ”において,

- 1) A と A' が単一化可能 or A が A' を意味的に包含
- 2) A と A' が条件 S のもとで左の関係



(順序ソート論理)

(拡張型順序ソート論理)

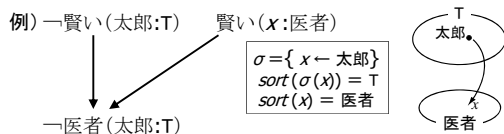
“ S ”はC.Beierle等の拡張型順序ソート論理で $SL(\sigma)$

46

拡張型順序ソート論理の考え方

$$SL(\sigma) = \wedge \{Si(\sigma(x)) \mid sort(\sigma(x)) \not\subseteq sort(x), \text{ where } x \in dom(\sigma), Si = sort(x)\}$$

- σ : 代入
- $sort(\sigma(x))$: 代入結果のソート(代入する側のソート)
- $sort(x)$: 代入される側のソート
- $\not\subseteq$: 左が右のサブソートでない
- $dom(\sigma)$: σ に関する変数のドメイン



47

◎ 拡張型順序ソート論理

[推論規則]

- EOS resolution rule (EOS): $\frac{\neg L_1 \vee A, L_2 \vee B}{\sigma(A \vee B) \vee \neg SL(\sigma)}$
 L_1, L_2 : 述語部が同じ
- Subsort resolution rule (SUBS): $\frac{\neg S_1(t_1) \vee A, S_2(t_2) \vee B}{\sigma(A \vee B) \vee \neg SL(\sigma)}$
 S_1, S_2 : 述語部がソート述語かつ $S_2 \subseteq S_1$
- Elimination rule (ER): $\frac{\neg S_1(t_1) \vee A}{sort(\sigma(t_1)) \subseteq S_1, \sigma(A) \vee \neg SL(\sigma)}$

[繰り延べ節]

- SL節: $SL(\sigma) = \wedge \{S_i \sigma(x) \mid sort(\sigma(x)) \not\subseteq sort(x), \text{ where } x \in dom(\sigma), S_i = sort(x)\}$

48

◎ 拡張型順序ソート論理の拡張

[推論規則]

- EOS resolution rule (EOS):
$$\frac{\neg L_1 \vee A, L_2 \vee B}{\sigma(A \vee B) \vee \neg SL(\sigma)}$$

 L_1, L_2 : 述語部が同じ
- Subsort resolution rule (SUBS):
$$\frac{\neg S_1(t_1) \vee A, S_2(t_2) \vee B}{\sigma_{op}(A \vee B) \vee \neg SL(\sigma_{op}) \vee \neg SLP}$$

 S_1, S_2 : 述語部がソート述語
 $S_2 \subseteq S_1$
- Elimination rule (ER):
$$\frac{\neg S_1(t) \vee A}{\sigma_{op}(A) \vee \neg SL(\sigma_{op}) \vee \neg SLP}$$

 $sort(\sigma(t_i)) \subseteq S_1$
 σ_p : ソート概念での代入

[繰り延べ節]

- SL節: $SL(\sigma_p) = \wedge \{S_i(\sigma(x)) \mid \sigma_p(sort(\sigma(x))) \not\subseteq \sigma_p(sort(x)), \text{ where } x \in dom(\sigma), S_i = \sigma_p(sort(x))\}$
- SLp節: $SLp = \wedge \{S_i(\sigma_p(z)) \mid \sigma_p(sort(\sigma_p(z))) \not\subseteq \sigma_p(sort(z)), \text{ where } z \in dom(\sigma_p), S_i = \sigma_p(sort(z))\} \wedge S_i'(t) \text{ where } S_i' = \sigma_p(g_b)$

49

[知識]

- 若者が先進国の資産家が営む企業で働いているなら、その若者は裕福である。
- 太郎は花子が営む会社で働いている。
- 花子は日本の財閥である。
- 日本は先進国である。
- 太郎は青年である。

[質問]

- 太郎は裕福ですか？



[知識]

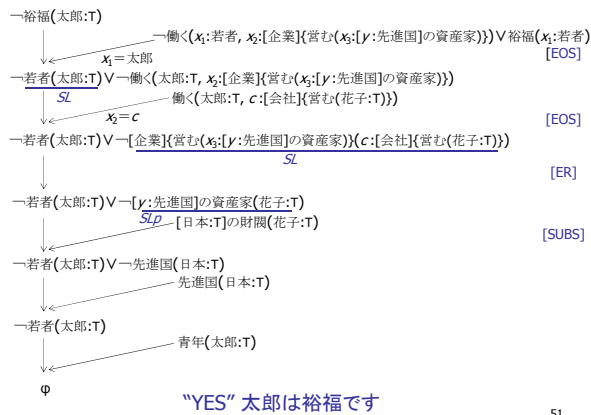
- 働く(x_1 :若者, x_2 :[企業]{営む(x_3 :[先進国]の資産家)}) → 裕福(x_1 :若者)
- 働く(太郎:T, c:[会社]{営む(花子:T)})
- [日本:T]の財閥(花子:T)
- 先進国(日本:T)
- 青年(太郎:T)

[質問]

- 裕福(太郎:T)？

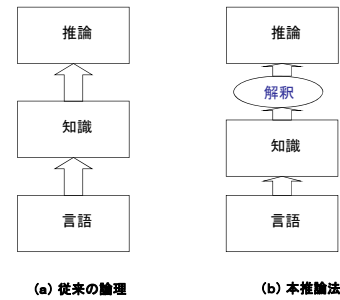


50



51

本推論法と従来法の本質的な違い

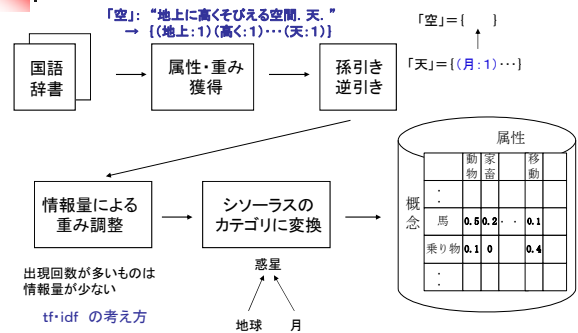


52

類似度算出ツール 共起情報DB 諸ツールの応用 (対話システム)

53

概念ベース



54

◎ 評価方法

類似度計算ツールに必要な特性

- ① 類似概念間の類似度と非類似概念間の類似度の差は大きい程良い

$$F_1 = (\bar{r}_1 - \bar{r}_2) / (1 + \sigma_1 + \sigma_2)$$
 $\bar{r}_1, \bar{r}_2 : r_1, r_2 \text{の平均}$
 $\sigma_1, \sigma_2 : r_1, r_2 \text{の標準偏差}$
- ② 類似概念間の類似の程度を正確に判定

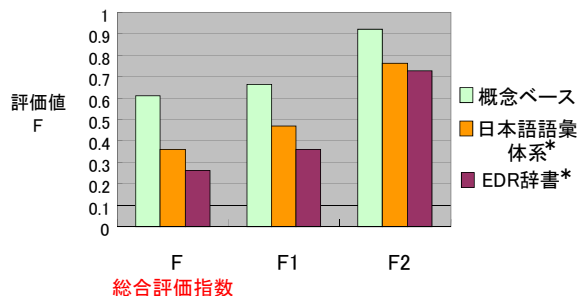
$$F_2 = m / N$$
 $m : r_1 > r_2 \text{となるデータ数}$
 $N : \text{全データ数}$

最終的な評価指数 $F = F_1 \times F_2$



55

◎ 概念ベースとシソーラスの性能比較



* 類似度: 共通上位ノードの深さ × 2 / 対象概念のノードの深さの和

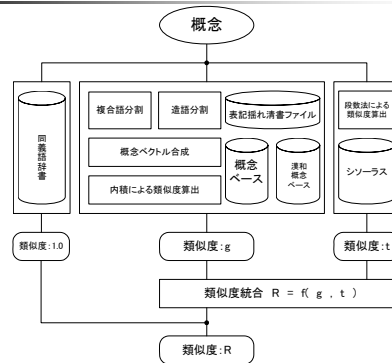
56

◎ 概念ベースの問題点

- 概念には様々な表記が存在する
 - ・ 表記揺れ、複合語、造語など
- 概念によっては、人間の感覚とかけ離れてしまう場合がある(機械的に作成しているため)

57

類似度算出ツールの全体構成



58

◎ あらゆる概念表記への対応

- 表記揺れ
 - ・ 1つの語に対して複数の表記が可能
 - 例) “くみあわせ” → “組み合わせ”, “組合わせ” など
 - 表記揺れ辞書の作成
- 複合語
 - ・ 複数の語を繋げた表記
 - 例) “整備工場” → “整備” + “工場”
 - 単語の分割とベクトル合成
- 造語
 - ・ 漢字の組合せで作る新語
 - 例) “激安” → “激” + “安”
 - 漢和GBの利用

59

◎ 表記揺れの形態

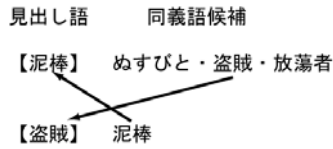
- 片仮名表記揺れ
 - 片仮名の部分的な違い
 - 例) “ハイオリン” → “ヴァイオリン”
 - 特定カテゴリに属す単語の片仮名表記化
 - 例) “桃” → “モモ”
- 漢字表記揺れ
 - 日常使用しない漢字の平仮名化
 - 例) “宝籤” → “宝くじ”
- 平仮名表記揺れ
 - 漢字で表記したものの平仮名化
 - 例) “山” → “やま”
- 送り仮名表記揺れ
 - 送り仮名の省略
 - 例) “組み合わせ” → “組合せ”
- 英字表記揺れ
 - 片仮名表記を英字化
 - 例) “エーエム” → “AM”

60

◎ 類似性判別能力の向上

・ 同義語辞書の構築・利用

相互に候補となる語を同義語とみなす



・ GBとシソーラスの類似度の統合

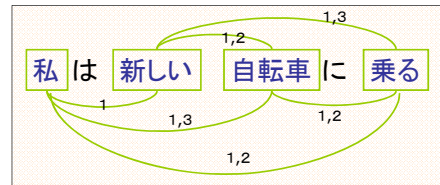
両者の差が閾値以上の場合(それ以外はGBの値)

61

共起情報DB

以下の共起情報DBを作成

- ① 単純共起DB : 1文中に同時に出現する単語
- ② 直接係り受けDB : 直接的に係り受け関係にある単語
- ③ 間接係り受けDB : 共通の係り受け関係にある単語を持つ単語



62

○ 情報源

- ・ 学研国語辞書の語義文
- ・ 新聞記事1年分(毎日新聞1994年)
- ・ google検索情報:

○ 総頻度数

頻度	1以上	5以上	10以上	20以上
直接係り受け	231,987,082	203,232,074	187,965,430	170,580,720
間接係り受け	67,840,004	53,162,873	44,637,904	36,752,581
単純共起	1,646,049,899	1,552,056,908	1,466,480,826	1,355,077,434

- 注1) 直接係り受けは、動詞-名詞、形容詞-名詞、名詞-名詞(複合語)間
 注2) 複合語と単語間の共起は、“複合語の構成単語間”、“複合語の主単語と他の単語間”
 例)「宝石店へ行く」の場合、“宝石-店”、“店-行く”
 注3) google検索の方法: 主要単語(約4万語)をgoogle検索エンジンにかけ、検索結果のすべての文から獲得。データの重複を避けるため同一のURLの検索結果は不使用。

63

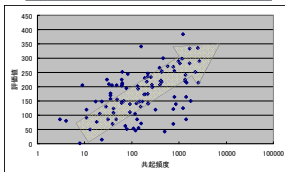
○ 共起概念の抽出例

・ 直接係り受け 概念: 自動車

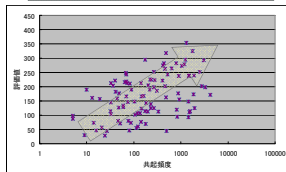
道	1451	運転	357	関連	219
保険	1122	産業	340	車	204
学校	1104	事故	328	北陸	199
メーカー	1023	販売	303	修理	197
電気	779	業界	281	バイク	166
用	674	税	281	中央	163
部品	569	会社	271	カタログ	163
免許	475	情報	238	化	160
自動車	466	普通	237	日本	156
整備	379	等	223	工場	154

64

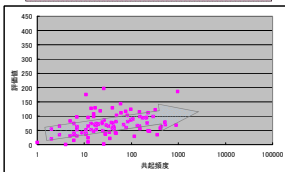
直接係り受けデータベース



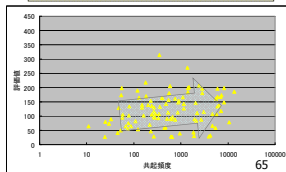
合成係り受けデータベース



間接係り受けデータベース

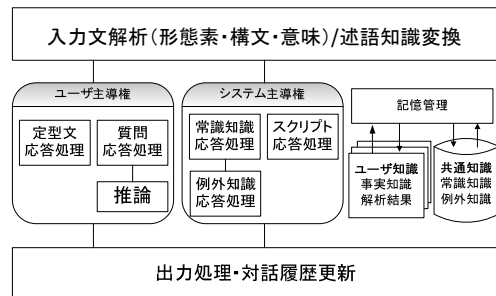


単純共起データベース



65

諸ツールの応用(対話システム)



66