

研究紹介：換喩解析，意味ソート，  
統語構造認識と  $7 \pm 2$ ，diffと言語処理

村田 真樹

通信総合研究所

(<http://www.crl.go.jp/jt/a132/members/murata/index.html>)

京都大学，集中講義

## テーマ

- 換喩解析
- 意味ソート
- 統語構造認識と  $7 \pm 2$
- diff と言語処理

## 換喩解析

- 換喩

例：「漱石を読む」

解釈：「漱石の小説を読む」

- 換喩解析 — 計算機により自動で換喩の解釈を求めること

## 換喩における関係とその例

参考文献：山梨 正明，比喩と理解，東京大学出版会，1988

換喩の関係	換喩の例	解釈
作者と作品	<u>漱石</u> を読む	漱石の小説
メーカーと製品	<u>フェラーリ</u> に乗る	フェラーリの車
容器と中身	<u>鍋</u> を食べる	鍋料理
主体と付属物	<u>詰め襟</u> が歩いて来る	詰め襟の学生
主体と手段	<u>白バイ</u> に捕まる	白バイ警官
主体と近接物	<u>ハムサンド</u> が勘定を払う	ハムサンドの客
地名と産地	<u>大島</u> を仕立てる	大島紬

## 種々の換喩解析の方法

- 人手で作成された換喩に関する特殊な知識に基づいて解析する方法 (Inverson 92, Fass 88)  
「漱石を読む」, 「作家 - 作品」
- コーパス中の言語表現を知識源とし, それを利用して換喩の解析を行なう (村田 2000, ここで解説する)  
換喩に関する知識をあらかじめ用意する必要がない  
「漱石を読む」, 「漱石の小説」

## コーパス中の言語表現を知識源とした換喩解析

利用する言語表現は以下の二種類

- 「名詞 A の名詞 B」

例：「漱石の小説」

換喩の解釈：「漱石を読む」 ⇒ 「漱石の小説を読む」

- 「名詞 A 名詞 B」

例：「鍋料理」

換喩の解釈：「鍋を食べる」 ⇒ 「鍋料理を食べる」

## 言語表現を知識源とした換喩解析の手順

### 1. 換喩になっている語の特定

換喩の入力文を格フレームと照合し，換喩になっている語を特定する．

### 2. 解釈候補の抽出

換喩になっている語(名詞A)を含む「名詞Aの名詞B」「名詞A名詞B」の言語表現を集めて，名詞Bを解釈候補として抽出する．

### 3. 解釈候補の絞り込み

候補の中から格フレーム，動詞との関連，頻度を考慮して換喩の解釈とする語を一つに決定する．

## 換喩になっている語の特定

1. 入力文と格フレームを対応づける．複数の格フレームを持つ動詞の場合は入力文と格フレームで格の種類がよく一致するものを用いる．
2. 格フレームの選択制限を満足しない格を換喩になっている語と特定する．

入力：「僕が漱石を読む」

格フレーム：{人}が {本, 新聞, 小説, 作品} を読む



## 解釈候補の抽出

- 換喩になっている語(名詞A)を含む、「名詞Aの名詞B」「名詞A名詞B」の言語表現を集めて、名詞Bを解釈候補として抽出する。

入力：

「僕が漱石を読む」

換喩になっている語(名詞A)：

「漱石」

「名詞Aの名詞B」「名詞A名詞B」の言語表現：

「漱石の小説」「漱石の作品」「漱石の遺族」「漱石記念館」

## 解釈候補の絞り込み

- 入力 : 「僕が漱石を読む」

格フレーム : {人} が {本, 新聞, 小説, 作品} を読む

解釈候補 : 「小説」「作品」「遺族」「記念館」

格フレームの選択制限により「遺族」「記念館」の候補が取り除かれる。

頻度の計算 : 「漱石の小説」11個, 「漱石の作品」3個

頻度の大きい「小説」を解釈と決定する。

## 換喩の解析結果

換喩ではないと判定された文 6/23

換喩であると判定された文 17/23

正しく換喩を解析できた文	41% (7/17)
解析できなかった文(解釈を行なわなかった文)	41%( 7/17)
意味的制約を満足する候補が残らなかった	14%(1/7)
動詞が格フレーム辞書にない	43%(3/7)
その他の格フレーム関係の問題によるもの	43%(3/7)
誤った解釈をした文	18%( 3/17)
候補の中に正解がない	100%(3/3)

## 解析結果の例

換喩の入力文	本来の正しい解釈
換喩ではないと判定された文(6文)	
<p><u>押入れ</u>を掻き回す</p> <p><u>川</u>が氾濫する</p> <p><u>庭</u>を掃く</p> <p><u>テーブル</u>を拭く</p> <p><u>頭</u>を刈る</p> <p>指揮者が<u>クラリネット</u>を笑う</p>	<p>押入れの<u>中身</u>を掻き回す</p> <p>川の<u>水</u>が氾濫する</p> <p>庭の<u>枯葉</u>を掃く</p> <p>テーブルの<u>ほこり</u>を拭く</p> <p>頭の<u>髪</u>を刈る</p> <p>指揮者がクラリネット<u>奏者</u>を笑う</p>

## 解析結果の例

換喩の入力文	本来の正しい解釈
換喩であると判定された文(17文)	
正しく換喩を解析できた文(7文) (ここの下線部は実際にシステムが出した解釈)	
<u>三島</u> を読む <u>詰め襟</u> が歩く <u>フォード</u> に乗る 僕が <u>鍋</u> を食べる <u>仮説</u> が原理を説明する <u>白バイ</u> が違反者を逮捕する <u>平安神宮</u> が満開だ	三島の <u>小説</u> を読む 詰め襟の <u>学生</u> が歩く フォードの <u>車</u> に乗る 僕が鍋 <u>料理</u> を食べる 仮説の <u>科学者</u> が原理を説明する 白バイ <u>警官</u> が違反者を逮捕する 平安神宮の <u>桜</u> が満開だ

## 解析結果の例

換喩の入力文	本来の正しい解釈
換喩であると判定された文(17文)	
換喩と認定するだけで解釈を行なわなかった文(7文)	
・ 格フレームを満足する候補が残らなかった(1文)	
顔を剃る	顔の鬚を剃る
・ 動詞が格フレーム辞書にない(3文)	
大阪がゲンナリしている 海が干上がる その皿を平らげた	大阪の人々がゲンナリしている 海の水が干上がる その皿の料理を平らげた
・ 格フレームの選択失敗(1文)	
湖が満ちる	湖の潮が満ちる
・ 複数の格が選択制限違反・換喩部分の特定の失敗(1文)	
理論がXと主張する	理論の提案者がXと主張する
・ 適合する格情報が格フレームにない(1文)	
私はどんぶりが好きだ	私はどんぶりの料理が好きだ

## 解析結果の例

換喩の入力文	本来の正しい解釈
換喩であると判定された文(17文)	
誤った解釈をした文(3文) (誤った解釈は矢印の後ろに記述)	
<ul style="list-style-type: none"> <li>意味制約を満足する候補の中に正解がない(3文)</li> </ul>	
<u>一升瓶</u> を飲む	一升瓶の <u>お酒</u> (⇒ラベル)を飲む
<u>霞が関</u> が動揺する	霞が関の <u>官僚</u> (⇒社会)が動揺する
<u>アデランス</u> が歩く	アデランスの(をしている) <u>人</u> (⇒社長)が歩く

## まとめ

- 「名詞 A の名詞 B」「名詞 A 名詞 B」の形式の言語表現を，換喩の解釈のための知識として利用．  
換喩に関する知識をあらかじめ用意する必要がない．
- 23 個の典型的な換喩の例文を入力として実験した結果，そのうち 17 個の換喩の文をシステムは換喩であると判定でき，さらにそのうち 7 個の換喩の文に対して正しい換喩の解釈を推定することができた．



## テーマ

- 換喩解析
- 意味ソート
- 統語構造認識と  $7 \pm 2$
- diff と言語処理

## 意味ソート

### 普通のソート

50音順，EUC漢字コード順

### 意味ソート

既存のシソーラスに基づき，意味の順番にソート

- 順番に意味がある．
- いろいろなときに便利．
  - － 辞書の作成
  - － コーパスの作成
  - － 情報検索

例 (行事の前に「Aの」という形でつく名詞群)

## 50音順のソート

行事 家元 祝い 王室 学校 官民 県 公式 皇室 恒例  
就任 全国 ソ連 寺 日本 農村 母校 祭り

## 意味ソート

行事 (人間) 皇室 王室 官民 家元  
(組織) 全国 農村 県 日本 ソ連 寺 学校 母校  
(活動) 祝い 恒例 公式 就任 祭り

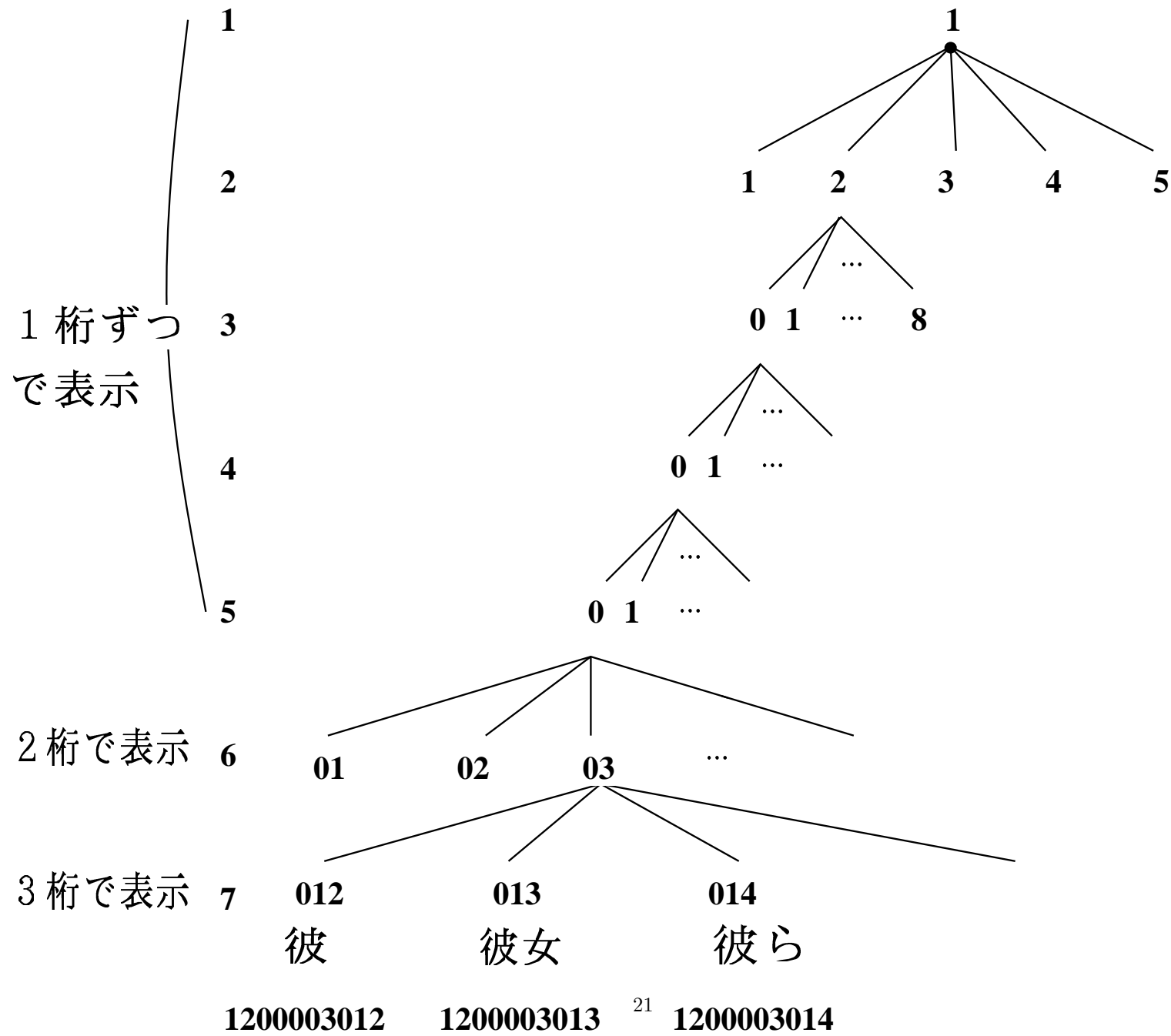
# 意味ソートの仕方 (1)

## 最も簡単な意味ソートの仕方

分類語彙表の分類番号を用いてその番号の順に並べる

1263005022	寺	1253007012	全国
1263005021	寺	1254006033	農村
1308207012	公式	1255004017	県
1311509016	公式	1263010012	学校
3101011014	公式	1336002012	祭り
3360004013	公式	1241023012	家元
1263013015	母校	1308205021	恒例
1331201016	就任	1231002013	官民
1210007021	皇室	1241101012	官民
1259001012	日本	1304308013	祝い
1259004192	ソ連	1336019012	祝い
1198007013	全国	1210007022	王室

レベル



## 意味ソートの仕方 (2)

### 分類番号の変更と意味素性の付与

分類語彙表の分類番号を意味素性 (IPAL 動詞辞書) の順番にふりかえる。

意味素性	分類語彙表の 分類番号	変換後の 分類番号
ANI(動物)	[1-3]56	511
HUM(人間)	12[0-4]	52[0-4]
ORG(組織・機関)	[1-3]2[5-8]	53[5-8]
PRO(生産物・道具)	[1-3]4[0-9]	61[0-9]
PAR(動物の部分)	[1-3]57	621
PLA(植物)	[1-3]55	631
NAT(自然物)	[1-3]52	641
LOC(空間・方角)	[1-3]17	657
QUA(数量)	[1-3]19	711
TIM(時間)	[1-3]16	811
PHE(現象名詞)	[1-3]5[01]	91[12]
ABS(抽象関係)	[1-3]1[0-58]	aa[0-58]
ACT(人間活動)	[1-3]58,[1-3]3[0-8]	ab[0-9]
OTH(その他)	4	d

## 意味ソートの仕方 (3)

### 分類番号の変更

5363005022	寺
5363005021	寺
ab18207012	公式
ab21509016	公式
aa11011014	公式
ab70004013	公式
5363013015	母校
ab41201016	就任
5210007021	皇室
5359001012	日本
5359004192	ソ連
7118007013	全国

5353007012	全国
5354006033	農村
5355004017	県
5363010012	学校
ab46002012	祭り
5241023012	家元
ab18205021	恒例
5231002013	官民
5241101012	官民
ab14308013	祝い
ab46019012	祝い
5210007022	王室

## 意味ソートの仕方 (4)

### 分類番号つきの意味素性の追加

5100000000	(動物)
5200000000	(人間)
5300000000	(組織・機関)
6100000000	(生産物・道具)
6200000000	(動物の部分)
6300000000	(植物)
6400000000	(自然物)
6500000000	(空間・方角)
7100000000	(数量)
8100000000	(時間)
9100000000	(現象名詞)
aa00000000	(抽象関係)
ab00000000	(人間活動)
d000000000	(その他)
5363005022	寺
5363005021	寺
ab18207012	公式
ab21509016	公式
aa11011014	公式

ab70004013	公式
5363013015	母校
ab41201016	就任
5210007021	皇室
5359001012	日本
5359004192	ソ連
7118007013	全国
5353007012	全国
5354006033	農村
5355004017	県
5363010012	学校
ab46002012	祭り
5241023012	家元
ab18205021	恒例
5231002013	官民
5241101012	官民
ab14308013	祝い
ab46019012	祝い
5210007022	王室



# 意味ソートの仕方 (5)

## 分類番号の値でソート

5100000000	(動物)
5200000000	(人間)
5210007021	皇室
5210007022	王室
5231002013	官民
5241023012	家元
5241101012	官民
5300000000	(組織)
5353007012	全国
5354006033	農村
5355004017	県
5359001012	日本
5359004192	ソ連
5363005021	寺
5363005022	寺
5363010012	学校
5363013015	母校
6100000000	(生産物)
6200000000	(動物の部分)

6300000000	(植物)
6400000000	(自然物)
6500000000	(空間・方角)
7100000000	(数量)
7118007013	全国
8100000000	(時間)
9100000000	(現象名詞)
aa00000000	(抽象関係)
aa11011014	公式
ab00000000	(人間活動)
ab14308013	祝い
ab18205021	恒例
ab18207012	公式
ab21509016	公式
ab41201016	就任
ab46002012	祭り
ab46019012	祝い
ab70004013	公式
d000000000	(その他)

## 意味ソート

### ソート後の名詞集合の整形

(人間) 皇室 王室 官民 家元

(組織) 全国 農村 県 日本 ソ連 寺 学校 学園 母校

(数量) 全国

(関係) 公式

(活動) 祝い 恒例 公式 就任 祭り

## 意味ソートの三つの利用例

### 1. 辞書の作成 : 「食べる」の格フレームの作成例

#### (a) ガ格の意味ソート結果

(動物) 牛 子牛 魚

(人間) わたしたち みんな 自分 乳幼児 親 妹 お客 日本人  
看護婦 作家

# 意味ソートの三つの利用例

## 1. 辞書の作成：「食べる」の格フレームの作成例

### (b) ヲ格の意味ソート結果

(動物)	動物 貝 プラントン
(生産物)	獲物 製品 材料 ペンキ 食べ物 えさ 和食 日本食 洋食 中華料理 おむすび 粥 すし ラーメン マカ ロニ サンドイッチ ピザ ステーキ バーベキュー てんぷら 空揚げ 穀物 米 白米 日本米 押し麦 キ ムチ カルビ 砂糖 ジャム 菓子 ケーキ ビスケット クッキー アイスクリーム
(体部)	遺骸 人肉 肝臓
(植物)	遺伝子 植物 牧草 ピーマン チコリ 桑 バナナ 松 茸 昆布
(現象)	珍味 雪
(関係)	中身
(活動)	朝食 昼飯 夕食 夕御飯 おやつ 塩焼き

## 意味ソートの三つの利用例

### 1. 辞書の作成：「食べる」の格フレームの作成例

#### (c) デ格の意味ソート結果

(人間)	自分
(組織)	事務所 レストラン ホテル
(生産物)	しょうゆ シャトー 楽屋 便所 荷台 食卓
(空間)	現地 全域 車内
(数量)	ふたり 割合 複数
(活動)	研究 会議

# 意味ソートの三つの利用例

## 2. タグつきコーパスの作成 : 「AのB」の意味関係の付与例

名詞A	名詞B	意味関係
パナマ	事件	場所
中学校	事件	場所
軍	事件	場所
アルバム	事件	間接限定
タンカー	事件	間接限定
最悪	事件	形的特徴
最大	事件	形的特徴
周辺	物件	場所
両国	事項	主体対象
文献	事項	分野限定
総会	事項	主体対象
上院	条項	分野限定, 主体対象
新法	条項	分野限定, 全体部分
条約	条項	分野限定, 全体部分
協定	条項	分野限定

## 意味ソートの三つの利用例

### 3. 情報検索での利用：キーワードによる情報提示例

#### 文書データベースAのキーワード群

検索 単語 文書 作成 候補 質問 数 キーワード 情報

#### 意味ソートによる情報提示

(数量)	数
(抽象関係)	候補
(人間活動)	検索
	文書 キーワード 単語 情報 質問
	作成

おわりに

意味ソート — 意味でソートする

- 意味ソートはいろいろなきに便利
- 意味ソートの利用による各研究の効率化を期待

その他のトピック

- EDR辞書での意味ソートの実現方法
- 複数の属性で単語を表現する辞書での意味ソートの実現方法



## テーマ

- 換喩解析
- 意味ソート
- 統語構造認識と  $7 \pm 2$
- diff と言語処理

## 統語構造認識と $7 \pm 2$

### 人間の短期記憶の容量

- George A. Miller –  $7 \pm 2$  程度のチャンク

### 本研究

- 文の理解過程 — 係り受け解析
- 係り先の未決定な文節  $\Rightarrow$  短期記憶

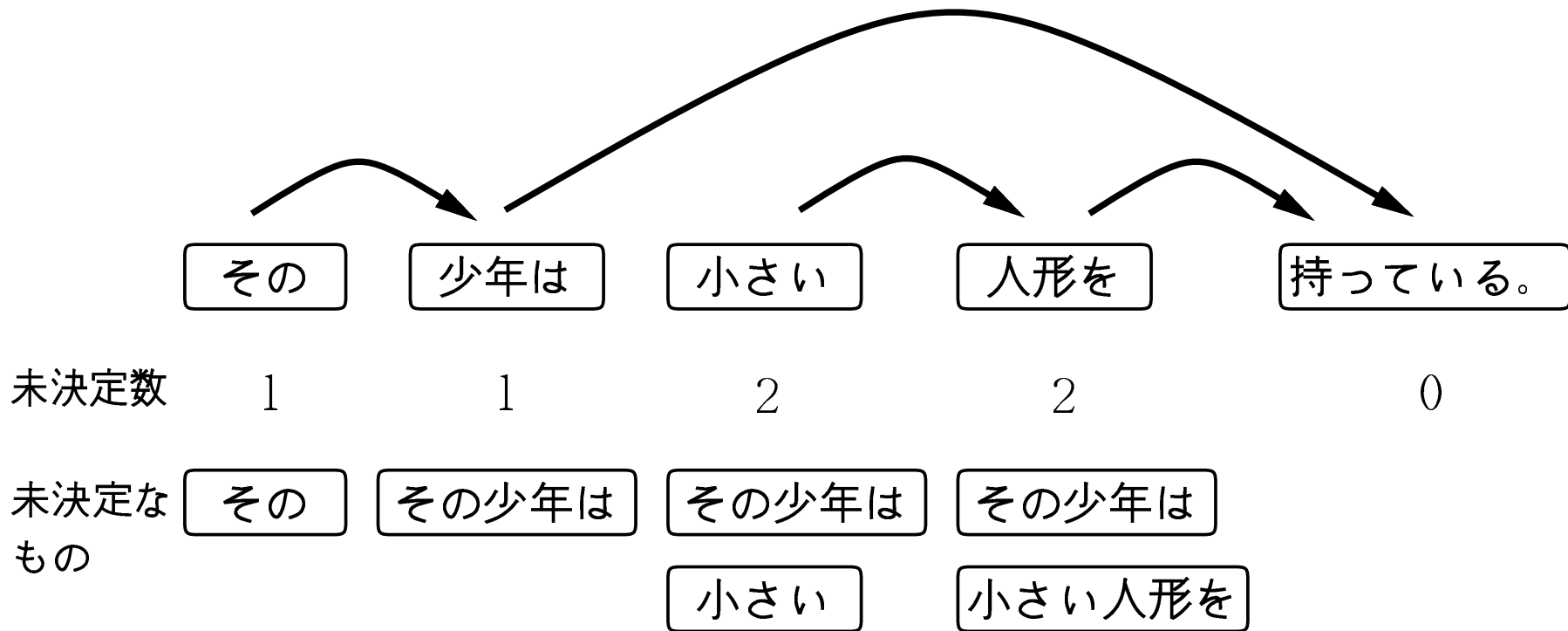
## 短期記憶と $7 \pm 2$

短期記憶の容量を計る，言葉，音感，味覚，視覚などを対象とした種々の実験結果が概ね  $7 \pm 2$

### $7 \pm 2$ の学問の応用

- カーナビなどで，一画面に表示する情報の制限
- 文のわかりやすさ，わかりにくさの指標

# 日本語文での調査報告(京大コーパス2万文) 係り先が未決定な文節の数を数える



# 日本語文での調査報告

未決定文節数	頻度	
	文節数	文数
0	19954	90
1	52751	1352
2	59494	5022
3	38465	6823
4	15802	4468
5	4488	1593
6	1143	480
7	195	102
8	47	17
9	10	5
10	3	2

# 日本語文での調査報告

## 未決定文節数が10の文

国はその後、このうち二十三点の公開は「やむを得ない」と認めたものの、主に電子機器などを置いてある地下部分の資料二十一点については「ASWOCはシーレーン防衛のための中枢基地であり、公開されると国防や警備上、重大な支障が生じる」などと主張、決定の取り消しなどを要求して争ってきた。

## おわりに

- George A. Miller –  $7 \pm 2$  程度のチャンク
- 日本語で文節を情報の認知単位と仮定，  
調査結果は Miller の  $7 \pm 2$  をほぼ満足する．
- 日本語で文節を認知単位と仮定すれば，  
Miller の  $7 \pm 2$  の理論と，言語解析・生成において短期記憶するものは  $7 \pm 2$  程度ですむという Yngve の主張を整合性よく説明することができる．
- 工学的立場からの効果  
言語解析・生成において短期記憶するものは  $7 \pm 2$  程度ですむことになる

## テーマ

- 換喩解析
- 意味ソート
- 統語構造認識と  $7 \pm 2$
- diff と言語処理



## diff と言語処理

- 差分検出を行なう diff コマンドは言語処理の研究において役に立つ場面が数多く存在する。
- 簡単に diff の説明を行ない、その後、diff を使った言語処理研究の具体的事例として、差分検出、書き換え規則の獲得、データのマージ、最適照合の例を示す。

## diff

- diff とはUNIX のファイル比較ツール diff のことである .
- このコマンドは , 与えられた二つのファイルの差分を順序情報を保持したまま行を単位として出力する .
- 実行例

ファイル1

```
今日  
学校へ  
いく
```

ファイル2

```
今日  
大学へ  
いく
```

diff の結果

```
< 学校へ  
> 大学へ
```

# mdiff

- -D オプションをつけて diff を走らせると，不一致部分だけでなく，一致部分も出力する．
- mdiff の m — merge の m
- 実行例

ファイル1

```
今日  
学校へ  
いく
```

ファイル2

```
今日  
大学へ  
いく
```

mdiff の結果

```
今日  
;  
学校へ  
;  
大学へ  
;  
いく
```

## 差分検出と書き換え規則の獲得

- 複数システムの出力の差分検出
- 差分の考察と書き換え規則の獲得

# 複数システムの出力の差分検出

- 複数システムの出力を融合する
- 実行例

## システム1

とと 助詞  
いった 言う 動詞  
こと こと 名詞

## システム2

とと 助詞  
いった 行く 動詞  
こと こと 名詞

## mdiff の結果

とと 助詞  
;  
いった 言う 動詞  
;  
いった 行く 動詞  
;  
こと こと 名詞

## 差分の考察と書き換え規則の獲得

- 話し言葉と書き言葉の diff を考える .
- 話し言葉 — 学会口頭発表の書き起こし
- 書き言葉 — 対応する論文のデータ

# 書き言葉と話し言葉のデータ

## 書き言葉データ

本  
論文  
で  
は  
意味  
ソート  
に  
ついて  
述べる  
。  
一般に  
ソート  
は  
50  
音  
順

## 話し言葉データ

今日  
は  
え  
意味  
ソート  
に  
ついて  
述べ  
ます  
一般に  
ソート  
って  
いう  
の  
は  
だいたい

## 書き言葉と話し言葉の diff の結果

；  
本  
論文  
で  
；  
今日  
；  
は  
；  
；  
え  
；  
意味  
ソート  
に  
(右欄につづく)

ついて  
；  
述べる  
。  
；  
述べ  
ます  
；  
一般に  
ソート  
；  
；  
って  
いう  
の  
；



## 差分部分の抽出と考察

書き言葉データ	話し言葉データ
本論文で  述べる。	今日 え 述べます っていうの

- 書き言葉と話し言葉の違いを考察できる。
- 書き言葉と話し言葉の変換規則が獲得できる。  
(例: 「述べる。」 ⇒ 「述べます」)
- 他の種類の差分検出にも使える。  
(例: 要約, 英文校閲など)

## 話し言葉と書き言葉からの差分表現の抽出

### ● データ

- 書き言葉データ — 論文データ (打ち込み, 82編, 352,660文字)
- 話し言葉データ — 開放融合のコーパスのうち, 上の論文データに対応するもの (330,679文字)

### ● 方法

- mdiffで差分表現を抽出 . 頻度の高い差分表現を考察する .

## 話し言葉と書き言葉から抽出された差分部分の分類

1. 表記のゆれ
2. 表記・読みを与えるもの
3. 同義関係のもの
4. 口語調のもの
5. 省略をしているもの
6. 補完をしているもの
7. コーパスの誤り検出に関わるもの

## 1 . 表記のゆれ

書き言葉	話し言葉
データ	データー
クラスタ	クラスター
パラメータ	パラメーター
モダリティ	モダリティー
データベース	データーベース
係る	掛かる
超える	越える
すべて	全て
ため	為
行う	行なう
言い換え	言い替え

## 2 . 表記・読みを与えるもの

書き言葉	話し言葉
=	は
2	二
零	ゼロ
g r a m	グラム
s	秒
S	S
H e b b	ヘブ

### 3 . 同義関係のもの

書き言葉	話し言葉
および	と
や	とか
論文	研究
・異なり	異なりで
各	それぞれ
i 番目のターム	ターム I
述語	動詞
識別	認識
異なれば、	違えば

## 4 . 口語調のもの

書き言葉	話し言葉
した。	という
、	いたしました
	ですね
	です
られる。	られます
	っていう
や	とか
	こう
	非常にこう
いる。	います
分かった。	分かりました
ない。	ません
。	訳ですが
	っていうの
れた。	れるんですが
であり、	であって
ことである。	訳ですけども

## 5 . 省略をしているもの

前方一致部分	書き言葉	話し言葉	後方一致部分
スムーズニング 各 スポーツニュース 平均時間が	処理 $C(V)_{\{k\}}$ における 11.25	の 11.3	を 素片 会話部分を 分まで



## 6 . 補完をしているもの

前方一致部分	書き言葉	話し言葉	後方一致部分
に対する損失の 会話に 音圧レベル	7 0 d B	値の 関しましては全然 七十五五デシベル	平均として 不便はない で提示

## 7. 誤り検出に関するもの

- 論文の誤植 (打ち込みエラーか?)

前方一致部分	書き言葉	話し言葉	後方一致部分
ニュース 日本語が 日本語複合	速記 述べ 名刺	速報 延べ 名詞構造	記事 178091語 解析法

- 書き起こしの誤り

前方一致部分	書き言葉	話し言葉	後方一致部分
	表層的な	表装的な	

- どちらかが誤り

前方一致部分	書き言葉	話し言葉	後方一致部分
マイクロホン 社会の	及び生体 死活	および声帯 生活	アンプ にかかわる問題

## データのマージ , および , 最適照合

diff では一致部分を最大にするように最適照合が実現される .

- 講演と予稿の対応づけ

## 講演と予稿の対応づけ

- 講演 — 口頭発表  
予稿 — 論文

- 講演と予稿が与えられたとき，講演の各部分と，予稿の各部分の対応がとれると，講演を聞いている時だと，それに対応する予稿の部分を参照できるし，予稿を読んでいるときだと，それに対応する講演の部分を参照できて便利である．
- ここでは特に予稿の各章が講演のどこの部分に対応するかを `mdiff` でもとめることにする．

## 予稿データの構成

```
<Chapter 1>  
(1章の内容)  
</Chapter 1>  
<Chapter 2>  
(2章の内容)  
</Chapter 2>  
<Chapter 3>  
(3章の内容)  
</Chapter 3>
```

## 予稿と講演の mdiff の結果

```
;  
<Chapter 1>  
(予稿のみの内容)  
;  
(講演のみの内容)  
;  
(共通する内容)  
;  
(予稿のみの内容)  
;  
(講演のみの内容)  
;  
(共通する内容)  
;  
</Chapter 1>  
<Chapter 2>  
(予稿のみの内容)  
;  
(講演のみの内容)  
;
```

## 講演データへの章の情報の挿入結果

```
<Chapter 1>  
(講演のみの内容)  
(共通する内容)  
(講演のみの内容)  
</Chapter 1>  
<Chapter 2>  
(講演のみの内容)
```

## 本講義の元となった論文

- 村田真樹，山本 専，黒橋禎夫，井佐原 均，長尾真．  
名詞句「AのB」「AB」の用例を利用した換喩解析  
人工知能学会誌  
2000年 5月，15巻，3号，p.503 ~ p.510
- 村田 真樹，神崎 享子，内元 清貴，馬 青，井佐原 均  
意味ソートmsort — 意味的並べかえ手法による辞書の構築例とタグつき  
コーパスの作成例と情報提示システム例 —  
自然言語処理 (言語処理学会誌)  
2000年 1月，7巻，1号，p.51 ~ p.66
- 村田 真樹，内元 清貴，馬 青，井佐原 均  
日本語文と英語文における統語構造認識とマジカルナンバー  $7 \pm 2$   
自然言語処理 (言語処理学会誌)  
1999年 10月，6巻，7号，p.61 ~ p.73
- 村田 真樹  
diff を用いた言語処理 — 便利な差分検出ツール mdiff の利用 —  
自然言語処理 (言語処理学会誌)  
2002年 4月，9巻，2号，p.91 ~ p.110