

用言等換言辞書を用いた 換言結果の考察

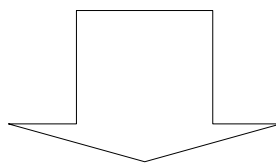
吉倉 孝太郎 山本 和英
長岡技術科学大学

研究の背景

1

換言処理の進歩は自然言語処理の進歩に寄与

現在は言語資源として辞書・シソーラスを主に利用
- これらは換言のために構築されたものではない



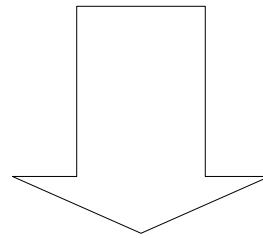
換言のための言語資源を構築したい

本発表の概要

2

本発表では

- 換言のための汎用性の高い言語資源を構築したい
- 先行研究として「**用言等換言辞書**」を構築



実際に新聞を対象に換言

用言等換言辞書

「用言等換言辞書を人手で構築しました」

- 山本、吉倉 [2013]

形態素解析器JUMANの形態素辞書を対象

人の感覚を直接反映した換言資源

用言等換言辞書の構築手法

4

JUMAN中の動詞・サ変名詞・形容詞・副詞を対象

例：動詞「和える」

①動詞を使った短文を人手で考える

- ごまと和える

②対象の動詞を簡単な語に置き換えて換言する

- ごまと混ぜる

和える-混ぜるの対を獲得

用言等換言辞書の構築手法

5

多義性のある語は複数の換言先を持つ

例: 仰ぐ

- 空を仰ぐ 空を見る
- 師を仰ぐ 師を尊敬する
- 指示を仰ぐ 指示を求める

仰ぐ-見る
仰ぐ-尊敬する の対を獲得
仰ぐ-求める

用言等換言辞書の構築手法

6

- その他条件
 - 辞書やシソーラスを参考にしない
 - 子どもや外国人に説明するつもりで換言
 - より一般的な語へと換言される
 - 換言しない語がある
 - 換言者が分からない場合(例: **かがる**、**笑み割れる**)
 - 適切な換言を思いつかない場合(例: **歌う**、**言う**)

用言等換言辞書の語数

7

品詞	換言対象語	換言作成	無記入
動詞	3,608語	3,206語	481語
サ変名詞	5,627語	4,496語	1,141語
形容詞	2,335語	1,851語	496語
副詞	1,243語	785語	463語
合計	12,813語	10,336語	2,585語

換言実験の条件

換言可能な「動詞」を1語のみ換言

多義性の解消・活用の変化・評価は人手で行う

形態素解析誤りは評価対象外

実際の換言

9

“二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」と意気込んでいる。

用言等換言辞書

履く-身につける

起こす-換言候補なし

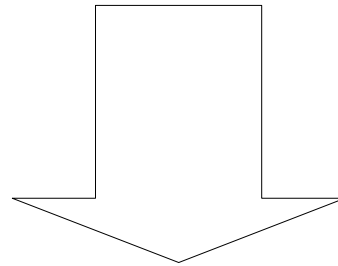
意気込む-やる気になる

履く、意気込むのどちらかのみ換言

換言手法

10

“二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」と意気込んでいる。

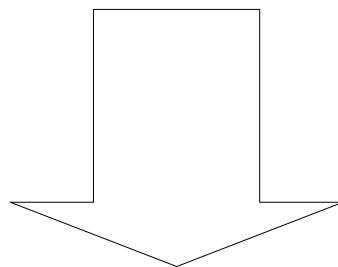


“二足のわらじ”を身につけて「関西にネットベンチャーの波を起こす」と意気込んでいる。

換言手法

11

“二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」と**意気込**んでいる。



“二足のわらじ”を履いて「関西にネットベンチャーの波を起こす」と**やる気**になっている。

評価手法

評価基準は4通り

換言可能

条件付換言可能

不自然な換言

換言不可能

換言可能

13

意味を等価に保つ換言

選挙で2大政党の候補を破ってミネソタ州知事に初
当選し..

破る-引き裂く、負かす

換言可能

13

意味を等価に保つ換言

選挙で2大政党の候補を破ってミネソタ州知事に初
当選し..

破る-引き裂く、負かす

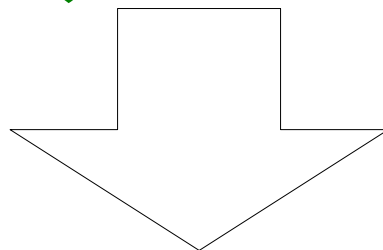
換言可能

13

意味を等価に保つ換言

選挙で2大政党の候補を破ってミネソタ州知事に初当選し..

破る-引き裂く、負かす



選挙で2大政党の候補を負かしてミネソタ州知事に初当選し..

評価基準-条件付換言可能

14

助詞、形式名詞を補正することで換言可能である

退任するのに**伴い**、参院自民党は..

伴う-**一緒だ**



評価基準-条件付換言可能

14

助詞、形式名詞を補正することで換言可能である

退任するのに**伴い**、参院自民党は..

伴う-**一緒だ**



退任するのに**一緒に**、参院自民党は..

評価基準-条件付換言可能

14

助詞、形式名詞を補正することで換言可能である

退任するのに**伴い**、参院自民党は..

伴う-**一緒だ**



退任するの**に****一緒**に、参院自民党は..

助詞が不自然である

評価基準-条件付換言可能

14

助詞、形式名詞を補正することで換言可能である

退任するのに**伴い**、参院自民党は..

伴う-**一緒だ**



退任するの**に****一緒**に、参院自民党は..



退任するの**と****一緒**に、参院自民党は..

評価基準-不自然な換言

15

文の意図はわかるが表現が不自然である

新たな仕事に一步踏み出した

踏み出す-はじめる



評価基準-不自然な換言

15

文の意図はわかるが表現が不自然である

新たな仕事に一步踏み出した

踏み出す-はじめる



新たな仕事に一步はじめた

- ※助詞や形式名詞の補正では換言可能でない

評価基準-換言不可能

16

文の意味が等価でなくなる

フセイン大統領が絡む「政治事件」..

絡む-巻きつく

評価基準-換言不可能

16

文の意味が等価でなくなる

フセイン大統領が絡む「政治事件」..

絡む-巻きつく



フセイン大統領が巻きつく「政治事件」

評価データ

- 構築したデータ
 - 毎日新聞99年版から無作為に6000文を抽出
 - 換言を1語のみ行うように8646文を構築
- 練習セット
 - 評価者間による差異を減らすため1000文を練習セットとして利用

評価データ

18

- 評価データ

練習セットを除く7646文から同じ換言対象語が
最大10回となるよう制限した**3500文**

述べる-言う(147文)などのように特定の語が何回
も同じ用法で出現するため

評価データ

19

- 評価者
 - 発表者を含む日本語を母語とする成人男性2名

- 評価者間の評価の一致
 - 構築した3500文の評価データから212文は2名で評価

評価者間の一致度

2名で評価した**212文**中の評価の一致度

- 212文中18文は一方でも評価しないとした文

	換言対象語数	評価一致数	一致度
文数	194[文]	166[文]	85.6[%]

- 二名間の評価の差異は大きくない

一致しなかった換言

21

- どのような換言が評価者間で一致しなかったか
 - 電話の向こうから喜びがひしひしと**伝わってきた**
伝わる-渡る
 - 電話の向こうから喜びがひしひしと**渡ってきた**
→換言不可能と不自然な換言で不一致
 - 「喜び」という感情が「渡る」は許容できるか否かで相違

目的語と換言元と換言先の関係による不一致

一致しなかった換言

22

- どのような換言が評価者間で一致しなかったか
 - 1機の価格が200億円を**超える**ため、代わりに**超える-越える**
 - 1機の価格が200億円を**越える**ため、代わりに
→換言可能と換言不可能で不一致
 - それぞれの語の細かいニュアンスによる差異

語への感覚が違うことによる不一致

換言の評価の分布

23

評価の種類	評価数	割合[%]
換言可能	1671	53.9
条件付換言可能	187	6.0
不自然な換言	402	13.0
換言不可能	840	27.1
全数	3100	
(評価しない)	(188)	

→**6割程度**は現状の用言等換言辞書を用いて自然な換言を構成することができる

換言不可能語を減らすために

24

- 換言不可能とされた語はどのような語か
 - 適切な換言候補がない
 - 分割すべきでない換言対象語
 - 複合動詞の一部
 - 慣用表現の一部

適切な換言候補を持たない例

25

フセイン大統領が絡む「政治事件」..

- 絡む-まきつく

..は拘束を解かれ..

- 解く-答える

適切な換言候補を換言辞書に追加することで対応可

複合動詞の一部を換言

- 複合動詞の一部を換言する
 - 時間が**過ぎ**去る
 - **過ぎる**-**経つ**、**通る**
 - 子どもを連れ**去**る
 - **去る**-**出る**
- ただし、**換言可能な例も存在**
- 花が**咲**き始める
 - **咲く**-**開く**

慣用表現の一部を換言

27

不要品を**引き取る** 換言可能

- **引き取る**-もらう

祖父が息を**引き取った** 換言不可能

- **引き取る**-もらう

→「**息を引き取る**」という句で「**亡くなる**」と換言可能

分割すべきでない換言対象語

28

- 複合動詞の一部
 - 過ぎ去る、連れ去るなど
- 慣用表現の一部
 - 息を引き取る、的を絞るなど

これらは分割して換言を扱うことが問題

句を換言対象として換言を行うべきである

今後の展望

- 用言等換言辞書の拡充
 - 適切な換言候補語の追加
 - 複数語からなる句をまとめて辞書に追加
- 動詞以外の換言の特徴の調査

まとめ

目的: 汎用性の高い換言処理用資源の構築

用言等換言辞書の動詞を用いて換言を実施

6割の語は現状で換言可能であるという結果

今後は不可分な句の換言を追加

ご清聴ありがとうございました。