

用言等換言辞書を 人手で作りました

山本和英 吉倉孝太郎

長岡技術科学大学

発表内容:

1. 換言処理の話
2. 作業内容
3. 結果と今後

1. 換言処理の話

研究の動機:

換言処理したいのに、
一般に利用可能な
汎用日本語換言辞書が
存在しない

なぜ換言処理？

表現の多様性は
自然言語の特質である。
これを正面から扱わない限り、
自然言語処理は進展しない。

理想

表層表現



理解

いおゆる意味表現



以後の
処理へ

現実

表層表現



理解



いおゆる意味表現

現実

表層表現



理解

いわゆる意

これで
いいのか？

提案

表層表現



換言処理

多様性を吸収した
言語表現



理解

いおゆる意味表現

表層表現



換言処理

多様性を吸収した
言語表現



理

いわゆる意

性能向上
するのでは？

主張: 換言処理は、

- 自然言語処理の性能向上に大いに寄与する。
- 表現平易化(やさしい日本語)や表現統一(産業日本語)などは社会での需要が大きい。

よって、換言処理の進展がどうしても必要だ。

2. 作業内容

作業の概要

- 廿変名詞、動詞、形容詞、副詞を対象
- JUMAN辞書項目が換言対象
 - 合計約12,000語
- 1名が手作業で換言
- 短く置換できる語を思い出す
- 平易に換言する(難しく換言しない)

言語資源を参照しない

作成時に国語辞典やシソーラスを一切見ないで作業を行う。直感勝負。

- 作業効率の向上
- 既存の言語資源と同じ情報がほしいのではない

無記入を許す

- 単語の意味が分からない
 - 笑み割れる、かかる
- 換言できる語を思いつかない
 - 投げる、言う、話す
- 換言可能でも長い説明を要する
 - 一進一退する

以上のいずれかに該当する場合には無記入を許す。

多義語は文脈を付与する

- 空を仰ぐ ⇒ 見る
- 師匠を仰ぐ ⇒ 尊敬する
- 指示を仰ぐ ⇒ 求める

多義語の場合は換言語と共に文脈(赤字)を記述する。

格変化は記述する

- 背く ⇒ 裏切る
- 上司に背く ⇒ 上司を裏切る

換言によって格が変化する場合は、
格の変化も記述する。

慣用的表現は まとめて換言する

- 入れる ⇒ ?
- 手に入れる ⇒ 得る

単語単位で換言できない(広い意味の)
慣用的表現はまとめて換言する。

意味の重複可能性も 付記する

- 片付ける ⇒ きれいにする
- 部屋を片付ける ⇒ 部屋をきれいにする

意味の重複可能性も 付記する

- 片付ける ⇒ きれいにする
- 部屋を片付ける ⇒ 部屋をきれいにする
- 部屋をきれいに片付ける
⇒ 部屋をきれいにきれいにする

意味が重複して表出する可能性がある
表現は括弧づけする。

3. 結果と今後

作業結果

品詞	換言対象	換言作成	無記入
動詞	3,608語	3,206語	481語
名詞	5,627語	4,494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

作業期間: 5ヶ月程度

辞書の分析

これからです。

- 国語辞典・シソーラスとの関係
- 日本語の基礎語彙との関係

辞書の利用

これからです。

- 人に向けた換言
(表現の平易化、規格化など)
- 計算機に向けた換言
(自然言語処理の性能向上)

名詞は換言しないのか？

- 名詞も換言辞書を作りたいが、用言等ほど換言は簡単ではない。
- 検討中。

発表まとめ

用言等換言辞書を
人手で作りました。