

自由記述答案の採点 —同意義文同定システム—

脇田早紀子

日本アイ・ビー・エム(株)東京基礎研究所

概要

国際医療福祉大学では、看護学の学習の際にパソコンを使った自学自習システムを取り入れている。そのシステムにおいて、看護学の学習のためには、選択肢による解答でなく、自由記述文による解答が不可欠という要望があった。そこで、日本語校正支援システムFleCSの「パターン照合」の仕組みを利用して、記述すべきポイントが自由記述文の中に含まれているかどうかを柔軟に判断する同意義文同定システムを作ることになった。

教育現場で用いる際、同意義文の定義が簡便に行えることが最も重要な条件であった。このために今回新しく「同義語機能」と「パターン生成機能」を付け加えたので、これらについて報告する。

Automatic assortment of answers written in Natural Language
- to find sentence which has similar meaning -

Sakiko Wakita

IBM Research, Tokyo Research Laboratory

Abstract

At International University of Health and Welfare, Self Learning System (on Personal Computer) is used in nursing class. This system is requested to process answers written in Natural Language, because it's important to describe "signs of patients" or "points of view in nursing" in their own words. So I decided to use 'pattern matching' mechanism which has used in Japanese Cretiquing System FleCS, to distinguish a sentence into a group of sentences which have certain meaning.

In this paper I describe 'words of same meaning' mechanism and 'pattern generation' mechanism which make definition of 'certain meaning' easier.

1. 背景

国際医療福祉大学では、看護学の学習の際にパソコンを使った自学自習システムを取り入れている。学生の入力をもとに飛び先ページを決定するにあたって、選択肢による指示・解答を求めるなら処理は簡単で確実だが、看護学の学習においては、患者の様態や看護の注意点などを自分の言葉で書き表すことが大切という考えが強く、当初から自由記述文で入力させるシステムを目指していた。

ただ、初めに作ったシステム（CAT-1）は、日本語処理を搭載していなかったため、文字列の完全な一致により処理しており、柔軟性に乏しかった。そこで、日本語校正支援システムFleCSの「パターン記法」を利用して、あらかじめ定義した記述すべきポイントが自由記述文の中に含まれているかどうかを判断するシステムを作ることになった。

2. 目的

すなわちこれは、自由記述文を入力とした同意義文同定システムである。教育目的なので、同意義文セットの定義は簡便に行えることを条件とする。

使用例

学生の入力

「呼吸が荒いようだ」
「息が荒かった」
「呼吸がとても苦しいらしい」
「肩で息をしている」
「息がひどく苦しそうだ」
「努力呼吸」などすべて

→「息が乱れている」という点についての記述があると判断

3. 自学自習教材の内容

3.1 前回の課題（CAT-1）

まず、一人の女性が救急で担ぎ込まれて来るところを映したビデオを見せる。その後、観察してわかった点について学生の入力を求める。看護する上で必要な点がもれなく列挙されていれば次のページへ進み、落ちている点があればビデオに戻る。

観察すべき点

「呼吸が速い」「息が乱れている」「顔が青い」

3.2 今回の課題（CAT-2）

まず、一人の男性の一日の生活（起床、満員電車で揺られて出勤、会議など）の様子を紹介するビデオを見せる。その後、学生の入力にしたがって、より詳しいデータを紹介するページに飛ぶ。入力された文に該当するページがない場合は「該当ページなし」を告げて再入力を促す。

こうして情報を収集したのち、この男性にする健康上のアドバイスを考える課題である。

飛び先ページ:情報の内容

- 2:妻の日常生活
- 3:妻から見た夫あるいは結婚生活
- 4:長男の進路
- 5:家族の関係
- 6:家（持ち家か、新築か）
- 7:結婚のなれそめ
- 8:妻から見た息子
- 9:妻から見た娘
- 10:妻の人生
- 11:老後
- 12:妻の趣味
- 13:結婚生活
- 14:長女の日常生活
- 15:長男の日常生活
- 16:食事の好み
- 17:排泄
- 18:健康意識
- 19:睡眠
- 20:家族構成
- 21:宗教
- 22:生い立ち
- 23:食習慣
- 24:生き甲斐
- 25:清潔
- 26:趣味
- 27:性格
- 28:アフター5
- 29:両親
- 30:休日
- 31:生化学検査
- 32:煙草
- 33:酒
- 34:接待
- 35:通勤時間
- 36:病歴
- 37:ストレス
- 38:兄弟
- 39:給料
- 40:近所づきあい
- 41:薬

- 43:起床から出勤まで
- 44:帰宅時間
- 45:バイタル(血圧、体温、脈拍、呼吸数)
- 46:身長
- 47:体重
- 48:レントゲン
- 49:心電図
- 50:血液検査
- 51:尿検査
- 52:人間ドック

4. 要求項目

今回開発する同意義文同定システムに求められる条件は以下の通りである。

- ・学生が多様な表現で書く文に対して、高い率で正しい飛び先ページの決定を行うこと。
- ・飛び先ページ決定のため用いる同意義文セットは、簡潔・簡便に定義できること。

以上2点である。

5. 既存の方法

5.1 文字列一致

CAT-1では、日本語処理の仕組みを搭載しておらず、考えつく表現をすべて登録して、文字列が完全に一致するものを選んでいった。

非常に原始的ではあるが、同意義文の定義方法がとてもしやすいところは利点である。単に可能な表現を列挙すればよく、その後実際に使用するとき、漏れていた表現に気付き次第追加していけば精度も向上していく。もちろん、向上するとはいっても融通がきかないので、助詞「の」を一つ挟んだだけで、あるいは過去形で表現しただけで一致しなくなるから、どこまで表記例を追加しても人手でチェックしなければならないものはかなり残ってしまう。

5.2 パターン照合

「パターン照合」は、日本語校正支援システムFleCSで用いていた仕組みである。これは文の特徴を字面・品詞その他の条件により記述する「パターン記法」[1]で書いた「パターン」に、形態素解析結果を照合することによって、その特徴をもった文かどうかを判定するものである。校正支援を目的として使用する場合、例えば、「務める」の変換ミスを探す「パターン」としてこのように記述する(注1)。

pattern: ["長"|"員"|"役"|"師"|"官"|"士"|"係"|"司会"|"理事"|"参与"|"大使"|"秘書"|"手"|"通訳"|"ター"|"教授"|"代表"|"など"]*["を"|"も"&助詞][!ANY]*["勤め"|"努め"&一段動詞]

[]で囲まれているのがユニット(文の部分:単語単位であってもなくてもよい)であり、ユニットの中に、そのユニットが満たすべき条件を記述する。「長」などは字面を表し、「助詞」などは品詞その他の条件を表す。「|」はOR、「&」はAND、「+」はユニットの一回以上の繰り返し、「*」はユニットの0回以上の繰り返しである。上記のパターンであれば、「*委員長を長年努め、」「*コーディネーターなども勤めた」などにあてはまる。

記述するものが、誤った文の特徴でなくても基本的には同じことである。例えば、「妻の日常生活」の同意義文の特徴を記述するならば、

pattern: ["妻"|"母"|"奥さん"|"奥様"|"の"|"は"&助詞][ANY]*["日常生活"|"一日"|"職業"|"専業主婦"|"仕事"|"何をし"]

のようにすればよい。

この方法では、ある程度柔軟な定義を、ある程度簡便に行うことができるが、同意義文同定という目的に使用することを前提に考えるなら、よりいっそう簡便な使い方を考えることができるだろう。

6. 今回の方針

日本語校正支援システムFleCSの仕組み(形態素解析+パターン照合)を基本に作成し、以下の二つの機能を付け加える。

- ・同意義語セットの定義、照合を簡単かつ柔軟に行う機能(同義語機能)
- ・簡単な定義から幾つかの定形的パターンを生成する機能(パターン生成機能)

7. 同義語機能

同じ人/こと/ものを指すのに多様な書き方がある。ユニットに直接書き並べるやり方でも、ある程度は対応できるが、同じ意味の言葉をまとめて指し示せるようにした方が便利である。その方が見た目もすっきりして理

解しやすいし、他のパターンでも同じ同義語セットを利用できる。

そこで、ユニットの中で使える *Dougigo* (“基本語”) という関数を用意した。この関数は、あらかじめ定義されている同義語セットに基づき「基本語」の置き換えを行い、現在位置からの文字列と一致させることができれば正、できなければ偽を返す。(注2)

同義語セットの定義方法は以下の通りである。

7.1 基本形

同意義文定義ファイルに、「#SYNONYM」で始まる行により同義語セットを定義する。同義語セット1つにつき1行で書く。

#SYNONYM 宗教 信心 信じ12 信仰

上記の定義は(半角数字は品詞の番号)、基本語の「宗教」は右側のいずれかの語に置き換え可能なことを表す。すなわち、*Dougigo* (“宗教”)は「宗教」「信心」「信仰」「信じる」「信じた」などについて正を返す。ユニットに[“宗教”|“信心”|“信じ-12”|“信仰”]と書いたのと同じ。

7.2 部分文字列の指定

#SYNONYM 妻 奥様 母 母親

#SYNONYM 夫 鈴木様 亭主

**#SYNONYM 鈴木 すずき スズキ 旦那 だんな
ダンナ**

#SYNONYM 様 さん さま 氏

7.2.1 部分文字列ごとの置き換え

同義語セットの中に「.」で部分文字列に区切られたものがある場合は、部分文字列をもとにさらに同義語セットで置き換えることができる。例えば、「妻」は「奥様」と入れ替えられるが、さらに「様」を置き換えて「奥さん」「奥様」でもよい(「奥氏」はないのだがあまり害がなさそうなので放しておく)。同様に、「夫」は「鈴木様」と入れ替えられるが、さらに「鈴木」と「様」をそれぞれ置き換えて「鈴木さん」「スズキ氏」「だんな様」などにマッチする。

7.2.2 部分文字列間に一文字挟む

さらに、「.」で区切られた部分文字列は、間に平仮名を一文字挟むことができるとする。

**#SYNONYM 食事生活 朝.食 食事 献立 食生活
いつも.食 一緒.食 食習慣**

この例でいうと、「食事生活」は「いつも食」という文字列と置き換えられて「いつも食べている」「いつも食卓に」などにあてはまる他に、「いつもの食べ方は」「いつもは食べているのか」などにあてはまる。

7.2.3 気がかりな点あれこれ

同義語とはいええないようなものにならないですが

また、「鈴木 すずき スズキ 旦那 だんな ダンナ」のようなものは普通、同義語ではない。これは、CAT-2の教材ビデオの中身の話という限定された世界(1つの「同意義文定義ファイル」中)なのでこのような同義語セットも有効なのである。

ただし、この世界の中でさらにあるパターン独自の同義語セットを定義したい場合は、パターンに直接書き込んで定義できる。

#PATTERN PATTERN-13 [夫婦]..

[生活/関係/会話/仲/間柄]

例えば夫婦の生活の様子についてのページ13は、「生活」のかわりに「夫婦の関係」「夫婦仲」「夫婦の間に会話はあるか」などでもよいが、そのほかのページ(2,14,15,22など)でいう生活は「間柄」などに置き換えられない。

複数の同義語セットが共通の語を含んでいるとどうなるでしょう

共通した語を持つ同義語セットでも、それぞれ独立に使えるように、必ず左端の語を右側の語いづれかに置き換える方向で使われ、逆は意味しないことにしている。この限定を付けておかないと、うっかりとんでもないものに置き換えられている事故が起こりやすい。

8. パターン生成機能

同意義文を探すにあたって必要な「パターン」をいくつかの型に分類し、「パターン記法」類似の単純な書き方の定義から自動で実際の「パターン」を展開するようにする。

以下、「同義語機能」によって同義と判定される語のグループを「語」と呼ぶ。

8.1 一つの「語」が出てくればよいもの例:ページ21の「宗教」、ページ32の「煙草」など

この場合、「同意義文定義ファイル」には

#PATTERN PATTERN-21 [宗教]

のように書けばよい。「パターン表記」としては、

```
#pattern "PATTERN-21"  
pattern:[Dougigo("宗教")]
```

と展開される。

これは単純かつ強力である。もちろん厳密には、「たばこを何本吸っているかというようなことはどうでもいいが、現在の血圧を知りたい」というような文を作ればだまることができるが、実用上はさしつかえない。

8.2 二つの「語」が出てくればよいもの (順序問わず)

例:ページ44の「帰宅時間」
この場合は、

```
#PATTERN PATTERN-44 [帰宅]..[時間]
```

のように書く。パターン表記として、

```
#pattern "PATTERN-44"  
pattern:[Dougigo("帰宅")]  
[ANY]*[Dougigo("時間")]
```

```
#pattern "PATTERN-44"  
pattern:[Dougigo("時間")]  
[ANY]*[Dougigo("帰宅")]
```

と二つに展開される。

「帰宅する時間は」「何時頃に帰宅するのだろうか」「いつもいつごろ家に着くのか」「家に帰る時刻は」など柔軟に対応できる。

8.3 ある「語」とある「語」の関係を示す必要があるもの

例:ページ8の「妻から見た息子」

```
#PATTERN PATTERN-8 [長男][に][思10]
```

ページ8は、「妻は長男のことをどう思っているのだろうか」に対する答えであり、例えば「長男は母のことをどう思っているのだろうか」などを排除しなければならない。簡便にそれを実現するため、上記の定義を、

```
#PATTERN PATTERN-8  
pattern:[Dougigo("長男")]  
["に"|"|は"|"|を"|"|へ"|"|を"] [ANY]*  
[Dougigo("思-10")]
```

のように展開する。

その他、

```
#PATTERN PATTERN-2 [妻][の][日常.生活]
```

は

```
#pattern "PATTERN-2"  
pattern:[Dougigo("妻")]  
["の"|"|は"|"|&isJoshi()"] [ANY]*  
[Dougigo("日常.生活")]
```

のように展開し、

```
#PATTERN PATTERN-10 [妻][は][人生][を][思10]
```

は

```
#pattern "PATTERN-10"  
pattern:[Dougigo("妻")]  
["の"|"|は"|"|から"|"|が"|"|&isJoshi()"]  
[ANY]*[Dougigo("人生")]  
["を"|"|の"|"|こと"|"|に"|"|対し"] [ANY]*  
[Dougigo("思-10")]
```

のように展開する。

C A T-2 のために作成した「同意義文定義ファイル」(一部)

//人

#SYNONYM 妻 奥.様 母 母親

#SYNONYM 夫 鈴木.様 亭主

#SYNONYM 鈴木 すずき スズキ 旦那 だんな
ダンナ

#SYNONYM 様 さん さま 氏

(略)

//食事

#SYNONYM 食事 食べ物 食 食べもの 甘いもの
辛いもの あまいもの からの味

#SYNONYM 好み 好き 嗜好 嫌い 好8

#SYNONYM 食事生活 朝.食 食事 献立 食生活
いつも.食 一緒.食 食習慣

#SYNONYM 朝 昼 夕 夜 外

#SYNONYM 食 ごはん ご飯 御飯

//一般

#SYNONYM 思10 考え12 感じ12 望8 満足 後悔
思い 望み 気づか 気遣 きづか

(略)

#ENDSYNONYM

```
#PATTERN PATTERN-2 [妻][の][日常.生活]
```

```
#PATTERN PATTERN-3 [夫][に][思10]
```

#PATTERN PATTERN-3 [結婚/夫]..[満足/後悔]
#PATTERN PATTERN-3 [からみた夫/から見た夫
/としての夫]
#PATTERN PATTERN-4 [長男]..[進路]
(中略)
#PATTERN PATTERN-50 [血液検査]
#PATTERN PATTERN-51 [尿検査]
#PATTERN PATTERN-52 [人間ドック]

9. 精度、使い勝手

現在、同意義文セット50個に対して343個の文例(国際医療福祉大学の教材作成者が準備した)をもとにして脇田が定義ファイルを作成し、それを用いて2481個の文例(学生のテストで収集)を分類すると精度は95%程度(程度、というのは手動での分類があいまいなため:下記の例を参照)であった。分類に失敗したうち、異なる同意義文に分類してしまったものは5個程度で、その他はいずれの同意義文セットにも分類されなかった。

異なる同意義文に分類した例

あいまいな分類「将来の生活に希望すること」→手動ではページ24「生き甲斐」に、システムはページ11「老後」に分類

いずれの同意義文セットにも分類しなかった例

非定形的な表現をされたもの「奥さん、鈴木さんに何か言いたいことありますか?」→手動ではページ3「妻から見た夫あるいは結婚生活」に分類
極端にかな表記されたもの「すずきさんとかぞくのかんけいは」→ページ5
未登録の表現「甘党ですか辛党ですか」→ページ16
未登録の表現「健康の為に何かしているのか」→ページ18
??「鈴木さんの経歴」→手動ではページ22「生い立ち」に分類されているが
??「生活の満足感」→手動ではページ24「生き甲斐」に分類されているが

定義ファイルを作成する作業については、始めに数少ない文例を見ながら書き下ろし、次に与えられた文例でテストしながら同意義文同定に漏れた表現を追加していく感じでスムーズに行えた。さらに今回分類に失敗したものをもとに「同意義文定義ファイル」を追加・修正する。

大学で実際に看護学を教えている人に定義ファイルを書いてもらうテストと、学生にCAT-2を使ってもらうテストは9月以降に

予定されている。発表当日間に合えばテスト結果について報告する。

10. まとめ

自学自習システム用に、学生の記述した自由記述文の中にあらかじめ定義した記述すべきポイントが含まれているかどうか判断する、同意義文同定システムを作成した。

柔軟な同定を行うことと、同意義文の定義を容易に行えることを条件として、日本語校正支援システムFleCSの「パターン記法」を基本に、「同義語機能」と「パターン生成機能」を追加して作成した。

(注1)ここに記したパターンは説明のためのものであり、実際に用いられているものとは細部において異なっている。

(注2)同義語機能の他の応用については参考文献[3]。

謝辞

題材の提供、テストなど全面的に協力いただいている、医療福祉大学の細井良三先生、金井P a k 雅子先生に感謝いたします。

参考文献

- [1]奥村ほか：日本語校正支援システムFleCSの新聞社における実用化、情報処理学会自然言語処理研究会92-NL-91, (1992)
- [2]奥村ほか：日本語校正支援システムFleCS-新聞社における実用化報告、情報処理学会第45回全国大会3F-5, (1992)
- [3]脇田：翻訳文における表記揺れの検出・訂正、情報処理学会第55回全国大会5J-03, (1997)