

SD式を用いた日本語文生成システムの試作

吉原 将太[†] 脇山 正博^{††} 河口 英二[†]

[†]九州工業大学 ^{††}北九州工業高等専門学校

自然言語の意味を計算機で扱う場合、意味表現に関する明確なモデルが必要である。そのようなモデルの一つとして、SD式(Semantic-structure Description Form)モデルが提案されている。本研究の目的は、このモデルに従い、SD式から日本語文を生成するシステムを実現することである。

本研究では、自然言語生成の第一段階である「一文一文生成」を当面の目的としている。また、生成する日本語文は、SD式で用いられるラベルが自然言語の単語(日本語や英語)そのものである場合を想定し、そのラベルをできるだけそのまま使用して出力する。

A Prototype System of Japanese Sentence Generation by Using the SD-Form

Shouta Yoshihara[†] Masahiro Wakiyama^{††} Eiji Kawaguchi[†]

[†]Kyushu Institute of Technology ^{††}Kitakyushu National College of Technology

The SD-Form is a type of interlingua proposed by the authors. This paper presents a prototype system of Japanese sentence generation which transforms a semantic input data expressed by the SD-Form into a Japanese sentence. The system at the moment works as an SD-Form monitoring system which we need when we are working on SD-Form writing jobs.

The system at present works in a "sentence by sentence" translation manner. This is a first-step toward the realization of a better system. The Japanese sentences are produced from the SD-Forms by using "labels" as the cores of the Japanese words.

1. はじめに

普通、自然言語処理といえば自然言語理解を指すことが多い。これに対して、自然言語の生成は、機械翻訳研究の一部として扱われてきたが、生成自身が興味の中心となることは少なかった。しかし、自然言語生成は、人間の高度な知的活動と密接に関連しており、また計算機と人間とのインターフェースの向上などの理由から近年急速に関心を集めつつある分野である。

計算機とは、与えられた規則に従って記号の操作を行うものである。しかし、自然言語を単に記号体系として計算機に与えるだけでは、計算機は、自然言語の意味を処理できない。そこで、自然言語の持つ意味情報を計算機で扱う場合、自然言語を中間言語に変換して扱うと都合が良い。中間言語は、自然言語の種類に依存しない定型的な意味記述の言語である。計算機は、中間言語から得られる自然言語の構造および意味情報をもとに種々の処理ができる。しかし、中間言語をどのように設計するかは非常に難しい。自然言語処理の研究において、この問題は重要である。

近年、自然言語の意味表現形式の一つとしてSD式(Semantic-structure Description Form)が提案されている[1]。SD式は、自然言語における個々の概念、陳述表現、感情表現、あるいはシステムに与える知識データなどを記述するための一種の中間言語である。その構文は、英語の基本文型を模しているため一般に受け入れやすく、自然言語の多様な表現にも対応できる。またSD式モデルでは、概念間の意味的な近さを定量的に扱うことができるという大きな特徴をもつ。

現在、SD式を用いた研究としては、

- ・英日会話文検索システム[2]
- ・都市環境に関する情報検索システム[3]

などがある。

著者等の研究グループでは、これらの研究で使用するSD式データを作成するために、SD式作成支援システムの開発を進めている。しかし、開発途上のシステムであるため、作成されたSD式が妥当なものとは限らない。そこで、逆にそのSD式を基にして日本語文を作り出せれば、元のSD式は正しく作成されていたと見なすことができるであろう。純然たる自然言語生成問題としての興味以外に、このようなSD式データ作成の支援の為にSD式から日本語文の作成システムを考えることは有用である。

本研究の目的は、自然言語生成研究の一環として、SD式から日本語文を生成するシステムを計算機上に実現することである。本研究では、まずは一文一文を生成の対象とするSD式からの(SD式概念ラベルを定める名詞や動詞などをそのまま使用して)日本語文生成を試みる。

以下、2章では本システムで用いるSD式を構文的に分類し、その記述例を示す。3章では日本語文生成システムの概要を示し、更に日本語文生成処理と生成例について述べる。最後に4章でまとめと今後の課題を述べる。

2. SD式の構文的分類

SD式は文脈自由言語であり、その構文は、SDG(SD-Form Generative Grammar)と名付けられた文法の生成規則を満たす。この文法に基づき、SD式の構文形式は以下の8種類に分類できる[4]。ただし、記号“/”を修飾子と呼び、この記号の右側で左側を修飾している。複数のSD式で並列的に修飾する場合は、結合子“para”を用いる。

(1) 変数概念ラベル (変数ラベル)

X, Y, Z, …

(2) 単純概念ラベル (単純ラベル)

本, 日本, トム, 買う, 赤い, 25, \$, ...

(3) パラメータ付きラベル

- ・ドル(100) : 100 ドル
- ・km(15) : 15km
- ・車(\$) : “ラベル(\$)” は関係代名詞的用法における先行詞

(4) 修飾形式

- ・本/歴史 : 歴史の本
- ・女性/(美しい)para(若い)
: 若くて美しい女性
- ・車(\$)/[s(トム), v(運転), α(\$)]
: トムが運転する車

(5) 規定子形式

- ・nega(買う) : 買わない (否定)
- ・pass(割る) : 割られる (受け身)
- ・assu([s(友美), v(来る)])
: もし友美が来れば (仮定)
- ・only(ポップ) : ポップだけ (限定)

(6) 結合子形式

- ・(男)plus(女) : 男と女
- ・(東京)equa(首都/日本)
: 東京は日本の首都

(7) 陳述形式

- ・[s(自分), v(テニス/時/毎日)]
: 私は、毎日テニスをする。
- ・[s(彼), v(名付ける/過去),
α(息子/所有/彼), α(健太)]
: 彼は、彼の息子を健太と名付けた。
- ・[s(相手), v(来る/(過去)para(時/何)
para(場所/終点/指示))]
: あなたは、ここへいつ来たのですか。
- ・[s(自分), v(思う), α(相手),
α([s(彼), v(nega(泳ぐ/mood/可能)))]])
: 私は、彼は泳ぐことができないと思う。
- ・[s(自分), v(依頼), α(相手),
α([s(相手), v(貸す), v(自分), α(ペン/指示)))]

: 私にこのペンを貸してくれませんか。

陳述形式の中で用いる陳述機能項目は、次の通りである。

- s: 主語項目 v: 述語項目
- o: 目的語項目 i: 間接目的語項目
- c: 補語項目 b: 行為者項目

(8) 感情形式

- ・[α(ジョン)] : ジョン (呼びかけ)
- ・[r(否定)] : いいえ (応答)
- ・[α([s(犬/遠指示), v(大きい)])]
: あの犬は、なんて大きいんだ! (感嘆)

感情形式の中で用いる感情機能項目は、次の通りである。

- a: 呼びかけ項目 r: 応答項目
- e: 感嘆項目

3. 日本語文生成システム

3.1 システムの構想

本研究は、SD式意味モデル研究の一環として、SD式からの日本語文生成システムの実現を目指している[5]。そこで本研究は、図1に示す日本語文生成システムを構想している。システム構成は、次の通りである。

- A. 入力SD式系列 B. 文脈センサー
- C. 日本語文生成制御部 D. 辞書データ部
- E. 書換え規則部 F. 知識データ部

図1のシステムは、複数のSD式、つまり文章の列を対象とした日本語文生成システムを示している。このような日本語文の生成過程においては、書換え規則を用いて多段のSD式変換を行う。そのときの中間形式をIM式(Intermediate-Form)と呼ぶ。文脈センサーは、照応表現等を処理するために文脈情報を蓄えるものである。日本語文生成制御部は、全体を統括し、辞書データ・書換え規則・知識データ・文脈情報を用いて適切な日本語文を生成する。

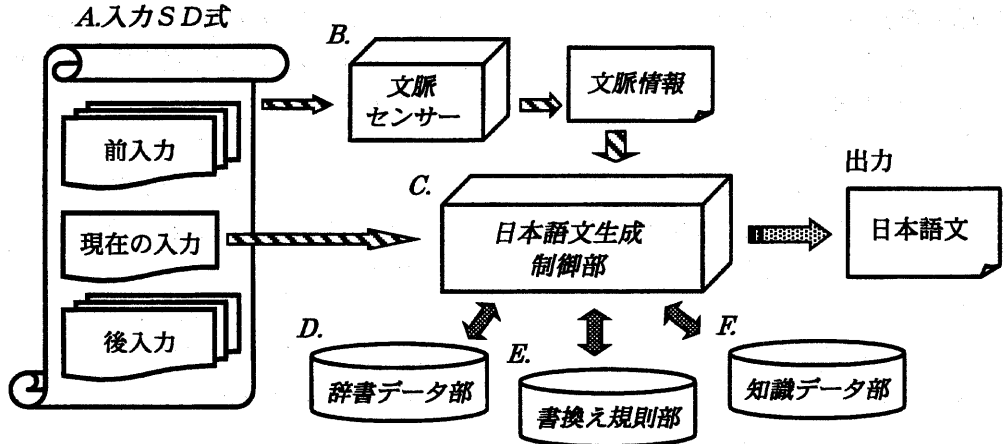


図1 日本語文生成システムの構想模式図

しかしながら、図1で示しているシステムは本研究の最終目標であり、現在は一文単位の生成を扱っている段階である。

3.2 試作システム

本システムは、日本語文生成制御部と辞書データ部・書換え規則部・知識データ部で構成している。使用したプログラム言語は、Prolog である。

・辞書データ部

辞書データは、SD式に使われている各々のラベルに対し、それぞれ名詞・動詞・形容詞・その他をあらかじめ登録している。ただし、動詞の形式は、

verb(①E,②J,③語幹,④未然形,⑤未然形2,⑥連用形,⑦終止形,⑧仮定形,⑨過去形,⑩相).
(E(J):英語(日本語)のSD式ラベル)

であり、形容詞の形式も同様である。

(例)

名詞: noun(BOOK,本). animal(DOG,犬).
person(TOM,トム).

動詞: verb(FIND,見つける,見つけ,見つけ、
見つけら,見つけ,見つける,
見つけれ,見つけた,見つけて).

形容詞: adjective(BLUE,青い,青,青く,0,0,
青い,青けれ,青かった,0).

その他: label(I,自分,私).

label(PLACE/DESTINATION,場所/終点,へ).

title(Teacher,先生). parameter(\$,\$).

・書換え規則部

入力SD式を次の形式の書換え規則を用いて書換える。

rule([ル-ルグループ],[書換え前のSD(IM)式]
:[条件部]->[書換え後のIM式]
>>[書換え後のル-ルグループ]).

(例)

陳述形式:

rule([begin],[[s(X1),v(X2)]]:[[]]
->[state([s(X1),v(X2)])]>>[state]).

ルールグループ“begin”に属するSD式[s(X1),v(X2)]は、ルールグループ“state”に属するIM式“state([s(X1),v(X2)])”へ無条件(条件部無しにより)に書換えられる。

結合子形式:

rule([pseq],[[X1,X2]]:[[]]
->[[initial(X1,2),。そして,initial(X2,2),。]]
>>[connect]).

感情形式：

```
rule([initial],[[s(X)]:[[form(X)]]
->[attention(X)]>>[attention]].
```

条件部の“form(X)”は、Xの構文形式を調べる。

・知識データ部

知識データは、書換え規則の適用条件・ラベルの置換など様々な場面で利用する。

```
(例) fact((生物)incl(動物)).
fact((動物)incl(人間)).
fact(assu(動物)incl(X)caus(animal(X))).
fact(assu(動物)incl(X)caus(person(X))).
```

```
incl: 包含関係を表す結合子
caus: 必然的な因果関係を表す結合子
assu(D): 概念Dの仮定を表す規定子
X: 変数
```

3.3 日本語生成処理

以下の①～⑤の手順(SD1,IM1～IM5)により、SD式から日本語文を生成する例を示す。ただし、入力は次の通りである。

```
[s(猿),v(見つける/(過去)para(場所/腰/桃太郎)),
α(きび団子)] …SD1
```

- ① 入力SD式内のラベルを、必要であれば適切なラベルに置換する。このとき知識データを用いている。入力例SD1の場合は、ラベルを置換する必要はない(SD1=IM1)。
- ② 入力SD式を再帰的に書換え規則と照合し、多段階の書換えを行う。この時、助詞も決定する。

```
[s(猿),v(見つける/(過去)para(場所/腰/桃太郎)),
α(きび団子)] …IM1
```

書換え規則(ア),(イ)と順に照合し、段階的に書換える(X1,X2,X3,Yは変数)。

```
(ア) rule([begin],[[s(X1),v(X2),α(X3)]:[[]]
->[state([s(X1),v(X2),α(X3)])]>>[state]].
```

```
state([s(猿),v(見つける/(過去)
para(場所/腰/桃太郎)),α(きび団子)]) …IM2
```

```
(イ) rule([state],[state([s(X1),v(X2),α(X3)])]
: [[X2=(Y/_),particle_s(X1,SJ),
particle_o(Y,X3,OJ)]]
->[[item(X1),SJ,
item-vo([item-v(X2),item(X3),OJ])]
>>[end]].
```

```
( X2=(Y/_): 述部が修飾形式である。
particle_s(X1,SJ): 主部X1に続く
助詞SJを決定する。
particle_o(Y,X3,OJ): 目的部X3に
続く助詞OJを決定する。)
```

```
[item(猿)は、item-vo([item-v(見つける/
(過去)para(場所/腰/桃太郎)),
item(きび団子),を))] …IM3
```

- ③ ②で得たリストIM3の先頭から順に、機能項目内(主語項目“item(猿)”は、“猿”の部分)を処理して日本語の単語に変換する。ただし、項目“item-vo()”は、述語項目の修飾部、目的語項目、目的部に続く助詞、述部の動詞の順で処理する。

a. 述語項目の処理

```
item-v(見つける/(過去)para(場所/腰/桃太郎))
```

“見つける”を修飾するSD式をリストとする。

```
[過去,場所/腰/桃太郎] …(4.1)
```

リスト(4.1)内のラベルを一つ一つ解析し日本語変換する。

→ [桃太郎,の,腰,に]

リスト(4.1)内に「時制」などを表すラベルがあるかどうかを検索すると、ラベル“過去”により、動詞“見つける”の活用は「過去形」と決定する。

→ [見つけた]

b. 主語項目, 目的語項目など(述語項目以外)の処理

述語項目の処理と同様に日本語変換する。ただし, 使用されているラベルが名詞であるため活用形を考える必要は無い。

item(猿) → [猿]
 item(きび団子) → [きび団子]

④ ②で得た *IM3* は, ③の処理により次のようになりリストに変換される。

[[猿],は, [桃太郎,の,腰,に, [きび団子],を],[見つけた]] … *IM4*

⑤ *IM4* の “[” や “] ” を除いて, 日本語文として出力する。

[猿,は, ,桃太郎,の,腰,に,きび団子,を, 見つけた,。] … *IM5*

↓ 出力

猿は, 桃太郎の腰にきび団子を見つけた。

3.4 生成例

実験では, 中学英語教科書に現れている英文を参照しながら作成したSD式を用いて, 種々の日本語文が生成できることを確認した。以下に生成例を示す(2節の例も参照のこと)。

- [s(自分), v(nega(興味/(状態)para(程度/大)para(対象/哲学)))]
 私は, 哲学にあまり興味を持っていない。
- [s(妹), v(pass(誉める/過去)), b(父)add(母)]
 妹は, 母のみならず父にも誉められた。
- [s(我々), v(洗う/頻度/回(1)/当たり/週(1)),
 α(自動車)]
 我々は, 週に1回自動車を洗う。
- [s(相手), v(会う/(過去)para(時/昨日)), α(誰)]
 あなたは, 昨日誰に会ったのですか。
- [s(自分), v(命令), α(相手),
 α([s(相手), v(開ける), α(窓)])]
 窓を開けなさい。

- [s(I), v(FIND/PAST), α(BOOK/
 (TALE)para(POSSESSION/DAUGHTER))]
 私は, 娘の童話の本を見つけた。
- [α([s(THAT), v(BE),
 α(fcus(FLOWER/BEAUTIFULE)))]]
 あれは, なんて美しい花なんだ!

4. おわりに

本システムは, 入力SD式を構文解析し, SD式のラベルが自然言語の単語(日本語や英語)であることを前提として日本語文を生成するものである。まずは自然言語生成の第一段階である「一文一文生成」を試みた。

今後は, 構想システムが可能としている処理, つまり文章を対象とし照応表現等を処理できるシステムを目指す。

[参考文献]

- [1] Eiji Kawaguchi, Masahiro Wakiyama : “The Semantic Metric Computation Scheme in the SD-Form Semantics Model”, Proc.PRICAI, pp.623-629 (1993)
- [2] Guifeng Shao, Koichi Nozaki, Sei-ichiro Kamata and Eiji Kawaguchi : “SD-Forms as Interlingua and Prototype of a Conversational-Text Retrieving System”, 人工知能学会, Vol.9 No.5, pp.684-693 (1994)
- [3] 江口 徹 : 「意味構造記述式を用いた都市環境に関する対話型情報検索システムの試作」, 九州工業大学工学部 平成6年度卒業論文
- [4] 河口 英二 : 「意味表現式データを用いた会話的英語文の検索, 生成と学習支援システムの基礎研究」, TELECOM FRONTIER SCAT TECHNICAL JOURNAL No.11, pp.16-28 (1992)
- [5] Masahiro Wakiyama, Shouta Yoshihara, Eiji Kawaguchi : “A Prototype of Japanese Sentence Generation System from SD-Formed Meaning Data”, PACLING'97 pp.333 (1997)