

フレームによる知識表現をもちいた
文章生成システム

5J-3

野口 喜洋 石原 正雄 下郡山 敬己 小川 東
榊CSK総合研究所

1. はじめに

特定の対象領域についての知識を用いて、報告書などの文章作成を支援するシステムをXEROX-1121上に開発している。対象領域は週間業務報告書の内容であり、図1に示すような表形式の入力に対して、数個の文からなる短い文章が出力として得られる。

・入力形式(網掛け部がユーザのタイプイン)

シーン: 学会発表	
学会名	情報処理学会
開催日	4月4日
開催場所	東京工業大学
参加者名	田中一郎
発表者名	
発表内容	文章生成システム

・出力文章

項目名: 情報処理学会
担当者: 田中一郎
内容:
4月4日、東京工業大学において情報処理学会が開催され、日本語処理研究室の田中一郎研究員が出席した。文章生成システムを発表した。

図1 文章生成システムの入力及び出力例

システムが文章生成に利用している知識には次のようなものがある。

- ・固有名詞に関する宣言的知識
- ・各名詞についての文生成の方法を決定する手続き的知識
- ・単文の意味表現から表層の文を生成するための知識(例えば、表層の動詞決定知識)
- ・報告すべき出来事の典型的な生起状況を記述したスクリプト的知識

これらの知識は階層的なフレームシステムを構成しており、報告文章の生成の際に使用される。また、出力文章の意味表現もフレームの形式で残り、要約文章の生成、英語などの他言語文章の生成など、他のプログラムで使用できる。フレーム操作システム自体は週報という対象

領域とは独立なので、知識ベースを入れ換えて他の領域の文章を生成することもできる。

2. システムの概要

文章生成システムのフレーム構成を図2に示す。

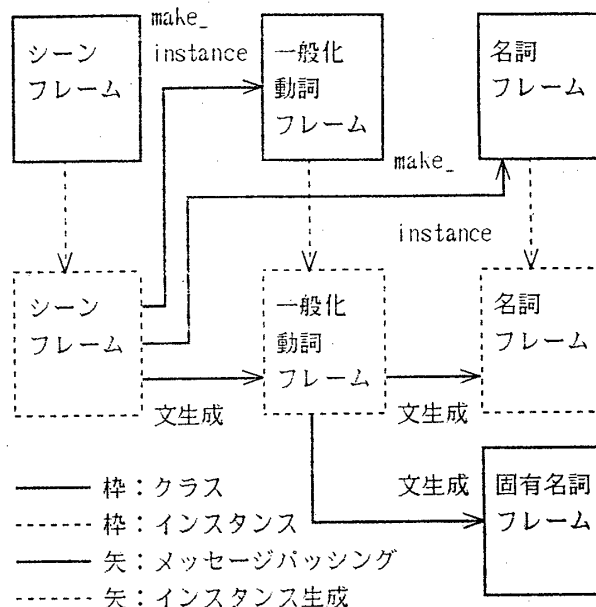


図2 文章生成システムのフレーム構成

全体は次の3つのレベルからなる。

- 名詞レベル 名詞句を生成する。
- 一般化動詞レベル 類似の動詞中から1つを選択し、単文を生成する。
- シーンレベル 数個の単文からなる1つの文章(週報項目)を生成する。

文章生成の手順は以下のとおりである。

ユーザが選択したシーンからインスタンスが1つ生成され、図1の入力形式で必要な知識を獲得する。その際必要な名詞レベルのインスタンスも生成され、知識を獲得する。十分な知識を獲得すると、シーンのインスタンスは単文の知識表現である一般化動詞にインスタンスを生成させ、それらに文生成メッセージを送る。

一般化動詞レベルのインスタンスは格スロットの値である名詞レベルのインスタンスまたは固有名詞フレームに文生成メッセージを送って名詞句を生成させ、適当な表層の動詞を選択して単文を生成する。生成された単文はシーンレベルからの制御によって適当な接続形態を持っているため、全体として文章の生成がなされる。

3. 名詞レベル

名詞レベルは、「人間」「株式会社」「時刻」など、さまざまな名詞概念についての名詞句生成メソッドを定義している名詞フレームと、固有名詞（人名、会社名など）についてのデータベースである固有名詞フレームからなる。実際の名詞句生成は、名詞フレームの下にインスタンスを作ってユーザからデータを獲得するか、固有名詞フレームのデータを用い、それらにクラスの名詞句生成メソッドを適用することによって行われる。メソッド定義は、固有名詞フレームおよびインスタンスに継承されるので、名詞句を生成するフレームにメッセージを送ることで、名詞句が値として得られる。例えば、図1の出力文章中、「日本語処理研究室の田中一郎研究員」は「田中一郎」固有名詞フレームの文生成メソッドの値である。メソッドを起動するメッセージは、一般化動詞フレームが単文生成時に格スロットの値である名詞フレームに送る。

4. 一般化動詞レベル

一般化動詞フレームは、表層で1つの単文に相当する知識を格納するためのフレームで、深層格を埋める名詞句を指定する格スロットおよび、法情報（時制、肯定／否定、能動／受動）、接続形態（終止／継続）などのスロットからなる。一般化動詞と表層の動詞は1対多に対応しており、例えば「#人の移動」という一般化動詞に対して、「行く」「訪問する」「来る」「来社する」「来所する」「来日する」などの表層の動詞が対応する。格スロットの値の有無及び内容によってこれらの内1つ表層の動詞を選択するメソッドを持っている。このような動詞のグルーピングの利点は2つある。

- ・出来事の意味に応じて、表層の表現を柔軟に変化させられる。
- ・同一、類似の出来事間の比較や検索が可能。

一般化動詞フレームには、必須格・任意格スロットが定められており、必須格のすべてが埋められた場合のみ、単文の生成が可能になる。一般化動詞フレームは、後述のシーンフレームから文生成メッセージを受けてメソッドを起動させる。自らの格・法情報・接続形態などの情報に基づいて、各名詞句に文生成メッセージを送り、表層の動詞を選んで形態素生成を行い、単文を生成する。

5. シーンレベル

複数の単文からなる文章（週報項目）の知識表現がシ

ーンである。報告書の報告事項などの記載には、典型的な流れに沿ったものが多い。この知識（シーン）はシーンフレームに格納される。シーンは時系列に沿って並べられた一般化動詞フレームのグラフ（Aグラフ）によって表現される。「\$学会」シーンのAグラフを図3に示す。

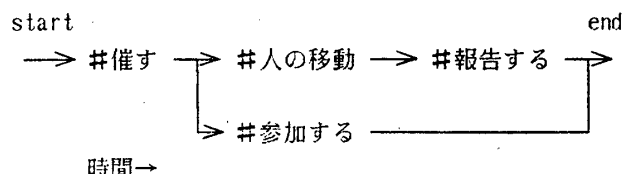


図3 「\$学会」フレームのAグラフ

Aグラフ中の一般化動詞フレームのスロットのいくつかにはすでに値が入っている。また、Aグラフ中の他のスロットと共通化しているものもある。たとえば、上のAグラフでは各一般化動詞フレームの動作主格は同一の変数で満たされている。

Aグラフ中のすべての一般化動詞フレームの必須格・任意格スロットが完全に埋まるために必要な項目を表にしたものがS表である。図1の「入力形式」にあるのが「\$学会」シーンのS表である。S表は文章生成に必要な知識を獲得するために使用される。ユーザがS表に入力した内容は、Aグラフ中の一般化動詞フレームのスロットに配分される。入力は、Aグラフ中の少なくとも1つのパス上の一般化動詞フレームの必須格スロットが完全に埋まった時点で打ち切られる。

この時点でパス上の各一般化動詞フレームでは単文の生成が可能であるから、システムはパス上の一般化動詞フレームを順に取り出し、一般化動詞レベルに新たなインスタンスを生成し、それぞれに文生成メッセージを送ることにより、文章を生成する。

6. 考察

一般化動詞及びシーンの適用によって、格構造からの単文生成を数個の文からなる文章の生成に応用することができた。しかし、意味素としての一般化動詞の網羅性及び独立性の点で、意味処理に適した知識表現になっているかどうか疑問がある。

また、文章の焦点や単文同志の接続形態、名詞句の省略などの制御は、現在シーンレベルから固定的に行っているが、今後長い文章の生成を行う際には、動的に制御する必要がある。

参考文献

- [1] Schank, "Dynamic memory", Cambridge University Press, 1982
- [2] Fillmore, 『格文法の原理』, 三省堂, 1975