

意味焦点駆動文生成

関洋平 飯島正 原田賢一

yohei@hara.cs.keio.ac.jp

慶應義塾大学理工学部

本論文は、文脈に即した新しい表層文生成技術の提案についての報告である。従来、入力として意味表現が与えられ、それに対応するひとつの文を出力する計算過程のことを表層文生成技術と呼び、テキスト中の流れや対人役割などの情報を重視する文生成技術と区別する。生成過程を構文解析過程の逆とみなす表層文生成技術のうち、効率および実時間性の観点から代表的な技術の一つとして意味主辞駆動生成がある。意味主辞駆動生成では、計算過程においては構文解析において使用されるのと同様の文法規則が用いられるが、規則の組み合わせによって複数の結果が得られる際には、それぞれの文のニュアンスが異なる場合、何らかの優先順位づけか規則の適用制限が必要となる。しかし、本提案により、入力として、論理形式に基づく意味表現に加えて、旧情報 / 新情報の区別に基づく名詞の優先順位つきリストを与えることで、必要以上に複雑な意味素性を与えずテキスト生成に応用できる表層文生成が可能となる。

Semantic Focus-Driven Sentence Generation.

Yohei SEKI, Tadashi IJIMA and Ken'ichi HARADA

Faculty of Science and Technology, Keio University

This paper describes about the proposal of surface-based sentence generation technique which utilizes contextual information. The surface-based sentence generation technique is regarded as the computational process from an input of a semantic representation to an output of the corresponding sentence, and such technique is distinguished from techniques based on deep textual or human communicative information. As representative of surface-based effective generation techniques, semantic-head driven generation was proposed. But if we can generate several sentences from one semantic representation because of syntactic rule ambiguity, we must use the preference constraint or selectional restriction. We propose semantic focus-driven generation, that adds preference list of nouns to input, which is based on functional given/new information, and this technique enables us to generate textual sentence based on contextual information without complex semantic features.

1 はじめに

自然言語処理には大きく分けて、既存の文書データを処理するための技術と、ユーザに対して何らかの自然言語の文章を提示する二種類の技術がある。後者の技術は文生成技術と呼ばれ、一般に命題的内容をどのような形の文で表すかという一文レベルの問題と、テキスト中の流れや対人コミュニケーション役割などに配慮した上でどのようなテキストを生成するのが自然かという問題に、更に細かく分類できる [1]。前者は一般に表層文生成技術と呼ばれ、論理形式や意味ネットワーク表現からどのように効率良く文章が生成できるかに焦点が絞られており、特に、ここ数年、構文解析で使われる文法規則に基づいたかたちでの生成に関する研究が行われている。

表層文生成技術の代表的なアルゴリズムとして、意味主辞駆動生成 [2] がある。意味主辞駆動生成はトップダウン生成とボトムアップ生成を融合したアルゴリズムであり、その効率性と実時間性という観点から、上記の研究の代表となっている。だが、入力として論理形式を与えている。

しかし、複数の文の候補が生成できる際に、それぞれの文のニュアンスが異なるとすると、候補を絞り込むための情報が入力の論理形式には不足していることが問題となる。そのため、何らかの優先順位づけか、もしくは意味的な選択制限に基づいた対処が必要となる。

後者の意味的な選択制限というアプローチには、いくつかの問題点がある。一つには、入力として論理形式以外の意味的な素性がどれだけ必要か、はっきりしないことがある。意味的な素性をすべて記述し、それぞれの情報を利用して選択器に基づいて生成候補を絞り込んだ後、表出制約に基づいて文を生成するというアプローチとして、システミック文法 ([3] 第 4 章参照) がある。しかし、計算機上で実現を考えた場合、入力情報を得るためのコストという点で困難がある。

本研究では、以上の考察から、入力の意味表現を複雑にすることなく、かつ文脈的な情報を反映した表層文生成技術として、意味焦点駆動文生成を

提案する。意味焦点駆動文生成では、従来の論理形式表現に加えて、談話中の出現という観点から、各名詞を古いものから新しいものへの順序で並べたリストを入力として与えて、語順に制約を与えることで、文脈に即した文の生成が可能となる。これは、従来の述語主導型の生成アプローチに対して、名詞主導型の生成アプローチを試みるものである。

本論文の構成は以下の通りである。まず、第二章で表層文生成技術の代表的技術である意味主辞駆動生成とその問題点について紹介する。第三章では、意味焦点駆動文生成を提案する。第四章では、入力となる優先リストの作成方法について説明する。第五章と第六章で簡単な評価と関連研究について示した後、第七章で全体をまとめる。

2 表層文生成技術と問題点

表層文生成技術にはさまざまなアプローチがあるが、本研究では、代表的なアルゴリズムの一つである意味主辞駆動生成 (Semantic Head-Driven Generation) [2] に絞ってその問題点を指摘する。

2.1 意味主辞駆動生成

意味主辞駆動生成では、論理形式の意味表現を入力として、自然言語の文を出力する。生成過程では DCG (確定節文法) 流の文法規則が使用され、[2] には Prolog に基づく実装の一部が示されている。意味主辞駆動生成は、おおざっぱに言って、以下のようなアルゴリズムに基づいている。

1. 最初に、ゴール範疇を開始記号 (S) とする。
2. 入力意味 (論理形式) の括弧の外側の原子的な意味素性 (述語論理表現に即していえば述語¹に対応、Pivot と呼ばれる) とアリティ (項の数) をキーにして、対応する語彙規則を検索し、前終端記号 (品詞) を判別する。この際、Prolog 上での実装では、non_chain_rule と呼ばれる規則が使用される。

¹ もしくは関数子 (Functor)

3. 入力意味に対応する品詞を右辺に持つ構文規則 (DCG rule) を検索する。ただし、左辺に来る非終端記号は入力意味と意味素性を共有する。この際、Prolog 上の実装では、chain_rule と呼ばれる規則が使用される。
4. 入力意味の argument(項) 部分の意味素性を、検索した構文規則の右辺の他の要素に伝搬し、それぞれの要素の品詞をゴール範疇、対応する意味素性を入力意味として 2 を再帰的に適用する。
5. 4 において構文規則のすべての右辺の要素が生成されれば、現在の左辺の品詞をキーにして、3 を再帰的に適用する。さもなければ、3 にもどって別の規則を検索する。ただし、現在の左辺の品詞が ゴール範疇であり、これ以上生成される要素が無い場合には、生成された要素を出力して現在の生成処理を終了する。

2.2 意味主辞駆動生成の問題点

意味主辞駆動生成の利点としては、語彙規則を拡張して、他に依存する補語要素や終端記号(語彙そのもの)を記述できることから、(call...up)のような熟語を生成するための規則を記述しやすことがあげられる。

問題点としては、生成の際のニュアンスの違いを表せるように動詞の形態を変えるためには、選択制限に用いられる意味素性が必要となるため、各補語名詞の語順の選択に適していない。次章では、この問題に対する解決策として、意味焦点駆動文生成を提案する。

3 意味焦点駆動文生成の提案

本章では、意味焦点駆動文生成を提案する。意味主辞駆動生成に、必須格要素の表層的な語順に関する制約を入力として与えることにより、文脈に応じた文生成が可能となる。

3.1 概要

意味焦点駆動文生成では、述語の必須格に対応する名詞を並べた優先リストを入力とすることで、表層的な語順に関する制約を与えている。この制約は絶対的なものではなく、生成過程で入れ替わることもある。

優先リストを入力として与えることにより、リストの先頭の要素を主語とみなして生成した後、主語の情報を参考にして動詞の形態を決定すること、優先リストの残りの要素の情報から、補語の形態を決定することが可能となる。以上の二点が、本研究の大きな特徴となっている。

動詞の形態を主語に応じて決定することの利点として、動詞の言い換えが可能となることがあげられる。すなわち、深層格役割を表している論理形式と、表層的な語順の制約を表している優先リストとの対応づけにより、受動化だけにとどまらずに、come & go, give & receive, enter & exit 等の言い換えを行っている。これにより受動態を避けたい場合の生成に対処できる²。

補語の形態に関しては、目的語に付加される前置詞と目的語間に関する語順について決定している。より具体的には、sold a car to Tom と sold Tom a car のどちらを生成候補とするかについて判断を行っている。

優先リストは、談話的な旧情報が英語では主語になりやすいという仮説に基づき、大雑把には、名詞を古い情報から新しい情報に並べている。優先リスト自体の作成方法については、次章で詳しく説明する。ただし、優先リストの作成方法と、本章での提案は独立しており、順序の与え方については問わないものとする。

3.2 意味焦点駆動文生成の実現

意味焦点駆動文生成は、意味主辞駆動生成を拡張することにより実現しているが、優先リストの要素を生成した後、関連のある主辞要素を生成している点が大きく異なる。すなわち、意味主辞駆

²英語で受動態を用いるか言い換えを行うかは客観性・視点など微妙な判断を必要とする

動生成の chain_rule の代わりに、補語と目標から主辞を検索するルールが必要となる。これを、便宜上 link_rule と呼ぶことにする。

link_rule は、語彙に依存しない構文規則に対して適用が行われる。また、入力として補語範疇だけでなくゴール範疇も与えることで、主辞範疇を検索する。現在のところ、語彙に依存しない構文規則は二分木で与えることにしており、検索に辺り大きな問題は生じていない。これらの構文規則は、これ以降構文主体規則と呼ぶ。

また、動詞の目的語の形態は、Verb Alternations[8]を参考にして、各動詞ごとに与えている。この規則は、これ以降語彙依存規則と呼ぶ。どの形態が選択されるかは、優先リストの情報に基づき決定される。語彙依存規則と構文規則を取り扱いのレベルで区別することにより、規則の無駄な適用を大幅に減らすことが可能となる。以下に、構文主体規則と語彙依存規則の具体例を示す。

%% 構文主体規則の例

```
g(s, Sem, S-[]) :- 1(np, Sbj, S-S1), 1(vp(Sbj), Sem, S1-[]).
```

```
g(np, Sem, S-S2) :- 1(det(N), Sem, S-S1), 1(n, N, S1-S2).
```

%% 語彙依存規則の例

```
1(vp(A), p(give, A, Obj, Obj2), ['give' | S1]-Y) :-  
  1(np, Obj2, S1-X), 1(pp, p(to, Obj), X-Y).
```

```
1(vp(Obj), p(give, A, Obj, Obj2), ['receive' | S1]-Y) :-  
  1(np, Obj2, S1-X), 1(pp, p(from, A), X-Y).
```

一方、non_chain_rule は、主語に応じた動詞の形態の決定を実現するために、動詞の場合とそれ以外の場合を区別している。また、動詞の補語形態と、主語以外の優先リストの要素を、依存要素生成述語 (generate_ds) に受け渡す。

また、入力の優先リストが空であれば、意味主辞駆動生成を起動する。意味主辞駆動生成において、どの規則が選択されるかは規則の記述された順序に基づいて行われる。

以下では、意味焦点駆動文生成について、単文の生成過程を示す。

1. 入力として意味表現と、述語の必須格要素に対応した名詞用優先リストを用意する
2. 優先リストの先頭の要素の一つ取り出して、ゴール範疇を {NP} として、主語を生成する
3. NP と ゴール範疇の S を入力として、link_rule を起動し、新たなゴール範疇として VP を導入する
 - (a) 意味表現を取り出し、既に生成された主語の情報に基づいて、non_chain_rule の適用を制限することで、動詞そのものの選択を行う
 - (b) 失敗した場合には、優先リストの要素を入れ替えて、主語の生成をやり直す
 - (c) 動詞の補語の生成
 - i. 動詞が決定されれば、次に、語彙依存規則に基づきゴール範疇を導入し、優先リストの残りの要素から、link_rule に基づき動詞の補語を順次生成する
 - ii. 失敗した場合には、優先リストの要素を入れ替えて、再度動詞の補語形態を決定し直す
4. 優先リストが空になっていれば、文を出力して生成過程を終了する

3.3 文生成の具体例

実際の生成の実現にあたっては、Prolog の fail 述語を用いることにより、バックトラック機構に基づき生成文を優先順位の高い順序で出力しているが、以下では、優先度が最も高い結果を出力しているものとして具体例を示す。

まず、入力として、give(i, tom, <birthday, present>) を意味表現、[i, <birthday, present>, tom] を優先リストとして与えた上で、「I give a birthday present to Tom.」という出力文を生成する過程について、以下の図によって示すことにする。

4.1 Prince の新 / 旧情報の取り扱いと主語との関わり

Ellen F.Prince は, [4] で, 新 / 旧情報の名詞句について, 新 / 旧情報を共有知識という観点から捉えた上で, 以下の分類 (前提となる親しさ (Assumed Familiarity) と呼んでいる) を示している。

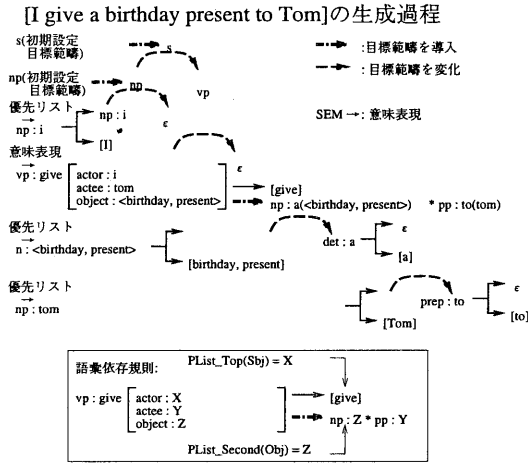


図 1: 「I give a birthday present to Tom」の生成過程

優先リストが [i, tom, <birthday, present>] であれば, 出力文として「I give Tom a birthday present.」が生成される。また, リストが [tom, <birthday, present>, i] であった場合には, 文「Tom receives a birthday present from me.」が生成されることになる。以上のように, 優先リストに応じて複数の出力文候補が選択できることにより, 文脈に応じた文生成が可能となる。

また, 上記の生成過程からわかることとして, 名詞が動詞の補語となる過程では, 名詞句や前置詞句をゴール範疇として, 冠詞や前置詞が付加されていく過程が示されている。この過程は意味主辞駆動生成とは逆の過程になっている。

4 優先リストの作成方法の提案

本研究では, 第 3 章で入力として与えた優先リストの作成方法について提案する。優先リストは, 言語によらない文脈の維持という概念に基づき作成される。本章では, まず Prince の新 / 旧情報に対する考え方と関連研究について紹介した後, 優先リストの作成方法について提案する。

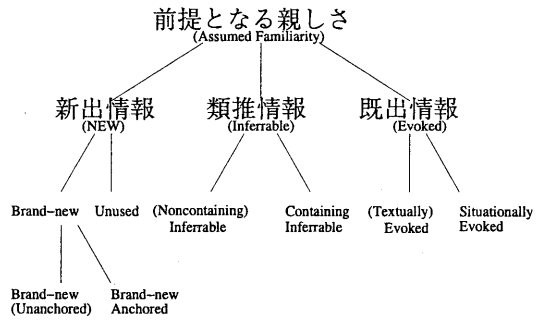


図 2: 新 / 旧情報と名詞要素の対応に関する分類

上記の分類における Evoked は既出の情報, Inferable は, 既出の情報から類推できる情報, New は新出情報に対応している。類推情報としてどのようなものを取り扱うかは難問であるが, 本研究では, 対象とする要素は [9] に基づいて規定することで, 述語と名詞の双方の制約から対象を絞り込む。また, 聞き手の推論の負荷がかからない順序という形で Familiarity Scale (親しさの度合) と呼ばれる以下の優先順位づけが成されている。

$$\{E, E^S\} > U > I > I^C > BN^A > BN$$

以上の議論は, 話し手に対する聞き手の状態を元にした情報の分類であったが, [5] では, 文において主語となる可能性は, 聞き手にとって新しいか古いかという情報よりも, 談話中で新しいか古いかといった情報が効いてくることを, 資金集めの手紙中の文章を評価することで示している。上記の分類はより簡略化され, 以下の三分類で主語となる可能性 (Subjecthood) を示している。

談話既出代名詞 > 談話既出非代名詞 > 談話新出名詞

4.2 優先リストの作成方法

本節では、英語の優先リストの作成方法を提案する。本提案は、普遍文法に基づいていると考えられており、表層的な手掛かりを除けば日本語へ応用することも可能であると考えている。

1. 対象文の必須格名詞要素のうち、前の文の優先リスト中の要素に含まれるものを、付属助詞に基づき、以下の規定により順位づける。前の文から補完できる省略要素は補完した上で、優先リストに追加する

(a) 主語 > 目的語 > 第二目的語 > 前置詞句

(b) 上記で規定できないものは、出現順で規定

2. 対象文の必須格要素のうち、1に含まれないものを、[9]に基づき、述語と名詞の双方の制約から対象を絞り込んだ結果、関連を持つ要素を次に順位づける
3. 同一文中で係り先動詞が異なる別の優先リスト中の要素により補完できる要素を次に順位付ける
4. 上記に含まれない現在の文の名詞句で、現在の文に出現したものを、以下の規定により上記の要素に続けて順位づけする

- 既出の普通名詞 > 固有名詞 > 既出の名詞または固有名詞を連体修飾要素または「～の」の前件部分に含んでいる普通名詞 > 代名詞 > 未出の普通名詞

5 評価

本研究に際して実験的に行った実装では、意味主辞駆動生成の chain_rule では、入力意味と品詞をキーにして構文規則を検索することで、左辺の品詞に到達するが、入力品詞の規則の位置に関する制約が無いことから、間違った規則を検索してしまう可能性が大きい。また、この品詞がゴール範疇に繋がられるかどうかという情報に関して制

約が無いため、規則を誤って選択した場合、そのまま規則の適用を進めていき深い段階まで到達した後に、バックトラックが起きてしまう。

本研究のアプローチでは、名詞を先に生成することにより、深さ優先的なアプローチに幅優先的な手法を取り入れていると考えられる。これにより、誤った規則の適用に基づく深いバックトラックを避けることができ、結果として実時間性が向上していると考えられる。また、構文主体規則と語彙依存規則をアルゴリズムのレベルで区別したことにより、構文的曖昧性は軽減されている。

6 関連研究

意味主辞駆動生成の深さ優先的な問題を解決する他のアプローチとして、ボトムアップチャート法に基づく並列文生成に関する研究がある [6]。この研究では、複数の生成結果をバックトラックを用いずに生成することと、GHC(Guarded Horn Clause)に基づく並列化が示されているが、主辞中心の生成手法であることに変わり無く、本研究とは区別される。

新 / 旧情報に基づく優先リストの作成方法の関連研究としては、ドイツの Michael Strube らが提案している S-list(a list of salience discourse entities)[7]がある。S-list は文内照応と文間照応の解決を目的として、リストの二文に渡るインクリメンタルな作成方法を取り入れている。だが、文生成に応用しようとする本研究とは目的が異なるため、作成方法に違いが生じる。

7 おわりに

本提案により、表層文生成技術の how to say 的な質の向上という課題について、入力として名詞の機能的情報に基づくリストを与えることにより、一つのアプローチを示すことができたと考えている。複雑な意味的制約よりも、前文などから与えられる語順の制約に基づき自然な文章を生成するという提案は、論理形式意味表現を出力として得るプラン認識の結果に対する文生成という応用の

立場からも期待が持てる。

今後の課題としては、優先リストの作成の自動化があげられる。また、述語をリスト化して優先順位付けを行うことによる関係節と連体修飾の生成についても詳しく検討したい。その他、構文解析用の構文規則を構文主体規則としてそのまま設定することは行わず、意味素自体を品詞として、生成に特化した規則を記述することにより、より効率的な実現を目指すことについて検討している。

謝辞 昨年度の卒業研究として、意味主辞駆動生成のプロトタイプシステム作成に御協力頂いた現トヨタ自動車の田中清司君に感謝します。また、関連研究として文献 [6] をご紹介くださった日本 IBM の武田浩一氏に深く感謝致します。

参考文献

- [1] Robert Dale, Barbara Di Eugenio, Donia Scott: "Introduction to the Special Issue on Natural Language Generation," *Computational Linguistics* Vol.24, Number 3(1998), pp.345-353.
- [2] Stuart M.Shieber, Fernando C.N.Pereira: "Semantics Head Driven Generation," *Computational Linguistics* Vol.16, Number 1(1990), pp.30-42.
- [3] 長尾真, 中川裕志, 松本裕治, 橋田浩一, John Bateman: 言語の数理, 岩波講座 言語の科学 8, 岩波書店 (1999).
- [4] Prince, E.F.: "Toward a taxonomy of given-new information," In P.Cole(Ed.), *Radical Pragmatics*, pp.223-255. New York, N.Y.:Academic Press(1981).
- [5] Prince, E.F.: "The ZPG letter: Subjects, definiteness, and information-status," In W.Mann & S.Thompson(Eds.), *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text*, pp.295-325. Amsterdam:John Benamins(1992).
- [6] 春野雅彦, 松本裕治, 長尾真: "文法主導による協調的文生成," 自然言語処理シンポジウム論文集, 電子情報通信学会 日本ソフトウェア科学会, pp.88-96(1992).
- [7] Strube, M: "Never Look Back: An Alternative to Centering," *Proceedings of COLING-ACL'98*(1998).
- [8] Levin, B. : *English Verb Classes and Alternations*, The University of Chicago Press(1993).
- [9] 関洋平, 飯島正, 原田賢一: "用言の意味マーカと名詞の意味素性の関係のギャップを part-of 関係および名詞の意味情報を用いて解決する手法," 言語処理学会第 4 回年次大会発表論文集, pp.614-617(1998). pp.63-64(1997).