

完全選択入力方式による意味確定手法に関する一提案

金久保 正明[†]

静岡理工科大学総合情報学部人間情報デザイン学科[†]

1. はじめに

自由に入力された自然言語の意味を確定するには、表記揺れ、語の多義性、係り受けの不確定性等が常に問題となる[1]。近年、セマンティック・ウェブの情報発信用オーサリングツール[2]のように、システム側が用意したコンテンツを選択入力させて、意味を確定させる方向性の研究もあるが、操作性にまだ難点がある。一方、日本語は係り受けを括弧で表現する事が容易であり、括弧付コーパスもあるが、入力段階で括弧を付すのは容易ではない。そこで、助詞等の特定の言葉や接続を数学の演算子と見做し、様々な優先順位を定義する事で、自然文から自動的に括弧を付すシステムを提案する。

2. 提案システムの概要

ユーザは数文字の平仮名しか入力出来ず、システム側が用意した単語群から、その文字列を読み先の先頭とするリストが表示され、選択する事で入力とする。読み以外に、システムに登録された接続規則に合致した単語しか入力候補とされない。多義語には意味の違いが付され、選択可能となる。入力途中でもシステムの規則に従った括弧付けが表示され訂正可能となる。

3. 演算詞の定義

自然文の中で、単語と単語を連結する単語を仮に明示的な演算詞と呼ぶ。助詞「と」は、左右に並ぶ単語の並列関係 (A と B、等) を示す演算詞 (A+B の+に「と」が該当している) と考えられる。一方、「美しい湖」等の修飾は、形容詞+名詞の+に該当するのは直接接続で、このような単語ではない演算詞を特に、暗示的な演算詞と呼ぶ。明示的演算詞は、「と」「か」のような連言・選言を示す言葉の他、「ならば」のような含意を示す言葉、「は」「を」のような格関係を示す言葉 (格助詞) 等が挙げら

れる。暗示的な演算詞として修飾関係の他、用言+助動詞、名詞+単位を示す名詞のような「一語を作るレベルの接続」が挙げられる。

4. 演算詞の強弱関係

乗除算が加減算に優先する、のような演算詞の強弱関係を、以下に定義した (番号が若い程、先に単語の結合 (括弧付け) が行なわれる強い演算詞となる。同格の演算詞が並ぶ場合は修飾関係の一部のみ、右優先で括弧が付される。

① 数字を作るレベル

数字 (小数点も含め) の他、「立方」「平方」のように数字しか左にとらず、右に単位を示す単語等が来る単語も数字とした。例えば、365.25平方、の入力ならば、まず(365.25平方)と括弧が付く。

② 一語を作るレベル

「各参加国」の「各」等は右に名詞が来る接頭名詞とし、「子供用」の「用」等は左に名詞が来る「接尾名詞」とした。「個」「メートル」等の単位を示す名詞を単位名詞とした。まず、接頭名詞+名詞、名詞+接尾名詞、数字+単位名詞を、一語 (名詞) を作る演算とした。

否定は、論理式でも単項に付き得るため、否定を示す言葉+単語も、一語を作るレベルと考えられる。名詞の否定は「非」「不」等の接頭名詞で表現出来る。用言の否定は「でない」等の助動詞になるため、用言+助動詞=用言も一語を作るレベルとした。

助動詞は否定以外にも様々な意味のものがあるが、いずれも接続する用言のバリエーションを作るという点で、用言+助動詞は一語を作るレベルと考えられる。なお、「れ・ない・そうで」のように助動詞同士が接続 (この場合は、受身・否定・伝聞) する場合も自然文では多いが、提案システムでは入力の利便性を向上させるため、予め可能な助動詞同士の結合を一語の助動詞として登録した。但し、あまり長いと意味が取りにくい悪文となるため、最大5つの助動詞の結合まで認める事としたが、用意された複合助動詞の総数は1500語を超える。

他に一語を作るレベルとして、以下を挙げた。

A sense resolution method based on complete select input system.

[†]Masaaki Kanakubo

Shizuoka Institute of Science and Technology

- ・ 名詞+断定の助動詞=用言 (A だ、等)
- ・ 名詞+助詞+断定の助動詞=用言 (A の ようだ、等)
- ・ 名詞+らしい=形容詞 (A らしい)

名詞+形容動詞語尾、名詞+する (サ変動詞)、名詞+名詞 (複合名詞) も一語を作るレベルであるが、予め可能な組合せをシステム側が一語として登録する事とした。

以上の暗示的演算詞は全て同格で、連続した場合は、左優先で括弧が付く。(((3 6 5 . 2 5 平方)メートル)) だ) となる (数字・単位名詞・断定助動詞の場合)。

③ 連言・選言を作るレベル

名詞は「と」が連言、「か」が選言の結合を作る明示的演算詞となる。動詞 (又は助動詞) 連用形+て (で) +用言で連言を表し、助詞 (て、で) が明示的演算詞となる (やってみる、等)。選言は連体形+か+用言となる (勝つか負ける、等)。形容詞、形容動詞、連体詞、副詞同士の連言は直接接続の暗示的演算詞となる。選言はやはり、「か」を挟む事とする。

④ 修飾関係を作るレベル

以下をこのレベルとして定義した。

- ・ 名詞+副助詞=名詞 (A だけ、等)
- ・ 連体詞+名詞=名詞
- ・ 副詞+用言=用言
- ・ (左が格助詞以外の) 用言連体形+名詞=名詞
- ・ 名詞+助詞+名詞=名詞 (A の B、等)

これらを連言・選言より弱いレベルとしたのは、(美しい (空と海)) のように形容詞が名詞の連言全体を修飾可能とするためである。最後以外は、暗示的演算詞となる。A の B、等の関係は「の」が明示的演算詞となり、他に「による」等の名詞で名詞を修飾させるための演算詞グループが考えられる。繰返し (形容詞+形容詞+名詞、等) も必要となるが、複雑な悪文の生成を避けるため、重複は三語までとする。

⑤ 修飾より弱い連言・選言を作るレベル

形容詞+複数の名詞で、形容詞を左から一定の名詞にのみ掛ける場合に使う連言として「及び」、選言は「若しくは」を定義した。美しい空及び海、の「美しい」は海には掛からないものとする。他の修飾関係に対する効果も同様とする。なお、用言も同様に「及び」「若しくは」で繋ぎ、「ゆっくり歩く及び休む」では、「ゆっくり」は「歩く」にだけ掛かる。

⑥ 複数の格関係を統合するレベル

「私は食事を～」のように名詞+格助詞+名詞+格助詞を、一つの名詞+格助詞に統合する。

これも悪文を避けるため、名詞+格助詞の繰返しは最大四回までに限定する。

⑦ 格より強い修飾を作るレベル

名詞+格助詞+用言連体形+名詞、のような場合 (「音楽を好む人」等)、名詞+格助詞+用言連体形が、右の名詞を修飾する関係となっている。格関係は修飾より弱いのを原則とするが、ここでは格助詞が修飾句に含まれているため、このパターンで配置されている格助詞に限り、通常の格助詞よりは先に括弧を作る強い演算詞としなければならない。なお、「景色が美しい町に住む人」のように、この修飾関係が重複する場合もあるが、意味の取りにくい悪文生成を避けるため重複は三回程度に制限する。

⑧ 格を作るレベル

動詞句と名詞句の関係を定める、主格や道具格等の格関係が文の骨格となるため、格助詞は①~⑥より弱い演算詞と考えられる。格文法では原因も格の一種なので、「ならば」等の含意関係も同レベルとする。また、副詞句も文全体を修飾する場合は格関係と同レベルと考えられ、例えば、用言+接続助詞+用言 (「読みながら食べる」等) の接続助詞も、このレベルの演算詞となる。但し、文+ならば+文等の表現を可能とするため、格よりも弱い⑦と⑧の中間レベルの含意を表す言葉 (格と同レベルと違う言葉になるように「だから」等) を用意する。

⑨ 文を受けるレベル

文間の係り受けを明確にするため、文+という事+を××とする、の形を用意し、一文の内容を××に代入された文字列で表し、他の文での使用を可能とする。この「という事」等が、文全体を受ける最も弱い演算詞となる。疑問文を作る「か?」「か?という疑問」等も、同レベルに用意する。また、「という事」等+格助詞+用言の形も用意する。

5. おわりに

今後の課題として、無作為抽出した文を対象に、提案システムの用意した単語、繰返し回数制限の下で、どの程度、十分な入力が可能であるか試験評価する予定である。使い勝手についても調査、改善したい。

参考文献

- [1] 「自然言語処理」、長尾真編、岩波書店
- [2] 「オントロジーと制約に基づくセマンティックプラットフォーム」橋田浩一、人工知能学会誌 Vol. 21, No. 6, pp712--717, 2006.