

# コンピュータの使う言葉

## Languages computers are using

宮澤 彰  
Akira MIYAZAWA

### どんな研究？

コンピュータの言語というと、通常プログラミング言語を考えるが、データやプロトコルメッセージ等は、人間とコンピュータ、コンピュータ同士で情報を伝達する言語と考えることができる。このように考えると、言語学の知見が応用できる。

### データは言語か

- プログラム相互、プログラムと人間との間で情報を伝達する、という点ではデータも言語と同様の機能を持つ。ただし、特定のプログラムにしか理解できないデータを言語と考えても、あまり意味はない。ペットと自分の間にしか通用しない言語があるといっているようなものである。
- 言語と考えて意味があるのは、ある程度以上の語彙、文法および話者のコミュニティーを持つときである。
- セマンティック・ウェブなどにより、データの公開が広がってきたことにより、言語として考えることに意味があるような、語彙、コミュニティーの広がりを持ったデータ群が、出現しつつある。

### データを言語と考えると

- もちろん、これは人工言語であるが、自然言語で見られるいくつかの現象は、データ言語でも現れることが予想される。
- どのような種類の現象が現れるか？また、どのような種類の現象は、現れないか？を明らかにすることにより、今後の処理系の設計や、起こりうる問題点の予想をつけることができる。

### データと言語のアナロジー

#### ■ セマンティック・ウェブのRDF

- ◆ RDF は、文法を制限し、文を述語論理の statement に還元して表現する。また、その statement の subject, object のみでなく、述語(property)も URI を用いて表現する点が特徴である。すなわち、RDF における URI は、自然言語における単語にほぼ相当する。
  - ◆ RDF における URI は、基本的には誰でも作成することができる。URI を用いてウェブにアクセスすることにより、その意味を自然言語で記述したものが得られるようにすることが求められているが、これは人間の利用のためであり、データ言語の上での意味を形式的に定義しているわけではない。
  - ◆ このことは、URI における意味の獲得は自然言語の場合と同じように、コミュニティーによる受容と、usage によって決まると考えられる。  
結果として、コミュニティー内での標準化と逸脱という自然言語と同様の過程が生じることが予想される。
  - ◆ さらに、その結果として、ひとつの URI に複数の意味が当てられる homonym や、同じ意味の語が複数存在する synonym などの現象も避けられないという予想もできる。
  - ◆ このことから、さらに予想されることとして、次の二つの方向があり得る。
- RDFを処理する処理系は、homonymの曖昧性解決や、synonymの正規化などの自然言語処理と同様の処理機能を要求されることになるだろう。
  - RDFの上に構築される言語においては、コミュニティー内でのusageの標準化努力が必要となるだろう。

### 上記予想から導かれる疑問

- それでも、これらの(RDFに基づいた)人工言語は自然言語を処理するのよりも有効であり得るか？  
(個人的な予想では回答はyesであるが、今後の展開とより詳細な検討を必要とする。)