

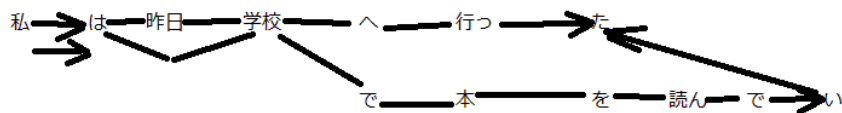
自然言語処理における、文法の獲得技術を考察します。

1. 文法獲得の問題点

文を単語に分解することと単語の意味を獲得することとは交差法で良いとして、それから単語表出の前後関係を単語間の線で結び、ネットワークにして、そのパスの出現頻度をカウントアップしていく・・・というのが文法獲得の基本でしょう。ですが、それまでに出現しない単語についても、前後関係を推論できるという、抽象化・・・完全な文法の獲得・・・は難しい。英語の冠詞を例にとっても、**an apple** はネットワーク法で獲得したとして、**orange** ではどうかといった問題は、もっと高い位置からの抽象化規則の獲得が必要ですね。単数という概念を獲得し、「存在」を表し、そのものという意味では **the apple** だとかを獲得する。多くの中にあればいいというものと、多くの中のこれというもの・・・そうした概念の獲得が必要です。文法にそれが関わってくる。たんなる確率統計ではなくなってくるのですね。日本語では、「登ったことのある山」という風に、「山」に動詞が掛かるときには、動詞の終止形が用いられますが、これは名詞に係るときには意味が完成しているからという風な概念の取得で説明できるとして、「登ったら」と言った場合、「ら」に「登った」という終止形が接続すると解釈したとして、「登るら」と言えるか・・・という問題が発生します「美しら」でなく、「美しかったです」です。「たら」が助詞なのだという根拠になることですが、それもあやしい。そんな概念の獲得が可能なのかということですね。なぜ、「ら」に終始形で動詞や形容詞が接続できないのか・・・過去形でなければダメなのか。文法の獲得はとにかく、難しいファクターがいっぱいあることは確かですね。

文：私は昨日学校へ行った。

私は学校で本を読んでいた。



単語出現ネットワーク

## 2. 文法獲得における基本事項

文法獲得において基本となる事柄をあげてみたいと思います。

### (1) 認識の基盤はコマンド

全ての自然言語の認識は（動作[、格、オブジェクト]・・・） [^/|(動作・・・)]・・・の形式のコマンドであること。「ある一つの」とか「すべての」もこのコマンド形式で表現します。特別扱いはしません。

### (2) 表層表現で文法に関係するものは次の通り

- ・ 文、句、単語の配置順序
- ・ 情報の区切り、塊としてのまとまりがあること
- ・ 修飾
- ・ 屈折
- ・ 特別語の挿入

これらいずれかの方法で文法事項は表層に現れてきます。

### (3)

表層表現に関与する大きな事項

- ・ 数
- ・ 性
- ・ 格
- ・ オブジェクトや行動の修飾
- ・ その他、属性

性については、客観性があるのは動物についてのみで、月や太陽などに性があるかどうかは言語によります。どちらかという、**der/das/die** とかの選択の違いとして言語毎に学習していくものになります。

### 3. 文法獲得の手順

#### (1) 単語を切り出す

文のビックデータを交差法により単語分割して単語を得て行きます。

#### (2) 意味を単語、句、文に与える

センサーやその他身体感覚、オントロジーを基盤にして、それと文との共起関係を記録して行きます。その記録の交差法によって単語、句、文の意味を推定して行きます。

(3) 単語の配置については、単語出現ネットワークから得られます。固定的なのか、自由な配置関係にあるのかといったことが、2つの単語関係から得られます。また修飾関係もこのネットワークから得られます。

(4) a や the みたいな特殊な単語が現れたら、その意味も交差法で同定できます。それは、こうした単語の出現によって「それは何かな」というような特別の関心によって注意して意味をすいていすることになります。屈折も同じです。

(5) 格は動詞の支配の領域・・・句として意味の交差法から得られます。

(6) 学習に使ったビックデータの単語以外の単語が出てきた時にどうするかという問題があります。それは、その単語の意味の最も一致する単語出現ネットワークに従うというのですが、あらかじめ文法規則として抽象化していくことも行われるでしょう。それは、単語の対について、得られた規則の交差法によって、行われるものです。同じ単語に幾つかの単語が前に来たとして、その前に来た単語の文法を交差法で、基本的なアンド条件で・・・共通な属性による文法決定要項を得て抽象化することができます。その要項を満足すれば、新しい単語でも同じ条件で後続の単語に接続して行くと推論できます。

### 4. おわりに

全て交差法で文法も獲得できることが示せました。これで、全ての自然言語の動きを学習するアルゴリズムができたと思えます。

おわり