

考察：「談話処理 ～「オブジェクト指向自然言語処理」の補稿」

前回の考察で、文章解析だけを対象に議論してきました。やはりなんといっても、談話は議論から落とせません。今回は談話処理について考えていたことを記述していきたいと思います。

談話は次のような特徴があります。

(1) 文法はあまり頼りにならない。

「その本取って」とか、助詞は大抵省略されますし、「山緑鶯なく」なんて、活用語尾がなかったり、同音異義語の処理がきつかったりと、処理を難しくします。

(2) 不要な語がいっぱいはいる。

「あー、うー」とか、「でね、でね」とか。これらをフォーカスを促す記号とか、感情をくみ取るキーにしていかなければなりません。

(3) 基本的に会話処理が必要

5W1Hとかを把握して、相手から足りない情報を聞き出すことが必要です。相手の価値観とかも推測していく必要があり、文章解析にプラスアルファな機構を作り込む必要があります。このプラスアルファがきつかったりする。

第1章 談話処理概要

例によって、システムのイメージを考えてみたいと思います。次にAとBの会話例を提示します。

【ある日の会話】

(1) A.長野へ行ってきた。

(2) B.長野？

(3) B.どうして？

(4) A.恵比須講だからさ。

(5) B.なんでいったの？

(6) A.歩いていった。

(7) A.あー、大変だった。

(8) A.妻と行ったんだ。

(9) A.妻は歩くのいやだといったんだが、我意を通した。

(10) A.歩くのはいいからね。

(11) B.体にかい？

(11) A.あーい、こっちだ、こっちだ。

(12) A.昨日の恵比須講の話をしていたのだよ。

(13) A.君もお茶にしよう。

この会話には A と B、それに A の妻が登場する。人は情報保持オブジェクトとして、フレームで表現できるでしょう。ここでは、3つのフレームが登場するということです。

この会話は A フレームと A.妻フレームの情報をいかに B フレームにコピーするかという問題を解くこと・・・と定義できます。むろん、A,A.妻そして B には常識という知識ベースが装備されているわけです。

以下に、フレームの動作を追ってみましょう。

【フレームの変化時系列】

(1) A の初期フレームで、B に話題のステージ情報を与えたというフラグが保持される。

A フレーム：ステージ（昨日、家から恵比須講会場（長野）への移動）

シーン：移動

歩く

with 妻

筋肉痛の感覚

アクター：私、妻

私：恵比須講に情熱

妻：歩くの嫌い。

同伴した。

シーン：出がけ

私：歩いての恵比須講見物を提案

妻：歩くのに反対

シーン：道中

私：歩く。

筋肉痛

妻：歩く。

B フレーム：ステージ（時間不明、長野）

(2) B はステージを確認。ここは、「いつ？」という問いが適切かも。

(3) B は時間よりも目的情報を重要と判断した。

(4) A は出がけシーンから目的が「恵比須講見物」であることを知る（意味解析が必要）。

B フレーム：ステージ（時間不明、長野）

シーン：道中

目的：「行く」の目的は恵比須講見物（買い物かもしれない、催しものがあるかもしれない・・・曖昧状況）

手段：不明

アクター：A

A フレーム：理由情報を与えたことをフラグする。

(5) B は目的の次に手段情報が重要と判断した。

(6) 「なんでいったの？」は手段を問うているとも、理由を問うてるとも受け取れるが、理由情報はもう与えてあるから、手段を問うてると判断。動作「歩く」を回答。

B フレーム：ステージ（時間不明、長野）

シーン：道中

手段：歩く。

アクター：A

A フレーム：手段情報を与えたことをフラグする。

(7) A は情報として、気まぐれに「大変だった」という属性情報を伝える。

B フレーム：ステージ（時間不明、長野）

シーン：道中

手段：歩く

アクター：A

属性：大変（ネガティブな感覚）

A フレーム：動作属性を伝えたことのフラグをあげる。

(8) A は情報として、アクターに妻があったことを伝える。

B フレーム：ステージ（時間不明、長野）

シーン：道中

手段：歩く with 妻

アクター：A、A.妻

A フレーム：アクターに妻もあったことを伝えたフラグをあげる。

(9) A は「歩く」前に諍いのシーンがあることを提示する。

B フレーム：ステージ（時間不明、長野）

シーン：出がけ

行動：争う

アクター：A、A.妻

A.妻：歩くのいや。

A：歩きたい。我を通す。->「我を通す知識フレーム」

A フレーム：出がけの状況を話したことのフラグをあげる。とくに、妻が反対したことについては伝えたことを示す。

以下省略。概要は把握できる例だったと思います。重要なことは、

(1) 5 W 1 H の把握

(2) 理由、原因、結果の把握

(3) 事象の連鎖の時系列の把握

(4) 伝えたことの記名

同じことを繰り返し伝えないようにする。

(5) フォーカスの管理

- ・話題の流れ
- ・話者の価値観
- ・相手の価値観の推定
- ・割り込み処理

第2章 談話文の解析手法

次のような例文を想定してみましょう。

(例文1) 山美しい登った。

(例文2) 山登った美しい。

いずれの例も何となくわかる文です。

(例文3) 山登る朝楽しく10時だった。

頭痛くなる文ですが、なんとなく雰囲気は伝わってきます。

2.1 処理イメージ

これは、文法が変でも、ある程度の情報は得られるものであることを示します。だから、談話処理では、組織だたない意味の羅列から、組織だったフレーム構造まで、複数のレベルも保持すべしということではないでしょうか。

一番上のレベルは、意味記号の単語対訳した羅列です。例文1ですと、

(意味解析1) mountain:(ND;+object,+location),beautiful:(AN;+favorite,+sight feeling;modify ND),clime:(VB;+move;+up on (object,location);verb frame);

なんて感じですが。これで十分、情報の検索に利用できそうです。

さらに深くですと、共起関係の連想で、修飾関係とか、動詞フレームの解決ができます。

(意味解析2) mountain (modify by "beautiful")

clime mountain

actor("私"; maybe by default)

このフレーム構造を作れるのは、共起情報(プライミング情報)と各単語の意味を解析して情報を得られることが前提となつてのことです。ここまでくると、情報の構造も解析できそうです。

こうした、厳密な文法・意味解析をしないレベルの処理というものを、厳密な処理と平行して実施していくのが正しい方法論といえます。

2.2 曖昧性処理

ただこの手法は曖昧性が大きくなります。そして、会話が進むほどに、意味フレームは厳密になっていきます。そういう手法を取らねばなりません。ステージとかシーン、アクターとかの動画システムを更新していく必要があります。デフォルト情報で始めは埋めておきますが、それはあくまでも推測、フラグ("maybe by default")をたてておきます。特

に、曖昧性が気になるころでは、曖昧性フレームにして、解決を鍵と鍵穴の手法を使って解決します。鍵、鍵穴もどんな情報を待っているかを示すフラグです。そうして、デーモンプロセスによって、鍵を与えられて、曖昧性を解決していくわけです。

1つの種類の鍵穴に対して、親デーモンを1つ生成させ、親デーモンは関連情報を持ったフレーズに遭遇すると、専門の解析ツールを次々に生成させていく。自律システムを目指していますので、デーモンは、また、意味プリミティブを定義するプロセスとして作り込み、これ自身も学習によって成長していくものでなくてはなりません。だから、プリミティブプロセスはパラメーターデータで制御できる構造になっているはずで

2.3 比喩表現

会話では比喩表現が頻出します。「お茶にしない?」とか、「花」を見に行くとか、「手を貸して」とか。「長野は美しかった」といえば、「長野の街の景色が美しかった」でしょう。「長野は美しかった」という会話があって、あとで「彼は街の景色はどうだと言った?」と問われて、「長野」のことかと判断して、「きれいだったと言っていたよ」と答えられねばなりません。

比喩には、つぎの3種があります。

(1)メタファー：ものの属性の類似性による連想(例:「月」と「女性」)

(2)シヌクドキー：同種の中の連想(例:「花」と「さくら」)

(3)メトニミー：時間・空間てきなまとまりによる連想(例:「手を貸して」は「体全体で手伝って」の意味)

こうしたものは、すべての関連情報を持たねばなりませんから、フレーム問題の困難に遭遇します。イメージを基本的な情報とする「文脈空間」を構築する必要がありますが、今はまだ実現できていないので、文の学習システムでこつこつと情報を収集することになります。

基本的処理としては、部分から全体を把握する機構を設けてのものになります。意味記号から全体枠(オブジェクト)を検索する問題です。これは議論済みですね。

第3章 まとめ

前回の「オブジェクト指向自然言語処理」では文法が正しい場合のみを扱いましたが、あとで、それではまずいと思い、考察を深めました。その結果をまとめてみましたが、どうでしょうか。本当の困難は、これらを作り込んでいくところで見えてくるのだと思います。まずは小さなシステムから作り込んでいきたいと思っています。無論、文法を厳密に処理する手法と、ここで述べたテンプレートマッチング手法を平行に実行するような、分散問題解決を実施するつもりです。

おわり