B 級機関
〜 コンピュータにおける「言語感覚」の目覚め 〜

金杉 友子
NTT アドバンステクノロジ
松澤 和光
NTT コミュニケーション科学基礎研究所

1　はじめに

「B 級機関」— 駄洒落を延々と出力し続ける永久機関 — は、98 年 2 月「ことば工学」を提唱する [3] にあたり、その「ことばの感覚面を工学的に扱う」コンセプトをシンボルするものとして制作された一種の “美術品”である (制作協力:(株) ボストーク)。したがって、このオブジェから何を見てとるかは、本来、観客 (?) の感性に委ねられているワケであるが、ここでは制作者側からのひとつの問題提起として、この制作・実演を通じて考えて来たことを述べてみたい。

2　B 級機関の仕掛け

B 級機関が吐き出す駄洒落には、Figure 1 に示すような 4 つのパターンがあるが [1]、どれも基本的には

- 似た音の 2 語の組み合わせ
- ある文の文末に似た音の語を無理矢理入れる

という形で出来ている。そこで、このメカニズムによって「駄洒落」が生成されるための必要条件について考えてみると、

A) 駄洒落自体に制約があること
　ひらがな表記、あるいは発音の一致 / 類似

B) 文として自然であるための秩序があること
　文法的な制約、品詞と助詞の接続関係、等

では、これらの条件を満たせば駄洒落が生成できるだろうか。実際、B 級機関の出力を見ると (Figure 2) 人間の考えるようなものとは全然違う変なものばかりでてくる。もちろん条件を満たすための技術が未熟なことも事実ではあるが、それだけでもないようだ。では何が違うのか?

パターン 1： 河馬（かば）の蒲焼（かばやき）。

パターン 2：「飴（あめ）のない旅に出る」なんてね。. . . はっ、「当て（あて）のない旅に出る」だろって. . .。

パターン 3： 一日の長↓
　　一日のチョーヨンピル

パターン 4：
B: まぁまぁ、落ち着きぃな。青い（あおい）から始めようや。
T: そうかー。ほな、いくか。
T: せやねぇ、まぁ、青い（あおい）と言うたら 未熟（みじゅく）やな。

B: それって、夜、野外で寝ること?
T: ちゃうちゃう。それ 野宿（のじゅく）やがな。
　　:
T: 君とはようつき合っておれんわ！

Figure 1: B 級機関が吐き出す駄洒落

3　コンピュータの言語感覚を考える

コンピュータが自然言語を扱う時、当然のことながら人間とはやはりズレがある場合がある。それを今まで「正しくない !!」として、規則や文法を直す方向ばかりを考えていたが、本当にそれだけでいいのだろうか。ここでは、新しく以下の 2 つの方向を提起したい。

(1) ことばの感覚的な側面も扱えるようにしてズレを直す

発音、社会的な意味合い (方言や立場によるつかいわけ)、文字の形、その語が含む 2 重 3 重の意味、ことばの連想など、従来の自然言語処理では扱わなかった感覚的な側面について考慮した処理が必要なのではないか。

「人間ってそういうところで、ことばかわしてるんだし。」

(2) コンピュータによるズレ自体に価値を見い出す

コンピュータが自然言語を扱う際の、我々が日常に行なっている言語活動とのずれ自体が面白いのでは?、という話。そもそも人間が (自身、よく理解できていないのに無理に) 規定した「ことばの枠組み」の中でことばを扱っている以上、人間が普段使うことばとは違って当然であり、だからこそその違い自体に価値を見い出すべきではないか。

上記 (1) は「ことば工学」が扱う主要テーマとして、既によく了解されていることかと思う。以下では (2) の視点について詳述したい。コンピュータによる「ことばの取扱い」について、ここでは特に 2 つの特徴に注目する。

「一（いち）が切れる」なんてね。. . . はっ、「息（いき）が切れる」だろって. . .。

「垢（あか）よろこび」なんてね。. . . はっ、「糠（ぬか）よろこび」だろって. . .。

「舌（した）不明金」。そりゃ「使途（しと）不明金」じゃないか. . . . はぁぁ。

Figure 2: B 級機関が吐き出す駄洒落

a) データの多量性

コンピュータであるから (外部記憶まで含めれば) とにかく無尽蔵と言ってよいくらい膨大なデータを扱える。もちろん、人間が実際、どのくらいの量のことばを記憶し扱っているかは不明ではあるが、B 級機関等で元としている国語辞書データのような細かな情報を一々取扱っているとは思えない。例えば、鋏と言うことばの特徴を表すデータ (属性) を辞書等から収集した場合を考えてほしい。「鉄」やら「切る」やらの属性に混じって、突然「足」と言う属性が収集されていたりする。これは実は「蟹のハサミ」が「甲殻類の足」に相当すると言う百科事典的な知識があったためであった。こうした (普段は余り人間の意識に上らないような) 細かなデータが複雑かつ律儀に絡み合う時、コンピュータが思いも寄らない振る舞いをしたりするのである。

b) 機械的な処理

これも当たり前のことであるが、コンピュータは人間が与えた秩序、それだけに従ってことばを処理しようとする。そして、もちろんそれは人間の言語活動をカバーできるほど緻密ではない。例えば B 級機関が出す駄洒落のパターンのひとつでは、「連想的な」機能を使っている (パターン 4:「青い」と言ったら「未熟」、等)。この機能は上述のような属性データを元に機械的な処理で近似的に実現したものであって、あくまで「連想的」、人間の連想そのものではない。場合によっては「青い」から「情事」という言葉が「連想」されてしまうこともある。

この 2 つの性質があいまって何か秩序に従っている結果を自然言語としてコンピュータが出すこと、えてしてそれは人間の考える正しい文章からずれたゆがんだものになることに多数の人が面白みを感じている。

例えば日常的に遭遇するワープロのおかしな誤変換 (「電線に雀が三羽止まっていた→伝染に雀がサンバと待っていた」のような) を収集する WWW のサイトは数多く存在し、中には 10 万アクセスを超える人気を誇るものもある[1]。また、宮沢は [2] 機械翻訳が英文を変換した際に出力される日本語が不思議と魅力的であることを記している[2]。

つまり、これは言葉をとらえる枠組自体が自然人間とは違うのだから、どこまでデータをつめていっても、ルールを仕組んでいっても所詮人間とはズレたものになるわけであり、「正しさの追求」とは別の角度からの研究が必要なことを示唆していると思われる。

[1] 代表的なものとして、世間虚仮 (http://www.wind.co.jp/sennen/)、笑える誤変換のページ (http://www02.so-net.ne.jp/~hoshi/henkan.html) を挙げておく。

[2] なお宮沢の主宰する劇団「遊園地再生事業団」のサイトには、機械翻訳ばりの日本語を駆使する人の目撃談が紹介されているので参考までに。(http://u-ench.com/b%7Cn/index04.html)

4　おわりに

今までの自然言語処理は人間の言語活動における「正しい文章」を考える方面ばかりに目をやっていた。ただ、ことばが文法や規則から離れた意味やニュアンスを扱いだすと何が正しいという「正しさ」さえ微妙になってくると思う。

工学としてこれはどこまで追求するものなのか不安はある。しかし、その得体のしれない感性的なものを、コンピュータに扱える話にしてデータとしてつめこむ、ルールを書くといったことをすると、「コンピュータによる人間からちょっとずれた奇妙な日本語」なんて新しいものがひょっとしてできちゃう可能性はないだろうか? そして、それがことばとして逆に魅力的に見えたりはしないだろうか?

もちろん以上の考察は、まだ問題提起以前、議論の取っ掛かりに過ぎない。言語感覚の何たるか、またそのコンピュータにおける実現と新しい利用価値を目指して長い道程が始まろうとしている。先輩諸兄の忌憚ないご意見を賜りたい。

References

[1] 松澤 和光, 金杉 友子 and 阿部 明典: コンピュータ上の言語感覚実現に向けて: *B* 級機関, 人工知能学会全国大会, *37-11, pp. 685–686* (1998)

[2] 宮沢章夫: 牛への道, 新潮文庫 (1997)

[3] 「ことば工学への招待」, コミュニケーションとメディアについて考える, 第 *1* ステージ, *ICC NEWSCHOOL'98* (1998)