



ことばのひびき～言語処理の音韻的側面

松澤和光（非文字資料研究センター研究員）

言語の研究者である私が、よりもよって「非文字」の研究に関わっている不思議はさておいて、いや、よく考えてみれば確かに言語の出自は文字ではない。原初人類の声による意思疎通が文字の形に定着したのは、とうぜん言語の誕生よりずっと後のことに決まっているが、何にせよ例外はある。

まず、私の本来の専門分野たるコンピュータ科学、そこにおける「プログラミング言語」はまさにその例外であろう。この言語は文字の姿で生まれ、音声による読み上げを想定していない。プログラミング言語は数多く開発されて来たが、発音記号付きの用語集やプログラムの正しい朗読法などと言ったものには、未だかつてお目にかかったことがない。しかし教育の現場では、学生に向けてプログラムを語って聞かせないわけにもいかぬ。そこで整数を意味する用語「int」を、さて「イント」と呼ぶか、integer の略なら「インテ」かしら、などと思ひ悩むことになる。（って私だけか？）

別の例外として現代日本語を挙げよう。もちろん、言語としての日本語は（成立の詳細はともかく）声から始まった筈であるが、問題が起こったのは明治維新であった。列強から怒涛のごとき新概念の侵略が起こったとき、維新の英雄たちはこれらを片っ端から新造語として国語に取り込んだ。その英断と努力は敬服に値するが、惜しむらくは新造語の構成漢字を意味だけから選んだこと。読みがどうなるかまでは手が回らなかったに違いない。さらに言語としての音素種の少なさも要因に加わって、日本語には大量の同音異義語が出現する。鈴木孝夫氏が様々な著作で提唱する「テレビ型言語」[1]なる日本語の特徴である。すなわち、例えば「こーこー」という発音を聞いたとき、それが高校・航行・孝行・煌々など数ある単語のどれなのか、私たち日本語話者は瞬時に漢字群を映像的に走査して判断しているのだという。

本来、音から発達した言語であれば、単語は全て異なる発音で区別されるべきだろう。現代日本語における同音語の多さは言語として異常とも思われる。そして、こ

の特徴は思わぬ副産物を産み出したのではないか。同音語を自在に利用した言葉遊び — 語呂合わせと駄洒落の文化の隆盛である。やれやれ、やっと私の研究を語る場所に辿り着いた。

白状しよう。冒頭で、言語の研究者とかコンピュータ科学が専門だとか色々気どってはみたものの、私はコンピュータによる駄洒落生成マシン～B級機関の製作者としての黒歴史？を負っている。石部金吉氏揃いの工学世界において、そんな怪しい研究者がどんな扱いを受けたか、シクシク……。いーんだもん、人工知能学会から賞まで貰ったもん、ベスト・プレゼンテーション賞だけど [2]（ま、駄洒落がプレゼンかは置いといて）。

さて、B級機関の雄姿を図1に示す。ピラミッド型の内部ではコンピュータがひたすら駄洒落を創り出し、これが永遠に吐き出され続けるという、永久機関ならぬB級機関。自動生成された駄洒落の一端は図2に示した。実のところ、駄洒落の生成自体には研究的意義はない。私の秘かな狙いは、駄洒落に代表される音韻的／感性的な日本語の言語感覚と、その工学的な取扱いの可能性を指摘することにあった。（すみません、今更ながらの言い訳です。）

ソシュールの言語哲学に拠れば、言語記号は表記（シニフィアン）と内容（シニフィエ）の2面を持つというのが定説である。「イヌ」という音韻表記は、「犬」が指し示すあの人懐っこい動物の概念内容とは独立しており、言語によってはその生き物を「ネコ」と呼ぶことも充分あり得るわけだ。言語をコンピュータ処理する場合も、犬は例えば N1002334 などと無機質に記号表現され、そこに本来の読み情報に関わる余地はない。「音声読み上げを想定しない」プログラム言語で操作される以上、イヌというシニフィアンが忘れ去られるのはある意味必然であったのかも知れない。こうして現代における人間言語のコンピュータ処理（「自然言語処理」と不自然に名付けられている）では、処理の早い段階で言語の音韻的側面は姿を消し、もっぱら無味乾燥な記号対象と

して検索やら翻訳やらの操作が施されている。言語で表現される本質がその概念内容にあるとすれば、シニフィエ偏重の処理も当然の帰結と言えるだろう。

しかし、人工知能の立場から人間の思考行為を担う基盤としての言語を考えると、その音韻的特徴を簡単に捨て去って良いのか疑問が生じる。イヌのことを考える際に、その音韻的類似性からフツと「居ぬ・去ぬ・稲」などの連想が働きはしないか。頭の中でネコが勝手に寝転んだりはしないか。語感から来る連想が、思考を発想豊かに展開し、推論を情感たっぷりに方向付け、判断を臨機応変に下す手助けとならないだろうか。同じ意味の文章でも、なぜ文豪の作品は心打つのか、名演説に心動かされるのか。息の合った友人らとのおしゃべりに、こんなにも心弾むのはなぜなのだろうか [3]。言語と思考において「音のクオリア」[4] が果たす役割の重要性を、私たちはもう少し認識し直すべきかもしれない。

以上のような考察を背景として、現在、松澤研究室では「駄洒連具」と名付けた言語処理ツールを開発している [5]。これは単語間の音韻的な類似度を計算するもので、既存の歌詞とよく似て聞こえる文章にニュース記事を整形したり、聞き間違え易い単語をチェックしたりといった応用を模索している。さらに、別途開発中の単語間の意味的な類似度を計算する「類神具」と組合せて、人間の感覚に則した言語処理を目指していきたい。こうした取り組みが、やや取り残された感のあるシニフィアンの言語処理の新たな発展に寄与できれば幸いである。これが単なるC級機関の作製で終わっては、洒落どころか「謝礼」にもならない。オアトガヨロシイヨウデ

参考文献

- [1] 鈴木孝夫：日本語教のすすめ、新潮新書、2009
- [2] 松澤・金杉・阿部：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、1998
- [3] 梅田規子：おしゃべりはリズムにのって、オーム社、1997
- [4] 黒川伊保子：怪獣の名はなぜガギグゴなのか、新潮新書、2004
- [5] 古田・天沼・松澤：ニュース記事を歌詞化するシステムの提案、第40回ことば工学研究会、2012



図1 B級機関

猫をかブルスウィリス
 好きこそもののジョニーデップ
 口は災いのモト冬樹
 憎まれっ子世にはばカルロスゴーン
 ーを聞いて十をシルベスタスタローン
 くさってもタイガーウッズ
 . . .

図2 自動生成された駄洒落群

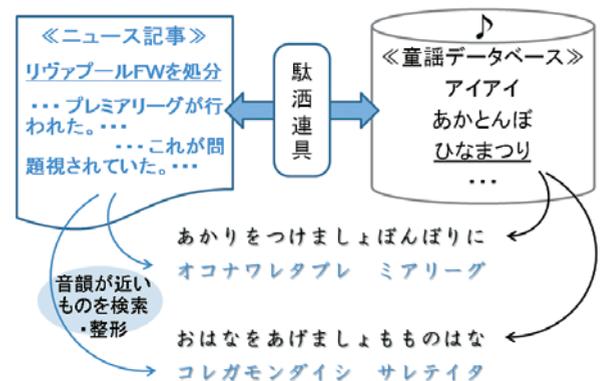


図3 ニュースの歌詞化システム