「言語意味成分の考察」

□「単語」の意味を調べる時に我々は辞書や辞典を紐解く。辞書等には「見出し語」があいうえお順に並んでおり、その見出し語には語釈文という文で意味を説明している。すなわち、「意味＝語釈文」は「単語で生成」されているので、「単語の意味」は「単語自身で階層化」されているということになる。単語の意味を「意味成分」という「意味素」で構成（生成）されているとする。この「意味素」をΔと記すと、このΔはｎ+１個の有限個軸を基底とする最小の凸多面体と定義でき、Δ^nと記す。

∴この意味素であるΔ^nは、

$\Delta^n = \{(x_1, x_2, \cdots, x_n) \in R^n \mid x_1, x_2, \cdots, x_n \geqq 0,\ \ x_1+x_2+\cdots+x_n \leqq 1\}$

と定義できる。

これを「n次元単体（略してn-単体)」といって、「意味素」も単語の一種であるので、この意味素の集合体である「単語」のことを「n次元位相多様体」とみなすことができる。

「位相」としたのは、「単語」を構成している「意味素」どうしが「開集合」として「どのような繋がり方」をしているか…であるから、

Δ^1_0とΔ^2_0を説明し易く、2 次元の（x, y）と（s, t）の座標軸をもっているとすると、$\phi : \Delta^1_0 \to \Delta^2_0$は、$\phi(x, y) = (s, t)$となり、$\phi$は「連続写像」であるから、「位相」構造が入る…といえる。

また、「近傍」と「開集合」とは、「同値」である。一般に「多様体」とは、次元の高い図形を考えると判り易いが、「体」の定義は複雑（公理系Ⅰ=加法に可換群，Ⅱ=乗法に可換群，Ⅲ=分配法則）なので簡単に、「四則演算で閉じている」集合である…と考えられ、ここでの四則演算とは、文法などの写像操作関数を示す。「単語」は「n次元位相多様体」と定義できましたが、「単語は何だかの階層化されている…」という感覚的な疑問が湧く。

すなわち、「単語の意味が階層化」されているとは、いわゆる言語系でいわれている「意味概念体系」のことで「シソーラス辞書」などのことを指し、「上位概念／下位概念」といわれている。「上位概念の単語は、下位概念の単語の意味範疇を含んでいる」単語$w_1 \supseteq$単語w_2と記述される。…から上位／下位といわれているので、上位概念の単語を何だかの関数で作用（微分…をイメージする）させると下位概念の単語が抽出される。

これをδと記すると、

$\delta(\langle A_0\ A_1 \cdots A_n \rangle) = \Sigma\ (-1)^i \langle A_0 \cdots A_{i-1} A_{i+1} \cdots A_n \rangle$

と定義でき、δ（単語$_1$）=単語$_2$となり、「単語$_1$」⊇「単語$_2$」である。

$A_0\ A_1 \cdots A_n$は、基底。

∴単語Ｃは、$\delta(C_n) = C_{n-1}$となり、

$C_n = z\Delta^n_1 + z\Delta^n_2 + \cdots + z\Delta^n_r$　と

$C_{n-1} = z\Delta^{n-1}_1 + z\Delta^{n-1}_2 + \cdots + z\Delta^{n-1}_k$　　（$z \in Z$）

　と書くことができる。
これは、単語Cの系列として、

$$0 \xrightarrow{\delta_{n+1}} C_n \xrightarrow{\delta_n} C_{n-1} \xrightarrow{\delta_{n-1}} C_{n-2} \xrightarrow{\delta_{n-2}} \cdots \xrightarrow{\delta_3} C_2 \xrightarrow{\delta_2} C_1 \xrightarrow{\delta_1} C_0 \xrightarrow{\delta_0} 0$$

となり、これが、「意味概念体系」となる。
最下位概念の単語の下位概念は「なし：０」であり、最上位概念の上位概念は
　$\delta_{q-1}(\delta_q(c))=0$ （$c \in C_q$）であるので、０である。
この意味は、2回操作すると０になるとは、$c \in C$が「閉じている」ことを意味する。
それでは、この定義された体系で、「意味の概念」をどのように表現するかというと、
「上位概念と下位概念との相関関係」を$\delta_n : C_n \to C_{n-1}$で群の準同型写像とすると、

$$Z_n(C_*)=\mathrm{Ker}\,\delta_n=\{x \in C_n \mid \delta_n(x)=0\}$$ （注：０とは「基底」でもある）

$$B_n(C_*)=\mathrm{Im}\,\delta_{n+1}=\{\delta_{n+1}(x) \mid x \in C_{n+1}\}$$

$$H_n(C_*)=\mathrm{Ker}\,\delta_n／\mathrm{Im}\,\delta_{n+1}=Z_n(C_*)／B_n(C_*)$$

として定義できる。
これは、其の単語の上位概念と下位概念が判れば、当該単語の意味概念は、
そこから抽出できることを意味している。
もうひとつは、「深層格」の対応付けです。これは、構文解析での「係り受け」の意味処理レベルでの拡張と考えてみることができる。研究されてきた「固有表現の抽出」での固有表現の種類は、IREXなどでは7種類とか8種類であったが、今では100種類とも云われている。これも「深層格」とか「関係子」とか云われているものと同値である。
では、「係り受け」をどのように定義するかというと…、
　単語w_1と単語w_2との「係り受け」は、$\Phi : w_1 \to w_2$とし、上記ｎ次元位相多様体同士の「準同型写像」と定義できる。
すなわち、「単語」である「ｎ次元多様体」の表現は、「ｍ×ｎ行列」として表現し、「単語」どうしの「係り受け」を「行列の積」として捉える。「係り側」の単語行例を$\{A_0 A_1 \cdots A_m\}^T$とし、「受け側」の単語行列を$\{B_0\ B_1 \cdots B_n\}$をした場合、単語行列空間は、「直行行列」であるので、$A^TA=I$,　$B^TB=I$　となり、　$D=A\Sigma B$　の特異値分解となる。（第１版）

⇒ *cTag*＞意味位相空間ページへ