1．Mathematical background and definition 数学的背景と定義

（1）言語空間Lと概念空間SCの基本定義

□言語空間L[1]を有限個の元m（形態素）を要素（変数）とする空間とし、文や文章、文書を構成する文法函数$G=\{g_1,\cdots,g_k\}$をそれの系列（曲線）とみなせば、意味ある文や文章、文書とは多変数多項式の斉次方程式

$g_1(m_1,\cdots,m_n)=\cdots=g_k(m_1,\cdots,m_n)=0$　の解となり、この全体を

$V(g_1,\cdots,g_k)=\{M=(m_1,\cdots,m_m)\in L^n \mid g_1(M)=\cdots=g_k(M)=0\}$とすると、この空間は代数多様体[2]AlgebraicVariety

となる。これは Syntax 上、正則な文集合である。また、この部分空間を複素空間C上と考えれば、この解の全体は無限遠点と極を有する genus 空間になり、文構成群の分類構造となる。

□$\Re$上の$F:L\to SC$で定義される意味概念空間SC：SemanticConceptualSpace は、$m\in L$の族集合となり、Bourbaki 位相が入り、意味概念位相空間になる。また同時に Hausdorff 位相からの視点も考慮され、mやK[3]の近傍で位相多様体 manifold の構造を持ち、特異点のない空間（超平面）となる。尚、上記の写像関数Fは、**Abelian 群**から Abelian 圏への導来関手とする。すなわち、以下の各関手の右導来関手で微分の概念を持つ長完全系列を生成するものである。

□形態素Mの族集合や Borel 集合族$family, Borel(M)\supset L$を考え、各クラスタ関数で、

$Morpheme(C):=\{m\}=M$	形態素　但し、C は文字、{ }は、集合を表す
$Phrase(M):=\{pr\}=PR$	文節
$Clause(PR):=\{cl\}=CL$	連文節
$Sentence(CL):=\{st\}=ST$	文
$Paragraph(ST):=\{pg\}=PG$	文章（段落）
$Document(PG):=\{dm\}=DM$	文書

という階層クラスタが生成される。

上記は$M\subset PR\subset CL\subset ST\subset PG\subset DM$という階層化の構造を明示的（半順序集合）に持つ。従って、SC は連続なので、微分概念作用素∂と積分概念作用素dを用いると、

$\partial(DM)=PG$	$d(M)=PR$
$\partial(PG)=ST$	$d(PR)=CL$
$\partial(ST)=CL$	$d(CL)=ST$
$\partial(CL)=PR$	$d(ST)=PG$
$\partial(PR)=M$	$d(PG)=DM$

という可換環でもあり、共役関係な階層でもある SC となる。

すなわち、

$$\phi\xrightarrow{\partial_\phi}DM\xrightarrow{\partial_{DM}}PG\xrightarrow{\partial_{PG}}ST\xrightarrow{\partial_{ST}}CL\xrightarrow{\partial_{CL}}PR\xrightarrow{\partial_{PR}}M\xrightarrow{\partial_M}0$$

$$0\xrightarrow{d_0}M\xrightarrow{d_M}PR\xrightarrow{d_{PR}}CL\xrightarrow{d_{CL}}ST\xrightarrow{d_{ST}}PG\xrightarrow{d_{PG}}DM\xrightarrow{d_{DM}}\phi$$

という長完全系列に拡張可能な循環系列になる。

これは完全系列なので、準同型定理によりホモロジーとコホモロジーが定義・構築される。

---

[1] 言語空間 L とは、表記言語の形態素空間のことで、文（又は文章、文書）は助詞などを含む形態素列の曲線となる

[2] 代数多様体とは、代数的演算を包含する多様体のことで、位相多様体 Manifold とは相違する

[3] K とは Knowledge 知識のことで、空間上の曲線で相や態の完結された文のことを知識という

□いわゆる、$\partial_{ST} : ST \to CL$ を準同型写像とすると、$ST / Ker\partial_{ST} \cong \mathrm{Im}\partial_{ST}$ が成立する。

この定義を説明すると、$Ker\partial_{ST} := (\{st_i\} \in ST \mid \partial_{ST}(st_i) = e_{CL} \in CL, st_i \in ST)$ であるので、$Ker\partial_{ST}$ は、ST の正規部分群になるので、$ST / Ker\partial_{ST}$ は、正規群（基）である剰余群（イデアル $Ker\partial_{ST}$ による商群）となり、$\mathrm{Im}\partial_{ST} \subseteq CL$ と同等になる。見方を変えれば、もともと $Ker\partial_{ST}$ の定義（上記）を見れば判るように、$e_{CL} \in CL$ に同等な要素を ST 上から選んで商群を生成しているので、Ker とは、$|ST| = |Ker\partial_{ST}| + |\mathrm{Im}\partial_{ST}|$ という次元の同一性（この場合は加法性）もあり、固有値問題 $Ax = \lambda x \to (A - \lambda E)x = 0$ と同等である Ker 値（ジョルダン標準形：$Ker(A - \lambda E)$）との意味合いも存在するので、これは固有値問題の一般化になっている。そのうえ、これは単射とのズレ度合でもあり、基近傍でもあると考えられので、ハウスドルフ空間の ε 近傍とはまったく違った近傍の概念でもあり、言語概念解析の数学的背景としているのが「ふきや理論」である。

□係り受け関係も各階層別に成立する。

$$F(M) := \{(m_i \to m_j) \mid m \in M\}$$

$$F(PR) := \{(pr_i \to pr_j) \mid pr \in PR\}$$

Λ …

$$F(DM) := \{(dm_i \to dm_j) \mid dm \in DM\}$$

上記、圏内の完全系列を導来関手 F によって意味概念空間 SC へ写像される。

□したがって、$F(DM) \to F(PG) \to F(ST) \to F(CL) \to F(PR) \to F(M)$ が SC 内で生成・構造化される。但し、$F = \{f_1, \cdots, f_k\}$ のように導来関手の集合になる。これが ST へ変換定義される。

□上記では、意味概念自体が位相構造を持ち、すなわち意味ある最小単位である形態素の「列」である意味概念が階層化された意味概念構造として構成される。これが知識 KE(超述語式 8 次元単体 Δ : Simplex)へと変換される。

□Syntax 圏から Semantic 圏への写像で Syntax タグから Semantic タグ（ST という）への変換が意味概念の一意性を表すことはできない。それは Syntax タグと ST とはまったく異次元の世界のものである為である。だから、従来の統計的言語意味解析であるディープラーニングなどでは精度が出なかった！Syntax | Context | Semantc は共に異次元。

◇しかし、異次元圏間での関手（導来関手も含める）を定義すれば、概念空間内での ST は、「意味の一意性」である対称性を詳細に明記する。意図 Intent が同じでも「表現の違い」や「ゆらぎ」「比喩」などで相違する文があるが、これらは同じ意味概念の ST で明確に表現される。これは意味解析上、重要なファクター（群の構造化）であるので、これを以下で証明する。

◇上記と同じく、形態素を対象 $object$ とし、対象同士の係り受けを射 $morphism$ とする圏 $Category$ を考える。そして Syntax 圏 τ の双対圏として τ° を定義する。

この τ° 圏内の形態素や文節（連文節）などを $U, V (U \subset V)$ とする。

この係り受け関係を包含射とすると $U \overset{\psi}{\to} V$ と表現する。

・完全な文の場合、形態素から文節、文節から連文節、連文節から文への射は、すべて全射になる。

・すなわち、形態素、文節、連文節、文の各集合への「係り受け」は、包含射と考えることができる。

Semantic 圏を仮にアーベル群Ωとし、そこへの関手F,Gとする。

・Semantic 圏内の完全文は、意味概念同士の結び付きによる「意味の一意性」（対称性）による群と考えられる。

すると、Semantic 圏Ω内は、$F(U)\overset{F\psi}{\leftarrow}F(V)$ と $G(U)\overset{G\psi}{\leftarrow}G(V)$という逆向きの係り受け（可換性）が成り立つことになる。但し、$\psi:V\to U$という単射である。また、

関手F,Gに関して、$F\overset{\varphi}{\to}G$を考えると、これは前層$P=\Omega^{\tau^o}$になる。φの核$Ker(-)$も前層Pに入る。

従って、Semantic 圏Ω内では、$F(U)\overset{\varphi U}{\to}G(U)$ と $F(V)\overset{\varphi V}{\to}G(V)$が成り立ち、そこで射$\varphi U$と$\varphi V$の核$Ker\varphi U\subseteq F(U)$と核$Ker\varphi V\subseteq F(V)$を上記のようにとると、$Ker\varphi U\overset{Ker\psi}{\leftarrow}Ker\varphi V$の射$Ker\psi$は、$Ker\psi(\alpha V)=F\psi(\alpha V)$となる。これは、表現可能関手であり、これで意味概念の一意性が証明される。

□今度は、Syntax 前層から Semantic 層への定義を考える。

◇Syntax 圏τ内の開集合 U に関し、イデアル$i\in I$が存在し、被覆$U=\bigcup U_i$とする。

Semantic 圏Ω内で、$s_i\in F(U_i)$とすると、$\rho_{U_i,U_i\cap U_j}(s_i)=\rho_{U_j,U_j\cap U_i}(s_j)$となる。

但し、$i\neq j$、$U_i\cap U_j$は共通開集合。$\rho(*)$は写像関数であり、ここではFを省略した。

これで「唯一」$F(U)=s$が存在し、$\rho_{U,U_i}(s)=s_i$ がすべての$i\in I$で成り立つことになり、関手$F\overset{\varphi}{\to}G$ が層 Sheaf になる。これが関手間や層間（層を圏として）の関手の層 Sheaves の定義である。この空中戦（汎関数の一般化）が言語概念空間の構造（しくみ）を間接的に、俯瞰的に解析できるしくみである。

□視点を変えて層の定義を考えると、

◇位相空間 X の要素であるαを含む開集合 U で$\alpha\in U\subset X$を考える。

茎 stalks を $O_\alpha=\{(f,U)\}$ とし、芽 germs を $(f_\alpha,U)\in O_\alpha$ とすると、

層 sheaves は、$O_X:=\prod_{\alpha\in X}O_\alpha=\{(\alpha,f_\alpha)\mid f_\alpha\in O_\alpha,\alpha\in X\}$ と定義できる。

◇開集合 U の要素αを中心とする等高線の断層を上から切断fで落とすと茎になり、その交点が芽である。

従って、層とは、芽$(f_\alpha,U\}\in O_\alpha$と要素$\alpha\in U\subset X$との直和集合である。上式∏パイは直和の記号です。

関数どうしの多峰性の偏差σを求める式の一般化（離散）と考えても良い。

□複素関数論の曲線論でいうならば、複素関数曲線Xの開集合U内の点pの近傍を正則とし、点pを高々m次の極とした場合、$F(U)$を層といい、$F(U)=O_X(mp)$と書くことができる。簡単な例でいうと、$f=\frac{x^2}{(x-1)^3}$ 2 次の零点と 3 次の極があり、無限遠点$x=\infty$で、$t=\frac{1}{x}$とすると、$t=0$になり、$f=\frac{x^2}{(x-1)^3}=\frac{t}{(1-t)^3}$なので、1 次の零点となる。従って、開集合 U の層間の層の定義を複素関数論（曲線論）でも表現できる。例題の二次の零点と一次の零点との違いは、複素空間だから相違することは判ると思うが、一次の零点になると虚軸ではなく、実数軸上の問題となることを意味している…これが重要です。

◇ちなみに切断は、$f:=(\varphi_1(\alpha),\varphi_2(\alpha),\Lambda\ ,\varphi_n(\alpha))$ というα での Section と定義できる。

この切断系列は、$0 \to F \to L_0 \to L_1 \to \cdots$ という軟弱 Flabby（injection）な完全系列 $H^0(U,F) = \Gamma(U,F)$ から

$0 \to \Gamma(U,L_0) \to \Gamma(U,L_1) \to \cdots$ となる。

この切断 Section は、語彙の意味概念のイデアル特定には重要なファクターで、語彙の意味概念である開集合の分布状況が明示的（等高線のように…）に判るものである。
その為に、層や茎、芽、スキームなどの手法が必要になり、これが意味概念層になる。