

形式言語における鳩ノ巣原理の探求

秋田大学・数理解科学コース 新屋良磨

Ryoma Sin'ya

Mathematical Science Course,

Akita University

ryoma@math.akita-u.ac.jp

1 はじめに

Ramsey 理論は離散数学における 1 つの花形的存在であり、種々の離散構造からある種の秩序 (単純な構造) を見出す技法・定理群から織りなされている理論である。Ramsey 理論の中に現れる定理 (以後、Ramsey 型定理と呼ぶ) の証明は、根本的にはほとんどが鳩ノ巣原理 (*pigeonhole principle*) の応用であり、実際、Prömel の書籍 [1] のページ xi においては “*To a certain extent, all results contained in this work can be viewed as generalizations of the pigeonhole principle.*” と評されている。Motzkin が “complete disorder is impossible” と評しているように、Ramsey 理論は様々な離散構造からある種の秩序 (単純な構造) を見出すことを目的としている。現在知られている Ramsey 理論に現れる種々の定理 (Ramsey 型定理) は多岐にわたり、例えば最も有名かつ古典的なグラフに対する Ramsey の定理、Euclid 平面の点集合に対する Erdős-Szekeres の定理、自然数の部分集合に対する Szemerédi の定理などがある。Szemerédi の定理は「自然数の部分集合 S のバナッハ密度が正ならば、 S は任意に長い等差数列を含む」と言ったシンプルな言明ながらも奥深い定理になっており、この定理を最初に証明した Szemerédi は Abel 賞を受賞している。また、Ramsey 型定理における最も深い結果の 1 つに、自然数に対する密度版 Hales-Jewett の定理 (以下、DHJ 定理と呼ぶ) がある。ここで言う密度とは自然数の部分集合に対して定義される「大きさ」を表す値 $\in [0, 1]$ であり、密度が 1 に近い集合ほど多くの自然数を含むことになる。DHJ 定理は、簡単に述べると、ある程度大きな密度を有する自然数の部分集合は長い “直線” (*combinatorial line*) を含むことを述べている (*cf.* [1] の 18 章)。DHJ 定理は重要かつ難解な定理として認識されており、Gower によって 2009 年に発足された (オンラインで) 多数の数学者とアイディアを交わしながら数学の問題を解く Polymath Project において最初に対象となった問題である。Polymath Project では DHJ 定理の証明の単純化に 40 名以上の数学者が参加し、最終的に単純化に成功して D.H.J Polymath (D.H.J は定理から取っている) という著者名で論文が arXiv にて公開された [2]。

本講究録では著者の今後の課題としての「Ramsey 型の定理や技法の形式言語理論への輸入・応用」について、これまでの研究の流れや今後の展望とともに簡単な解説を行う。

2 形式言語における鳩ノ巣原理

前節で Ramsey 理論にグラフ・自然数に対するいくつかの定理を紹介してきたが、形式言語理論における研究対象は言語 (文字列の集合) であり、言語に対する Ramsey 型理論の存在は自然な疑問となる。特に、形式言語理論においては正則言語、文脈自由言語、文脈自由言語などのある程度の構造を持った言語を主な研究対象とするため、それぞれの言語クラスにおいて Ramsey 型定理を探求することが「形式言語理論の Ramsey 理論」としては自然と言える。

	言語クラス	ある程度の大きさ	秩序 (単純な部分言語)
Pumping 補題	正則言語	無限濃度	uv^*w
Pumping 補題	文脈自由言語	無限濃度	uv^*wx^*y
既存成果 [6]	正則言語	密度 1	A^*wA^*
課題 1	正則言語	正密度	?
課題 2	文脈自由言語	密度 1	?
予想 1	文脈自由言語	正密度	無限濃度の正則言語 (よって Pumping 補題より uv^*w)

表 1. 形式言語における Ramsey 型定理

形式言語理論においても密度の研究は古くから存在しており、アルファベット A 上の言語 L の密度 (density) とは、次の極限 (極限が存在しない場合密度は未定義)

$$\lim_{n \rightarrow \infty} \frac{\#(L \cap A^n)}{\#A^n}$$

を表す (ここで $\#S$ は集合 S の濃度を、 A^n は A 上の長さ n の文字列全体を表す). 言語の構造を (正則や文脈自由などに) 制限することで、どのような密度版 Ramsey 型定理が現れるか? となる.

本課題の目的は、形式言語理論における密度版 Ramsey 型定理の発見と関連するいくつかの未解決問題への応用にある. 例えば、著者は [6] において、「アルファベット A 上の密度 1 の正則言語 L は、 A^*wA^* という単純な形の言語 (つまり「 w を含む A 上の文字列全体」を含む)」ということを示したが、これは本課題で探求する密度版 Ramsey 型定理の、すでに発見されている具体例となる (表 1 上段). 実は、良く知られた Pumping 補題もある視点からは形式言語理論における Ramsey 型定理と見ることができる. 正則言語 L の Pumping 補題は、 L に属するある程度長い文字列は、(pump して) L に属したままどんどん長くしていけることを保証する補題であるが、簡単な考察によって「 A 上の正則言語 L が無限濃度 (無限個の文字列を含む) であれば、適当な $u, w \in A^*, v \in A^+$ で $uv^*w \subseteq L$ が成り立つ (ここで A^+ は A 上の長さ 0 でない文字列全体)」ということが読み取れる.

Pumping 補題における「ある程度の大きさ」とは単に無限濃度であるが、無限濃度の中でもより細かい「大きさの尺度」として密度がある. 本課題で求めるものはまさに密度版 Ramsey 型定理であり、具体的には

- 言語クラスを「正則言語」 or 「文脈自由言語」に
- 大きさとして「密度 1」 or 「正密度」を

考えたときに現れてくる単純な部分言語の形 — 秩序 — を明らかにすることを目的とする (図 1 下段). また、pumping 補題と同様に、密度版 Ramsey 型定理は与えられた言語が正則言語 (または文脈自由言語) でないこと (以降、否定命題と呼ぶ) を示すための強力な道具となり得る可能性がある. 実際、著者は [5] において、図 1 の密度 1 の正則言語に対する結果を用いた否定命題の証明技法を提案しており、その有用性についても論じている. 本課題では密度版 Ramsey 型定理の否定命題への応用の有用性についても研究を進めていく. 一般に、文脈自由性の否定命題は難しいものが多く、例えば「原子語 (primitive words) 全体の集合は文脈自由ではない」という Dömösi-Horváth-Ito 予想は形式言語理論における大きな未解決問題である (cf. [3]). 本課題によって、正則言語での場合 ([5]) のように、文脈自由性の否定命題に有用な技法を発見することができればその効用は大きいと考えられる.

3 関連研究と展望

言語の密度に関する研究については歴史が古く、「正則言語の密度は(定義されるならば)常に有理数, 無意味文脈自由言語の密度は代数的数」という Berstel の定理を始めとした, 密度そのものに関する研究は多い. しかし, 密度を用いた Ramsey 型定理の考察を行っている研究は, 密度 1 の正則言語の関する(著者が知る限り)著者の結果 [6] しか無い. 著者はここ数年の研究の中で, (広く知られている)無限の猿定理と呼ばれる定理に対する理解を深めてきた (cf. [7] の第 5 章). 無限の猿定理とは「どのような長い文章でも, 猿が十分長い時間ランダムにキーをタイプすれば, いずれはその文章がタイプされる」としばしば比喩的に述べられる有名な定理であるが, 形式的に次のように述べることができる:

どのような文字列 w でも, A^*wA^* の密度は 1 になる

表 1 の [6] にあたる結果は, 正則言語のみを考えるならば, 無限の猿定理の逆も成り立つということ述べている. Ramsey 型の諸定理は「ある単純な構造が必ず現れる」という類の言明に抽象化できるが, 一方の無限の猿定理は「ある特定の構造が確率 1 で現れる」という言明に抽象化できる. この視点から, 著者は「無限の猿定理と Ramsey 理論に繋がりはないか?」という疑問をいだき, 最終的に形式言語理論における密度版 Ramsey 定理の探求という課題に至った.

言語の密度に関する研究は, 上に挙げた Berstel の定理を筆頭に, 言語クラスと密度の有理性・代数性・超越性に関する結果は一通り揃っている. また, 代数的符号理論の文脈においては, 言語 L の密度と L の代数的性質の関連が良く研究されている. 密度よりも“粗い”大きさの尺度である無限濃度に関するものとしていろいろな結果があり, 例えば種々の言語クラスの**免疫性** (immunity, cf. [4], 例えば言語 L が「正則免疫性を持つ (regular-immune である)」とは L が無限濃度の正則言語を部分集合に持たないことを意味する)に関する結果がいくつか知られている. 本課題では密度 1 や正密度であるような言語に対する免疫性を考えていると言えるため, 密度と免疫性の研究を融合・昇華したような議論展開や発想が必要となる場合が考えられる.

参考文献

- [1] “Ramsey Theory for Discrete Structures”, Hans Jürgen Prömel, Springer, 2013 edition.
- [2] “A new proof of the density Hales-Jewett theorem”, D.H.J Polymath, arXiv:0910.3926, 2015.
- [3] “Context-Free Languages and Primitive Words”, Pál Dömösi, Masami Ito, World Scientific, 2015.
- [4] “Immunity and pseudorandomness of context-free languages“, Tomoyuki Yamakami, Theoretical Computer Science, Vol. 412, Issue 45, pp. 6432–6450, 2011.
- [5] “言語の測度に基づく非正規性の証明技法”, 新屋良磨, コンピュータソフトウェア, Vol. 34, No. 1, pp. 119–124, 2017.
- [6] “An Automata Theoretic Approach to the Zero-One Law for Regular Languages: Algorithmic and Logical Aspects”, Ryoma Sin’ya, In Proc. GandALF’15, EPTCS 193, pp. 172–185, 2015.
- [7] “オートマトン理論再考”, 新屋良磨, コンピュータソフトウェア, Vol. 34 No.3, pp. 3–35, 2017.