

統計解析の生物学的手法によるテキストの系統樹作成 ——蔵訳『般若心経』を中心に——

石 田 勝 世

1. はじめに

文献学の分野では写本や版本などテキストの時系列的な伝承過程を系統樹で表現することが多い。系統樹は、一般にテキストの内容分析による「定性的手法」により作成されるが、本稿では系統樹を統計解析により作成する「定量的手法」について述べる。まず、代表的な定量的手法を紹介し、その先行研究を述べる。次に、蔵訳『般若心経』に適用した結果を報告し、定量的手法の適用可能性について検討する。

2. 系統樹を作成する定量的手法

系統樹¹⁾を統計解析により作成する場合の代表的な手法を紹介する。(1) クラスタ分析 (統計学の手法) 性質の似たものどうしを集めてグループを作り分類する手法であり、テキスト分類に適用できる。結果はツリー構造で表現される。(2) 系統ツリー推定手法 (生物学の手法) 生物進化の系統樹を統計解析により推定する手法である²⁾。「写本伝承」における過誤が「生物進化」における過誤に対応づけられることから、写本の系統推定に適用できる。結果はツリー構造で表現され、コンタミネーション (混成) があっても、その存在は表現されない。(3) 系統ネットワーク推定手法 (生物学の手法) 系統ツリー推定手法と基本的に同じであるが、本手法ではコンタミネーションを扱うことができ、結果はネットワーク構造で表現される。

3. 定量的手法による系統樹作成の先行研究

主要な先行研究を紹介する³⁾。(1) 梵文『法華経』クラスタ分析 (逢坂・山崎 1999) 対象: 「方便品」のネパール系梵文写本。方法: 異読によるテキスト間類似度にもとづく分類。結果: 写本間の系統を表現しているか不明。(2) 漢訳『般若

(146) 統計解析の生物学的手法によるテキストの系統樹作成 (石 田)

心経』クラスタ分析 (師 2002) 対象: 漢訳テキスト. 方法: N グラム分析によるテキスト間類似度にもとづく分類. 結果: 小本 (羅什訳, 玄奘訳) と大本 (その他) を分離. 原テキストの系統を反映した分類かどうかは不明. (3) 蔵訳『無尽意菩薩経』系統ツリー推定 (Maas 2008) 対象: カンギュル・テキスト. 方法: PAUP * による系統ツリー推定. 結果: 東西の系統が分離, コンタミネーションが存在するか否かは不明.

4. 蔵訳『般若心経』への適用

定量的手法の系統樹作成への適用可能性を探るためにクラスタ分析と系統ネットワーク推定の2通りの手法を適用してみた⁴⁾. (1) 定性的分析の結果 (図1) Silk 1994 掲載の内容分析による系統樹である. 「古ナルタン (Old Narthang) からツェルパ系 (東の系統) とテンパンマ系 (西の系統) の2つに分離」「プダク (S) は独立」などが分かる. コンタミネーションは点線で示されている. (2) クラスタ分析の結果 (図2) 「テンパンマ (MLN) 北京 (BKFP) ((HR) リタン ((JC X) (UD)) 独立 (S) の4グループに分類される」「X (校訂テキスト) はCやJに近い」「H (ラサ) とR (トクパレス) は東の系統に分類される」などが読み取れる. (3) 系統ネットワーク推定手法 (図3) 「東 ((UDCJ) X (BKFP)) 西 (LNM, R, H) 独立 (S) の3グループに分類される」「H (ラサ) とR (トクパレス) は東の系統にも近い」「X (校訂テキスト) は (UDCJ) や (BKFP) のグループに近い」などが読み取れる.

5. 定量的手法の評価

系統樹を定量的に作成する3手法を評価すると次のようになる. (1) クラスタ分析 (統計学的手法) 長所: テキストが分類され, 大まかな傾向を知るには便利. ツリー構造で表現されるので分かりやすい. 欠点: テキスト分類にすぎず (系統は不明), 系統ツリー作成には向かない. 生物学では使用されない傾向にある (分類思考から系統樹思考へ進化). (2) 系統ツリー推定手法 (生物学的手法) 長所: ツリー構造で表現されるので分かりやすい. コンタミネーションがない場合には有用. 欠点: コンタミネーションが扱えない (有無も判断できない). (3) 系統ネットワーク推定手法 (生物学的手法) 長所: コンタミネーションが扱え, その有無が判断できる (田の字型構造の有無). 校訂作業で祖本推定に利用できる (ネットワーク中心部分が祖本候補とみなせる). 欠点: ネットワークで表現されるので分かり難い.

以上から, 現状では3手法のうち系統ネットワーク推定手法が最良と考えられ

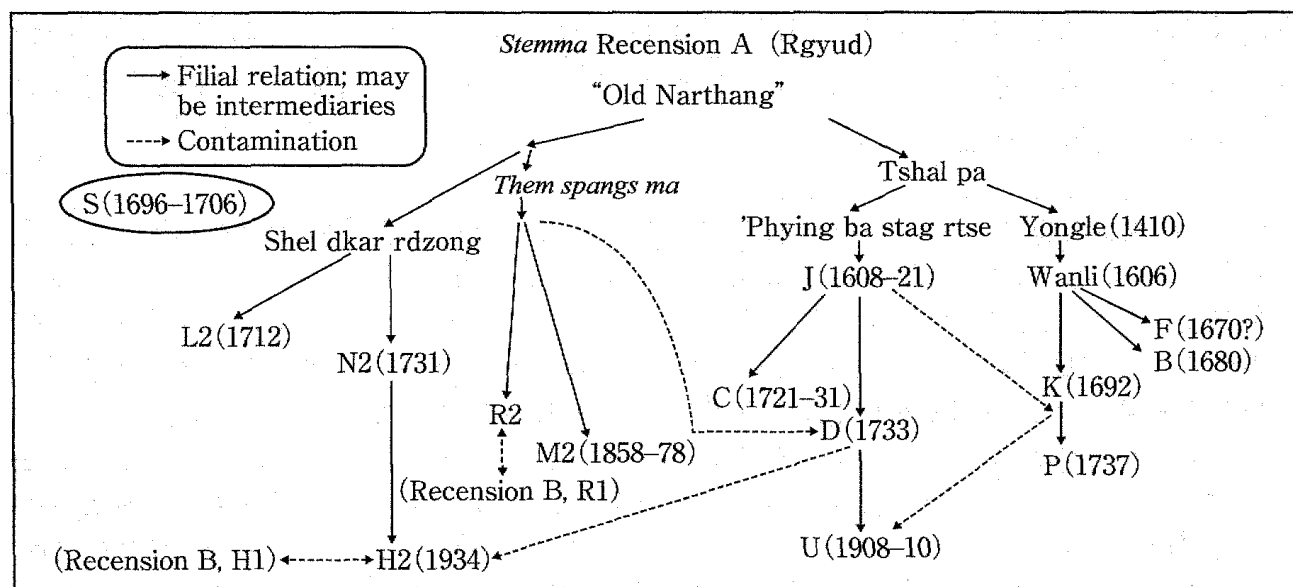


図1 蔵訳『般若心経』系統樹 出所: Silk 1994 (p. 21)

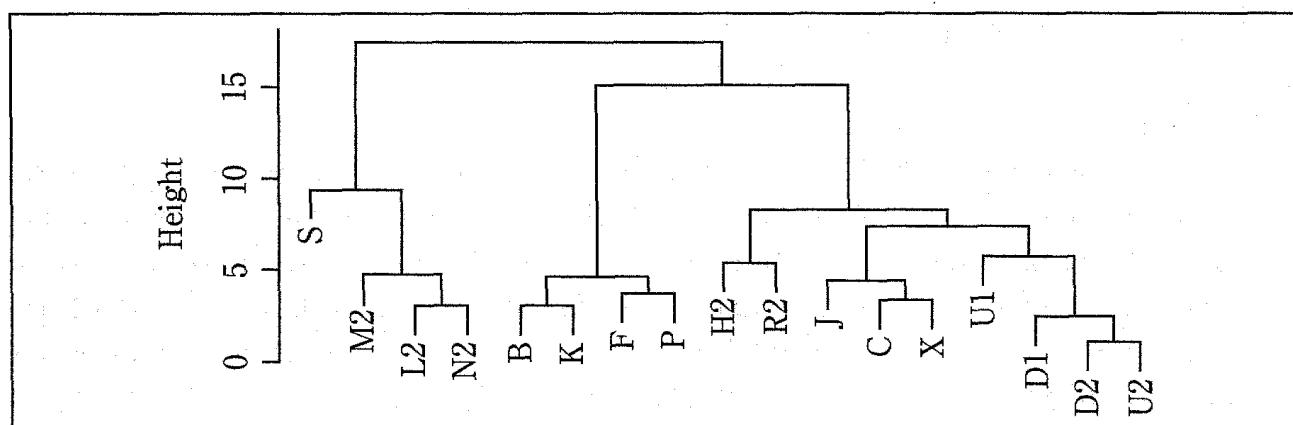


図2 蔵訳『般若心経』クラスタ分析 (Rで作成)

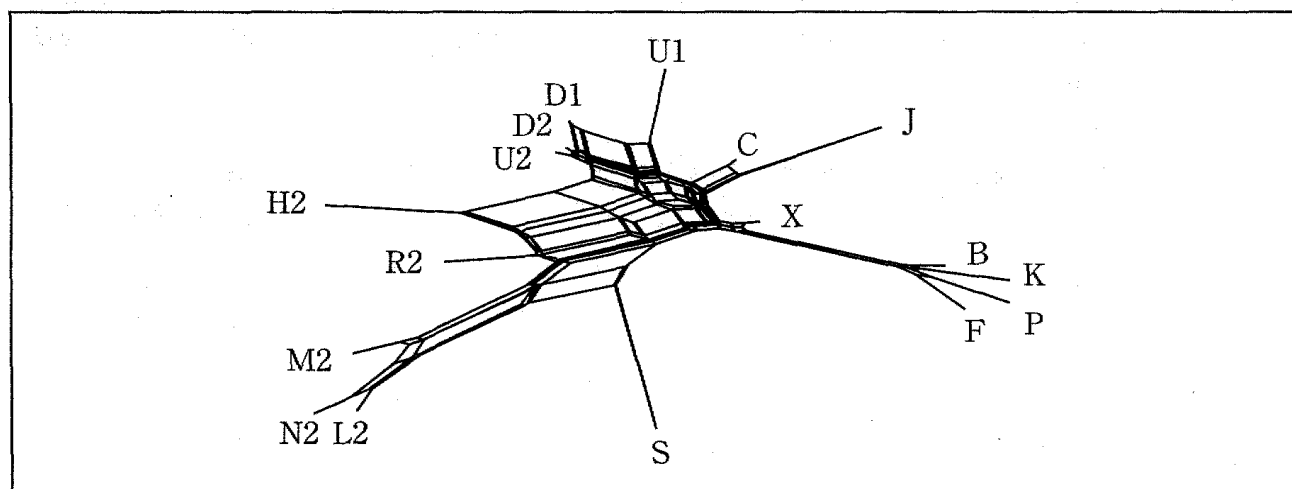


図3 蔵訳『般若心経』系統ネットワーク (SplitsTree4で作成)

(148) 統計解析の生物学的手法によるテキストの系統樹作成 (石 田)

る。定量的手法を利用する場合は「推定自体はアブダクションであり、正しい結果ではなく、最も確からしい結果が導かれること」などに留意し定性的な内容分析との補完利用が望ましいであろう。

1) 「系統樹」は、ルート（根）から枝分かれして伸びていき枝の融合は認めない「系統ツリー」、枝の融合を認める「系統ネットワーク」に分類できる。 2) 生物学の系統推定は三中 1997 が分かりやすい。 3) いずれの結果もツリー構造で表現されるが、紙幅の都合上、省略した。(1)(3)は、底本を決めて他のテキストと比較し異読を作成してから異読の情報のみを利用し系統樹を作成している。(2)は、テキストごとの「N文字の漢字パタンの出現頻度」をもとにクラスタ分析を行いテキストを分類している。 4) 対象は次の蔵訳『般若心経』17テキストである（異読は Silk 1994 の校訂研究を利用）。B: ベルリン写本 C: チョネ版 D1, D2: デルゲ版 F: 台北写本 H2: ラサ版 J: ジャンサタム／リタン版 K: 北京版（康熙帝）（1962） L2: ロンドン写本 M2: 東洋文庫写本 N2: ナルトン版 P: 北京版（乾隆帝）（1737） R2: トクパレス写本 S: プダク写本 U1, U2: ウルガ版 X: 校訂テキスト（Silk 1994）。添え字の1は般若部、2は密教部に属することを示す。

〈参考文献〉

Daniel H. Huson *et al.* (2006) "Application of Phylogenetic Networks in Evolutionary Studies," *Molecular Biology and Evolution*, 23(2): 254–267. Philipp Maas (2008) "A Phylogenetic Approach to the Transmission of the Tibetan Kanjur — The *Akṣayamatīnirdeśa* Revisited," *Indica et Tibetica* 36: 229–243. 三中 信宏 (1997) 『生物系統学』, 東京大学出版会. 師 茂樹 (2002) 「N グラムモデルとクラスター分析を用いた漢文古典テキストの比較研究—『般若心経』の異訳の比較を例に」, 『京都大学大型計算機センター第 69 回研究セミナー「東洋学へのコンピュータ利用」予稿集』. 逢坂 雄美, 山崎 守一 (1999) 梵文『法華経』の統計解析, 人文科学とコンピュータシンポジウム. Jonathan A. Silk (1994) "The Heart Sutra in Tibetan: A Critical Edition of the Two Recensions Contained in the Kanjur," *Wiener Studien zur Tibetologie und Buddhismuskunde*, vol. 34. Vienna: Arbeitskreis für Tibetische und Buddhistische Studien, Universität Wien

〈キーワード〉 系統樹, 統計解析, 生物系統学, カンギュル, 『般若心経』

(九州大学大学院)