

大規模仏教文獻群に対する確率統計的分析の試み

師 茂樹

はじめに

問題の所在

近年、中国学においては『二十五史』『四庫全書』などの大規模な漢字文献のデータベースが、台湾、中国を中心に次々に公開されるようになった。仏教学においても『大正新脩大藏經』（以下『大正藏』）をはじめとする各種叢書の電子化が着実に進んでいる。電子テキストは、これまで多くの時間を必要とした用語の検索、用例調査などにおいて劇的な効率化をもたらした。したがってこのような発展は、東洋学の従来の研究方法の延長線上にあり、これからの中洋学において必要不可欠なものであるという意見もある¹。

確かに急速な電子テキストの開発・普及によって、文献学者に多くの利便性、特に研究の効率化がもたらされることは期待できる。しかし、それと同時に、これまであまり意識されてこなかった問題が露呈してきたことにも注意しなければならない。例えば、次のようなことがある。

1. 資料の配列や選択などにおける先見性の問題
2. 大規模資料群に対する仮説形成の困難さ

まず 1 について言えば、例えば筆者が所属する大藏經テキストデータベース研究会²においては、『大正藏』電子化（SAT プロジェクト）の初期の段階において、阿含部から始まり日本撰述部に終わる『大正藏』の配列についての議論がなされた³。『大正藏』以前の大藏經は、それまでの仏教史観を反映して華嚴部から開始するが、西欧の近代的実証主義的な文献学を導入した『大正藏』においては、阿含部から始まるように改められた。このような文献の配列は、単に便宜的なものではなく、我々のモチベーションや解釈を少なからず方向付けている可能性がある。コンピュータ上であれば、『大正藏』の配列との継続性を保ちつつ別の配列（例えば、最新の仏教学の研究成果を反映した配列など）を採用することも可能であるので、『大正藏』そのままで入力する必然性も相対的に小さくなる。電子化が『大正藏』の配列の相対化を促したのである。仏典ではないが、大英図書館の International Dunhuang Project⁴においては、従来の文献目録に加えて、Map Search という従来の目録とは全く異なる発想によってテキストの分類を試みている。これも従来のテキストの直線的で階層的な分類法、配列などに対する根本的な反省から来るの

¹ 陳弱水（山下一夫訳）「デジタルアーカイブと東洋学 —中央研究院歴史語言研究所の経験から—」（『漢字文献情報処理研究』5、2004）

² <http://www.l.u-tokyo.ac.jp/~sat/>

³ 結局のところ『大正藏』の通りという結論になった。なお、中華電子仏典協会（CBETA）が公開する『大正藏』データベースにおいては、『大正藏』とは異なる配列を採用している実装もある。

⁴ <http://idp.bl.uk/>

ではないかと考えられる⁵。

次に 2 については言えば、文献学は厳密な資料批判と網羅的な資料収集によって成り立つ客観的、実証的な学問であるとされているものの、そのためにはあらかじめ何らかの仮説が必要である。たとえ無意識的にでも、ある研究者が無数にある文献の中からどれかを選び出し読み始めるとき、その選択の背景には文献に対する知識や先行研究、研究者の関心などに基づいた何らかの仮説が形成されているはずである。電子テキストに対する検索もまた、テキストの内容や表現形式をあらかじめ知っていなければ、適切な検索語を決定することができず、有意な結果が出にくい場合が多い。しかし、近年の電子テキストの爆発的な増加は、仮説を前提とする従来の方法が通用しなくなっているのである。

以上のことから、大規模な電子文献群を扱うための方法論の検討が、大きな課題となっているのではないかと考えられる。

本研究の目的

本研究は、大規模文献データベースを効果的に扱い、またその大規模性を従来の文献学的研究に活かすための方法論を探求する一環として、大規模文献に対する仮説形成を確率統計的な分析を用いて試みる。

仮説生成とは推論の一種（その一部は *abduction* とよばれる）で、例えばジグザグしながらも概ね右肩上がりの平均気温の折れ線グラフから、気候の変動を統一的に説明するための仮説（二酸化炭素排出量の増大が原因ではないか云々）を得る、というようなことである。土橋喜氏は、仮説生成による推論の特徴として、

仮説生成は蓋然的な推論方法であり、結論をそのまま真実であると認めるることは適切ではない。しかし、ほぼ正しいであろうと認めることができ、ほとんど間違いない命題を引き出す推論方法である。そのため結論が絶対的な確実性を持つものではなく、厳密には得た結果に対して検証が必要になる⁶。

と述べ、さらに、

学問の研究や日常の生活のさまざまな場面において、仮説生成が果たしている役割は演繹推論に劣らないものがあり、その推論機能を無視することはできず、逆に支援することによって、人間の問題解決能力を拡大することが考えられる⁷。

と言う。

情報科学の分野では、大規模なデータベースの中から、確率・統計やパターンマッチなどの技術によってパターンを発見し、発想や仮説生成を支援する技術であるデータマイニングが注目されている。データマイニングとは「大規模なデータから思いがけない (unsuspected) パター

⁵ 明星聖子「文献学はコンピュータに何をたくすべきか」(『情報処理学会研究報告』Vol. 2004, No. 58 (2004-CH-62), 2004) では、文献学のアポリアを克服するためにコンピュータが利用できないか、という問題提起がなされており、興味深い。

⁶ 土橋喜『情報視覚化と問題発見支援 問題構造の可視化による仮説生成』(あるむ、2000)、p. 17。

⁷ 土橋前掲書、p. 16-17。

ンを発見する」技術であり、もともとは「適切な仮説を前もってたずね、しらみつぶし、デタラメにパターンを探すこと」というような否定的な意味で用いられてきたが、仮説を立てることすら困難になるほどデータベースが大規模化した今日、その評価が逆転したのである⁸。

本研究もまた、東洋学、文献学においてテキストデータベースを真に有効活用し、さらには新たな問題意識を喚起するための仮説形成を、大きな目標としている。具体的には、石井公成氏が N グラムモデルに基づき提唱した文献比較方法である NGSM (N-Gram based system for Multiple document comparison and analysis⁹) を用いて、玄奘三蔵が翻訳したとされる仏典（一部外典を含む）を網羅的に比較することを試み、それを翻訳スタッフとの関係を考察するための仮説の形成を試みることで、大規模電子文献を用いた文献学に資する仮説形成が可能かどうかを検証する。

研究の方法と題材

N グラムモデルによる文献比較

N グラムモデルとは、確率・統計的自然言語処理の分野で広く用いられている言語モデルである。大雑把に言えば、N グラムとは「n 個の文字列または単語列」をすべて数えあげ、その発生する確率を計算することで、ある文章、あるテキストの特徴を割り出そうという方法である。例えば、「摩訶般若波羅蜜多心經」という文章を 3 グラムで（3 文字ずつ）分解すると、

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 摩 | 訶 | 般 | | | | | | | |
| | 訶 | 般 | 若 | | | | | | |
| | | 般 | 若 | 波 | | | | | |
| | | | 若 | 波 | 羅 | | | | |
| | | | | 波 | 羅 | 蜜 | | | |
| | | | | | 羅 | 蜜 | 多 | | |
| | | | | | | 蜜 | 多 | 心 | |
| | | | | | | | 多 | 心 | 經 |

という 8 つの文字列がそれぞれ一回ずつあることがわかるが、それはつまり「摩訶般」も「羅蜜多」も同じ確率で発生し、「摩訶經」というような組み合わせの文字列が発生する確率はゼロ、ということを意味する。同様に日本語の文章で「ばらき」という音の列は「茨城」「原木中山」などの地名、「薔薇・木」「バラキ（マフィア映画のタイトル）」など、単語レベルでもそれなりに高い確率で発生するが、順番を並び替えただけの「らきば」という音の列は、「これから木場へ行く」など複合的な例はあるものの、「ばらき」と比べて発生する確率はかなり低いと言える。この確率の差や分布によって、文章や言語の特徴を記述しようというのが N グラムというモデルである。すなわち、単語や文字（「アイテム」と総称）の生起が直前のアイテムのみに依存するという一方向的・線的性質なものとして言語をモデル化し、確率・統計的な処理の俎上にの

⁸ 福田剛志・森本康彦・徳山豪『データ・マイニング』（データサイエンスシリーズ③、共立出版、2001）

⁹ Ishii, Kosei. "NGSM and Cluster Analysis: Its Usage in the Digitization of Variant Texts in the SAT (Taisho Daizokyo Text Database)." *Proceedings of PNC Annual Conference and Joint Meetings 2002*. 2002.

せたものである。N グラムによるテキストの比較分析は、北研二氏による多言語コーパスの分類¹⁰をはじめ研究が積み重ねられており、古典文献の分析においても、国文学・国語学¹¹における応用を皮切りに、仏教学¹²、中国古典¹³などにも広がっている。

このようなモデルに対して、我々が言語活動に対して持っている印象と比べて単純すぎるのではないかという批判も予想される。実際、この言語モデルによる分析については多くの欠点が指摘されている。その中でも大きなものとしては、上の例であげた「羅蜜多」のように単語として意味をなさない文字列と「波羅蜜」のような単語とを同列に扱うことへの批判である。実際、従来行われてきた統計モデルによる文献分析（計量文献学）においては、形態素解析が不可欠の作業とされており、実際、品詞をマークアップしたテキストデータベースによる研究によつていくつかの注目すべき成果が上がっている¹⁴。

しかし、近藤みゆき氏は、形態素解析に基づくデータ処理には次のような問題点があると指

¹⁰ 北研二「確率的言語モデルに基づく多言語コーパスからの言語系統樹の再構築」（『自然言語処理』Vol. 4, No. 3, 1997）

¹¹ 近藤みゆき「平安時代和歌資料における特殊語彙抽出についての計量的研究と利用ツールの公開 一古今和歌集の歌語と表現のジェンダー性についてー」（『科学研究費特定領域研究 人文科学とコンピュータ研究成果報告書 一コンピュータ支援による人文科学研究の推進ー』、1999）、近藤みゆき「和歌とジェンダー」（『国文学』〈特集・和歌の脱領域〉平成 12 年 4 月号、学燈社、2000）、近藤みゆき「n グラム統計処理を用いた文字列分析による日本古典文学の研究 一『古今和歌集』の「ことば」の型と性差ー」（千葉大学『人文研究』第 29 号、2000）、近藤泰弘「コンピュータによる文学語学研究にできること 一古典語の「内省」を求めてー」（全国大学国語国文学会夏季大会シンポジウム「情報技術は文学研究をいかに変えるか」発表レジュメ、2001）、近藤みゆき「n-gram 統計による語形の抽出と複合語 一平安時代語の分析からー」（『日本語学』Vol.20、2001 年 8 月号）、近藤泰弘・近藤みゆき「平安時代古典語古典文学研究のための N-gram を用いた解析手法」（言語情報処理学会第 7 回年次大会『発表論文集』2001）、近藤泰弘・近藤みゆき「N-gram の手法による言語テキストの分析方法 一現代語対話表現の自動抽出に及ぶー」（『漢字文献情報処理研究』2、2001）、谷本玲大「曖昧検索性を持たせた N-gram サーチの手法 一『新撰萬葉集』と菅原道真の詩の比較を例にー」（『漢字文献情報処理研究』2、2001）など。

¹² 石井公成「N-gram 利用の可能性 一佛教文献における異本比較と訳者・作者判定ー」（『漢字文献情報処理研究』2、2001）、師茂樹「XML と NGSM によるテキスト内部の比較分析実験 一『守護国界章』研究の一環としてー」（『漢字文献情報処理研究』2、2001）、Ishii, Kosei. "Classifying the Genealogies of Variant Editions in the Chinese Buddhist Corpus," 『電子佛典』第 3 輯、東國大學校 EBTI、2001、師茂樹「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——『般若心經』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、2002）、石井公成「仏教学における N-Gram の活用」（東京大学東洋文化研究所附属東洋学研究情報センター編『明日の東洋学』、2002）、師茂樹「N グラムによる比較結果からの用例自動抽出 —— 横宗系の偽經を題材に」（『東洋学へのコンピュータ利用第 14 回研究セミナー』予稿集、2003）、角田泰隆「異本処理システムによる道元禪師関係文献の書誌学的研究（序）—真字『正法眼藏』による試みー」（『駒澤短期大学研究紀要』31、2003）、石井公成「敦煌發現之地論宗諸文獻與電腦自動異本處理」（『戒幢佛學』2, 2003）、石井公成「『大乘起信論』の用語と語法の傾向 —NGSM による比較分析ー」（『印度学仏教学研究』52-1、2003）、師茂樹「N グラムと文字データベースによる漢字佛教文献の分析」（『情報処理学会研究報告』Vol. 2004, No. 7 (2004-CH-61)、2004）など。

¹³ 山田崇仁「『國語』韋昭注引系譜資料について—N-gram 統計解析法による分析ー」（『立命館史学』22、2001）、秋山陽一郎「『老子』傳奕本来源考 テキスト処理による「項羽妾本」介在の検証」（『漢字文献情報処理研究』4、2003）、山田崇仁「歴史記録としての『春秋』—N-gram モデルと統計解析法による分析ー」（『中國古代史論叢』、2004）、齊藤正高「趙岐『孟子章句』の特徴 —「注釈」と本文の語彙頻度比較を通じた問題発見ー」（『漢字文献情報処理研究』5、2004）、山田崇仁「中国戦国期の語彙量について —N-gram とユールの K 特性値を利用した分析ー」（『漢字文献情報処理研究』5、2004）など。

¹⁴ 村上征勝『真贋の科学 計量文献学入門』（朝倉書店、1994）、同『文化を計る 文化計量学序説』（朝倉書店、2002）等参照。

摘する¹⁵。

- 1語の単位を認定する基準が一通りではない。
- 複合語や強い共起性のある単語群（連語、慣用句など）の分析に不利。

一方、N グラムの場合、単語の区切りなどを問題にすることなく網羅的に数え上げるため、大量のノイズが含まれる可能性がある反面、上記のような問題点を克服しうるデータを得ることができるのである。また、大規模な電子テキストに対する統一的な形態素解析やマークアップなどには多くの困難が予想されるし、そもそもその行為自体が文献学的研究になるため、本研究の目的である文献学的研究に入る前の仮説形成とは領域の異なる方法である。

ところで、上に例に挙げた「羅蜜多」には、本当に分析上の価値がないのだろうか。これに関して、Anthony Kenny 氏が次のように述べている点は注目される。

文体に指紋があるとすれば、それはどのようなものだろうか？ それはおそらく、ある著者の文体的な特徴——例えば ‘such as’ の生起度数といった、まったく取るに足りないと言ってもよいような特徴を組み合わせたもの——であって、指紋と同様にその人に特有のものであろう。文体上些細で取るに足りぬ特徴だからといって、文体分析に利用しない理由にはならない。指先にある渦巻や輪が我々の容姿においては大切でも目につくわけでもないが、指紋が一生涯変わらないように、そういったものこそが著者の叙述において変化することのない特徴となるはずであり、他の書き手には見られないその人だけのものとなるはずであろう¹⁶。

また、近藤泰弘氏は、コンピュータを利用したテキスト研究の期待されることとして、次の 2 点をあげている。

1. 徹底的に網羅的な研究（すべての単語・すべての文字の単位にまで網羅性を及ぼすことが可能になる）。
2. それによって現代人には通常認知できないデータの構造性や規則性を探り出す。それは、現代人の古典語に対する「内省」(introspection)（語感）の欠如を補うことができ、文学研究に貢献する。なぜなら、古典文学の正しい読みにとって、「内省」（文法的直観と言語外知識など）の欠如は大きな障害のひとつだからである。

これは語学・文学研究をテーマにしたレジュメからの引用であるが、もちろんテキストに関する多くの研究にも同様に当てはまるだろう。

NGSM とクラスタ分析

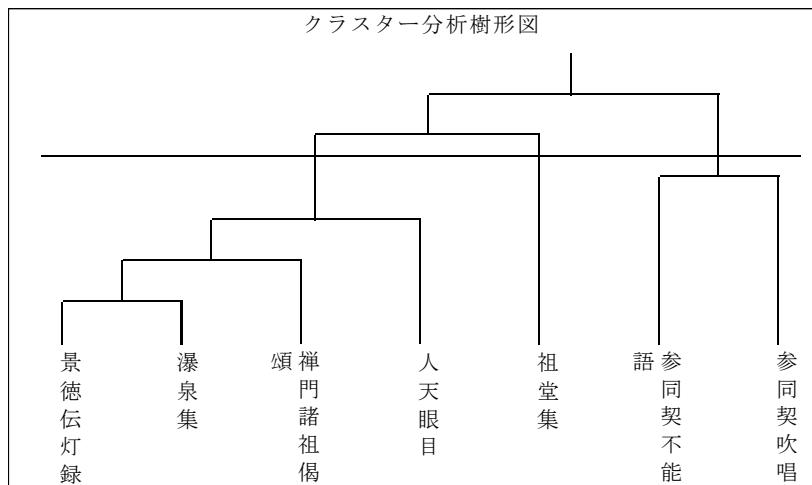
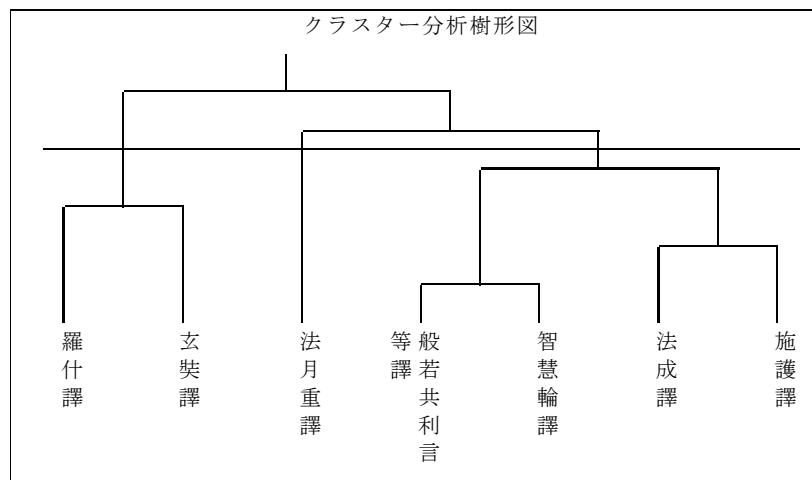
石井公成氏の提唱する NGSM は、文献ごとの N グラム結果を比較するための方法である。例えば、5 つの文献に対して、文字列 X、Y ごとに頻度をまとめると、次の表のように表すことができるとする。これによって、例えばこの表であれば文字列 X は文献 C と文献 D に多いので、両者が文体的に近い関係にある、というような分析が可能になる。これが NGSM である。

¹⁵ 近藤みゆき前掲「n-gram 統計による語形の抽出と複合語 —平安時代語の分析から—」参照。

¹⁶ Kenny, Anthony. *Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. 1982; 吉岡健一訳『文章の計量 文学研究のための計量文体学入門』(南雲堂、1996)、p. 24。

| | 文献 A | 文献 B | 文献 C | 文献 D | 文献 E |
|-------|------|------|------|------|------|
| 文字列 X | 1 | 0 | 7 | 5 | 1 |
| 文字列 Y | 2 | 3 | 0 | 2 | 1 |

さらに、このような文献の集合について、文字列 X、Y の頻度をそれぞれ x 軸、y 軸上の座標の値と見なせば、各文献間の距離を計算することができる。それらの近いものどうしを順次クラスタとしてまとめていくのが、クラスタ分析である¹⁷。NGSM によるクラスタ分析では、従来の文献学的研究と相似する結果が出ており（下図は上が『般若心経』異訳間の比較¹⁸、下は『参同契』異本間の比較¹⁹）、分析モデルとしてある程度の妥当性を持っているものと考えられる。



¹⁷ この方法は、キーワードの頻度をベクトルの座標と見なす G. Salton, A. Wong and C. S. Yang. "A Vector Space Model for Automatic Indexing." *Communication of the ACM*, Vol. 18, No. 11. 1975. を参考にした。

¹⁸ 師茂樹前掲「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」参照。

¹⁹ 石井公成前掲「敦煌發現之地論宗諸文獻與電腦自動異本處理」参照。なお、本図については、石井公成氏より直接頂いたものである。

玄奘訳の問題点

さて、本研究の対象は、玄奘三蔵が翻訳したとされる經典である。玄奘の翻訳については、これまで多くの研究が積み重ねられてきた。

全体的な研究としては、玄奘の翻訳時期に明確な区分が見られ、それが玄奘の佛教觀と相応することから、翻訳全体に玄奘の意図が貫かれていたことを見いだす研究²⁰がある。また、訳風について言えば、伝統的には他の漢訳よりも概して直訳的であることが述べられる反面、個別の文献について見れば、例えば『俱舍論』の場合「真諦訳の方がむしろ生硬で直訳的である²¹」という指摘があるなど、様々な問題がこれまで指摘してきた。

その中、特に『成唯識論』の場合に顕著なように、翻訳スタッフの介在が指摘されているものもある。宮元啓一氏によれば「玄奘は、とりわけ重要な仏典の漢訳に精を出し、比較的重要度の低いものは、ほかの訳経僧にまかせ、いわゆる監修者の立場に立ち」、さらには『(勝宗)十句義論』の場合は、どうやら、玄奘みずからではなく、玄奘の右腕といわれた慈恩大師窺が監修して終わりとしたようであるという²²。

幸い、玄奘の訳場で誰がどのような役割を担ったのかについては、よくわからないことが多い他の訳経僧の場合とは異なり、ほとんどの文献について知ることができる。諸資料中、ほぼすべての文献について、筆受を担当した者(『大菩薩藏經』のみ証文をした道宣)の名前を挙げているのが『開元釈教錄』卷八(大正 55, 555b-557b)であるが、一部の文献の識語等に記録された情報と比較すると、ごく一部の人間の名前しか記録されておらず、多少の食い違いがあるものもある。

以下の表は、各文献における『開元釈教錄』の記録と識語とを対照させたものである。

| No. | 巻数 | 題名 | 『開元釈教錄』筆受 | 識語 |
|-----|-----|-------------|-------------|---|
| 124 | 1 | 縁起經 | 神咬(昉?) | |
| 220 | 600 | 般若波羅蜜多經 | 大乘光・大乘欽・嘉尚等 | 筆受: 大乘欽・大乘基・大乘光・慧朗・嘉尚・道測・神咬 綴文: 玄則・神昉・靜邁 証義: 通・神泰・慧景・慧貴 |
| | 1 | 能斷金剛般若波羅蜜多經 | 杜行顥 | |
| 251 | 1 | 般若波羅蜜多心經 | 知仁 | |
| 289 | 1 | 顯無邊佛土功德經 | 大乘雲 | |
| 310 | 20 | 大菩薩藏經 | 智証 | 証義: 靈潤・文備・慧貴・明琰・法祥・普賢・神昉・道深・玄忠・神泰・敬明 綴文: 栖玄・明濬・辯機・道宣・靜邁・行友・道卓・慧立・玄則 字学: 玄応 証梵語: 玄暮 |
| 367 | 1 | 称讚淨土佛攝受經 | 大乘光(詢?) | |

²⁰ 吉村誠「玄奘の大乗觀と三転法輪説」(『東洋の思想と宗教』16、1999)

²¹ 桑山正進・袴谷憲昭『人物 中国の佛教 玄奘』(大蔵出版、1981 [1991 新訂])、p. 303。

²² 宮元啓一『牛は実在するのだ! インドの実在論哲学『勝宗十句義論』を読む』(青土社、1999年10月)、p. 33。

| | | | | |
|------|-----|------------------|---------|---|
| 390 | 1 | 仏臨涅槃記法住 經 | 大乘光 | |
| 411 | 10 | 大乘大集地藏十 輪經 | 大乘光等 | 筆受：大乘光・神昉・道觀 証文：行友・道智・玄忠 正字：明濬・玄應 証義：文備・神泰・道深・法祥・慧貴・ 道明・道洪（大正 13, 728a） |
| 436 | 1 | 受持七仏名号所 生功德經 | 大乘光 | |
| 450 | 1 | 藥師琉璃光如來 本願功德經 | 慧立 | |
| 476 | 6 | 說無垢稱經 | 大乘光 | |
| 515 | 1 | 如來示教勝軍王 經 | 大乘光 | |
| 592 | 1 | 天請問經 | 辯機 | |
| 648 | 1 | 寂照神變三摩地 經 | 大乘光 | |
| 676 | 5 | 解深密經 | 大乘光 | |
| 680 | 1 | 仏說仏地經 | 辯機 | |
| 689 | 1 | 甚希有經 | 大乘欽 | |
| 691 | 1 | 最無比經 | 大乘光 | |
| 714 | 1 | 緣起聖道經 | 大乘光 | |
| 717 | 2 | 分別緣起初勝法 門經 | 大乘詢 | |
| 765 | 7 | 本事經 | 靜邁・神昉等 | |
| 840 | 1 | 稱讚大乘功德經 | 大乘光 | |
| 918 | 1 | 諸仏心陀羅尼經 | 大乘雲 | |
| 1034 | 1 | 呪五首 | 大乘光 | |
| 1071 | 1 | 十一面神呪心經 | 玄則 | |
| 1094 | 1 | 不空羈神呪心經 | 大乘光 | |
| 1162 | 1 | 持世陀羅尼經 | 神察 | |
| 1360 | 1 | 六門陀羅尼經 | 辯機 | |
| 1363 | 1 | 勝幢臂印陀羅尼 經 | ？？？ | |
| 1365 | 1 | 八名普密陀羅尼 經 | 大乘雲 | |
| 1395 | 1 | 拔濟苦難陀羅尼 經 | 大乘光 | |
| 1499 | 1 | 菩薩戒羯磨文 | 大乘光 | 「其證義、證文、正字、筆受、義業、沙 門明琰等、二十許人」（大正 24, 1106c） |
| 1501 | 1 | 菩薩戒本 | 大乘光 | |
| 1530 | 7 | 佛地經論 | 大乘光 | |
| 1536 | 20 | 阿毘達磨集異門 足論 | 弘彥・釡詮等 | |
| 1537 | 12 | 阿毘達磨法蘊足 論 | 大乘光等 | |
| 1539 | 16 | 阿毘達磨識身足 論 | 大乘光等 | |
| 1540 | 3 | 阿毘達磨界身足 論 | 大乘基 | |
| 1542 | 18 | 阿毘達磨品類足 論 | 大乘光等 | |
| 1544 | 20 | 阿毘達磨發智論 | 玄則等 | |
| 1545 | 200 | 阿毘達磨大毘婆 沙論 | 嘉尚・大乘光等 | 筆受：嘉尚・海藏・神昉・大乘光 証義：明珠・惠貴・法祥・慧景・神泰・ 普賢・善樂 |

| | | | | |
|------|-----|-----------|------------|--|
| | | | | 綴文：括玄・靜邁・慧立・玄則 正字：義褒・玄應 執筆：神察・辯通（大正 27, 4c-5a） |
| 1554 | 2 | 入阿毘達磨論 | 訖詮／嘉尚 | |
| 1555 | 2 | 五事毘婆沙論 | 訖詮等 | |
| 1558 | 30 | 阿毘達磨俱舍論 | 元瑜 | |
| 1560 | 1 | 阿毘達磨俱舍論本頌 | 元瑜等 | |
| 1562 | 80 | 阿毘達磨順正理論 | 元瑜 | |
| 1563 | 40 | 阿毘達磨藏顯宗論 | 慧（惠？）朗・嘉尚等 | |
| 1570 | 1 | 廣百論本 | 大乘謹 | |
| 1571 | 10 | 大乘廣百論釋論 | 敬明等 | |
| 1578 | 2 | 大乘掌珍論 | 大乘暉 | |
| 1579 | 100 | 瑜伽師地論 | 靈會・朗（明？）濬等 | ○卷一後序 筆受：靈會・靈雋・智開・知仁・玄度・道卓・道觀・明覺 証梵語：玄薈 正字：玄應 証義：道洪・明琰・法祥・惠貴・文備・神泰・道深 証文：道智（卷 1-10）・行友（11-20）、玄口（21-29）・玄忠（30-34）・靖邁（35-50）・辯機（51-80）・處衡（81-84）・明濬（85-100） ²³ （大正 30, 283c） ○卷百 筆受：知仁・靈雋・道觀・明覺 証文：辨機・靖邁・行友・道智・玄忠 正字：明濬・玄應 証梵語：玄謨 証義：文口・神泰・道深・法祥・慧貴・明琰・道洪（同 881c） |
| 1580 | 1 | 瑜伽師地論釋 | 大乘暉 | |
| 1585 | 10 | 成唯識論 | 大乘基 | |
| 1586 | 1 | 唯識三十論頌 | 大乘光 | |
| 1590 | 1 | 唯識二十論 | 大乘基 | |
| 1594 | 3 | 攝大乘論本 | 大乘巍 | |
| 1597 | 10 | 攝大乘論釋 | 大乘巍等 | |
| 1598 | 10 | 攝大乘論釋 | 大乘巍・大乘林等 | |
| 1600 | 3 | 辯中邊論 | 大乘基 | |
| 1601 | 1 | 辯中邊論頌 | 大乘基 | |
| 1602 | 20 | 顯揚聖教論 | 智証 | |
| 1603 | 1 | 顯揚聖教論頌 | 辯機 | |
| 1605 | 7 | 大乘阿毘達磨集論 | 大乘光・大乘雲等 | |
| 1606 | 16 | 大乘阿毘達磨雜集論 | 玄頤等 | |
| 1609 | 1 | 大乘成業論 | 大乘光（雲？） | |
| 1612 | 1 | 大乘五蘊論 | 大乘光等 | |

²³ 「本地分中、五識身相應地意地・有尋有伺地・無尋唯伺地・無尋無伺地、凡十七卷、普光寺沙門道智、受旨證文。三摩呬多地・非三摩呬多地・有心無心地・聞所成地・思所成地・修所成地、凡十卷、蒲州普救寺沙門行友、受旨證文。聲聞地初瑜伽種姓地盡第二瑜伽處、凡九卷、玄法寺沙門玄口、受旨證文。聲聞地第三瑜伽處盡獨覺地、凡五卷、汴州真諦寺沙門玄忠、受旨證文。菩薩地・有餘依地・無餘依地、凡十六卷、簡州福衆寺沙門靖邁、受旨證文。攝決擇分、凡三十卷、大總持寺沙門辯機、受旨證文。攝異門分・攝釋分、凡四卷、普光寺沙門處衡、受旨證文。攝事分、十六卷、弘福寺沙門明濬、受旨證文。」

| | | | | |
|------|---|----------------|---------|--|
| 1614 | 1 | 大乗百法明門論 | 玄忠 | |
| 1615 | 1 | 王法正理論 | 大乗林 | |
| 1624 | 1 | 觀所緣縁論 | 大乗光 | |
| 1628 | 1 | 因明正理門論本 | 知仁 | |
| 1630 | 1 | 因明入正理論 | 知仁（明濬？） | 筆受：明濬 証文：明濬 証梵語：玄謙 正字：玄應 詳証大義：道洪・明琰・慧貴・法祥・文備・道深・神泰 |
| 2030 | 1 | 大阿羅漢難提蜜多羅所說法住記 | 大乗光 | |
| 2031 | 1 | 異部宗輪論 | 大乗基 | |
| 2138 | 1 | 勝宗十句義論 | 靈雋（携？） | |

これを見る限り、筆受が最も代表的な役職であることが推察される。袴谷憲昭氏も、筆受は「原語や漢語に通じていることはもとより、広く仏教の教義全般に通曉した人でなければなら」ず、「したがって、翻訳の第一段階における重要な役目である」と述べられている²⁴。

『開元釈教録』がどのような基準で筆受者の名前をこのように限定しているのかは定かではないが、このような場合最も代表的な人物をあげていると考えるのが妥当であろう。本発表の目的は全文献の比較研究であるため、網羅的な『開元釈教録』記載の筆受を中心として分析を行い、適宜識語等を参照する。

NGSM とクラスタ分析による玄奘訳経論の分析

使用ソフトウェアと分析方法

N グラム解析のためには、独自に開発した morogram²⁵を用い、NGSM 結果の出力には近藤泰弘氏作 ngmerge.pl²⁶を一部改造して使用した。また、クラスター分析には、STATSOFT 社の STATISTICA、フリーソフトウェアの R²⁷等を使用した（ただし、本稿における出力結果は STATISTICA による）。

今回の分析は、

1. 各経論を巻ごとに分解し、
2. 各巻ごとに 3 グラムの頻度を採取し、
3. それを長さが 100 のベクトルとなるよう正規化し、
4. 複数巻ある文献については、各ベクトルの平均値によってまとめ、
5. 全 74 文献を NGSM にまとめ（1,625,154 個の異なり文字列）、
6. 文献間のクラスタ分析（ウォード法、ユークリッド距離）を行う。

という手順で行われた。

分析結果と評価

樹状図のごとく、クラスタは大きく二つに分けられる。便宜上、上から一番目のクラスタを

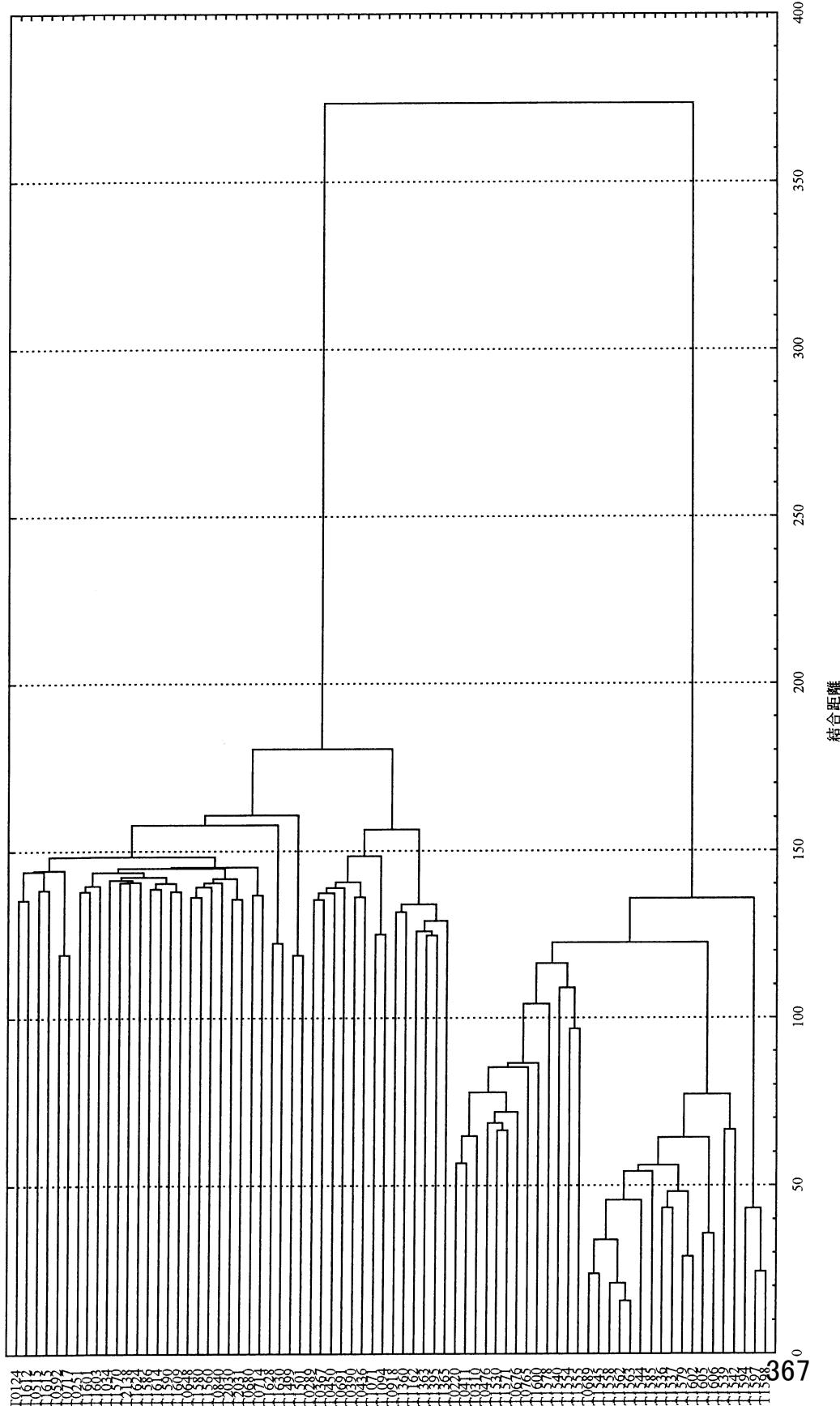
²⁴ 桑山・袴谷前掲書、p. 299。

²⁵ <http://sourceforge.jp/projects/morogram/>にて共同開発を行っている。

²⁶ <http://klab.ri.aoyama.ac.jp/tool/index.html>

²⁷ <http://www.r-project.org/>

樹状図 74 変数
ヨーロッパ距離



結合距離 150 のラインで 5 つ (1a～1e) に、2 番目のクラスタを 120 のラインで 3 つ (2a～2c) に分類したものが下表である。これを見る限り、文献の内容的な分類と各クラスタが概ね一致していると言つてよいと思われる。

| クラスタ | 文献 No. | 備考 |
|------|---|------------------|
| 1a | 124, 1612, 515, 1615, 592, 717, 251, 1601, 1603, 1034, 1570, 2138, 1624, 1586, 1614, 1590, 1609, 648, 1580, 1560, 840, 2030, 2031, 680, 714 | |
| 1b | 1628, 1630 | 因明論書 |
| 1c | 1499, 1501 | 菩薩戒本 |
| 1d | 289, 367, 450, 691, 390, 436, 1071, 1094 | 大乗經典、密教經典 |
| 1e | 918, 1360, 1162, 1363, 1395, 1365 | 密教經典 |
| 2a | 220, 411, 310, 476, 1530, 1571, 676, 765, 1600, 1578, 1540, 1554, 1555 | 大乗經典、大乘論書、阿毘達磨論書 |
| 2b | 689, 1545, 1558, 1562, 1563, 1544, 1585, 1536, 1537, 1579, 1602, 1605, 1606, 1539, 1542 | 甚希有經、阿毘達磨論書、対法論系 |
| 2c | 1594, 1597, 1598 | 攝大乘論系 |

さて、各クラスタと、筆受との対応についてみて見ると、まずクラスタ 1、2 の両方にわたっている筆受については、次の表のようにまとめることができる。

| 筆受 | クラスタ |
|-----|--|
| 大乗光 | 1a, 1c, 1d, 2a, 2b, 2c (ただし 2a～c の対応文献のうち、476, 1530, 676 を除いたものは「大乗光等」となっている。また 476, 1530, 676 は 5～7 巻あり、他の例から見て一人で筆受をしたとは考えづらい) |
| 大乗欽 | 1a, 1d, 1e, 2a (2a は大般若經) |
| 大乗雲 | 1a, 1d, 1e, 2b (ただし 2b は 1605 のみで大乗光等と連名) |
| 知仁 | 1a, 1b, 2b (2b は瑜伽師地論のみ) |
| 大乗基 | 1a, 2a, 2b (1a の文献は短いものばかり) |
| 元瑜 | 1a, 2b (1a は阿毘達磨俱舍論本頌〔しかも元瑜等とある〕のみ) |
| 大乗暉 | 1a, 2a (いずれも小編) |
| 神皎 | 1a, 2a (1a は縁起經、2a は大般若經) |

これ以外の筆受については、クラスタ 1 に慧朗、辯機、大乗詢、靈雋、玄忠、慧立、玄則、神察らが、クラスタ 2 には、嘉尚、道測、智証、敬明、靜邁、神昉、釈詮、海藏、玄頤、大乘巍、靈会、智開、玄度、道卓、道觀、明覺らがそれぞれ該当する。

興味深いことに、いずれの筆受についても、大乗暉などの場合を除いてクラスタ 1、2 に満遍なく属している者はなく、各クラスタとの間におおまかな対応関係を見いだすことができる。上表については今後の厳密な検討が必要であろうが、例えば大乗光はクラスタ 1 に、大乗基はクラスタ 2 に属すると見なすことができよう。ここから、玄奘訳の仏典においては、文

体上、筆受の影響が少なくなく、また大きく 2 つのグループに分類することができる、という仮説を形成することができるのではないだろうか。

まとめと今後の課題

以上、方法論的展望と、実際に玄奘訳全文献を用いた仮説形成について、一定の成果をあげることができたのではないかと思われる。今後は、この仮説を下敷きに、特に思想的にこのような筆受の影響が見いだしうるのか、文献学的に厳密な検討を行う必要がある。それによってこそ、仮説形成の意義があるからである。

今後の課題としては、

- 同じタイプの文献であるにもかかわらず、異なったクラスタに分類される例（1558 と 1560 など）の検討。
- 卷ごとのクラスタ分析。
- 筆受以外の役職との対応関係の検討。
- 処理の高速化、省メモリ化（NGSM 处理中、数十 GB の中間ファイルが生成され、その処理に 20 時間ほどかかる²⁸ため、気軽にトライ＆エラーを試みるのは難しい）。

などがあげられよう。諸賢のご叱正を請う次第である。

謝辞

本報告は、科学研究費補助金・若手研究 B「N グラムモデルを用いたクラスタ分析による大規模漢字文献分析の基礎的研究」（課題番号 15700215、研究代表者：師茂樹）による成果の一部である。

²⁸ 主な実行環境としては、Apple Power Mac G5 (CPU: Dual 2GHz PowerPC G5, RAM: 4GB), Mac OS X 10.3.5, Perl 5.8.1 など。

