chapter.03

文字検索のさらなる地平に向けて

――文字列の散在的一致を網羅するために

船山 徹

1. 序説

原典の電子化と通常型文字検索は、人文学原典研究の作業効率を飛躍的に向上させた。原典が文献名を明示せずに引用する文言を同定・特定することは、いまや通常検索の範囲で誰でも行える。さらに、電子化された文献を対象として、任意の語句ないし文を検索することにより、そこに含まれる電子資料にその語句が見いだせるか見いだせないかも断定可能となった。

ただ、知りたい文字列が検索対象とする全電子資料のうちに何カ所あるかを結論づける環境は依然として整っていない。すなわち通常方式で文字検索するだけでは、“完全に一致する文字列”を知るにすぎない。文字列の一部が異なる場合は完全一致から排除されるから、通常方式の検索の限界が現れる。この点において、漢語仏教叢書である大正新脩大蔵経電子版を検索するSATの2018年版に“曖昧(あいまい)検索”が付加されたのは大きな進歩である。

筆者は電子プログラム開発に携わる者ではない。古典的な従来型文献研究のために電子資料を用いる、単なる一ユーザーにすぎない。ただ、かなりの程度でヘビーユーザーである。それゆえ、従来型の通常文字検索の恩恵に浴する場合が多い一方で、その限界を痛感することも多い。本稿は、伝統的文献学研究者としてのユーザー目線から、近い将来に開発されることを切望する検索プログラムの具体的な一事例を取り上げる。

2. 文字検索の価値

筆者は前近代の漢語仏教文献を主な資料として中国中世の仏教における思想的展開と文化史的意義を研究対象とする。まず、わたしの研究領域において文字検索の価値を示そう。文字検索で知られることは多いが、とりわけ価値の高い事柄は、ほぼ以下の二点に集約できると感じる。

第一は、文字検索（語彙検索）は、原典資料に含まれる引用文献や、著者が想定する対論者を同定可能となることである。文字検索の結果から、当該文献著者が誰を、あるいは何を論敵として論述しているかを定めることができる場合がある。さらには当該文献の著者の心の中にある著作動機がわかる場合もある。被引用文献の年代特定によって、当該文献の年代的上限 *terminus post quem* を知ることもできる。当該文献を引用する後代諸文献を特定することで、当該文献の年代的下限 *terminus ante quem* を知ることもできる。仏教文献には成立年代の不明な場合が多いので、年代の上限と下限は思想史研究にとって意義深い。

第二に、直接的引用とは異なるが、当該文献の原文中が引用であることをあえて明示せず、先行文献を暗黙裏に使用することが多く、その場合、暗黙に用いられている文献は、当該文献作者の主張にとっての典拠となる場合も少なくない。前近代アジアでは現代のいわゆる著作権がなかったため、他者の説を無言で使用する事例はおびただしい。他者の言葉を暗黙に使用するのは現代いうところの“剽窃”とは異なる。無言であることが信頼性を表すことさえある。例えば師の教えを用いて何かを論述するとき、それが完全に著者自らの血となり肉となっているならば、師の名を明記して引用したりはしない。むしろ著者の言葉の地の文に師の説を取り込む。引用の形をとらない暗黙裏の文言使用はそれに対する信頼を含意することがある。そうした事例の解明は、著者の学識・帰属学派・思想的系譜・上限年代などを知る上で必須である。

3. 文字検索の二種

ある任意の文献と、その前後の年代に編まれた文献を対象とする文字検索に

は大別して二種類があると言ってよいだろう。

第一種は、文字列の“完全一致”による検索である。これがいわゆる通常の文字検索である。文献研究のかなりの部分はこの検索で事足りる。しかし通常検索で得られる完全一致の文字検索結果だけでは研究に使えない。著者が同じ意味で同義語・類義語を用いる、表記を意図的に変えるなどによって、論述の単調さや語彙が貧困している印象を避ける場合もある。それゆえ、完全一致型の文字検索では研究に必要な全事例を抽出できない。

第二種は、文字列の“部分的一致”である。これは、直前に記述したような事例を知る上で効果がある。例えば著者が原文を韻文で著し、それを散文で自ら注釈する場合、著者は韻文の語とはあえて別の語を用いて言い換えて注解する。

この第二の場合について従来型の通常検索しか使えないとすれば、現時点で研究者はどのような検索で対処するかと言えば、最も一般的方法は、「当たり」を付けて少しずつ文字列を変え、意味的には同義であるが文字的には異なる文字列を複数種想定し、そのすべてを逐一検索するという方法である。こうして異なる文字列を自ら作り出し、何度か通常検索することを繰り返せば、完全一致方式であっても、検索結果の数量を増やせる。だがその結果はやはり満足からほど遠い。あらゆる可能な文字列を網羅的に検索したかどうかという点でまだ欠陥があるのではという不安を払拭できないからである。

4. 現時点での対策

前節末に示した不安が自らの研究の過程で生じた場合、われわれはどう対処すべきか。わたしはコンピューター技術の基本的知識と最新の知識とを欠くので間違っていたらすぐ潔く撤回するにやぶさかでないが、筆者が思うに、前節に示したような完全一致型文字検索で問題が残る場合に対して、現時点で可能な対策をもしあげるとするなら、大別して三種あるのではないだろうか。

対策第一は、すでに述べた「曖昧検索」である。SAT2018 の場合に即せば、検索オプションの項目に、①「通常検索」（文字数は無制限）、②「正規表現」（8 文字まで）、③「曖昧フレーズ検索」その一「8 文字までの一次違い」、④「曖

昧フレーズ検索」その二「5-8 文字の 2 文字違い」の四種から一つを選択できる。曖昧フレーズ検索の二種③④は、想像するに、さまざまな事例で実験した結果、最も現実的効果の高い二種をあげたのであろう。しかしそうであっても、文字数に制限を設定することは、それ以外の方法ではこのウェブサイトの曖昧検索が不可能なことを意味するから、完全無欠の曖昧検索をしたくても SAT2018 のインターフェースでは実行不能である。

対策第二は、SAT2018 にも含まれる正規表現 regular expression である。すでによく知られている通り、正規表現は汎用度の高い、有効な手段である。例えば「X.*Y」という文字検索は、文字 X を含み、その後 0 から無限数までの文字を経て文字 Y が現れる全結果を表示する。しかし実際には、文字 X と文字 Y のあいだに 100 字が挟まれるような場合は通常の電子データと通常の GREP（global regular expression print）ソフトとでは検出できない場合が多い。つまり X と Y に介在する文字数にはおのずと上限があるという現実問題がある。それゆえ、正規表現を駆使して複数回の検索を繰り返しても、見過ごした重要テキスト・文字列がまったくないとは決して言い切れない。

対策第三は、N-gram に基づく分析であろう。これを仏教文献に応用する方法を述べた早期の概論として石井（2002）がある。同様に N-gram に基づく漢語大蔵経分析ソフトウエアとして、Jamie Norrish と Michael Radich が共同開発した TACL が無料公開されている。TACL の恩恵に浴した自らの研究に船山 2016, 2018 がある。普段筆者は通常検索をするにすぎないが、通常検索では検出できなかった事例を TACL は示してくれたのだった。

N-gram は、意味と無関係に、文字列を任意の文字数ずつ機械的に区切り、そのすべてを他文献と比較して一致・不一致を結果として示す。文字数を何字と設定すれば最も効果的か等において使用者側に選択の余地がある。そうした中で N-gram の最大の功績は、文字列の文字数と一種に限定せず、二文字〜十文字（あるいはそれ以上）に変更して繰り返し検索すれば、完全一致する文字列の全体を遺漏なくすべて獲得できるという点である。

しかしながら、N-gram はこれまで不可能だった文字検索をさまざまな形で可能にした大きな功績と同時に、N-gram に特有の大きな問題をも生み出した。それは N-gram による任意の文字数の文字検索で得られる結果は膨大な数に上

り、そのすべてから自らにとって有意味の結果を過不足なく選択しなければならないのである。言い換えれば、N-gram では巨大な“ジャンクデータ”が発生するため、それらを無意味と排除し、データ数を大幅に絞り込むことによって、残った極少数の有意味データを求める必要がある。ところがこの絞り込み作業を機械的に行うことは、現時点で達成されていない。絞り込み作業は、膨大なジャンクデータを含む何千、何万という事例から、それを機械によってでなく、人間の目によって“マニュアル”で遂行しなければならない。この点は TACL も同様である。筆者は完全に無意味と断定できる結果を機械的に除去するソフト――ただし有意味のデータまで誤って含まないよう、やや絞り込み度の甘いソフト――を開発して欲しいと開発者に繰り返し要望した経緯があるが、残念ながら現時点では「ジャンクデータ」を除去するアプリケーションを作ることに TACL はまだ成功していない。

網羅的かつ客観的に文字列の一致する全データを獲得しても、その中から本当に必要なデータを見つけるには「慣れ」と「勘」に頼るしかないというのが現状なのである。これでは、何を「ジャンク」と見なし、何を「有意味」と見なすかの区別は、限りなく主観的判断から逃れられない。機械による解析を、最後は完全手動の勘で処理するという方法である限り、もし分析する人が異なれば、分析結果も異なる可能性は極めて高い。これは客観的かつ網羅的な検索結果を得られるかどうかという点で大問題である。この状態では、巨大なジャンクデータからほんの一、二例の有意義なデータを探し出す過程で自らが下した結論が唯一の正解か、それともほかの可能性もあったのか、まったく別の人が絞り込み作業を行ったとしたらまったく別な結果となったのか等の点で確信が持てないままとなってしまう。

5.「散在的一致」――文字検索が現在不可能な事例

以上に略説した事例とは別に、本稿で特に紹介し、近い将来に信頼し得る検索結果を得たい種類の原典資料がある。それを理論的に言葉で十分説明することは筆者の能力を超えているため、端的な例を示しながら説明したい。

筆者が主に扱うのは 5 ～ 8 世紀の中国仏教である。王朝名で言うと南北朝時

代（特に南朝）・隋代・唐の初唐～盛唐時代である。この時代における南朝仏教の大きな転回は南斉（479 ～ 502）と梁（502 ～ 557）に起こった。

具体的には、中国歴代皇帝の中で最も仏教を篤く信仰した皇帝が梁の武帝（位 502 ～ 549）だったことは非常によく知られているが、武帝が仏教信仰を確立した直接原因を形成した人物として、南斉の武帝の第二子だった蕭子良（460 ～ 494）の名をあげることができる。

蕭子良は若くして逝去したにもかかわらず膨大な著作を行った。その中でもとりわけ同時代と直後の時代に最も大きな影響を与えた書は、蕭子良撰『浄住子』20 巻であった。ところがこの書は量的に大きかったことが災いして、歴史の中で比較的早期に散逸し、現在のわれわれが 20 巻本の全貌をくまなく知ることはできない。現在残っているのは初唐の道宣（596 ～ 667）が 20 巻本を二十分の一の分量に縮小した節略本『統略浄住子浄行法門』1 巻のみであり、664 年に道宣が編纂した『広弘明集』巻 27 におさめられている。筆者はかつて『統略浄住子浄行法門』の校訂訳注研究として船山（2006）を公表した。その際、さまざまな関心の中でも特に常に知りたいと願ったのは、散逸した蕭子良『浄住子』20 巻の具体的な形式ならびに内容であった。ところがそれを知るために、種々の方法で大蔵経中を渉猟したが、具体的に 20 巻原本の内容を確定することは、ほとんどできなかった。

ただ、そうした作業の中で注目すべき文献がいくつかあることには気づいた。それらのうち、第一の種類は、『浄住子』からの引用であることを明示しながらも、その文言を道宣の『統略本』に同定できない事例だった。それらが道宣の節略本からでなく、蕭子良の原本から引用した可能性は大きいと推定されたので、それら引用断片を網羅的に蒐集して示した。

さらに第二の種類は、『浄住子』からの引用であることを一切示していないにもかかわらず、道宣の『統略本』が記述内容と記述順序に一致する文献であった。その中でも特に注目すべき仏教書として梁初頃に原形が編纂され、その後さらに一部を補足して現存形となった撰者不明の『慈悲道場懺法』10 巻がある。以下に例示するのは、『慈悲道場懺法』と『統略浄住子浄行法門』のあいだに見られる文字列の一致である。

わかりやすくするため、筆者の注目する結論を先にあげておこう。まず、文

献は以下の順で成立した。

1. 南斉・蕭子良『浄住子』20巻　西暦490年成書　散逸（現存せず）
2. 梁・撰者不明『慈悲道場懺法』10巻　500年代前半　現存
3. 唐・道宣『統略浄住子浄行法門』1巻　664年　現存（1の縮小版）

このような成立順序を頭に入れた上で、2『慈悲道場懺法』の文言を精査すると、その中には3道宣『統略』本と一致する語句が、まったく同じ順序（！）で現れる箇所が複数ある。しかし2における文字を見ると、3と一致する文字列は出現する順序がまったく同一であるにもかかわらず、3と一致する文字列が2には散在的にしか現れないのである。

“散在的な一致”を別の表現で言い直せば、3と一致する文言が2では同じ順序で“飛び飛びに現れる”のである。“文字列が緩やかに斑状（まだらじょう）に一致する”と表現してもよいかもしれない。

では2と3はどのような関係にあるかと言えば、2は年代的に先行するから、3を素材として編纂したと考えることはできない。2は3を知らずに撰述されているにもかかわらず、3と散在的に一致するのである。

ところで一方、2『慈悲道場懺法』は1『浄住子』原本の直後に編纂されたから、1を素材として2を編纂した可能性が否定できない。むしろその可能性が極めて高い。ⓐ1の原文が散逸し現在残っていないにもかかわらず道宣『統略本』に二十倍する分量であったこと、ⓑ2と3（＝1の縮小版）に文字列の一致が散在的に見られること、ⓒ一致する文字列のあいだに、内容を補填する説明語句が多くの字数で挟まれていること――以上の3条件を満たす文献として2を扱い、その内容を3と比較すると、3と一致する文字列に挟まれた、一致しない2の文字列は、1蕭子良『浄住子』20巻本からの直接的引用を多く含む可能性が高いことを、論理的に、矛盾なく理解できる。以上に列記した事柄の例として次の表を見てほしい。『慈悲道場懺法』と『統略浄住子浄行法門』のあいだで一致する文字列を**太字**で示す。さらに、それら完全一致型文字列のあいだに挟まれた、内容的に連続する2の文字列を<u>下線</u>で示す。現時点の仮説として、<u>下線</u>は蕭子良『浄住子』20巻から引用した文字を極めて多く含んでいるので

はないかと筆者は推測している。

【成立順 1】現存 撰者不明『慈悲道場懺法』10 巻 最古層は 6 世紀前半頃（後代さらに補足）	【成立順 2】現存 道宣『統略浄住子浄行法門』 1 巻　664 年
又願今日道場同業大衆，各自發如是願。 尋**衆惡所起，皆緣**六根。是爲六根衆禍之本。雖爲禍本，亦能招致無量福業。故『勝鬘經』言，「**守護六根**，浄身口意」。以此義證生善之本故，於六根**發**大誓**願**。	原**衆惡所起，皆緣**意地貪，瞋，癡也。自害害他，勿過於此，故『經』號爲根本三毒。能煩能惱，勞擾身心，於緣起惡，三三九種。然此九種，義通善惡。三善根生，名善業道，三不善根生，名惡業道。是故行人常一其心，不令動亂。微塵起相，見即覺察，**守護六根**，不令塵染。常**發**弘**願**，以自莊嚴。
(1) 初發眼根願。　**願**今日道場同業大衆，廣及十方四生，六道**一切衆生，從今日**去，**乃至菩提，眼常不**見**貪**欲無厭詐幻之色，不見諂諛曲媚佞會之色，不見玄黄朱紫**惑人之色，不**見**瞋恚**，鬪諍，**醜状**之色，不見打扑苦惱損他之色，不見**屠裂**傷毀衆生之色，不見**愚癡**無信疑闇之色，不見無謙無敬**驕慢**之色，不見九十六種**邪**見**之色**。眼常不見如是一切衆惡不善之色。 **願**眼常**見一切十方常住法身**湛然**之色**，常見三十二**相**紫磨金色，常見八十種**好**隨形之色，常見諸天諸仙奉寶來獻散華之色，常見口出五種色光説法度人之色，常見分身散體遍滿十方之色，常見諸佛放肉髻光感有緣來會之色。又願眼常見十方菩薩，辟支，羅漢衆聖之色，常得與諸衆生及諸眷屬觀佛之色，常見衆善無教假色，常見七覺浄華之色，常見解脱妙果之色，常見今日道場大衆歡喜讚法頂受之色，常見四衆圍繞，聽法，渴仰之色，常見一切布施，持戒，忍辱，精進之色，常見一切静默，禪思，修智慧之色，常見一切衆生得無生忍，現前受記，歡喜之色，常見一切登金剛慧，斷無明闇，補處之色，常見一切沐浴法流不退	(1) **願一切衆生，皆從今日，乃至菩提，眼常不**看**貪**婬邪艶**惑人之色，不**看**瞋恚，醜状，屠裂，愚癡，闇鈍，倨慢邪**衆**之色**。 **願見一切十方常住法身之色**，菩薩下生八相之色，如來**相好**，聖衆和會善集之色。

之色。已發眼根願竟，相與至心五體投地，歸依世間大慈悲父。　……	
(2) 次發耳根願。又願今日道場同業大衆，廣及十方四生，六道一切衆生，從今日去，乃至菩提，耳常不聞啼哭愁苦悲泣之聲，不聞無間地獄受苦之聲，不聞鑊湯雷沸振響之聲，不聞刀山劍樹鋒刃割裂之聲，不聞十八地獄間隔無量苦楚之聲。又願從今日去，耳常不聞餓鬼飢渴熱惱求食不得之聲，不聞餓鬼行動節間火然作五百車聲。又願從今日去，耳常不聞畜生身大五百由旬爲諸小虫絢食苦痛之聲，不聞抵債不還生綾駝驢馬牛身常負重鞭杖楚撻困苦之聲，耳常不聞愛別離怨憎會等八苦之聲，不聞四百四病苦報之聲，不聞一切諸惡不善之聲，不聞鐘鈴螺鼓琴瑟箜篌琳瑯玉珮惑人之聲。 唯願一切衆生，從今日去，耳常得聞諸佛説法八種音聲，常聞無常苦空無我之聲，常聞八萬四千波羅蜜聲，常聞假名諸法無性之聲，常聞諸佛一音説法各得解聲，常聞一切衆生皆有佛性法身常住不滅之聲，常聞十地菩薩忍音修進之聲，常聞得無生解善入佛慧出三界之聲，常聞諸法身菩薩入法流水眞俗並觀念念具足萬行之聲，常聞十方辟支羅漢四果之聲，常聞帝釋爲諸天説般若之聲，常聞十地補處大士在兜率宮説法不退轉地行之聲，常聞萬善同歸得佛之聲，常聞諸佛讚歎一切衆生能行十善隨喜之聲。 願諸衆生常聞諸佛讚言「善哉是人，不久成佛」之聲。已發耳根願竟，相與至心五體投地，重復歸依世間大慈悲父。　……	(2) 願一切衆生，耳常不聞悲啼，愁歎聲，地獄苦楚聲，餓鬼，畜生受苦聲，八苦交對聲，四百四病起發聲，八萬四千塵勞聲。 願耳常聞諸佛説法八音聲，八萬四千波羅蜜聲，三乘聖果，十地功德，如是等聲。
(3) 次發鼻根願。又願今日道場同業大衆，廣及六道一切衆生，從今日去，乃至菩提，鼻常不聞殺生，滋味，飲食之氣，不聞畋獵，放火，燒害衆生之氣，不聞蒸煮，熬炙衆生之氣，不聞三十六物革囊臭處之氣，不聞錦綺羅鮎惑人之氣。又願鼻不聞地獄剥裂燋爛之氣，不聞餓鬼飢渴，飲食，糞穢，膿血之氣，不聞畜生腥臊不淨之氣，不聞病	(3) 願一切衆生，鼻常不聞殺生，滋味，飲食之氣，三十六物革囊之氣，發欲羅綺脂澤之氣，五辛能薰九相尸氣。 願鼻常聞十方世界諸樹草木之香，五戒，八戒，十善，六念諸

臥床席無人看視瘡壞難近之氣，不聞大小便利臭穢之氣，不聞死屍斡脹虫食爛壞之氣。 唯願大衆六道衆生，從今日去，鼻常得聞十方世界生頭旃檀無價之香，常聞優曇鉢羅五色華香，常聞歡喜園中諸樹華香，常聞兜率天宮説法時香，常聞妙法堂上遊戯時香，常聞十方衆生行五戒，十善，六念之香，常聞一切七方便人，十六行香，常聞十方辟支，學，無學人衆德之香，常聞四果，四向得無漏香，常聞無量菩薩①歡喜，②離垢，③發光，④焔慧，⑤難勝，⑥遠行，⑦現前，⑧不動，⑨善慧，⑩法雲之香，常聞衆聖戒，定，慧，解脱，解脱知見五分法身之香，常聞諸佛菩提之香，常聞三十七品，十二縁觀，六度之香，常聞大悲，三念，十力，四無所畏，十八不共法香，常聞八萬四千諸波羅蜜香，常聞十方無量妙極法身常住之香。已發鼻根願竟，相與至心五體投地，歸依世間大慈悲父。……	功德香，學，無學人，十地，五分，十力，八萬四千諸度無漏之香，十方諸佛説法之香。
(4) 次發舌根願。　又願今日道場同業大衆，廣及十方四生，六道一切衆生，從今以去，乃至菩提，舌恒不嘗傷殺一切衆生身體之味，不嘗一切自死之味，不嘗生類血髓之味，不嘗怨家對主毒藥之味，不嘗一切能生貪愛煩惱滋味之味。 願舌恒嘗甘露百種美味，恒嘗諸自然飲食之味，恒嘗香積香飯之味，恒嘗諸佛所食之味，恒嘗法身戒定慧之所熏修所現食味，恒嘗法喜禪悦之味，恒嘗無量功德滋治慧命甜和之味，恒嘗解脱一味等味，恒嘗諸佛泥洹至樂最上勝味之味。已發舌根願竟，相與至心五體投地，歸依世間大慈悲父。……	(4) 願一切衆生，舌恒不嘗衆生有命身肉雜味，能生煩惱滋味。 願舌恒嘗甘露不死之味，天自然食在其舌根變成上味，諸佛所食之味，法喜禪悦之味，解脱泥洹最上勝味。
(5) 次發身根願。又願今日道場同業大衆，廣及十方一切衆生，從今日去，乃至菩提，身常不覺五欲邪媚之觸，不覺鑊湯，爐炭，寒氷等觸，不覺餓鬼頭上火然烊銅灌口焦爛之觸，不覺畜生剥裂苦楚之觸，不覺四百四病諸苦之觸，不覺大熱大寒難耐之觸，不覺蚊蚋蚤蝨諸虫之觸，不覺刀	(5) 願一切衆生，身常不覺邪婬細滑，生欲樂觸，不覺鑊湯寒氷之觸，餓鬼畜生諸苦惱觸，四百四病，寒熱風霜，蚊虻蚤虱，飢渇困苦等觸。

杖毒藥加害之觸，不覺**飢渴困苦**一切諸**觸**。 **願身常覺**諸天妙衣之觸，常覺自然甘露之觸，常覺**清涼**不寒不熱之觸，常覺不飢不渴無病無惱休**強**之觸，常覺**臥安**覺**安**無憂無怖之觸，常覺十方諸佛淨土微風吹身之觸，常覺十方諸佛淨國七寶浴池洗蕩身心之觸，常覺無老病死諸苦之觸，常覺**飛行**自在與諸菩薩聽法之觸，常覺諸佛涅槃**八自在觸**。已發身根願竟，相與至心五體投地，歸依世間大慈悲父。	**願身常覺清涼強**健，心悟**安**隱，證道**飛行**，**八自在觸**。
(6) 次發意根願。又**願**今日道場同業大衆，廣及十方**一切衆生，從今日**去**乃至菩提，意常**得**知**貪欲瞋恚愚癡爲患，常知身殺盜婬妄言綺語兩舌惡口爲患，常知殺父害母殺阿羅漢出佛身血破和合衆是無間罪，常知謗佛法僧不信因果人死更生報應之法，常知遠惡知識親近善友，常知諮受**九十六種邪師之法**爲非，常知三漏五蓋十纏之法是障，常知**三途可**畏，**生死**酷劇**苦**報之處。 **願意常知一切衆生皆有佛性**，常知諸**佛**是大慈悲父無上**醫王**，一切尊**法爲**諸衆生病之**良藥**，一切賢聖**爲諸衆生看病**之母。願意常知歸依三寶應受五戒，次行十善，如是等法能招天上人中勝報，常知未免生死應修七方便觀煖頂法等，常知應行無漏苦忍十六聖心，先修十六行觀，觀四眞諦，常知四諦平等無相，故成四果，常知總相，別相一切種法，常知十二因緣，三世，因果輪轉，無有休息，常知修行六度，八萬諸行，常知斷除八萬四千塵勞，常知體會無生，必斷**生死**，常知十住階品次第具足，常知以金剛心斷無明闇得無上果，常知體極一照萬德圓備累患都盡成大涅槃，常知佛地十力四無所畏十八不共無量功德無量善法。已發意根願竟，相與至心五體投地，歸依世間大慈悲父。 (大正 45,963c-966a)	(6) **願一切衆生**，皆**從今日**，**乃至菩提，意常覺知**九十八使，八萬四千塵勞之法，十惡五逆，**九十六種邪師之法，三塗可**厭，**生死**大**苦**。 **願意常知一切衆生皆有佛性，佛**爲**醫王，法爲良藥**，僧爲**看病**者，**爲諸衆生**，治**生死**患，令得解脱，心常無礙，空有不染。 (大正 52,321ab)

一覧表は以上の通りである。**太字**で示した文字列が、左欄に示した『慈悲道場懺法』においては“散在的一致”を示している様子を理解いただけるに違い

ない。紙数の制約のため、上記原文に現代語訳を付して内容理解の便を図る余裕がないことを遺憾とするが、**太字の語**に挟まれた下線文字が、**太字文字列の内容**をさらに詳しくつなげていることがわかる。それらは決して**太字**の注釈や言い換えではない。むしろ**太字**では示しきれない事柄を下線部は示す。

本来の相当に長い文章——蕭子良『浄住子』20巻——から、唐の道宣は特に重要な語をピックアップして圧縮的に用いた結果が3の文献とりわけ**太字の部分**であり、そして、それらのあいだに介在していたさらに細かな説明—— 2の下線部すなわち蕭子良『浄住子』20巻の語句——を割愛することによって、内容的には蕭子良自身の語彙をそのまま用いながら、分量を二十分の一に縮小する作業を実現したと、このように想定すると、具体例を最も自然に、無理なく、論理的に解釈できると筆者はいまのところ考えている。

6. 提案したい結論

前節に示した“散在的一致”が『慈悲道場懺法』に確認できるということは、実は拙稿（2006）においてすでにしていた指摘であり、それ自体は最近見いだした新視点ではない。しかしながら筆者は、このような“散在的な文字列の一致”を大蔵経全体の中から網羅的に探そうと切望しつつ伝統的な通常形式の文字検索を、文字列を少しずつ変更しながら、幾度となく試みてきた。事実(2006)の原稿作成時にその作業はやり尽くしたつもりになっていた。しかし昨年、一年かけて“散在的な文字列の一致”を再検討しようと思い立ち、同じ作業を再び行ってみた。すると、2006年までの時点ではまったく見つけられなかった“散在的な文字列の一致”をさらに多く発見し、補足する結果となった。このことは、第五節に例示したような“散在的な文字列の一致”を完全網羅的に行い、信頼に足る結果を得ることは、現時点の検索においては実現不可能な、未解決の問題であることを如実に示している。

このような散在的な飛び飛びに一致する文字列を有する文献を大蔵経から網羅的に抽出することは、現在の“手動型”の検索では扱いきれない、限界領域である。しかしながら、現存しない蕭子良『浄住子』20巻を内容を文献学的に満足できる形で遂行するためには、今後開発すべきツールとして、《多くの

文字列を介在しながら散在的に一致する文字列を含む複数文献》を網羅的に知るためのソフトウエアが是非とも必要であることを提言する。

略号と先行研究

SAT2018　SAT 大正新脩大蔵経テキストデータベース 2018 版 .
http://21dzk.l.u-tokyo.ac.jp/SAT2018/master30.php

TACL　https://tacl.readthedocs.io/en/latest/

石井（2002）　石井公成「仏教学における N-gram の活用」、『明日の東洋学』8, 東京大学東洋文化研究所附属東洋学研究情報センター報、2002、pp. 2-4.

船山（2006）　船山徹『南斉・竟陵文宣王蕭子良撰『浄住子』の訳注作成を中心とする中国六朝仏教史の基礎研究』、科研報告書基盤研究（C）、2006, 324p.

同（2016）　同「《大方便仏報恩経》編纂所引用的漢訳経典」、方廣錩（主編）『仏教文献研究』2、広西師範大学出版社、2016、pp. 175-202.

同（2018）　同「梁の宝唱『比丘尼伝』の定型表現――撰者問題解決のために」、『東方学』135、2018、pp. 36-53.