

# ばつくとどうびばすとその五四

## 禁断の？「くずし字AI-OCR」による

### 会計史料の翻刻実験

#### 和式簿記法研究の現状

会計学の研究領域に「会計史」があり、江戸時代の商家による帳合（和式簿記法）を解明する研究が確立されています。そして、『江州中井家帖合の法』（ミネルヴァ書房、1962年）をはじめとする我らが小倉榮一郎先生による中井家帳合法を解明した一連の研究成果は、本学の会計教員による研究成果のなかで最も誇るべきものといつてよいはずで

すが、現在、若手・中堅層の会計研究者が和式簿記法研究を志すことが難しい状況に陥っています。「ジョブマーケットのニーズ」と「くずし字リテラシー」が、その要因の最たるものといえるでしょう。後者について、一説によると「くずし字が読める人」の割合は、0.01%だそうです（<http://codh.rois.ac.jp/nimo>）。そうすると、日本の会計研究者のほとんどすべてが会員となる日本会計研究学会の会員数は1694名（2022年3月31日現在）であり、「くずし字がきちんと読める」会計研究者の数は一名を割る計算になります（実質的には意味のない数字でしょう）。これまで本学にご縁のあった会計教員たちが個々のライフワーク的な研究テーマに加えて中井家帳合法研究に新規参入した形跡がまったくないのも、「くずし字リテラシー」が大きな障壁となったためであるように思われます（御多分に漏れず、筆者もくずし字の学習経験はありません）。

#### 「くずし字AI-OCR」の解禁

近年、「くずし字AI-OCR」が開発され普及しつつあります。「Deed」や「Grammarly」などを駆使して英語論文を執筆することへの抵抗感がなくなりつつある現代において、「くずし字リテラシー」を有しない

会計研究者の和式簿記法研究への参入を促すべく、「くずし字AI-OCR」を解禁する機運が高まったもよいように思います。そのきっかけとして、幸いなことに、今般、史料館とTOPPAN株式会社のご協力・ご支援を賜り、「中井源左衛門家文書」を対象としてくずし字AI-OCR「ふみのはせみ」による翻刻を行い、その精度を検証する実験を行う機会を頂くこととなりました。

往々にして、新しいことへの挑戦は、ふとしたきっかけで始まるものです（小倉先生による一連の研究も、江頭恒治先生や原田敏丸先生からの働きかけによって始まったとか）。今回の実験も、無碍に断りにくい？人たちからの打診が事の始まりです。ちなみに、「ふみのはせみ」を史料の解読に積極的に利用した先行研究は、筆者の知る限り、一編（近世経済史料研究会「貞享五年・伊豆蔵五兵衛『店法度・作法并異見之事』―伊勢商人の江戸店の店則―」『三井文庫論叢』第56号、2022年、313―348頁）しかありません。しかも、その一編は会計史料を対象としたものではありません。しかも、その対象とする今回の実験は本邦初となります。

#### 実験の詳細

「ふみのはせみ」による翻刻の手順を大まかに整理すると、①OCR自動処理（フルオートOCR）↓②フルオートOCRで認識されなかった文字枠の追加↓③ここまでで認識された文字枠の誤りの修正↓④翻刻誤りの修正↓⑤並び順の修正の順に行われます。「くずし字リテラシー」を有しない会計研究者が和式簿記法研究に参入することを促すためには、①の段階で相応の精度が担保された翻刻が提供される必要があります。そこで、今回の実験は、①の時点における翻刻を精度検証の対象とすることとしました。また、比較対象とすべき解答例を別途用意する必要がありますが、先行研究で翻刻されたものを拝借することが効率的です。そこで、今回の実験は、小倉先生による一連の研究において翻刻された中井家本家、仙台店グループ（元方・見世・質店）、押立店、香良洲店の会計史料を対象とすることとしました。

1 『令和四年度日本会計研究学会会報』日本会計研究学会、2023年、18頁。

2 「ふみのはせみ」は、TOPPANホールディングス株式会社の登録商標で

あくまでも個人としての実感ですが、「ふみのはせみ」には長所も短所もあるといったところです。素人でも容易に解読できそうな文字を誤って翻刻している箇所もあれば、目視では見落してしまいそうな小さな文字も認識し、正確に翻刻している箇所もありました。さらには、後述するとおり、先行研究による翻刻の誤りや創作を発見するきっかけになることもありました。

なお、「ふみのはせみ」は、会計史料の学習を行っていないそうです。そこで、

一金〇〇兩〇歩ト

摘要書

〇〇匁〇分〇厘

という会計史料に固有のフォーマットを学習させれば、精度が向上するはずですが。例えば、「ふみのはせみ」は、史料によっては「兩」と「匁」をうまく区別することができませんでした。一行目には「兩」、二行目には「匁」が書かれることを学習させておけば、位置関係から「兩」と「匁」を明確に区別して翻刻できるようになるでしょう。

また、各種翻訳ツールを用いた英語論文の執筆に際しては、最後にネイティブチェックを依頼するはずですが。同様に、「くずし字AI-OCR」による翻刻を行う際にも、最終段階で「くずし字がきちんと読める人」に気兼ねなくチェックをお願いできる環境が不可欠です。

### ― 思いがけない発見 ―

「ふみのはせみ」による翻刻を行った一次史料（写真を参照）と解答例とを対照表示した資料（合計190頁）を成果として公表すべく、一次史料と小倉先生による翻刻を照合する作業を進めていくうちに、小倉先生による翻刻には、①翻刻誤り、②複数項目の金額を合算した翻刻、③さらには一次史料にはない文字が盛り込まれていることが判明しました。①については、計算誤りも含め、存命の論者がすでに詳らかにしている箇所もありますが、今回の作業をつうじてそれら以外にも誤りがあることが確認されました。

史料館規程により、実験対象とした史料データを「ふみのはせみ」の再学習に利用しないよう、使用契約の締結に先がけて TOPPAN 株式会社との間で調整を行っています。

いうまでもなく、「くずし字 AI-OCR」による翻刻の精度は、個々の史

この事実は、とりわけ「再現性」に照らして、研究環境が現代とは異なる萌芽期における翻刻や計算結果の正確性に対する疑念を拭うことができず、一次史料をもとにゼロベースで地道に確かめる作業が不可欠であることを示唆しています。畏敬の念を抱きつつも、「小倉先生がそのように翻刻・指摘しているから絶対に正しい」と妄信し、先行研究を所与とすることは禁物です。

そうすると、よい意味で余計な感情が介入することのない「ふみのはせみ」をはじめとする「くずし字 AI-OCR」が、今後の検証作業においてよきパートナーとなってくれるのかもしれない。



押立店「成年店卸勘定帳」  
（寛政3（1791）年1月）  
（史料館請求番号：11821）

（経済学系教授 赤塚尚之）

料の状況や状態にも依存します。実験対象とした史料には、上記フォーマットの派生型で AI-OCR にとって読取り（とくに並び順の認識）が困難なもの、（手代の）筆跡に癖があるもの、虫損が激しいものがあり、文字枠を適切に認識できないことも多々ありました。