

オピニオンペーパーNo.58「デジタル時代におけるシチズン・サイエンス—市民社会と協働した人文・社会科学研究の可能性—」（2021年8月4日発行）

正誤表

箇所	誤	正	訂正日
3.市民科学の新しい潮流 過去に眠るビックデータを蘇らせる	一方で、文系・理系の垣根を越えた人文・社会科学の事例として紹介されたのが国立情報学研究所の北本教授が取り組む「みんなで翻刻」プロジェクトだ。	一方で、文系・理系の垣根を越えた人文・社会科学の事例として紹介されたのが、 <u>国立歴史民俗博物館の橋本雄太助教、東京大学地震研究所の加納靖之准教授らを中心に、同博物館、同研究所、京都大学古地震研究会などが協力して構築を進めるプロジェクト「みんなで翻刻」だ。</u>	2021年8月31日
3.市民科学の新しい潮流 過去に眠るビックデータを蘇らせる	だが実のところ、参加者の大半は古文書の解読に初挑戦するいわば素人だった。そこで開発されたのが「くずし字学習支援アプリ」だ。前出のモニタリング調査においても、市民が調査員として活動する下準備として、ゲーム形式で昆虫や植物の分類するトレーニングが行われていた。また、 <u>翻刻プロジェクトでは、ウェブ上で参加者が相互に添削しあうことでも正確性を高めていった（Wikipediaと同様の方式）。</u>	驚くべきは、参加者の多くは古文書の解読に初挑戦するいわば素人だということだ。素人でも参加できるのは、 <u>大阪大学文学研究科が中心に開発した「くずし字学習支援アプリKuLA」を初学者用の学習コンテンツとして提供し、さらに国文学研究資料館などが公開する「くずし字データセット」を教師データとして訓練したAIをもちいて、参加者の解読をサポートしているからだ。この自動くずし字認識ソフトは、人文学オープンデータ共同利用センター（CODH）と、凸版印刷株式会社が、それぞれ独自に開発した。加えて、ウェブ上で参加者が相互に添削しあうことでも正確性を高めていった（Wikipediaと同様の方式）。前出のモニタリング調査においても、市民が調査員として活動する下準備として、ゲーム形式で昆虫や植物の分類するトレーニングが行われていた。</u>	同上
3.市民科学の新しい潮流 市民の解読データがAIの“燃料”になりうる	市民の解読データがAIの“燃料”になる	市民の解読データがAIの“燃料”になりうる	同上
3.市民科学の新しい潮流 市民の解読データがAIの“燃料”になりうる	実は、 <u>翻刻プロジェクトには続きがある。Kaggle（カグル）のコンペティションに出しているのだ。Kaggleとは世界中の機械学習エンジニア、データサイエンティスト約40万人をつなぐプラットフォームで、政府や企業などが提示した課題に対し、最も優れた分析モデルの開発者に賞金を与えるコンペが行われている。つまり今回の場合は、5千人の目によって解読された600万字の翻刻データを「教材」としてAIに学習させ、自動認知を可能にすることでさらに膨大な資料の解析を進めようというわけだ。</u>	実は、 <u>くずし字の自動認識AIの開発については続きがある。2019年に、CODH、国立情報学研究所（NII）、国文学研究資料館（国文研）がKaggle（カグル）でコンペを開催したのだ。Kaggleとは世界中の機械学習エンジニア、データサイエンティスト約40万人をつなぐプラットフォームで、政府や企業などが提示した課題に対し、最も優れた分析モデルの開発者に賞金を与えるコンペが行われている。全世界的なコンペを通して、くずし字認識手法の開発を加速させようというわけだ。このコンペでは、「くずし字データセット」がコンペ用に改良され提供されたが、AIが学習する「教材」が増えるほど、AIの解析の精度は高まっていく。上記の「みんなで翻刻」によって解読された翻刻データも、AIの教材になることが期待される。</u>	同上