

講義 1 「日本語研究に役立つ情報資源の効果的な使い方」岡田一祐氏へのご質問

話し言葉のコーパスは、どれほどの頻度でデータが更新されるのでしょうか？  
特に流行語などは移り変わりが激しいため気になっております。

(講師からの回答)

国立国語研究所の「日本語話し言葉コーパス」(CSJ)を含め、同所の現代語コーパスで、継続的な更新のされているものはありませんというのがお答えになります。ご関心に沿って補足すれば、「日本語話し言葉コーパス」は、文体としての「話し言葉」ではなく、文字ではなく音声によって実現された言葉という意味で名付けられており、流行語の検出にはそもそも向かないという事情があります。その点では、多種多様な資料を集めた「現代日本語書き言葉均衡コーパス」(BCCWJ)のほうが流行語を調べるという目的には沿いますが、残念ながら構築が完了してデータ追加が行われていないこと、サンプルコーパスであるため、流行語の捕捉率はそれほど高められないという恨みがあります。

その点では、新聞全文データベースや Google Trends、国立国会図書館の Ngram viewer などのほうがご関心には沿ったデータを調べられるのかと思います。

講義 2 「デジタルアーカイブを活用した近現代史研究の実践」河西秀哉氏へのご質問

もしも違いがあるのであれば、国立公文書館デジタルアーカイブの横断検索と、日本外交文書デジタルコレクションや宮内公文書館のデータベースからの検索との違いをご教示いただければ幸いです。

(講師からの回答)

国立公文書館と宮内公文書館のものは基本はかなり似ていると思います。キーワードを入れたり、組織から検索する機能など、近年のデジタルアーカイブは国立公文書館の機能に準拠して作られている場合が多いかと思います。

一方、日本外交文書デジタルコレクションはかなり特殊で、基本的には本として刊行されたものが PDF で見られるようになったと考えてよいかと思います。キーワードの検索機能はなく、年ごと項目ごとに PDF ファイルがアップロードされているという形です。調べたい年がわかっている、調べたい事象がわかっているという時にはこちらの方が使いやすいかもしれません。

知りたい情報に辿り着く際には検索ワードをより適切なものにすることが重要だと思いますが、そのような検索ワードを考える上でのコツなどがありましたらご教示いただきたいです。

(講師からの回答)

なかなか難しい問題ですね。まずはいろんなキーワードを入れてみて、どのくらいの結果が出て来るのか、どんな結果が出てくるのかを、いろいろと知ってみるとというのが重要だと思います。そうすると、こういうキーワードならばあまりにたくさん出てくる、こういうキーワードだと逆にほとんど出てこないなどの感触がだんだんとわかってきます。ある種の熟練技のような部分があります。

もう一つは、論文などではその論文に関する 5 つほどのキーワードが書かれているケースがあります。こうしたものをチェックしていると、検索ワードとして適切なものは何なのか、これまただんだんとわかってくると思います。

日本は予算が少なく、海外に比べるとデジタル化が進んでいないと実感しています。貴重資料をデジタル化するために、予算を要求しても「ほんの少数の研究者しか利用しない。費用対効果がない。」と言われ、どうしたらよいか、ご教示ください。確かに貴重資料は一般受けする資料ではないので、デジタル化しても、閲覧回数は増えるとは思えません。

(講師からの回答)

たしかに貴重資料は一般受けするわけではないところもあるので、閲覧回数が増えないように思うかもしれません。ただ、デジタル化してネット上で閲覧できるようにすると、こちらが想像することができなかった関心からその史料が閲覧されるケースがあります。全国各地、そして海外からも見られるからこそ、そうした可能性が高まるという経験が私の近くにもありました。そして、そのデジタル化された史料を閲覧した人が、周辺史料を探しに、館へ足を運ぶケースもあります。そうすれば、館の利用者数も増えるという効果があると思います。

また、デジタル化は貴重資料の保存という点でも有用です。デジタル化せずに（数は少なくても）研究者などに利用されれば、それだけその史料が痛む可能性が高まります。デジタル化しておけば、そのリスクが減るという問題はあります。予算との兼ね合いは大変難しいのですが、デジタル化は何十年というスパンでの費用対効果を考えると有用であるとも言えるかもしれません。