

# 職業自動コーディング

## 巻頭言

3月下旬、仙台市内で数理社会学会大会が開かれたのを機会に、久しぶりで出席して司会も務めた。その中で印象に残ったのは、高橋和子氏（敬愛大学）による「確信度付き職業・産業コーディング自動化システムの開発と公開」という報告であった。

調査における職業分類は実際にはどう行われてきたか。私も携わってきた社会階層と社会移動全国調査（SSM調査）を例にとってみると、従業上の地位・従業員数・役職名が選択回答形式で、従業先の名称・従業先の事業内容・仕事の内容などが自由回答形式で質問されてきた。これらの回答をもとに（コードブック+総合的判断によって）コーディングが行われ、約200種類に分類される。

高橋氏は、コンピュータを用いてこれを自動的に行おうとしている。「言語」に対する工学的アプローチである自然言語処理研究の成果を取り入れて開発されたシステムで、データにあたるのが上記回答のテキストファイル、辞書にあたるのはコードブックおよびこれまでの調査における回答とそのコーディング結果の膨大な記録である。

高橋氏の永年にわたる研究を、私は高く評価している。学会報告のたびに判別の精度が上がっていくのは楽しみだ。国勢調査などに実際に用いられれば、時間的・経済的コストの節減に大きく貢献することになるだろう。また、コーダー自身、あるいはコーダー間の判断基準のブレという問題もクリアできる。

放送大学宮城学習センター所長 原 純 輔

外国でも研究が進んでいるようであり、多分、基本的な趨勢なのだろう。しかし、学術調査の場合もそれでいいのか、という思いがある。

SSM調査でも、「職業」がこの調査のキー概念であることにもよるのだが、コーディングのコストは膨大である。現職、父職、母職、さらには初職から現在までの職歴を漏らさず尋ねるから、一票あたりの作業量は国勢調査よりずっと多し、苦労も大きい。しかし、作業に携わることで得られる情報量も、小さくないはずだ。従業先・事業内容・仕事内容についての記述、また、その他の選択式項目あるいは調査地点（従業地）・学歴・収入などとの組み合わせによって、回答者の職業活動像が具体的に浮かび上がってくる。さらには、両親の職業、学歴、職歴などを通して、回答者の生活歴についても同様だ。それらは、データ分析の際に引用されるような性質のものではないとしても、その背景として大きな力になるのではないか。

コーダーの判断のブレを最小限に抑えるために、SSM調査では、全メンバーによる一週間ほどの合宿形式で集中的に作業が進められてきた。その中でさまざまな意見交換や新人の教育が行われて、知識や情報の共有と伝達が図られてきた。また、協同作業を通して、その後の分析に向けたエネルギーの醸成に役立ってきたように思われる。このような場を残してほしいと思うのは、あながち懐旧的心情とばかりはいえないだろう。