

NRDF 作成過程におけるコーディングチェックの役割

The Role of Coding-Check in the Making of NRDF

板垣 直之
北海道大学理学部物理

Naoyuki Itagaki
Graduate School of Science Hokkaido University

1 データのチェックの必要性

日本荷電粒子核反応データファイル (NRDF) の作成には、論文の収集、論文のコーディング (データのコード化) とグラフの読み取り、データファイルへの登録の、3 段階の作業が必要である。その中で、この第 2 段階にあたるコーディング作業の後、作成されたコーディングファイルのチェックをもう一つの相補的な過程として注意深く行い、データの信頼性を高める努力を行っている。そもそも NRDF の主旨は、日本国内における荷電粒子を入射した原子核反応 (電子散乱を除く) 全てに対して責任を持ち、IAEA のデータベースである EXFOR にデータを供給することであり、そのチェックはデータベースの質の維持の上から必要不可欠の作業である。

チェックには、1) コーディングファイルの中で辞書に登録されている用語が使われているかのチェック、及び 2) コードの文法的なチェックの 2 つが従来から行われて来たが、昨年度から、加えて 3) コーディングファイルの物理的な内容のチェックも行うことになった。ここでは、その「物理的な内容のチェック」の意義と、それを通じて明らかになった NRDF コーディング作業の現状と課題を振り返ってみる。

2 チェックの内容

「物理的な内容のチェック」とは、コーディングが文法的には正しいが、それ以外の入力の間違いによる誤った物理的情報を発見し訂正することである。それは大きく 2 つに分類される。

1 つは多くの間違いがそうであるように、論文のタイトル名のスペルミスであるとか、物理量の数値や単位の入力ミス、あるいは論文の文章の意味の取り違いに起因するものなどで、これらは単純なミスと呼べるものである。しかし、残念なことにこれらは機械的な処理では発見できないミスであり、1 つ 1 つのコーディングに対する手作業による緻密なチェックが必要である。

もう 1 つは、コーディング者の物理的解釈が含まれる箇所に対しても、チェックの過程でさらに深い議論が必要となることがある。これらの中にはミスとは言えないものも含まれるが、コーディング者の解釈が NRDF の共通の解釈と符合するように修正を行うことによって、データベースとしての統一性を確保することがある。そのためには、まずコーディングされた内容のうち、どの項目がコーディング者の解釈に依存する可能性が高いかを系統的に把握し、それに対するコーディング者共通の方針を確認し、個人差を埋める必要がある。

特にコーディング者によって差が生じるのは反応タイプ (RTY) の項目である。例えば直接反応には DIRECT-RCT のコードが割り当てられているが、NRDF の辞書を有効に利用すれば、さらに細かい直接反応の指定が可能であり、この点に関しては個人差が生じる。ピッ

クアップ反応 (PKUP) やストリップ反応 (STRP) にはそれぞれ対応したコードがあり、特に重陽子については重陽子ストリップ反応 (D-STRP) も辞書登録されている。それらをコーディングする際には、辞書にある限り細かい指定をする態度が望ましいであろう。反面、 (p, n) 反応や (p, π) 反応、 (π, K) 反応など、日常多用する用語に対応したコードが今のところ反応タイプの分類の中には存在しない。そのため、これらは直接反応のコードにコメントを付けて言及するなどの処置が取られることがあるが、この処理を行うかもコーディング者に依存している。さらに、移行反応に関しては核子移行反応 (N-TRNSF) のみが辞書中にコードとして存在し、 α 粒子などを含めた一般の移行反応に対するコードが、反応タイプの分類の中には存在しない。このように適切なコードが辞書に存在しない場合にもっともコーディングの個人差が生じるが、その中で非弾性散乱 (INEL-SCATT) を使って分類するコーディングが散見されるが、これは正しくない。NRDF においては非弾性散乱は「弾性散乱以外の全ての反応」という広い意味ではなく、入射粒子と標的の間でエネルギーの受渡しがおこる散乱という狭い意味に限定して使うべきであり、この点に関してもコーディングチェックの段階で一貫性を持たせるべきである。

次にコーディング者によって個人差が出る項目は、ディテクターシステムを列挙する箇所である。一般に実験系出身者がコーディングを行った際に丁寧な記述得られ、時にはコメントをおり混せて説明される傾向がある。そのため、理論出身者に対して、ディテクターシステムの理解を促進する、説明のような機会を持つことが望まれる。

さらに、はるかに重要な問題として、論文中のどのデータを NRDF のデータとして採用するかがコーディング者に依存するために、チェックの段階での議論が必要になる。たとえば、原子核の構造に関するデータである、細かい γ 線の測定による核の 1 つ 1 つの状態の分析を、「核反応のデータ」とみなして NRDF に採用するかどうかという問題である。NRDF の本来の定義に従えば、荷電粒子を用いた核反応の結果として得られた γ 線はコーディングの対象となる。これらはデータの数が多いことがしばしばあるために、データを収集する側の一致した認識が必要である。また、クーロン励起のような方法によって核の励起状態を生成し、そこからの γ 線を測定するような実験の場合、これを核反応とみなすか、あるいは NRDF のデータとして採用しないかという議論も必要であろう。最近増加している不安定核の実験の中には、不安定核の生成という部分には核反応が関与しているが、論文の主たる関心はそこからの崩壊である場合が多い。このような場合、これを荷電粒子を入射した核反応と見なすことができるかについても、個人の判断にまかせるのみならず、なんらかの基準を設けるべきであろう。

3 コーディングを改善するためにはどうすれば良いか？

以上、コーディング作業のチェックを通じて明らかになった問題点を列挙したが、最後に NRDF がさらに質の高いデータベースとして利用されていくために、必要とされる改善点を述べてみたい。

まず、最近コーディングの仕方でも新しく問題になるのは、 K 粒子ビームや不安定核ビームなどの 2 次ビームを用いた実験の書き表し方である。現在の方法では、生成された 2 次ビームを用いた核反応のデータに主眼がおかれ、2 次ビームの生成方法自体についてはコメントを付けることで説明されている。しかし、これら 2 次ビームを用いた原子核実験は最近の研究の大きな潮流であり、このことを単なるコメントではなしに、何らかの方法で NRDF の検索にかかるように処理されることが望ましい。

次に、従来通りのコーディングの範囲内で期待される改善点をあげる。まず、NRDF の辞書に登録されているコードを見直し、いくつか足りないものを加えて行くことである。物理量としては、例えば「テンソル偏極量」のようなこれから使用頻度の増加が見込まれるものに対応するコードが無い。また、物理量の単位に関するコードが足りない場合が多く、こ

れらは現在コメントを付けることで処理されている。コーディングを行う側からの、積極的な新しいコードの申請が望まれる。

最後に、データの質の向上という面から最も重要な、生データの必要性について述べたい。NRDF のデータ収集が、論文からデジタイザーによって取り出す方法ではなく、実験家から直接生のデータを取り寄せることによって行われるシステム作りが必要である。このことには、データの精度が向上することのみならず、実験の細かい情報を直接確認できるメリットがある。論文の中では、例えばデータが実験室系で書かれているのか重心系で書かれているのか明言されていない場合もあり、この確認はデータベースの信頼上、絶対に必要な点である。