

(言語学者による) 容認度評定の認証システムを試作する構想

入念に設計された日本語文の容認度評価データベースに基づいて

黒田 航¹ 阿部 慶賀² 横野 光³ 田川 拓海⁴ 小林 雄一郎⁵ 金丸 敏幸⁶ 土屋 智行⁷ 浅尾 仁彦⁸

どうして？

言語学の文献は、信頼性の異なる証拠の玉石混淆で、“玉”と“石”を区別したい

何がしたいの？

個々の言語学者の容認度判定の信頼性を定量評価したい！

どうやって信頼性を測るの？

個々の言語学者 x_1, x_2, \dots, x_n の容
認度評定が、非専門家の判断/反応
 $R = \{r_1, r_2, \dots, r_m\}$ (= 集合知) と
一致する度合い $D = \{d_1, d_2, \dots,$
 $d_n\}$ を測って、それを個別に評価
する。

実例サンプリング

実態把握のため、言語学者が出版物中で提示した例文と評定値の実例を、書籍毎にサンプリングして、データベース化する作業が進行中。

現時点でランダムに選定した専門書 6/123 中にある4658例の容認度判定の結果をデータ化

想定

0. 容認度のような性質は集合知としてのみ記述可能 (言語知識の社会性を考えれば当然).

1. 評定者 x_j が事例 s_i 与えた容認度 a_{ji} と r_i との差 Δa , 予想した分散クラスの差と実際のクラスの差 ΔC が小さい程, x_j の判定は集合知との一致度が高く, その分だけ信頼できる.

2. x_j の Δa と ΔC 食い違いの総計が一定の範囲内にない場合、その判断は信頼性が低い。

どうすれば D を測れるの？

1. 理論的バイアスのなるべく少ない日本語の刺激文の集合 $E = \{e_1, e_2, \dots, e_N\}$ を用意 (N は千から万の桁).

2. E を非専門家 $R = \{r_1, r_2, \dots, r_m\}$ に独立に
 評価してもらい, ExR の反応行列を得る (m
 は千の桁).

3. E の部分集合 s_1, s_2, \dots を評定者 (e.g., 言語学者) x_j が評定する。この際、 s_i に容認度 a_{ij} を与える他に、その分散のクラス (大, 中, 小) を予想させる。

4. 差分の合計の量によって評定をクラスに分類する.

この手順を実装し、言語研究者の容認度評価の
認証システムを開発する

それなりの規模の ExR データの構築

i. D の測定に必要な ExR の評価データは、より一般的な価値をもつ、標準化データであり、それ自体で構築する価値がある。

ii. D の構築は社会調査の一つである事が望ましい (が、本気でそれを実施するには予算がおそらく足りない…).

お知らせ & 宣伝

刺激文集合 E の選定と構築でプロジェクトに参加していない方々 (特に言語学者や心理学者) を対象に希望調査を行う予定です。Web調査を考えています。関与を希望する方は第一著者に連絡を下さい。

[illegible]

本研究は一部、日本学術振興会から科学研究費助成（課題番号16K13223）を受けて実施されている。