Prediction of CDI Words & Sentences values from Words & Gestures

Trevor K. M. Day

February 18, 2023

1 Introduction

The Baby Connectome Project (BCP) collected MacArthur-Bates Communicative Development Inventorys (CDIs) beyond the typical normed range of 8–18 months. This means it is unclear how to interpret scores beyond this range.

2 Methods

I acquired CDI Words & Sentences (WS) data from Wordbank (Feb. 15. 2023), comprising $n = 7905$ unique individuals with an age field between 16–30 months. The Wordbank data were divided into a training set (75% of the data at each age) and a testing set (the remaining 25%).

All Words & Gestures (WG) items appear on WS, with the exception of "in" and "inside," which is two items on the WG form and one on WS. For the WG score, an endorsement of the "inside/in" item was scored as two *words produced*.

Within each category, I regressed the difference against age (in months since 18 months, the youngest age that WS is normed for) and the calculated WG score, and the interaction between them.

For each participant's inventory score (e.g. *Sound Effects and Animal Sounds*, *Animals*, ...), I calculated the total number of items endorsed, and the total number of items endorsed, counting only those that also appear on the WG form, see Figure 1. The lowest R^2 of the quadratic regression of WS on WG was *Helping Verbs* at 0.694, but the rest exceeded 0.890.

Figure 1: True WS scores (y) against score using only WG items (x). R^2 values: action_words: 0.993; animals: 0.996; body_parts: 0.985; clothing: 0.980; connecting_words: N/A; descriptive_words: 0.989; food_drink: 0.973; furniture_rooms: 0.992; games_routines: 0.966; helping_verbs: 0.694; household: 0.993; locations: 0.911; outside: 0.973; people: 0.974; places: 0.942; pronouns: 0.947; quantifiers: 0.951; question_words: 0.99; sounds: 1.00; time_words: 0.976; toys: 0.899; vehicles: 0.966.

An additional key problem is that no *Connecting Words* items appear on the WG form. Secondly, *Sound Effects and Animal Sounds* is identical between the two forms. These two categories are ignored for the total-score analyses to be more conservative in estimates of error rate.

3 Results

The Pearson correlation between the true WS score and the calculated WG score for each individual/category was $r = 0.962$ (Spearman $r = 0.968$). Mean differences by category across ages are shown in **??**. Notably large differences emerge in *Food and Drinks*, *Action Words*, and *Descriptive Words*, whereas statistically no differences emerge in *Sound Effects and Animal Sounds* and *Question Words*. Despite the lack of *Connecting Words*, statistically significant differences don't seem to emerge until 23 months, and even then never exceed a mean underestimate of more than 2 words.

3.1 Total Score

Within each category, I regressed the ground-truth WS score on age, WG score, the age–score interaction, as well as age^2, WG^2 and the interaction of the quadratic terms. The quadratic terms were included because of the clear nonlinearity of the age effect in development, as well as the nonlinear effect of WG score on true WS score (Figure 1).

Figure 2 shows the error rate among the test sample. As can be seen in Figure 2a, most participants are estimated to within 50 words of their true valeu. Furthermore, the percent error is small, with the exception of participants with a very low ground-truth WS score Figure 2b. Using 50 words as a cutoff (part of the Delay 3+ criterion), percent error rates are typically within 15% (Figure 2c).

(a) Error in words between WS and ŴS.

(b) Percent error for all test participants, based on true WS score.

(c) Percent error for all test participants, who had more than 5 words.

Figure 2: Results of the modeling indicating total predicted WS score across all categories (except *Connecting Words*).