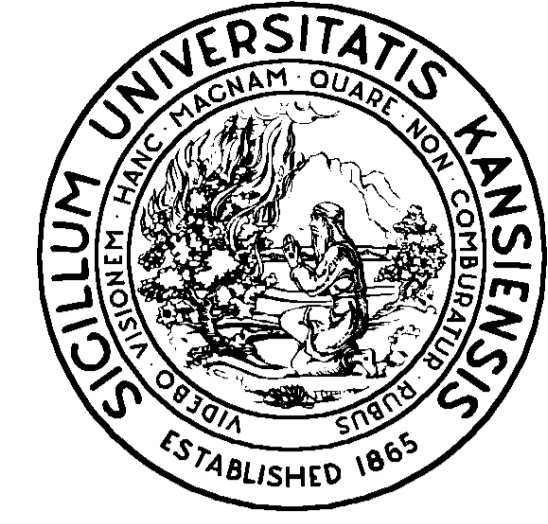# MLU-based age recentering: a tool for studying developmental trajectories

Filip Smolík (`smolik@ku.edu`)
Child Language Doctoral Program, University of Kansas

Summary

The study presents alternative ways to conceptualize age in longitudinal studies. The proposed procedures use a predefined MLU value as a reference point for setting up the age scale, therefore reducing the variability in MLU that is due to the differences in onset timing of language development. The method is tested on two large corpora, and is shown to successfully reduce unwanted variance in the relationship between MLU and age. The potential usefulness for this method as a substitute or complement of MLU-matching procedures is discussed, as well as implications the study has for evaluation of MLU validity as a developmental index.

Background

Developmental studies of language usually use chronological age as the time scale of development. However, raw chronological age may not be the most appropriate way study developmental trajectories. If the interest is the unfolding of a certain characteristics over time, *time from onset* may often be more appropriate measure.

- there is substantial variability in the absolute, chronological timing language development (this was one of motivations for the use of MLU in Brown, 1973)
- using chronological age to compare children is essentially a cross-sectional strategy
- as such, it may overestimate the error variability because it does not take into account interindividual differences in absolute timing
- overestimating error variability in a measure leads to underestimating the validity of such measure
- the shape of individual growth curves can be similar in different individuals but the growth process may be shifted in time

Objectives

Longitudinal data make possible to adjust the time variable so that it may better reflect the growth process. This study adopts this approach to:

- examine the commonalities in language development when the onset-related variability is removed/reduced
- explore how much variability in cross-sectional data is due to the differences in developmental timing
- test if limited reliability of MLU reported in literature may be due to the cross-sectional nature of available data
- explore if MLU-matching is supported by the observed properties of MLU growth

Data and Method

Two longitudinal corpora of spontaneous language transcripts, available from CHILDES (MacWhinney, 2000), were used (Manchester and Wells corpus, see Data section).

Calculation of MLU-referenced age

The adjusted age variable was calculated separately in each corpus:

- time 0 for each child was set to the time point where this child's observed MLU-value was closest to the reference MLU
- the reference value was the lowest MLU available for all children in each corpus (i. e. maximum of individual minimal MLU values)
- the time unit remains unchanged but the time scale is reset so that for each child, it is as close to the onset of language development as possible with the available data
- days were used as the age/time unit throughout the study

In Manchester corpus, the reference MLU was 1.63; the median chronological age of achieving this value was 741 days (min. 622, max. 896). The MLU reference value in Wells corpus was 1.77, achieved at median age 801 days (min. 561, max. 1174).
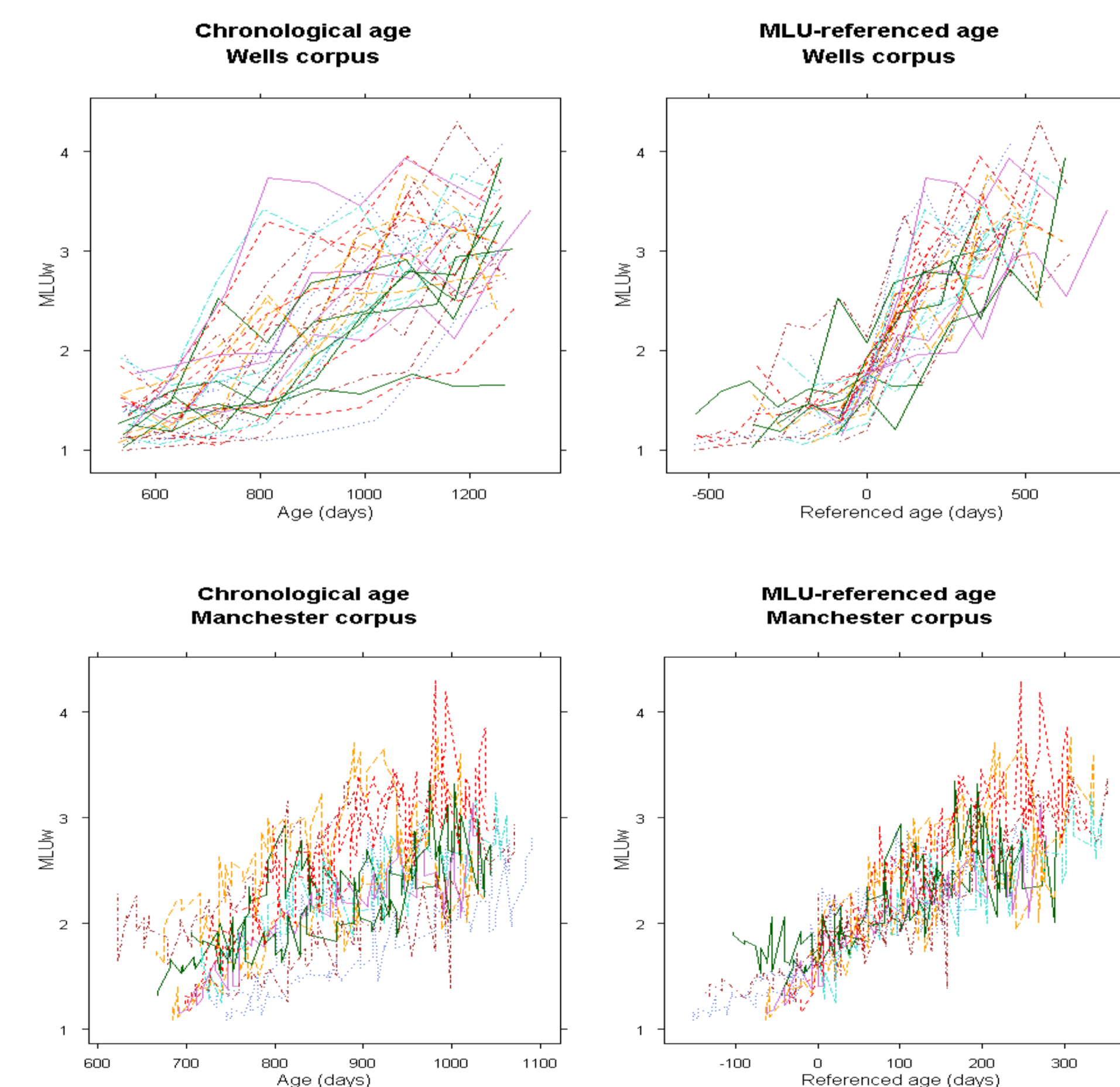

Figure 1: Comparison of chronological and MLU-referenced time scales.

Data

- Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001) – a total 798 transcripts from 12 children aged 20 – 36 mo. used for this study
- Wells corpus (Wells, 1981) – total of 279 transcripts from 32 children were used here; age range 17 – 43 mo.
- MLU *in words* was calculated from each transcript using custom-made Perl routines; MLU in words used to avoid problems with defining morpheme productivity; however, correlations of MLUw with MLUm in the samples were over 0.99 most of the time

Analytic method

- mixed-effects growth curve modeling
- best linear models predicting MLU from age were sought for each corpus
- modeling procedure performed twice for each corpus, using chronological and MLU-referenced age

Results

It is apparent that using the referenced time scale results in less variability at a given time point (see Fig. 1)

- variability is reduced not only on the reference point (referenced time 0) but along the whole time axis
- this suggests that individual developmental trajectories are more similar than can be observed in cross-sectional research
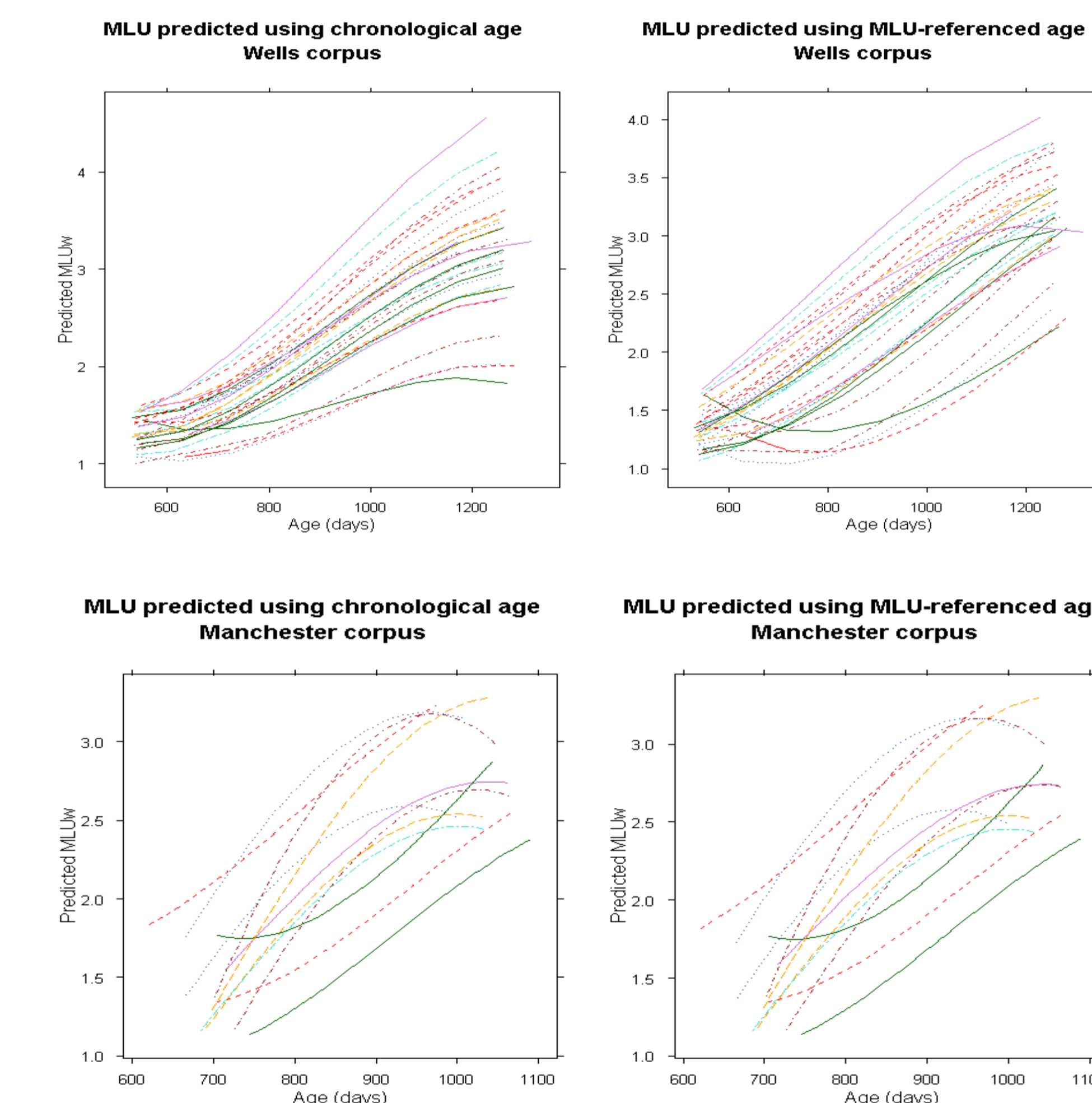

Figure 2: Individual developmental curves as predicted by models using chronological and referenced age.

Modeling results

Growth curve modeling revealed somewhat different best models for Manchester and Wells data.

- in both corpora, there were significant linear, quadratic and cubic fixed effects
- significant random intercept and linear random effect present in both corpora; significant quadratic random effect only present in Manchester corpus
- all models were fit with heteroscedastic error variance increasing with increases in predicted MLU

Modeling results contd.

- using chronological age in model fitting led to the same best models as the use of MLU-referenced age; yet the t−statistics for most effects was lower in the chronological age models
- final models using MLU-referenced age fit better than those using chronological age ($BIC = 0.2$ vs. $BIC = 29.2$ in Manchester corpus; $BIC = 272.1$ vs. $BIC = 314.1$ in Wells corpus)

Models for Manchester data predict more variability in individual growth curves than models for Wells corpus (see Fig. 2).

- Manchester data were collected over a shorter period of time (ca. 1 year); the modeling procedure highlights the individual differences in growth curves
- Wells corpus data show that when observed during a longer period (ca. 2 years), MLU growth shows more similarities across children.
- larger amount of data available in Manchester corpus allows more precise estimation of individual-specific growth components (random effects): reflected by the low values of within-person residual MLU variance ($\sigma = 0.07$)
- this indicates that when MLU individual differences in MLU development are accounted for (by means of random effects), precise estimation of MLU is possible

For Manchester corpus, the predictions from referenced age are very close to those from chronological age. In Wells corpus, the model with chronological age predicts more divergence in MLU growth curves with increasing age. Since practically all children sooner or later arrive at similar ceiling MLU values (cf. Leadholm & Miller, 1992), the predictions from referenced age model appear closer to the true developmental process.

Conclusions

Time from onset, as approximated by the MLU-referenced age variable, reveals structure of MLU growth curves better than chronological age.

Models using referenced age show substantially better fit to the data, and in the Wells corpus, the models using chronological age produced clearly inappropriate predictions. Matching of populations on the basis of MLU gains some support from these findings because it suggests that after the differences in the onset time are reduced in the data, individuals show more similarities in their developmental curves.

References

Brown, R. (1973). *A First Language: The Early Stages.* Cambridge, Mass.: Harvard University Press.

Leadholm, B. J., & Miller, J. F. (1992). *Language sample analysis: The wisconsin guide.* Madison, WI: Wisconsin Department of Public Instruction.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Mahwah, NJ: Lawrence Erlbaum.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language,*, 28, 127–152.

Wells, C. G. (1981). *Learning through interaction: The study of language development.* Cambridge, UK: Cambridge University Press.