Longitudinal L2 Development of the English Article in Individual Learners

Akira Murakami, Theodora Alexopoulou
{am933, ta259}@cam.ac.uk
Department of Theoretical and Applied Linguistics
University of Cambridge

1. Introduction

- Since the early days of second language acquisition (SLA) research, a large number of studies have investigated the learning curve of second language (L2) development.
- However, a number of issues remain unaddressed. Most notably, previous studies tend to be small-scaled and/or only exploit cross-sectional data.
- The lack of longitudinal information makes it impossible to study the relation between individual learning curves and the averaged cross-sectional pattern.
- This has been an important limitation in SLA research in view of evidence from psychology that the averaged pattern can conceal individual trajectories (e.g., Heathcote, Brown, & Mewhort, 2000).
- However, the absence of sufficient amounts of longitudinal data for individuals is a practical obstacle.
- We exploit a recently developed large-scale longitudinal learner corpus and track individual learners.
- We focus on the accuracy in the use of the English article by L2 learners.

2. Method

2.1 Corpus

- EF-Cambridge Open Language Database (EFCAMDAT; Geertzen et al., 2014)
- Writings submitted to *Englishtown*, the online school of EF Education First
- 16 teaching levels of 8 teaching units each, covering A1-C2 in the Common European Framework Reference (CEFR) levels
- Error corrections by teachers were used to calculate accuracy.
- EFCAMDAT is publicly available at http://corpus.mml.cam.ac.uk/efcamdat/.

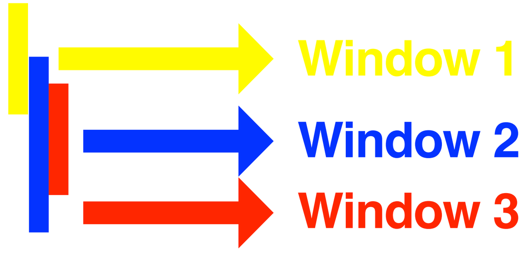
2.2 Target linguistic feature, L1 groups, and accuracy measure

- Linguistic feature: both definite (*the*) and indefinite (*a*, *an*) articles
- We used the country of residence of learners as the closest approximation to their native language (L1).
- We targeted 10 L1 groups and classified them into two L1 types, PRESENT and ABSENT, depending on whether or not they have an article.
  - The PRESENT group included L1 Brazilian, German, French, Italian, and Spanish.
  - The ABSENT group included L1 Chinese, Japanese, Korean, Russian, and Turkish.
- 140,000 writings consisting of 10 million words
- Accuracy measure: Target-like use (TLU) score = $\frac{\text{\# of correct suppliance}}{\text{\# of obligatory contexts + \# of overgeneralization errors}}$

2.2 Tracking development through moving windows

- Individual writings are too short to provide enough obligatory contexts to reliably calculate the TLU score.
- We thus computed TLU scores over multiple writings (= window).
- Each window included at least 10 obligatory contexts (OCs).

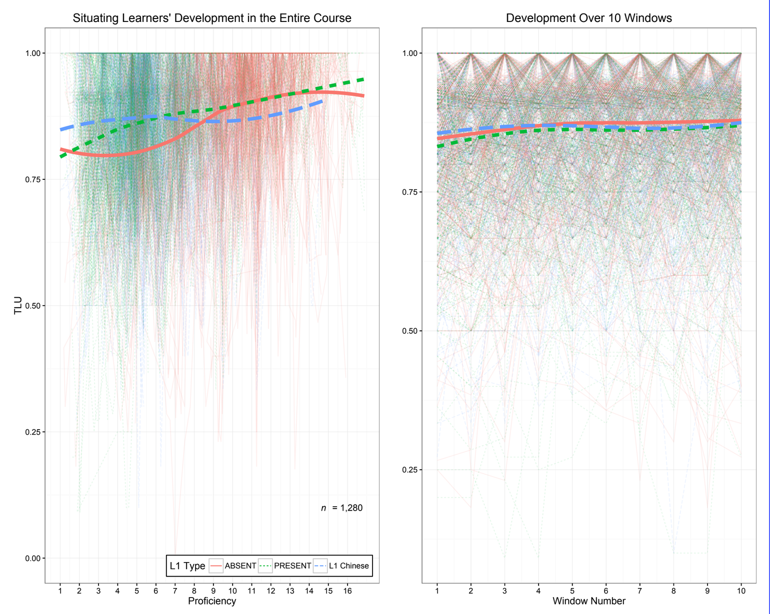
Writing 1: 6 OCs
Writing 2: 6 OCs
Writing 3: 3 OCs
Writing 4: 7 OCs
Writing 5: 2 OCs



- 70,879 TLU scores (windows) in total by 20,394 learners. Out of them, 1,280 (6.8%) had 10 or more windows.

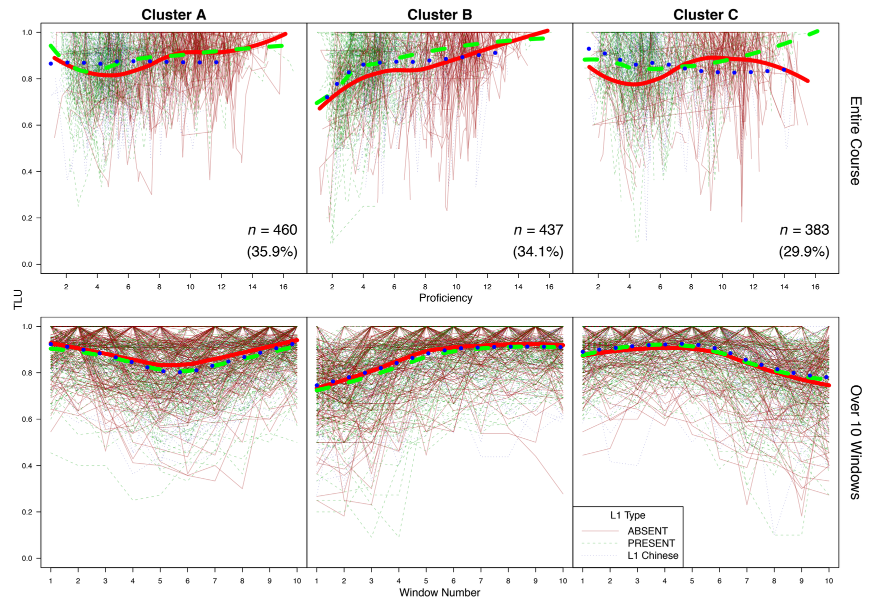
3. Longitudinal View of Development

- The figure shows the development of 1,280 learners
- Thin line: Development of an individual learner
- Thick line: Trend line (LOESS)
- Left panel: Development across Englishtown levels
- Right panel: Development over 10 windows
- Overall, there is little change in the mean accuracy across longitudinal development.
- Large individual variation is present in the developmental pattern.
- There can be a difference between the averaged longitudinal pattern and the learning curve of individual learners.



4. Clustering Learners According to Their Developmental Shapes

- Identified clusters of learning curves in a bottom-up manner through k-means clustering.
- Input data: Learner-mean-centered first 10 TLU scores of each learner
  - Neutralised between-learner differences in absolute accuracy
- Decided on $k = 3$ after inspecting how the shapes of emerging clusters change between $k = 2$ and $k = 10$.
- The figure shows the clusters of article development.
- Upper panel: Learners are situated in the entire Englishtown course
- Lower panel: Development over 10 windows
- Cluster A: U-shaped development
- Cluster B: Accuracy increases until Window 7 and then levels off
- Cluster C: Accuracy decreases from Window 5 onwards
- These three patterns differ from the aggregated pattern.



5. Cluster Validation

- We empirically tested whether the clusters capture true underlying learning curves or random noise.
- Logic: We compute the null distribution of the metric that measures goodness of clustering based on random data, and if the value of the metric in the observed data falls outside of its 95% range, we consider it as the evidence that the observed clusters are too good to be derived from random data and conclude that our clusters indeed reflect the difference in the learning trajectory.
- The metric: the mean silhouette value (Rousseeuw, 1987) whose value is large if learners within each cluster have similar developmental trajectories and those in different clusters have different trajectories.
- Null distribution:
  1. Randomised writings within individual learners.
  2. Computed the mean silhouette value in 1.
  3. Repeated the above 1,000 times and inspected the distribution of the mean silhouette values.
- The upper bound of the 95% range of the distribution in 3 (0.144) was below the observed mean silhouette value (0.151), so we consider the clusters we observed to be non-random.

6. Discussion

- We identified systematic learning trajectories in the accuracy of the L2 English article.
- There can be differences between average longitudinal development of a group of learners and the individual learning curves that constitute the group.
- This suggests that we cannot necessarily infer the development of individual learners based on the longitudinal data aggregated over multiple learners.
- The obvious question now is why learners show these different developmental patterns.
- Admittedly, we can offer no insight at this point to this question.
- Two potential sources of variation; (i) changes in the internal knowledge of learners and (ii) individual variation.
- Though the interpretation of such results is very challenging, we believe that our findings show the need for investigating individual patterns in a more comprehensive way.

References

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2014). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In R. T. Millar et al. (Eds.), *Selected proceedings of the 2012 Second Language Research Forum. Building bridges between disciplines* (pp. 240–254). Cascadilla Proceedings Project.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185–207.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.