

# A bootstrap-based reanalysis of Zamma (2013)

Shigeto Kawahara

Keio University

The Institute of Cultural and Linguistic Studies

Hideki Zamma passed away on March 22nd, 2016, a week after he was involved in a car accident, at the age of 46. In his recent book “Patterns and Categories in English Suffixation and Stress Placement: A Theoretical and Quantitative Study” (Zamma, 2013), he identified four classes of English suffixes in terms of the (i) root attachment behaviors and (ii) stress patterns, instead of the more traditional “Class 1 vs. Class 2” distinction (Siegel, 1974). He showed that these four types of suffixes are not evenly distributed in the English lexicon, and their distributions are affected by whether the suffixes are light or heavy (taking into consideration the “extrametricality” and other effects). On page 131 of the book, he summarizes the distribution of these suffixes in the English lexicon.

Table 1: The distributions of the four types of suffixes identified in Zamma (2013)

	Class 1	Class 2	Class 3	Class 4
L	24	16	18	0
H	37	3	18	3

This is the core descriptive finding of his book. He went on to argue that with the theory of unranked constraints developed by Anttila (2002), we can predict the distributions in Table 1. Given his assumptions about the constraint set and which constraints are unranked with respect to each other, he argued that the predicted distributions should be as follows (p. 133):

Table 2: The predicted distributions

	Class 1	Class 2	Class 3	Class 4
L	20.7	10.8	19.6	6.9
H	28.4	5.9	19.2	7.5

He used a Chi-square test to compare Tables 1 and 2 and to examine the goodness of fit between them. This approach has a few problems: (i) we would have to rely on null results, (ii) we cannot figure out which classes fit the prediction and which classes do not, and (iii) we do not know the properties of the distribution underlying this sample. I suggested to him in the summer of 2015 that he should instead run a bootstrap resampling test.

Now he has left us, before implementing this bootstrap analysis together, so I would like to do it here as a belated tribute to him. Based on the data in Table 1, I created 50,000 new samples using resampling with replacement (Efron & Tibshirani, 1993). We can calculate 95% percentiles over these samples to get 95% confidence intervals. I used R to implement this bootstrap method (the code is available upon request).

Table 3: The 95% confidence intervals calculated by the bootstrap method

	Class 1	Class 2	Class 3	Class 4
L	17-31	10-23	11-25	0
H	29-44	0-7	11-25	0-7

Class 4 in the L category does not exist in the English lexicon, so let us leave that aside. The question is whether each value in Table 2 fits within the 95% confidence intervals calculated in Table 3. And they generally do. The H-suffixes for Class 1 and Class 4 are a bit off, but the rest fit in the predicted ranges, and even those that do not fit are very close.

I am not an expert in this area, but am glad to find that his results are as good as this. I would like to dedicate this result to his memory.

## References

- Anttila, Arto (2002) Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory* **20**(1): 1–42.
- Efron, Bradley & R. J. Tibshirani (1993) *An Introduction to Bootstrapping*. Boca Raton: Chapman and Hall/CRC.
- Siegel, Dorothy (1974) *Topics in English Morphology*. Doctoral dissertation, MIT.
- Zamma, Hideki (2013) *Patterns and Categories in English Suffixation and Stress Placement: A Theoretical and Quantitative Study*. Tokyo: Kaitakusha.