

**Feeding, breeding, extinction, and river length:
Problems with treating words like species,
Comments on Meara and Olmos Alcoy (2010)**

John P. Racine
Dokkyo University
Japan

In their paper, Meara and Olmos Alcoy (2010) attempted to find a means of estimating productive second language (L2) vocabulary size based on the premise that many known lexical items simply do not appear in learner-produced texts. To do so, they borrowed an ecological model, in which a capture-recapture formula, the Petersen estimate (Petersen, 1896), is used to estimate the number of animals existing in a given environment. As the authors are probably the only vocabulary acquisition researchers searching for applicable models in the 19th century swimming habits of plaice in the German Sea, they should indeed be commended for their original approach. There are, however, a number of aspects to this approach that should be reconsidered before this model can accurately be applied to vocabulary.

The authors listed a number of assumptions that must be made before Petersen's formula can be used to estimate population size. Two of these deserve further examination (Meara & Olmos Alcoy, 2010, p. 226):

1. The population needs to remain constant. The fish must have an equal chance of being caught at Time 1 and at Time 2.
2. The means of collection must be reasonable. That is, we must have a trap that catches the fish we want to count, and the area in which we catch the fish must be somewhat representative of the river as a whole.

In the words of the authors: "If these assumptions do not hold, then the model will not work" (p. 226). Unfortunately, in applying the model to productive vocabulary, the authors seem to have ignored their own stipulation. This is the case in both the design of the experiment and in the interpretation of the results.

In terms of experimental design, participants were asked to perform an identical task in both writing sessions. This is surely in violation of the first assumption. For fish to have an equal chance of being caught on separate occasions, nothing can have affected the population in the meantime (e.g., disease or poaching). More importantly, nothing about the trap can have affected the likelihood that the same fish will be re-caught. By assigning the same task at Time 2, the researchers have essentially fed the fish, increasing the likelihood that they will return to the net at Time 2. They have primed the words in the first narrative and have increased greatly the likelihood that they will be used again. Perhaps any writing elicited under identical experimental

conditions to a previous task will show some effects of priming, but these effects would certainly have been minimized if the cartoon about a dog, an umbrella, and the sea had been replaced by, for example, a photo of an otter, a fish, and a linguist.

The same assumption—that the population must remain constant—is violated by the use of intermediate-level participants. There is a growing body of research into interlanguage in general and into L2 lexical acquisition in particular (e.g., Bell, 2009; Fitzpatrick, 2007) that points to the instability of representations in the emerging L2 mental lexicon. Just as some fish join the school while others become extinct, new L2 vocabulary items come in and out of productive use in the minds of learners. Others are forgotten shortly after being learned, existing in the lexicon for only a brief period, never to be fully consolidated in memory. The researchers are somewhat baffled by the increased production of the intermediate group at Time 2, but I am suggesting that lexical activation and output may fluctuate wildly until a relatively large and stable L2 lexicon is formed. It is interesting that the researchers did not frame this result in these terms, as this is one of the basic findings of Meara's own recent attempts to simulate vocabulary network activation using computer models.

On the other hand, if future research shows that this effect is stable (that intermediate learners consistently produce greater token and type output at Time 2), I would suggest that the experimental methodology is again altering the population in violation of the assumption above. By its very nature, the task may be inducing learning. Lexis that might otherwise be unstably represented in the learner lexicon may be further consolidated by the writing task itself. Simply recalling and organizing the words into sentences may strengthen connections in the network and increase the likelihood that new words will also be incorporated. In other words, the trap may be causing the fish to breed. This effect may be subject to a kind of ceiling effect in that the task does not affect learners to the same degree after a certain level of proficiency has been attained. This would account for the fact that advanced learners showed no change in production at Time 2.

Finally, there are a number of areas in which our fishing analogy becomes a little bit fishy. That is, although the capture-recapture model may still prove to be somewhat useful in estimating productive vocabulary size, there are at least three areas in which the analogy does not fit well. First of all, one of the essential calculations involved in extrapolating from the Petersen formula to an estimate of an entire population has been excluded in the current application. That is, if we apply our net to a 10-mile stretch of river, assuming that this stretch is not unduly different from the rest of the river, we can simply multiply the product of the formula (e.g., by 10 for a 100-mile long river) to get an estimate of the river's entire population. The researchers, however, have made no attempt to estimate the multiplier for their findings. Indeed, how does one estimate an entire productive vocabulary from the lexis produced in writing the dog and umbrella story? How long is the River Lexicon? The authors concede: "the elicitation instrument needs to be aware of the size of the productive vocabulary that we think our participants have at their disposal" (p. 231). Of course, if we were already aware of the vocabulary size, we would not need to estimate it.

Second, the authors have chosen not to tackle (pun intended) this final calculation necessary for extrapolating their "ridiculously low" (p. 231), estimates to reasonable estimations of the

learners' complete productive vocabularies. Instead, they suggest that their findings may reflect the "relative sizes" of the participants' vocabularies, and that there may be a "fairly straightforward relationship between each participant's Petersen estimate and their actual productive vocabulary size" (p. 233). I suspect that this relationship is far from straightforward. Productive vocabulary, by its very nature, violates the second assumption I've listed above. Unlike a river that remains relatively uniform from its source to its delta, productive vocabulary knowledge consists of areas of dense knowledge and other areas of relatively little knowledge. Depending on the specifics of the elicitation test, any attempt to extrapolate to the entire productive lexicon may prove inaccurate.

Finally, another area in which the capture-recapture analogy breaks down concerns the impartial nature of word knowledge. Fish are caught whole, but vocabulary may not be. Lexis for which the meaning is known, but the spelling is not, may never show up on tests of production. This kind of productive vocabulary test makes no distinction between depths of individual word knowledge.

Although I've indicated some premises of the current study that should be given further examination, as the authors suggest, a modified application of Petersen's estimate may still yield a bountiful catch in terms of successfully estimating vocabulary size.

References

- Bell, H. (2009). The messy little details: a longitudinal case study of the emerging lexicon. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners* (pp. 111–127). Bristol, UK: Multilingual Matters.
- Fitzpatrick, T. (2007). Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics*, 17, 319–331.
- Meara, P. M., & Olmos Alcoy, J. C. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, 22(1), 222–236.
- Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Linsfiord from the German Sea. *Rep. Dan. Biol. Stn*, 6, 5–84.

About the Author

John Racine is a lecturer in the Interdepartmental English Language Program at Dokkyo University, Japan. His research interests include L2 vocabulary acquisition and word association. Email: racine@dokkyo.ac.jp