

Peer Commentaries

Cite this article: Meara P (2022). Modelling L2 vocabulary acquisition: The devil is in the detail. *Bilingualism: Language and Cognition* 25, 226–227. <https://doi.org/10.1017/S1366728921000407>

Received: 14 July 2021
Accepted: 15 July 2021
First published online: 22 November 2021

Address for correspondence:

Paul Meara p.m.meara@gmail.com

Anyone who has ever worked seriously with models will tell you that this type of research is not nearly as easy as it looks. You have to work very hard to avoid falling into the trap of a plausible metaphor which seems to explain the phenomena you are interested in, but in reality just describes them in terms of other phenomena. The way to avoid this trap is to turn your model into a computer program, run it as a set of simulations and discover exactly what kind of behaviours it exhibits. Work of this type usually involves a long period of reflection during which you make your assumptions absolutely explicit, and test out how the parameters you have built into your simulations interact with each other. It is surprising how often this process leads you to completely re-evaluate the way you think about the problem you are working on.

I did try to implement this approach with the Ontogenesis Model described in Bordag, Gor and Opitz's paper. Greatly simplifying things (always a good idea, in my experience), the Ontogenesis Model consists of only three basic *dimensions*, a linguistic dimension, a mapping dimension and a network dimension, so it should be an easy matter to write a program that illustrates how it works. However, on closer examination, the three dimensions are described in a somewhat loose kind of way, and this makes it very difficult to predict what the model actually does.

By way of illustration, let us consider the network dimension. The paper assumes that words in an L2 vocabulary are “connected”. Some words have few connections to other words, while some words have more connections. The connections can vary in strength. The paper also assumes that there is an optimal number of connections that a word should have. At first sight, this seems like a plausible set of assumptions, and easy to simulate. A simple approach would be to suppose that each newly learned word starts off with one connection (or no connections?) and gradually increases the number of connections it makes with other words in the vocabulary until the optimum level of connectivity is achieved.

At this point, we run into two problems. The first problem is that some of the key ideas that characterise the network dimension - *connection*, *optimum* and *activation* - are just not well-defined. What exactly is a *connection*? Are there different types of connection? Are connections fixed or variable? What does a connection do? How is a connection made? When it comes to *activation*, the text is vague. We don't know what activation consists of, where it comes from, or how it works. And we don't know how it affects any individual word. Similarly vague is the idea of an *optimum* level of connectivity for individual words. The text is unclear as to what properties this optimum has. Is it the same for all words? Probably not. Is it a fixed characteristic of a word. Maybe. Is the optimum context dependent? What drives a word towards its optimum level? What changes when a word reaches its optimum? Uncertainties of this sort have to be resolved before you even think about coding the model.

The second problem with the network dimension is that it is unclear whether the Ontogenesis Model sees connectivity as a property of individual words or as a property of the network as a whole. Figure 3 implies the former, but the text implies the latter. It's easy to see how characteristics of a whole network might be cast as a dimension, and it's easy to see how some simple properties might be used to map the way a network grows and its complexity develops. Networks are typically described in terms of their *DIAMETER*, for example, and changes in a network's diameter can easily be mapped onto a single dimension. It's much less obvious how this approach can be used to describe the development of individual words in a lexicon. Counting the number of connections individual words have, as the paper does in figure 3, doesn't give us a figure that can be mapped onto a single dimension. At best, what we get is a distribution of connectivity scores which needs to be interpreted in the light of the overall size of the vocabulary. Different interpretations of these features give us models with very different types of behaviours.

Finally, it's worth pointing out that the paper assumes that the network of connections linking words in an L2 lexicon is basically co-terminous with the networks that we can construct from word association data. Again, this is not an obvious assumption to make: word association networks are usually based on word association norms, a compilation of responses collected from large groups of test-takers. But we know that any individual's association network

can differ quite markedly from these norms, and it's not clear how the Ontogenesis Model would handle this. Going further, and pushing the network argument harder, it's not actually obvious that word association behaviour is a fundamental characteristic of a vocabulary network: it could just be an emergent behaviour that is the result of some other, more fundamental characteristic.

The caveats that I have raised here in respect of the network dimension of the Ontogenesis Model also apply to the linguistic dimension and the mapping dimension of the model. In both

cases, the vaguely defined terms, and the uncritical approach to the model's fundamental assumptions are significant problems – the mapping dimension is particularly complex and difficult to characterise concisely.

It seems to me that the main strength of the Ontogenesis Model is that it highlights some important issues in the way we currently think about L2 lexicons. It doesn't provide any solutions for these issues, but it can perhaps help us to identify some crucial areas where we need to make our thinking much less fuzzy and much more precise.