Mitigating catastrophic interference in models of bilingual lexical acquisition

James S. Magnuson[1,2,3], Marian Simarro González[1], Thomas Hannagan[3], Nils Beck[4], & Manuel Carreiras[1,2,5]

[1]BCBL: Basque Center on Cognition Brain and Language; [2]Ikerbasque: Basque Foundation for Science; [3]Universityof Connecticut; [4] University of Stuttgart; [5] University of the Basque Country (UPV/EHU)

INTRODUCTION

- Catastrophic interference (CI)[1] is the (partial or complete) overwriting of previously learned information when learning new information in a neural network
- Complementary Learning Systems[2,3] theory proposes a division of labor between neocortical and hippocampal pathways that could provide a solution
  - Long-term memories are stored in cortex
  - New learning relies heavily on the hippocampus, and is influenced by prior learning without disrupting it
  - Gradual consolidation allows new learning to integrate with long-term memory
- CLS has not been formally applied to late second language acquisition (to our knowledge)
- This project simulated learning of L1 (English) and later acquisition of L2 (French) using a simplified approximation of CLS

METHODS

- **Models**
  - Feedforward networks trained to activate the phonological form of a word given phonological features or semantic features as input (phonological 'autoencoding' is meant to emulate generating heard patterns)
  - Two hidden layers (*Hidden$_c$* and *Hidden$_h$*, meant to emulate cortical and hippocampal pathways respectively), each with 30 nodes, allow exploration of *weight reservation* (see below)
- **Phonetic representations**
  - 16 slots, each with 28 features (so each slot = 1 phoneme)slot-based, each phoneme represented by 28 features
  - Allows words up to 16 phonemes long (16 · 28 = 448 total inputs)
  - Inputs are aligned to random initial positions on each training trial to prevent learning of position-specific representations
- **Semantic representations**
  - 300 features based on WordNet, plus 'language' and grammatical gender (303 total)
- **Lexicon:** ~300 words
- **Training**
  - Within a language, model is first trained on phonological inputs and then semantic inputs (with phono and sem training alternating)
  - 1 epoch = 1 pass through all words at random alignments (never using test alignments)
- **Testing**: 1 token of each word at an alignment never used in training; a trial is scored as accurate if the output vector is closer to the target vector than any other word vector
- **Weight reservation as an approximation of CLS**
  - Base model uses both hidden layers all the time (Sim 1)
  - Weight reservation models divide the hidden layer into a pseudo-cortical pathway (Hidden$_c$) and a pseudo-hippocampal pathway (Hidden$_h$), with all H pathways reserved (Sim 2) or only subsets (Sims 3-5)
  - Then we combine the best reservation model with interleaving to emulate consolidation (Sim 6)
  - Aims: determine whether weight reservation is feasible, and which pathways are critical

HYPOTHESES

- Weight reservation should significantly mitigate catastrophic interference (without strong predictions for which architecture will be best)
- An additional CLS assumption of interleaving (which could happen in-the-moment or during sleep-based consolidation) should further mitigate catastrophic interference

REFERENCES

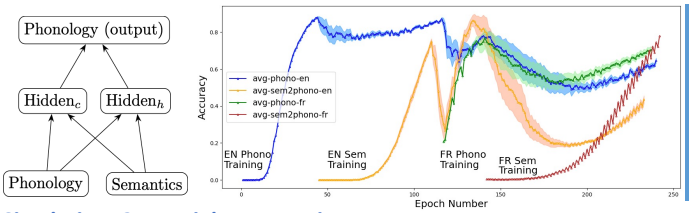
[1] McCloskey, Michael; Cohen, Neal J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychology of Learning and Motivation. Vol. 24. pp. 109–165. doi:10.1016/S0079-7421(08)60536-8. ISBN 978-0-12-543324-2.

[2] McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological review, 102(3), 419.

[3] Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends in Cognitive Sciences, 20(7), 512-534. https://doi.org/10.1016/j.tics.2016.05.004

SIMULATIONS

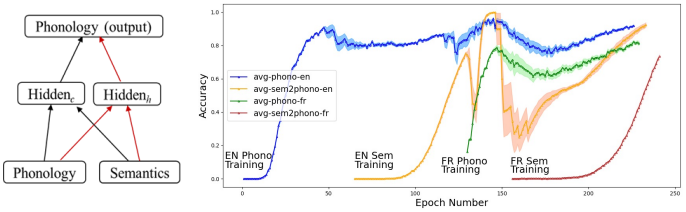
Simulation 1: Base model. *2 hidden layers function like a single hidden layer, since all connections operate feedforward and all connections are trained for both L1 and L2.*



- Each new mapping caused substantial interference, with the FR sem-phono mapping causing massive interference for EN sem-phono.
- Implausible for L2 to wipe out L1 in humans

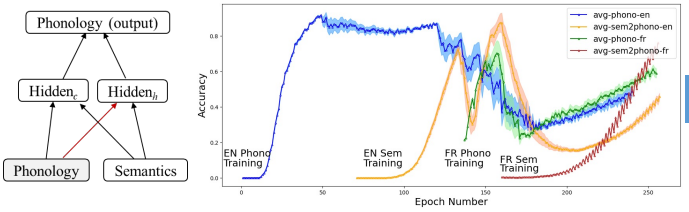
Simulations 2-5: weight reservation. *Only pathway(s) drawn in red were trained during FR training. We start with 'maximal' reservation and then assess which pathways are critical by reserving one at a time.*

Simulation 2: Maximal reservation



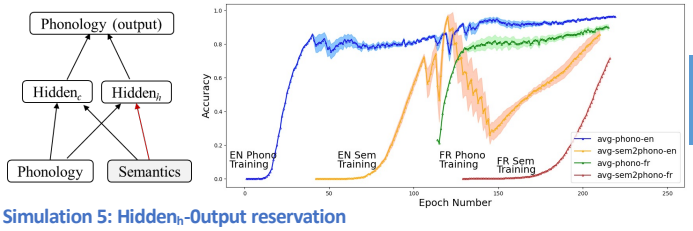
- Substantial improvement vs. base model

Simulation 3: Phono(input)-Hidden$_h$ reservation



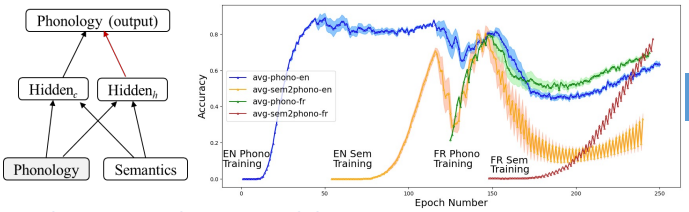
- Greater catastrophic interference

Simulation 4: Semantic-Hidden$_h$ reservation



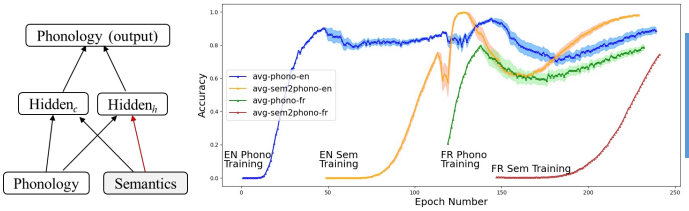
- Substantial improvement vs. base model, even better than 'maximal' reservation

Simulation 5: Hidden$_h$-Output reservation



- Minor improvement vs. base model

Simulation 6: Emulating consolidation. *A formal CLS approach would explicitly model (sleep-based) consolidation. Here, we emulate this by simple interleaving of the 2 languages. Rapid interleaving (alternating epochs; not shown) avoids catastrophic interference altogether, but is implausible (e.g., for immersion in L2). When weight reservation (Semantics-Hidden$_h$) is combined with interleaving, slightly larger 'runs' of each language (here, 2 epochs each) are possible. A formal CLS approach is likely needed.*



- "Consolidation" plus weight reservation robustly mitigates catastrophic interference

CONCLUSIONS

- Our approximation of CLS significantly mitigated catastrophic interference in late L2 lexical acquisition
  - Which weights are reserved matters: weak mitigation from *Phono-Hidden$_h$* or *Hidden$_h$-Output*; *Semantic-Hidden$_h$* is where weight reservation provides the most mitigation (better than 'maximal reservation')
  - Weight reservation provides a minimal separation that, when combined with interleaving, robustly mitigates catastrophic interference
- Next steps: extend to a formal CLS model for understanding late second language acquisition