

# A modular toolkit for semantic model implementation and research

Johnathan E. Avery, Indiana University  
Michael N. Jones, Indiana University



## Purpose

We present a generalized, modular toolkit that implements often overlooked models of semantic memory. The toolkit fosters an exploration of the continuous parameter space between existing models of semantic memory

- System implements:
  - LSA<sup>1</sup>, HAL<sup>2</sup>, GloVe<sup>3</sup>, COALS<sup>4</sup>, & BEAGLE<sup>5</sup>

## Motivation

Many streamlined software packages exist

- GenSim<sup>6</sup> Implements CBOW, skip-gram, LSA, LDA, and td-idf transformations
- TensorFlow<sup>7</sup>: Implements CBOW, skip-gram

They are well equipped to execute a **specific set** of **non-generalized** models of semantic memory

To compare model predictions, must use numerous software suites

There is **continuity** between many models of semantic memory that is **underexplored**.

### For Instance:

*HAL vs LSA*

If we change the matrix structure of LSA to  $W \times W$ , we have HAL with a window=length of document

*HAL vs GloVe*

GloVe is HAL plus gradient descent

*LSA vs Random Vector Accumulation*

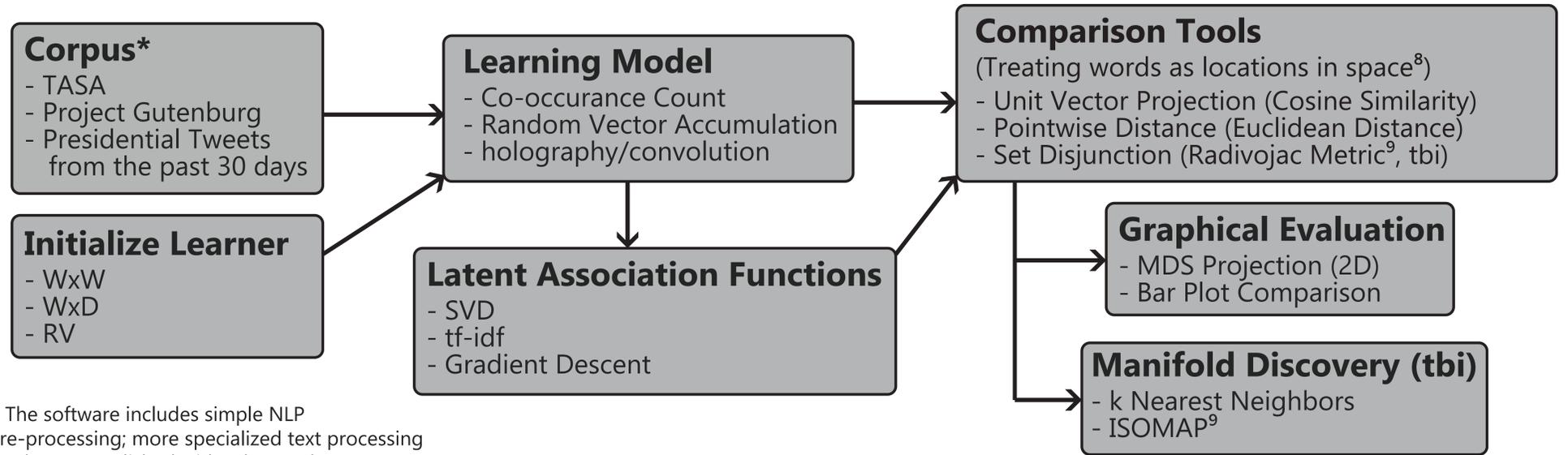
RVA is the dot product of  $W \times W$  matrix and set of random vectors

### Abridged References:

<sup>1</sup>Landauer et al (1997); <sup>2</sup>Burgess et al (1997a); <sup>3</sup>Pennington et al (2014); <sup>4</sup>Rohde et al (2004); <sup>5</sup>Jones et al (2007); <sup>6</sup>Rehurek et al (2010); <sup>7</sup>Martin et al (2015); <sup>8</sup>Osgood et al (1957); <sup>9</sup>Yang et al (2018); <sup>10</sup>Tenenbaum et al (2000); <sup>11</sup>Jamieson, Avery, et al (2018)

<https://github.com/masterccoli/Phonomaxia>

## Software Flow Diagram

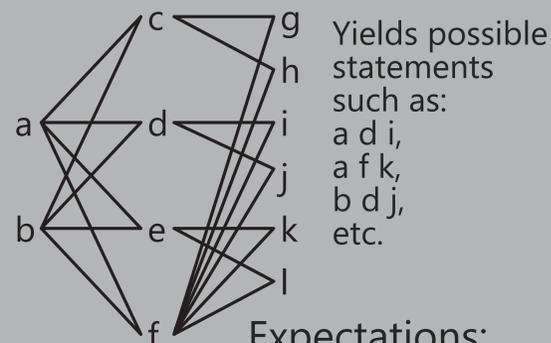


\* The software includes simple NLP pre-processing; more specialized text processing can be accomplished with other packages

## Implementation Demonstration

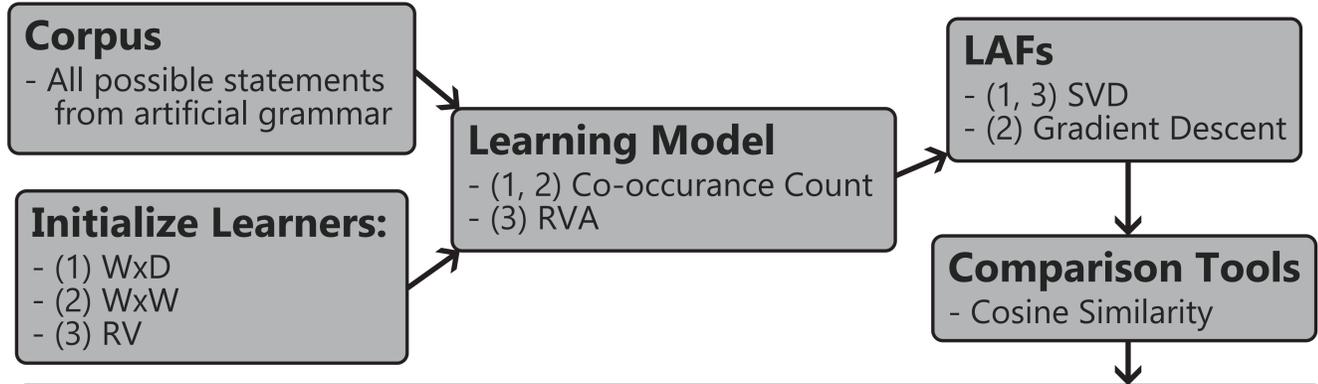
- (1) LSA
- (2) GloVe
- (3) Random New Model (RNM, RVA->SVD)

### Artificial Grammar<sup>10</sup>:

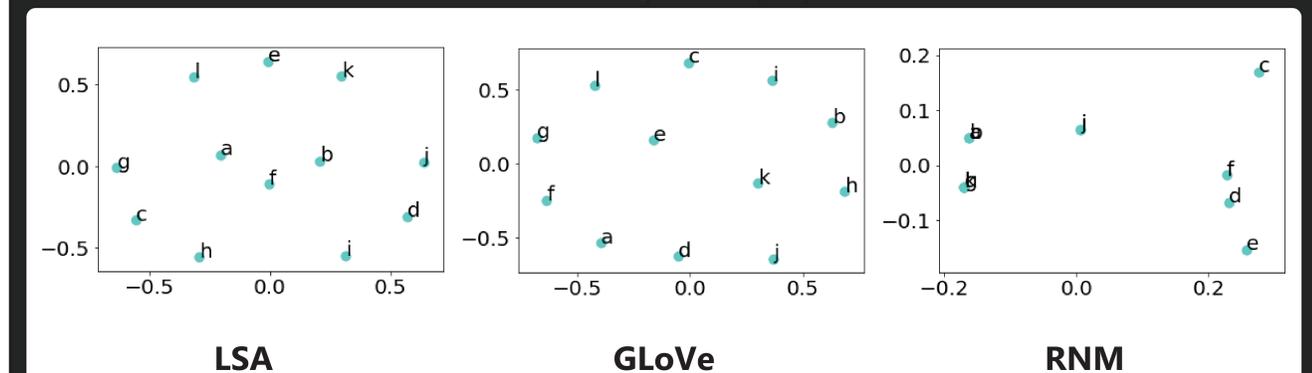


### Expectations:

a & b are similar  
f is similar to c, d, & e  
c, d, & e are not very similar etc.



## Graphical Evaluation



<https://github.com/masterccoli/Phonomaxia>