Models of bilingualism

Why model?

- Because we are scientists.
- Because it forces theories to be more precisely specified.
- Because it allows us to explore complex interactions between components.

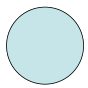
Some caveats…

- Understand *why* the model behaves the way it does.
- Beware hidden assumptions.
- Simplification doesn't always *simplify*.

Connectionist models

- Inspired by neural processing
- May be implemented at different levels of abstraction
- Components:
  - Units (activation)
  - Connections (weights)
- Change:
  - Activation update function
  - Weight update function
- Characterizing models
  - State (Activation and Weights matrices)
  - Task (Input vectors, maybe Target vectors
  - Dynamics (Updating functions)

Units



- Activation can be discrete or continuous
- Activation is typically *squashed* (0, 1) or (-1,1).
- Some have other characteristics like *phase, resting activation, bias*.
- Update can be *feedforward* or *recurrent*.

Connections

- Weights encode the patterns in the input (learning).
- Weights can be unidirectional, bidirectional, or symmetric
- Weight update may be incremental or one-pass.
- Learning can be supervised, unsupervised or reinforcement learning.

Representation

- Localist representations: each thing being represented is assigned a single unit.
- Distributed representations: each thing being represented involved more than one unit (may be features, random pattern, thermometer).
- **Tradeoff*:*** ease of understanding vs. generalization power

- Models can use a mixture of localist and distributed representations.
- Representations can be local or distributed to different degrees.
- Localist networks can learn, and distributed networks can be hardwired, but in general if you are going to hand-wire them, it's easier if they are localist to some degree.

The localist/adult-state models

- Bilingual Interactive Activation (BIA)
- Semantic, Orthographic, and Phonological Interactive Activation (SOPHIA, BIA+)
- Bilingual Interactive Model of Lexical Access (BIMOLA)

BIAM

- Extended IA model (word processing)
- Layers for: **Features** (graphical), **Letters**, **Words**, and **Language Nodes**
  - Features - Letters: forward activation, top-down inhibition
  - Letters - Words: bidirectional activation, top-down inhibition
  - Words-Words: lateral inhibition
  - Words-Language: forward activation, cross-lang top-down inhibition
- Assumptions:
  - Integrated lexicon
  - Nonselective access

BIA

- Language nodes can be used as bias/switch:
  - Language tags
  - Global lexical activation
  - Language filter
  - Contextual pre-activation
- Word frequencies -> unit resting activations
- Proficiency can be modeled as low frequency.

BIA models:

- Neighborhood effects
  - Neighborhood density
  - Shifting neighborhood
- Masked priming effects
  - Effect of proficiency
- Interlingual homographs effects
- Language of previous items effects

Limitations of BIA

- There are no phonological or semantic representations
- Representation of interlinguial homographs and cognates is underspecified
- Representational and functional aspects of language nodes are confounded
- Limited account of how non-ling and linguistic contexts affect word recognition
- No detailed description of task -- word identification vs. lexical decision

BIA+

- Adds a task schema system and a semantic system
- Makes a distinction between word identification and task/decision.
- Task schema includes:
  - Specific processing steps for task at hand
  - Continuous input from identification system
  - Decision criteria to determine when a response is made based on relevant codes

SOPHIA

- The implemented version of BIA+ (no surprise!)
- Adds phonology and semantics
- Layers for
  - Letters & Phonemes
  - (Clusters, Syllables, Words) x Orthography & Phonology
  - Semantics, and Language Nodes
- Accounts for more data than BIA

BIMOLA

- Based on IA model TRACE (auditory word recognition)
- Layers for **Features**, **Phonemes**, and **Words**
- Assumption: differences in modality yields differences in architecture

BIMOLA

- Differences with BIA:
  - Phoneme and word levels are organized by language
  - Separate lexicons - only within language competition at the word level
  - No language nodes, emergent top-down Language effects from inhibition, not activation

The distributed/developmental models

- Bilingual Single Network (BSN)
- Bilingual Simple Recurrent Network (BSRN)
- Self-Organizing Model of Bilingual Processing (SOMBIP)

BSN

- Extended word recognition/reading distributed model
- Cares about emergence of representations and development
- Layers for:
  - Orthographic input (features)
  - Semantic output (features)
  - Language context
- Task: mapping word orthography to word meaning.

BSN models

- Emergence of language separation and shared semantic space.
- Effects of balanced vs. unbalanced bilinguals
  - Language representation (well delineated or not)
  - Interlingual homographs

BSN limitations

- No way to implement lists or task demands explicitly
- Unrealistic training input
- Training on both languages is intermixed, otherwise catastrophic forgetting

BSRN

- Simple Recurrent Network (Elman Net)
- Layers:
  - Input - Language Alpha/Beta x Nouns/Verbs
  - Output - Language Alpha/Beta x Nouns/Verbs
  - Hidden, Context
- Task: given simple sentences, predict next word

BSRN models

- Emergent separation of two lexicons based on:
  - Word order
  - Word-word co-occurrences
  - Phonology
- Predicts that switching must occur with less than .1 probability for languages to be segregated.

SOMBIP

- Self-organizing maps + Hebbian learning
- Unsupervised learning by competition, similarity-based clusters
- Layers:
  - Phonological Map (English/Chinese)
  - Semantic Map (English/Chinese)
- Realistic input - training set from bilingual child corpus
- Word meaning represented as word co-occurrence (LSA)
- Task: Learn associations between phonological and semantic maps

SOMBIP models

- Emergence of differentiated semantic and phonological representations
- Distinct patterns of bilingual lexicon without use of language nodes or language tags
- Priming and interference effects
- Effect of proficiency and working memory capacity on the representation of the lexicon

SOMBIP limitations

- No orthography
- Two distinct semantic reps
- No language tags BUT
  - Phonology for Chinese but not English has "tonality" vector
  - Word meaning representation forces distinct lexicons AND semantic reps

Conclusions

- Modeling is good!
- Localist/adult-state models account for more data BUT distributed/developmental models have more potential

- Aren't these models just mapping input/output and therefore pointless?
- How biologically plausible are these models?
- Do they model the tasks realistically?
- Do they model input/learning environment realistically?
- Why can't we just run developmental models to get and end-state grown-up model?
- One lexicon or two lexicons, enough already! (but what about semantic stores?)
- Tags bug. Do we need them?
- How do these models do the switch?