Using SOMs to Gain Insight into Human Language Processing

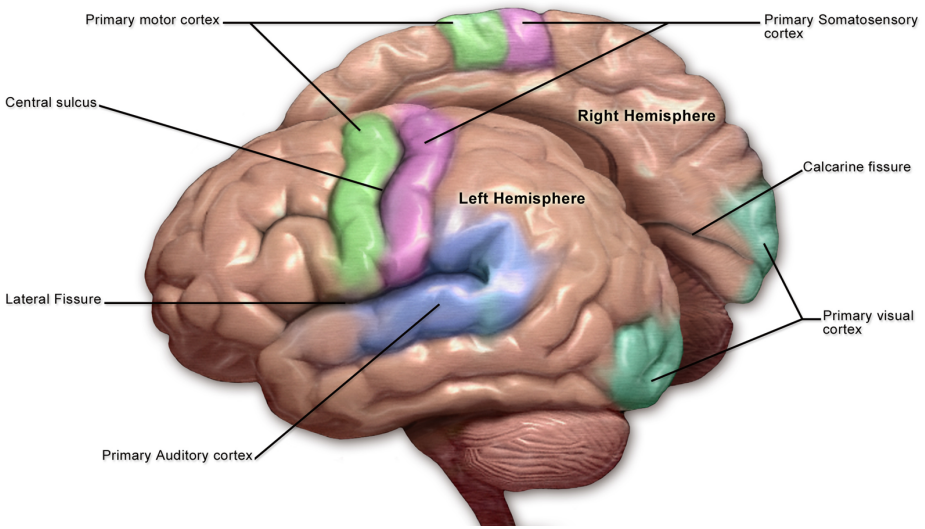
Risto Miikkulainen

The University of Texas at Austin

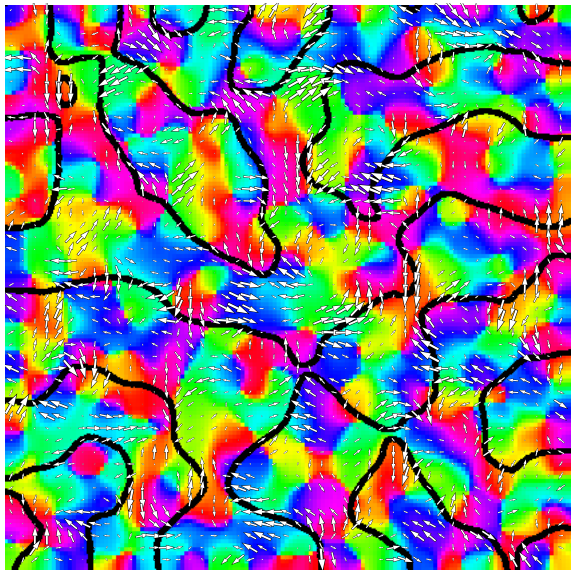
risto@cs.utexas.edu

Joint work with Uli Grasemann, Swathi Kiran, Marty Mayberry, Chaleece Sandberg
Supported by NIH R21-DC009446, NIH R01-MH066228,
IARPA FA8650-14-C-7357

Why Use SOMs?



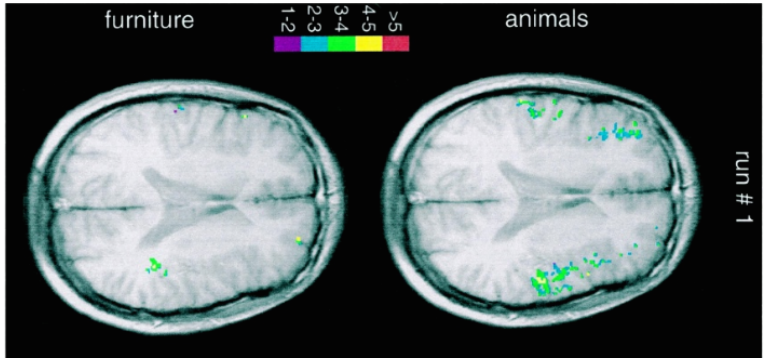
(Wikipedia, 2016)



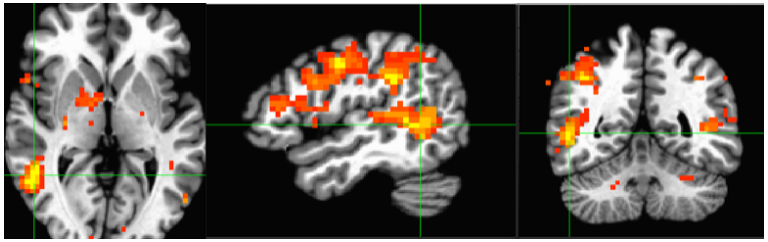
(Miikkulainen, Bednar, Choe & Sirosh 2005)

- SOMs commonly used in engineering applications
  - Data visualization
  - Preprocessing for categorication
- Original motivation was brain maps
  - Visual, auditory, somatosensory maps
  - Self-organization of perceptual pathways
- Can be used as a model of perceptual maps
  - E.g. LISSOM model of the visual cortex

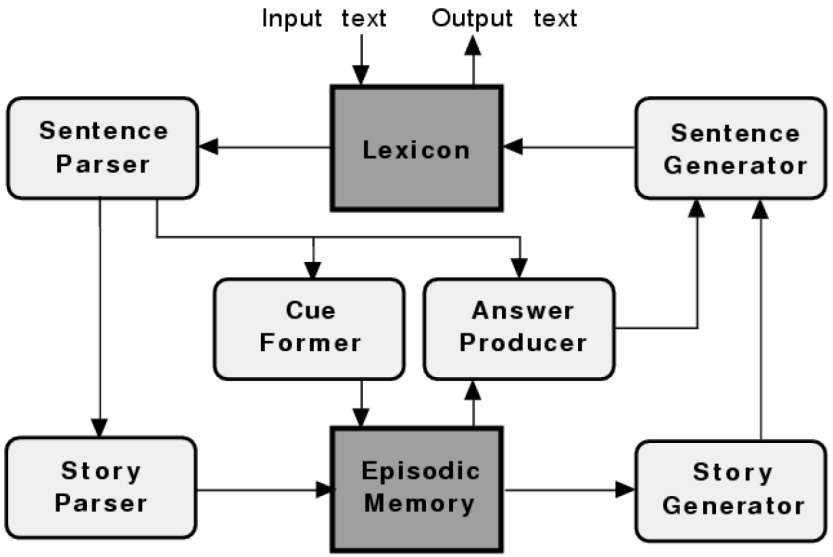
Why Use SOMs for Language Processing?



"furniture", "animal"(Spitzer et al. 1998)



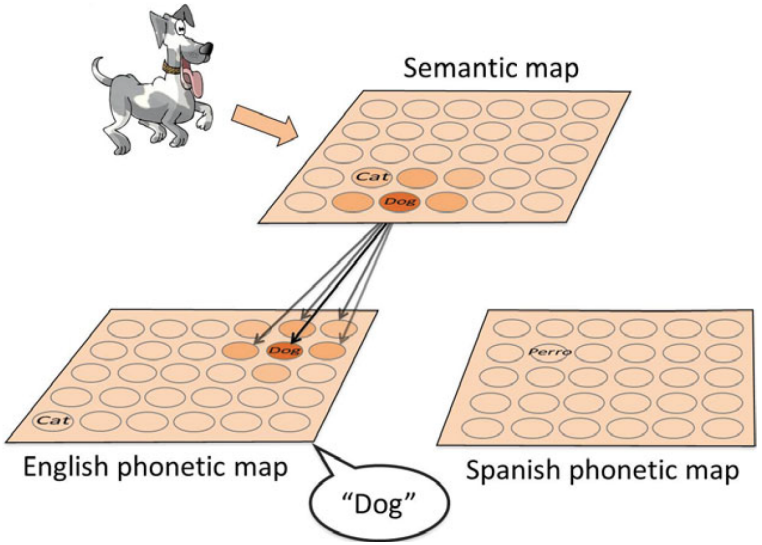
"face" (Binder et al. 2014)



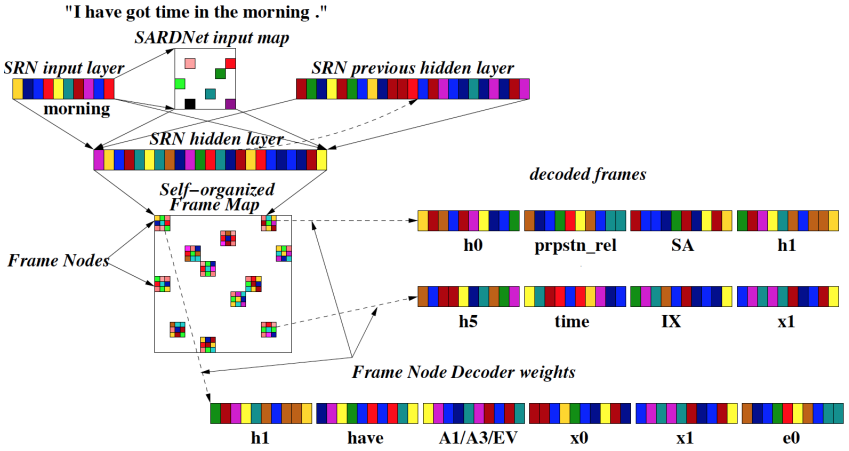
(Miikkulainen 1993)

- Maybe there are maps at a higher level as well?
  - E.g. fMRI of word semantics
  - fMRI extends to sentence semantics?
  - Computationally: Episodic memory as well?

Focus of this talk



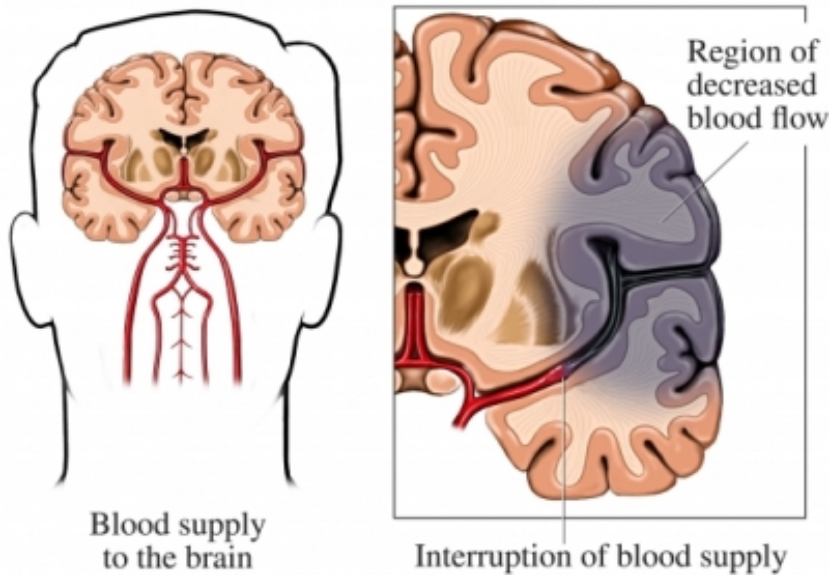
(Kiran, Grasemann, Sandberg, & Miikkulainen, 2013)



(Mayberry & Miikkulainen 2003)

- Models motivated by recent fMRI:
- Bilingual lexicon
  - Impairment and rehabilitation of individuals
- Semantic parsing
  - Graded sentence semantics

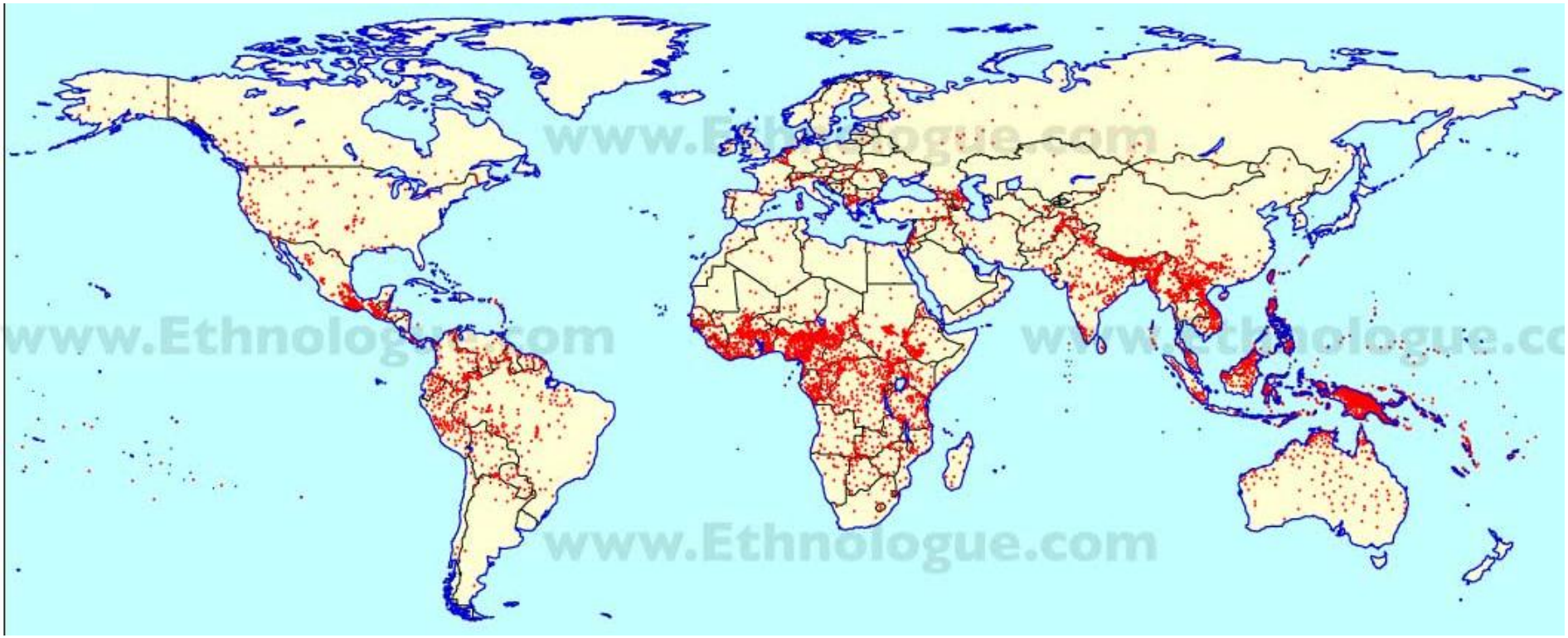
Why Model Individual Subjects?



Goals for Modeling

- Understand development, organization, performance
  - General models are informative
- Understand breakdown, rehabilitation
  - Modeling individual performance is crucial

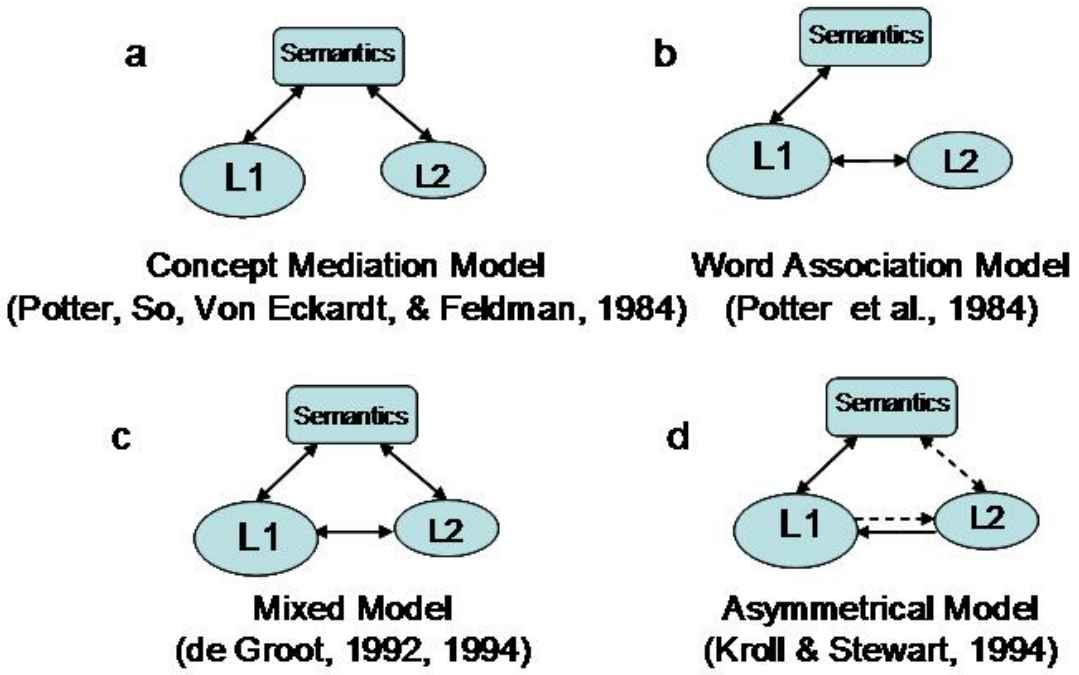
Why Model the Bilingual Lexicon?



Each dot represents geographic center of a language

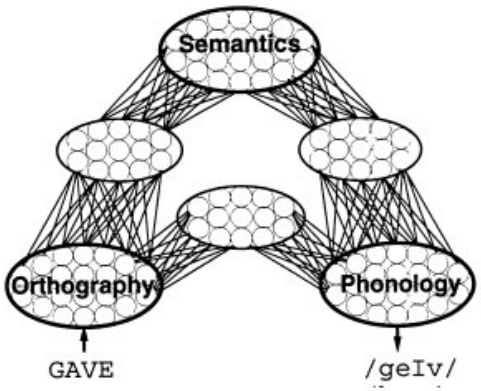
- Most people are bilingual/multilingual
  - More than 50% worldwide; 44% in CA, 35% in TX
  - However, bilingualism is not well understood
- Current treatment methods monolingual
  - Could potentially be improved

Bilingual Lexical Processing

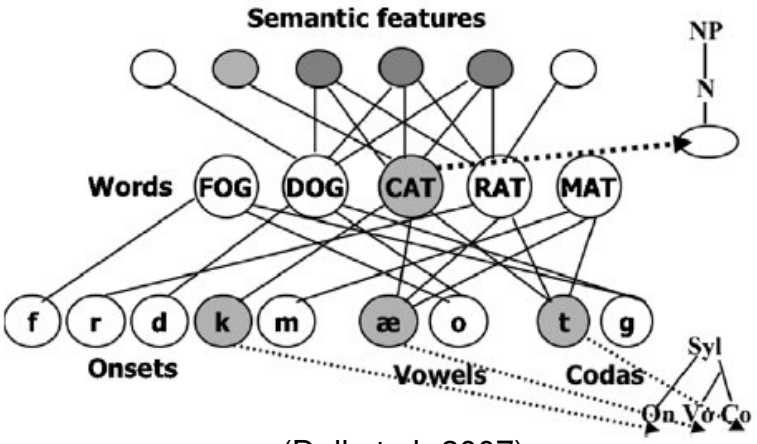


- A shared semantic system
- Separate lexica for the different languages
  - Both languages activated at once?
- Proficiency, age-of-acquisition (AoA) important

Computational Models of the Lexicon



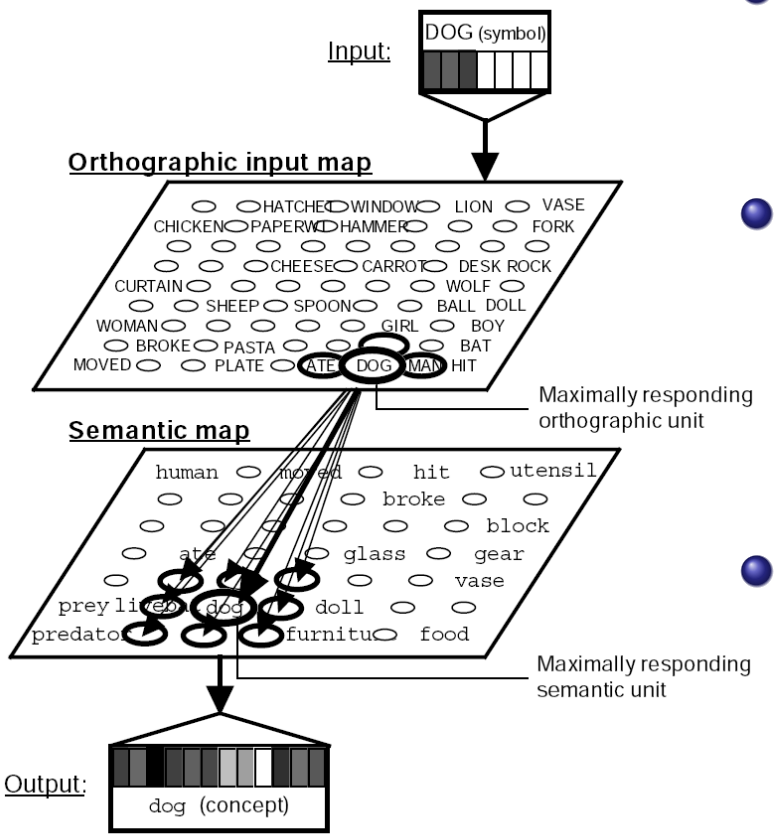
(Seidenberg & McClelland 1989, Plaut 1999)



(Dell et al. 2007)

- 20 years of ANN models
  - Different approaches: backprop, attractors, deep learning...
  - Capture avg. human performance (big data!)
  - Detached from physical organization
- Physical models can do more
  - Can be trained even with little data
  - Can model impairments and rehab

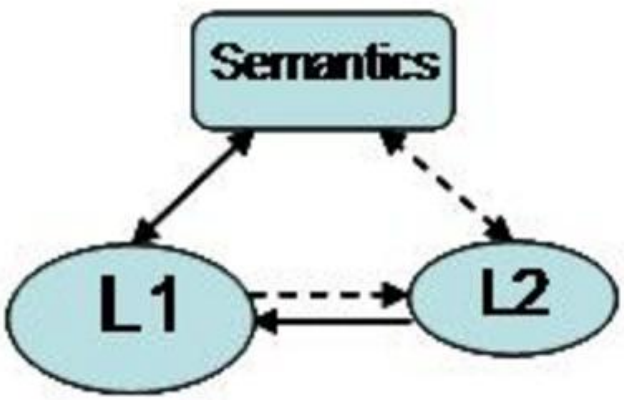
The DISLEX Model



(Miikkulainen 1997; Li et al. 2005)

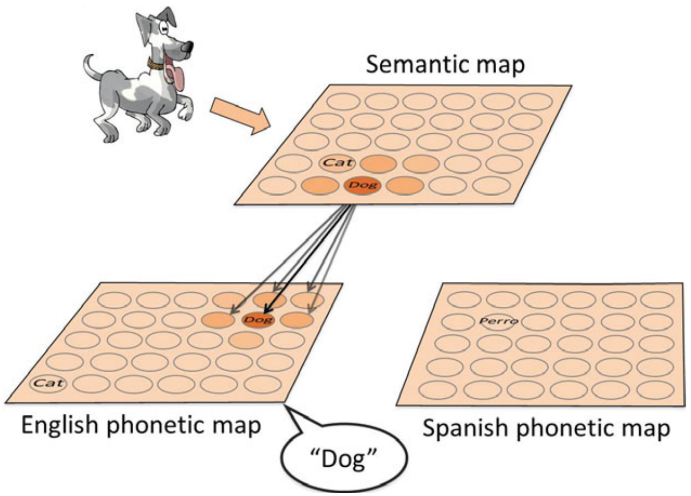
- Each component is a SOM
  - Semantic, orthographic, phonological; different languages
- Connected by associative connections
  - Production, comprehension, and transfer
  - Normalized Hebbian adaptation after propagation
- Dyslexic and aphasic impairments
  - Noisy mapping, propagation; Damage
  - Ball->doll, lion->tiger, sympathy->orchestra
  - Category-specific impairments

Bilingual DISLEX



- Extended to bilingual lexicon
  - Phonological L1 and L2
  - Different exposure to each language
  - Different AoA
- How does it develop?
  - Not a concern here (see DEVLEX; Li 1999-)
- How damaged and rehabilitated?
  - The main focus here

Experiments

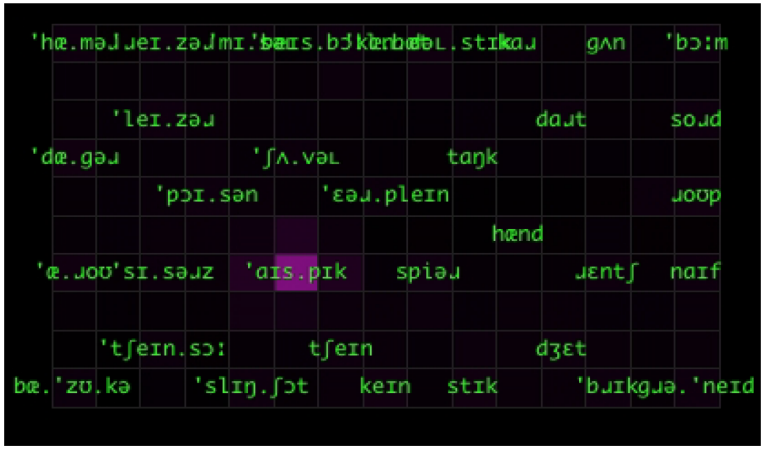


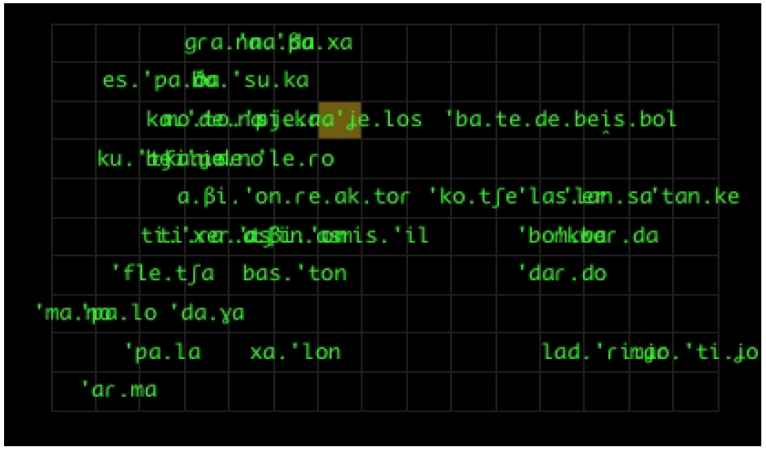
- Goal: Can fit model to any patient
  - Data from 19 patients at UT Austin and Boston Univ.
- L1w Spanish, L2d English
  - 300 words from real treatment sessions
  - 261 semantic features from treatment
  - 168(S) / 120(E) phonetic features (24 IPA in 7/5 syllables)
  - 30$\times$40 SOMs
- 64 possible patient models
  - 4 age levels: -45/60/80/80+ yrs $\rightarrow$ 800/1100/1400/1700 epochs
  - 4 AoA levels: -8/18/30/30+ yrs $\rightarrow$ 0/100/180/300 epochs

Semantic Map

- Organized by semantic similarity
  - Here a subarea “can be a weapon”

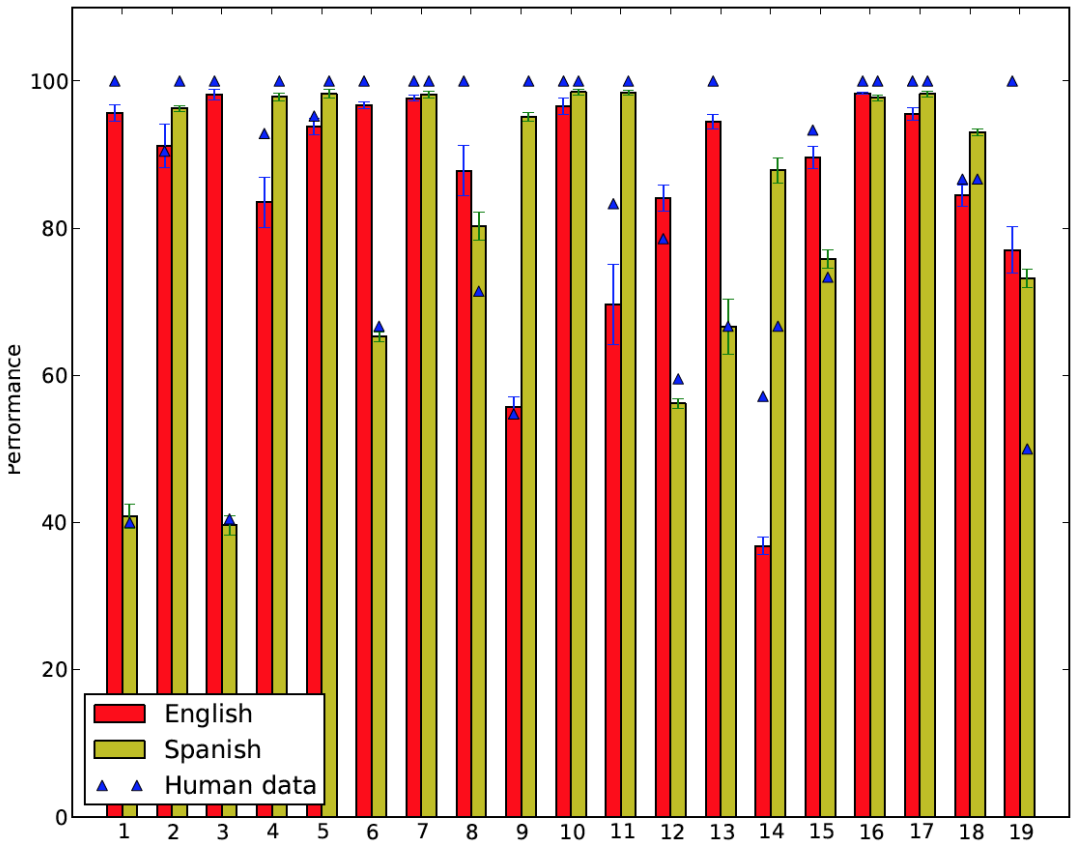
Phonological Maps





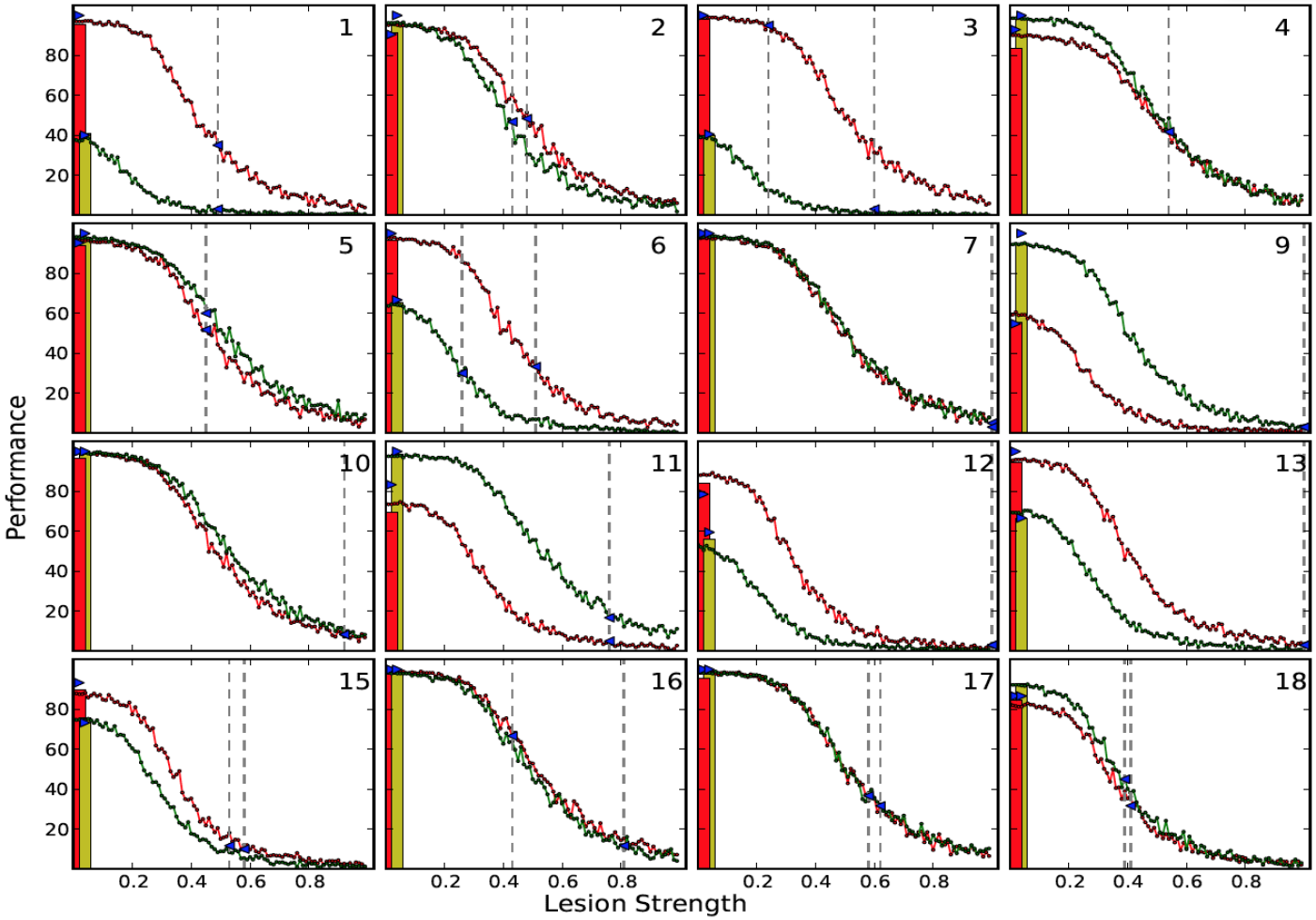
- Organized by word length and similarity
- Late AoA results in poor organization (maps and connections)

Matching Prestroke Performance



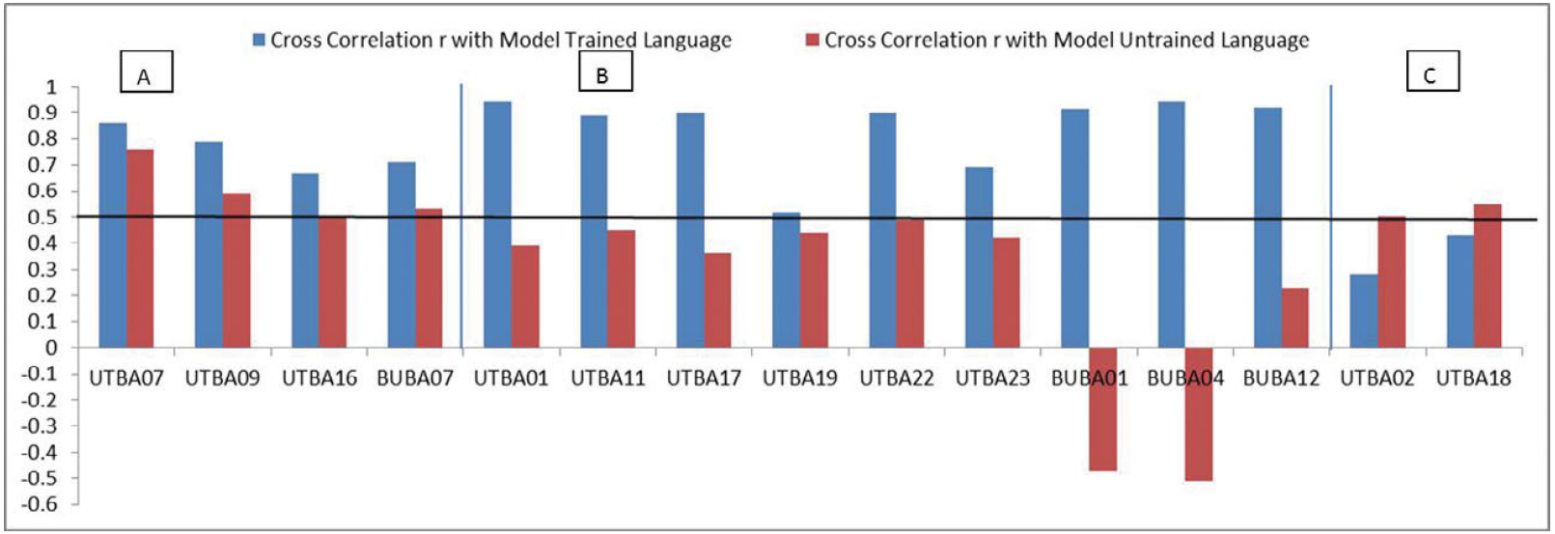
- Can model almost every patient
  - Except one with “irregular” patient data

Matching Poststroke Performance



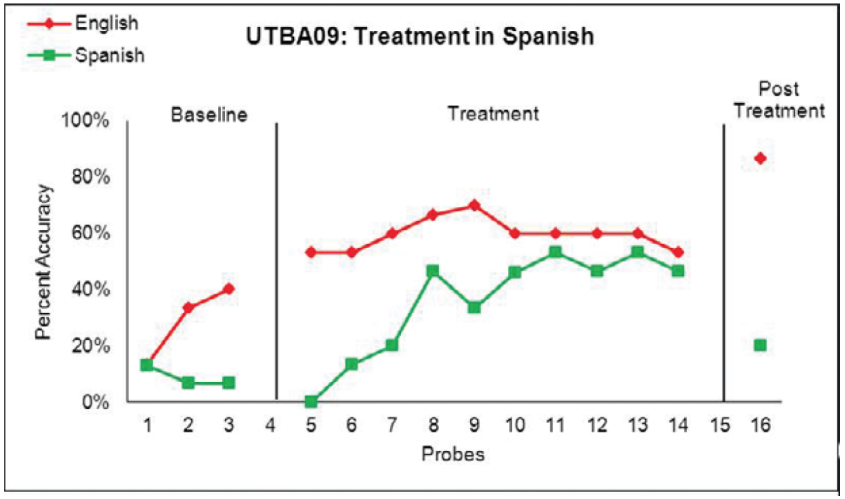
- Adding noise to associative connections
- Equal damage to both languages is mostly sufficient
  - Differential damage in three cases

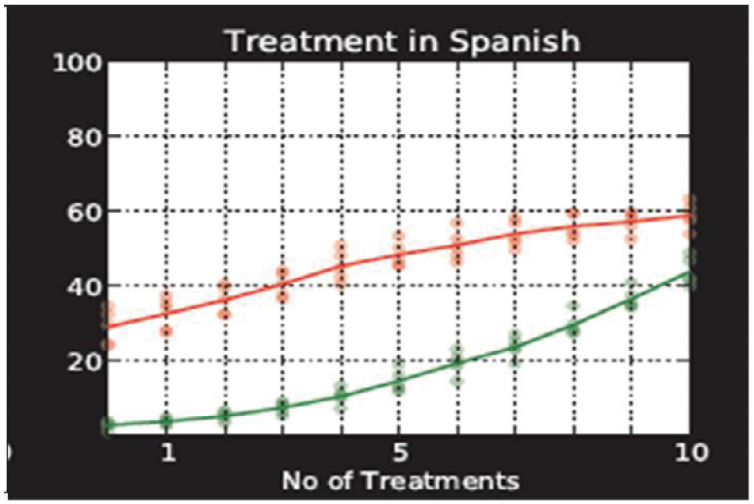
Modeling Rehabilitation



- 10 patient sessions $\rightarrow$ 10 epochs with 0.01 η
- A good match in all but two cases
  - Severely damaged patients

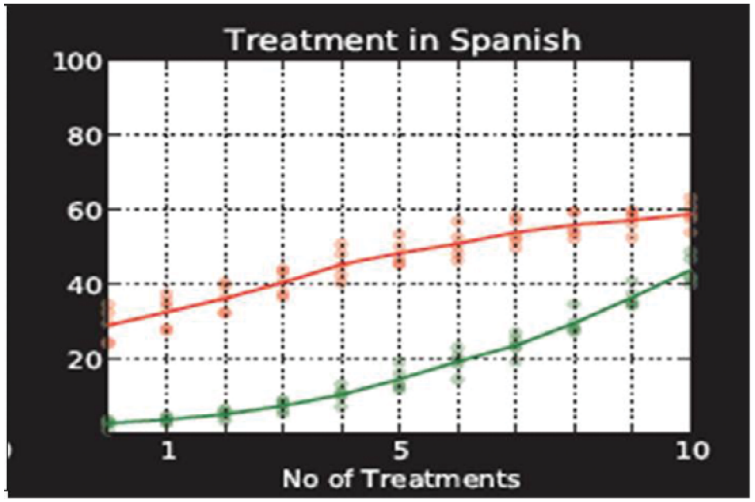
Modeling Transfer

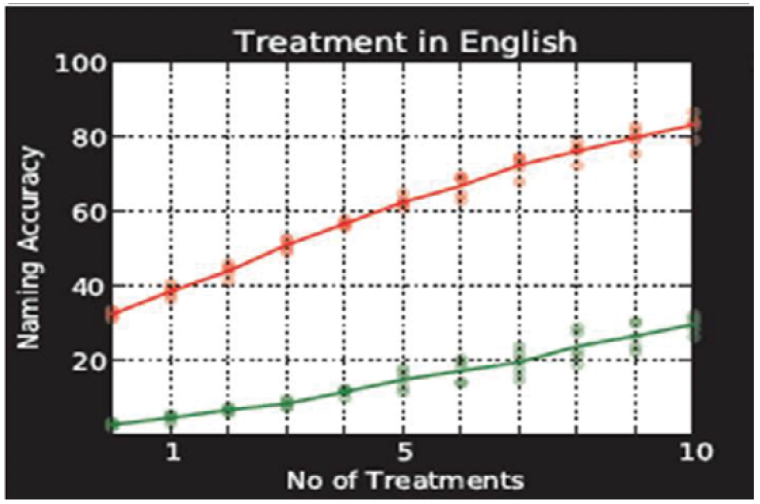




- Untrained language (E) improves as well
  - Stronger when trained with non-dominant language (S)
  - Explains the clinical observation

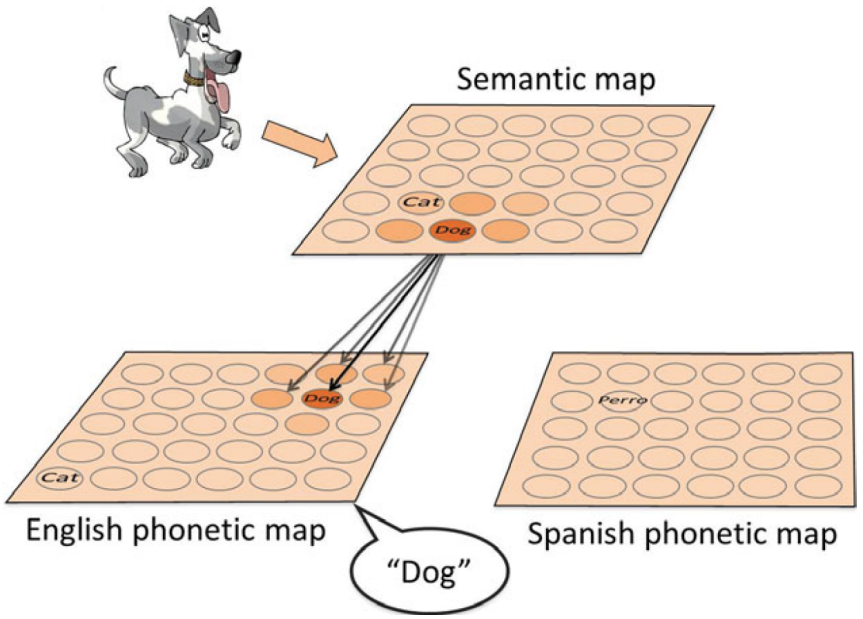
Recommending Treatment





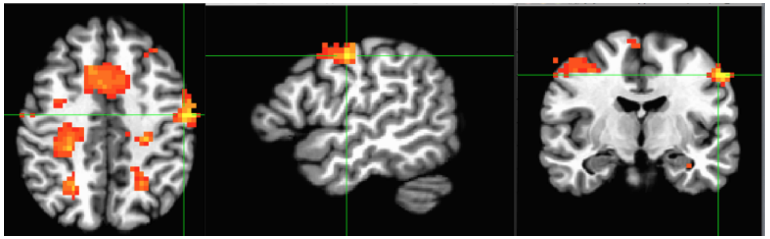
- Model was trained separately in both languages
- In some cases, the actual rehab was in wrong language!
  - Model predicts better recovery with the other language (E)
- Can be used to recommend treatment

Bilingual Lexicon Conclusion

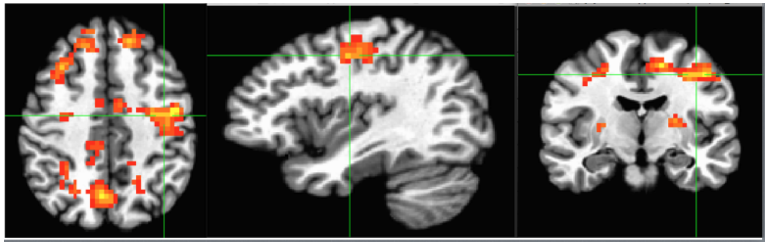


- With SOMs possible to model individual subjects
  - Physical model: available data is enough
- Can be used to predict treatment outcome
  - And thereby recommend treatment
  - Currently no guidelines exist
- Ready for a clinical trial!

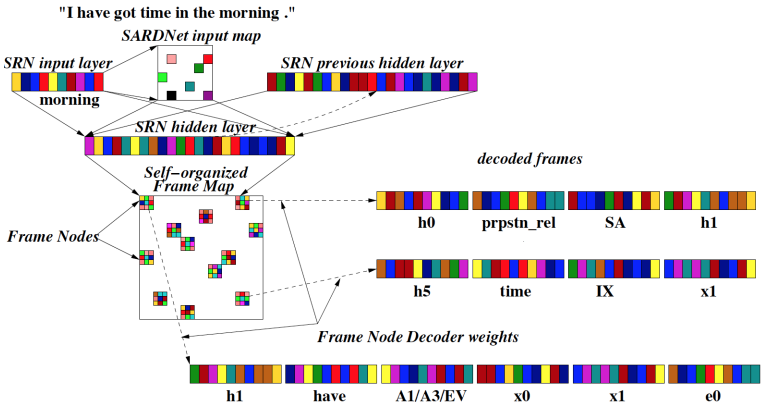
Extending to Sentence Semantics



“manipulate” (Binder et al. 2014)



“touch” (Binder et al. 2014)

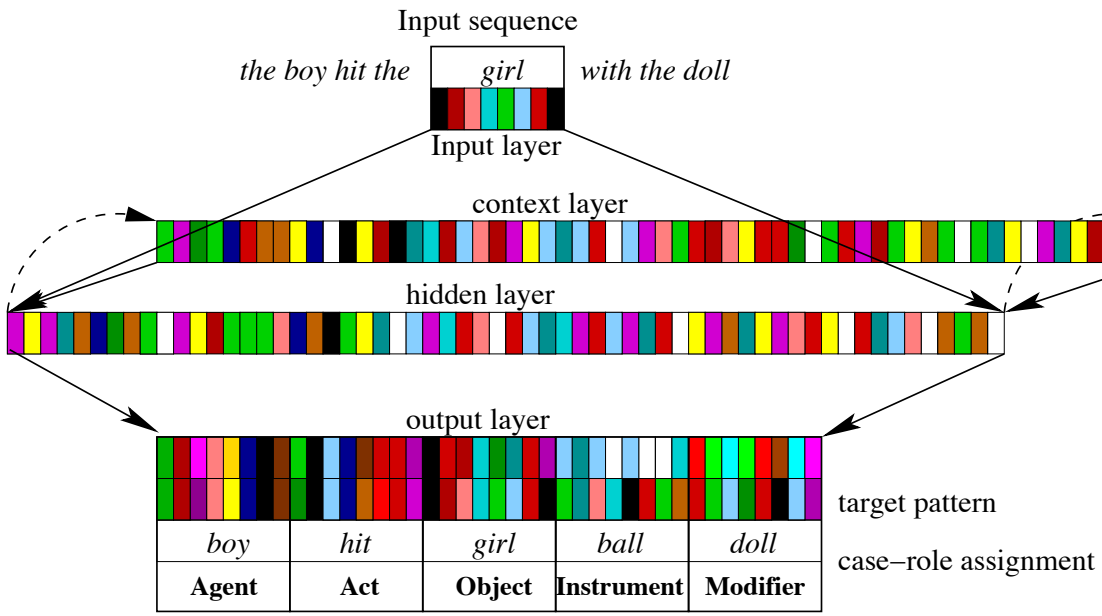


(Mayberry & Miikkulainen 2003)

- If word semantics are on a map, maybe also sentence semantics?
  - As a collection of mapped components
- INSOMNet runs with this idea, resulting in:
  - Parsing a real language corpus
  - Robust parsing
  - Cognitive performance

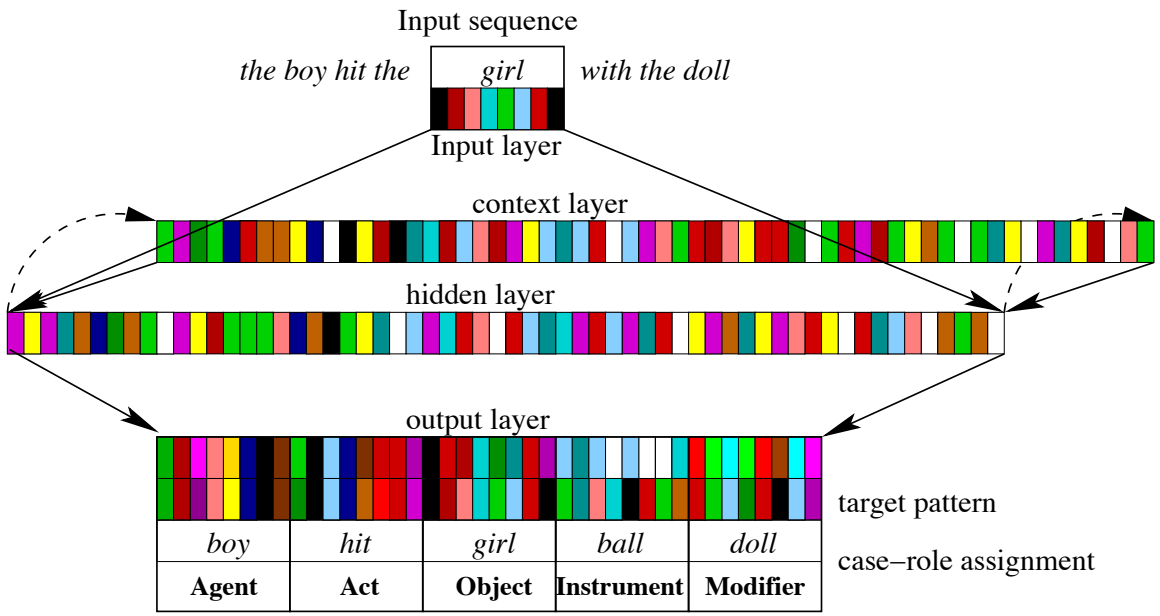
Related Work

Simple Recurrent Network (SRN; Elman 1990)



- ⋆ General architecture with simple memory
- ⋆ Parsimonious account of language processing
- ⋆ Sentence mapped into a case-role representation
- ⋆ Incremental development of interpretation
- ⋆ Interpretation actively revised as more context is read

BUT ...

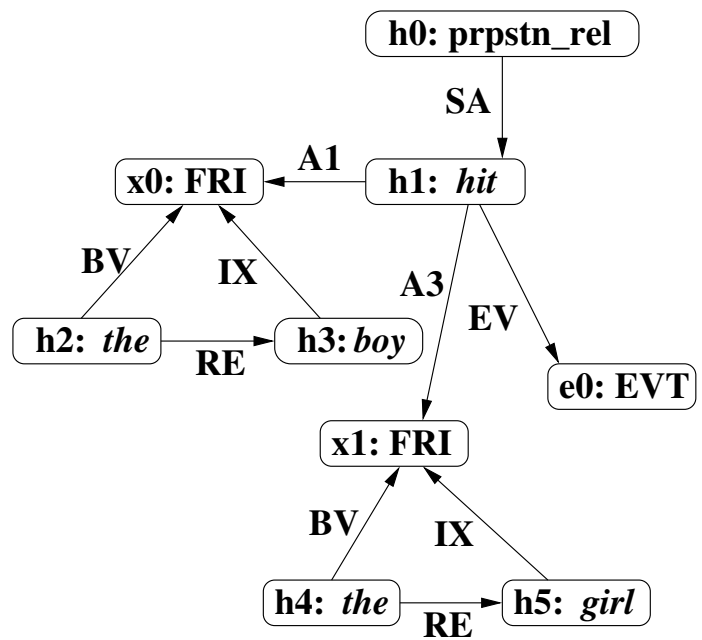


- ⋆ Fixed output
- ⋆ Prespecified role assemblies
- ⋆ Multiple case-role frames? Binding?
- ⋆ Embedded clauses? Multiple phrases?
- ⋆ **Cognitively appealing but hard to scale up.**

Minimal Recursion Semantics (MRS)

A. Copestake, D. Flickinger, & I. Sag

the boy hit the girl



- ⋆ **Predicate Calculus**
  hit(the(x_0,boy(x_0)),the(x_1,girl(x_1)),e_0)

- ⋆ **Minimal Recursion Semantics**
  h1: hit(x_0, x_1,e_0)
  h2: the(x_0,h3)
  h3: boy(x_0)
  h4: the(x_1,h5)
  h5: girl(x_1)

Use of *handles* yields

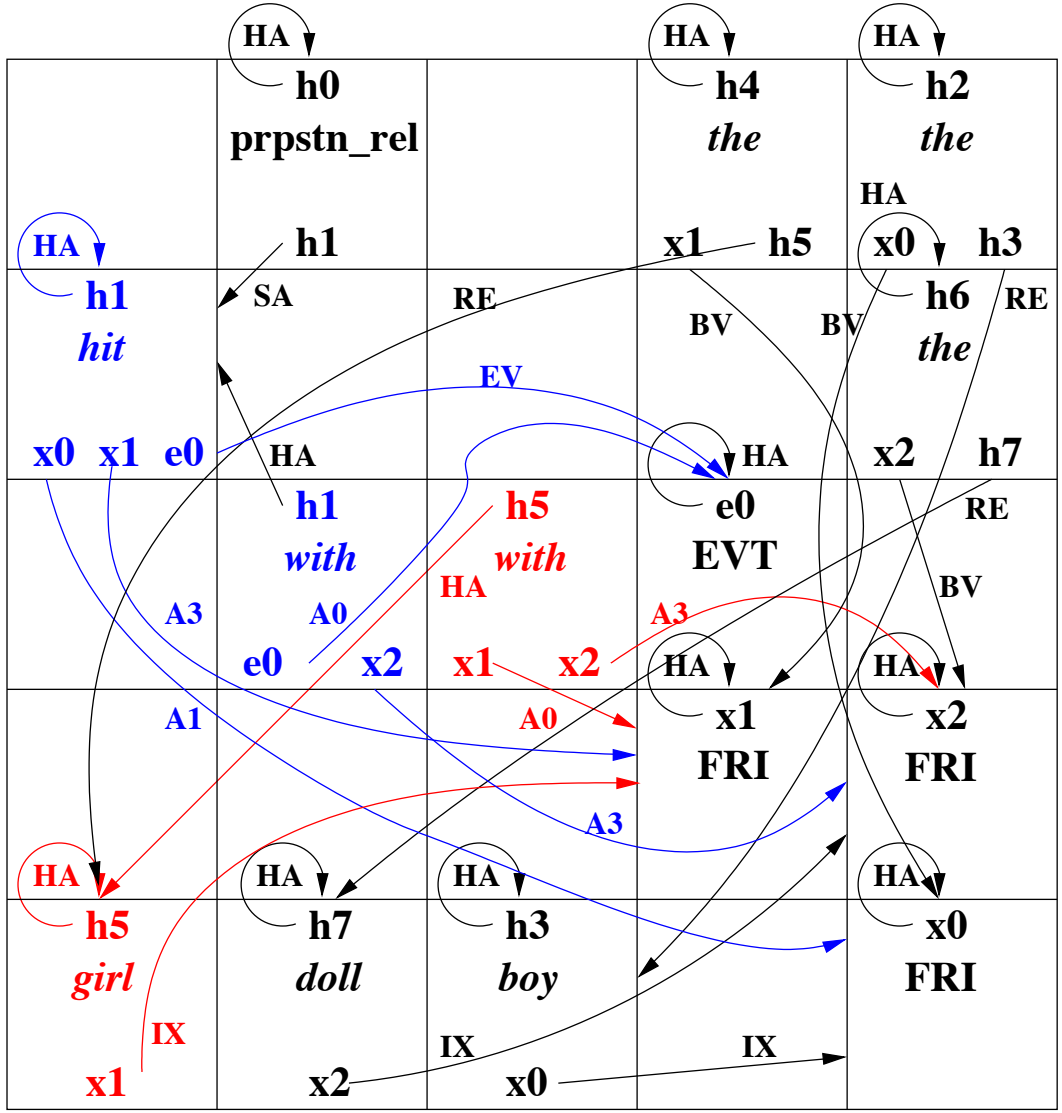
- ⋆ Flat semantics
- ⋆ Logical equivalence
- ⋆ Computational tractability

MRS Graph Representation (1)

	h0 prpstn_rel h1		h4 *the* x1 h5	h2 *the* x0 h3
h1 *hit* x0 x1 e0				h6 *the* x2 h7
	h1 *with* e0 x2	h5 *with* x1 x2	e0 EVT	
			x1 FRI	x2 FRI
h5 *girl* x1	h7 *doll* x2	h3 *boy* x0		x0 FRI

Grid cells encode MRS frame components

MRS Graph Representation (2)



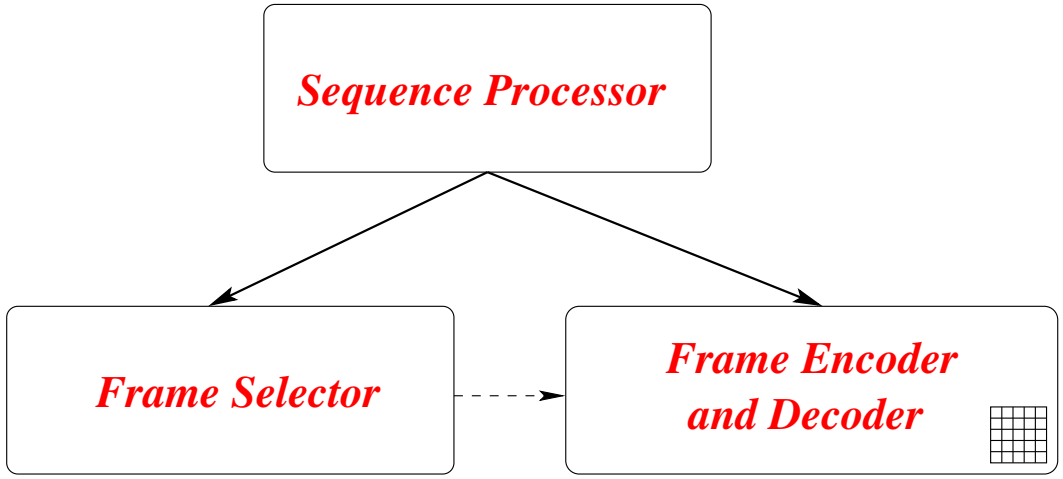
Handles are addresses of grid cells

MRS Graph Representation (3)

	h0 prpstn_rel SA h1		h4 *the* BVDMRESC x1 h5	h2 *the* BVDMRESC x0 h3
h1 *hit* A0A1A3DMEV x0 x1 e0				h6 *the* BVDMRESC x2 h7
	h1 *with* A0A3DMEV e0 x2	h5 *with* A0A3DMEV x1 x2	e0 EVT DVASMOTN	
			x1 FRI DVGNPTPN	x2 FRI DVGNPTPN
h5 *girl* A3IX x1	h7 *doll* IX x2	h3 *boy* A3IX x0		x0 FRI DVGNPTPN

Labeled arcs encoded as subcategorization type

INSOMNet Overview



Sequences?

Incremental sentence processing.

Targets?

MRS frame representation.

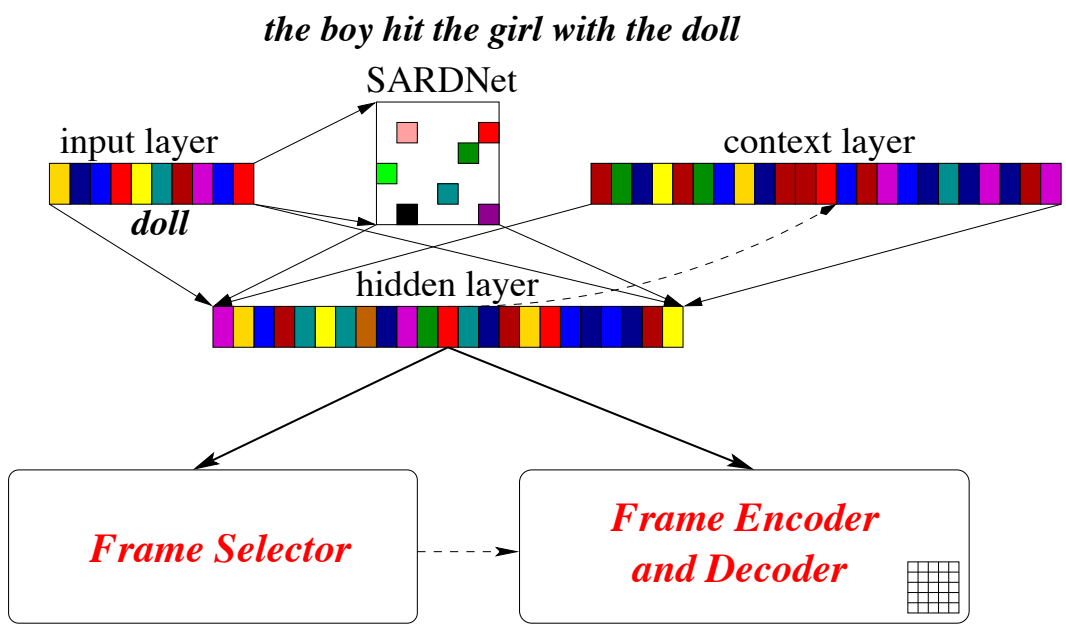
Parsing?

Graded frame selection.

Bindings?

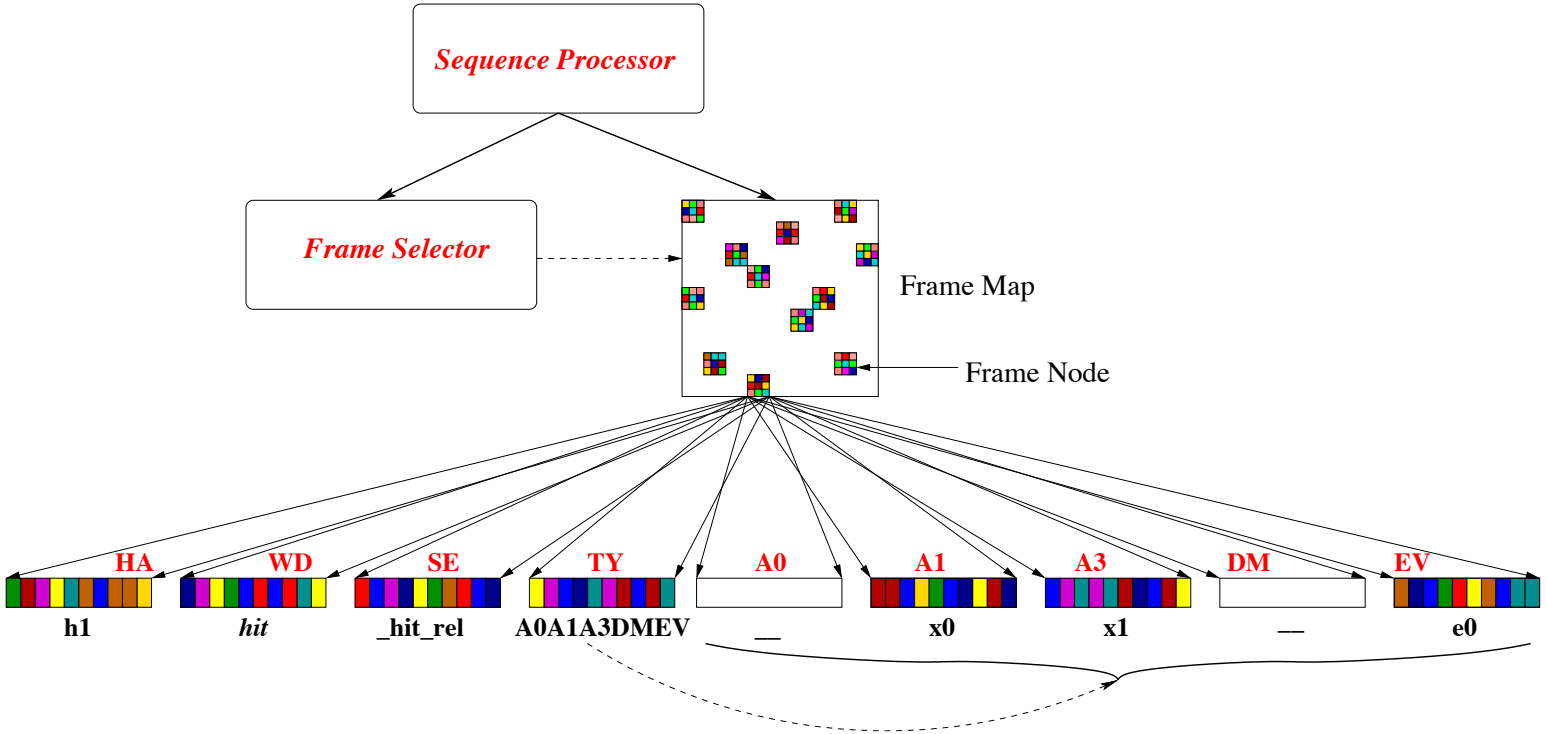
Represented using handles as role fillers.

Sequence Processor
(Activation)



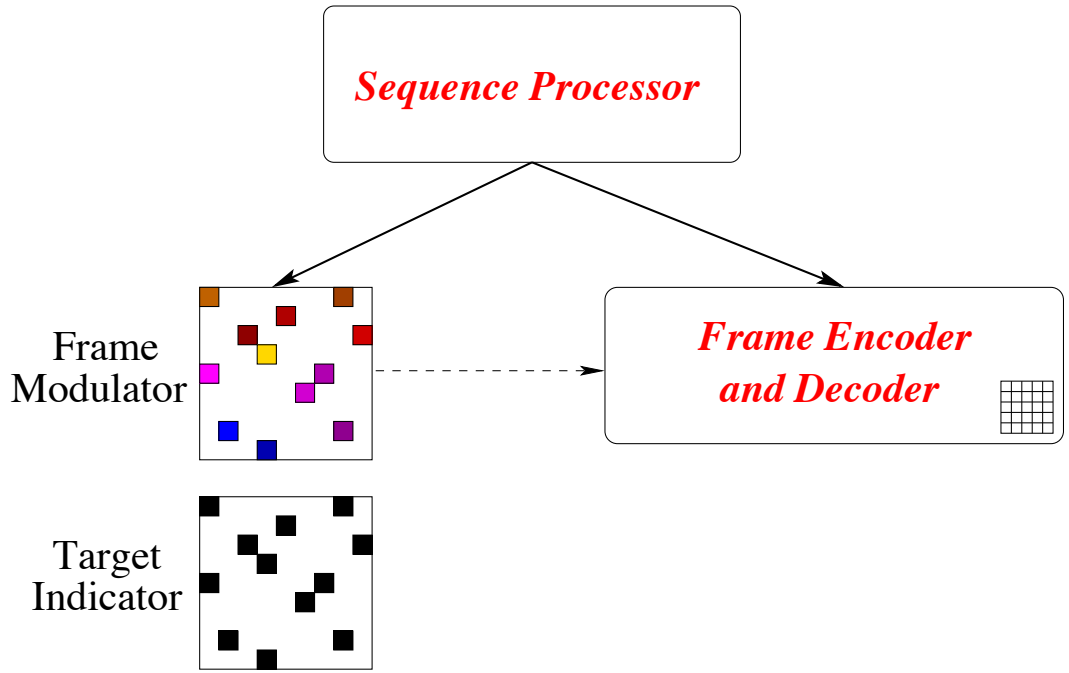
- ⋆ SRN
  - basis for INSOMNet
  - processes sentence incrementally
  - produces explicit semantic representation (vs. query-based models; e.g. Rohde, 2002)
- ⋆ SARDNet helps retain long-distance dependencies

Semantic Frame Encoder/Decoder
(Activation)



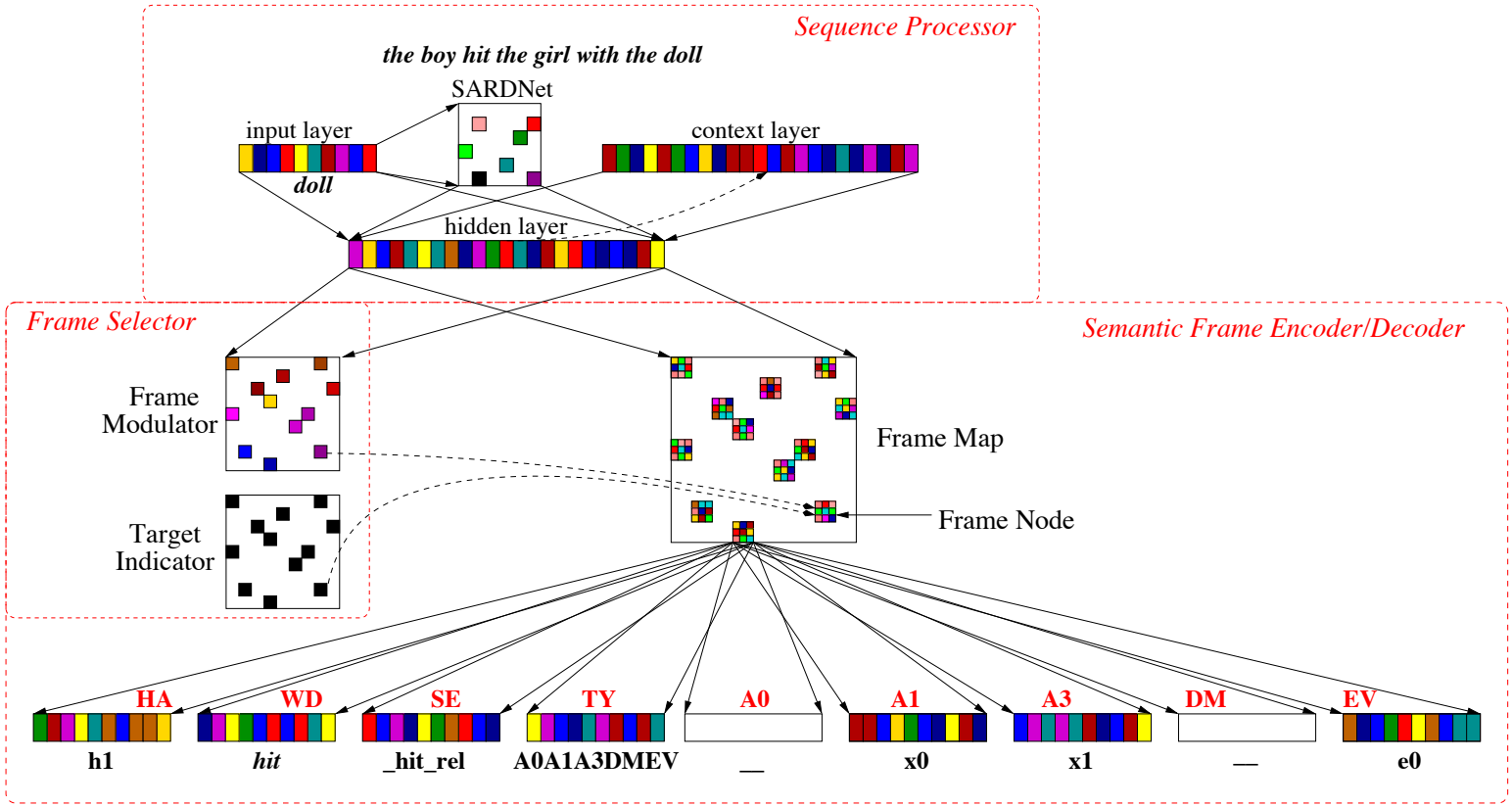
- ⋆ SRN output is map of encoded MRS frames
- ⋆ Dedicated links for specific argument roles
  - All frames have HA, WD, SE, TY links (handle, word, semantic relation, and subcat type)
  - Other links based on subcategorization type (e.g., A0, A1, A3, DM, and EV above)

Frame Selector
(Activation)



- ⋆ Frame Modulator Map
  - One-to-one correspondence with Frame Nodes
  - Gives graded frame selection
- ⋆ Target Indicator Map
  - One-to-one correspondence with Frame Nodes
  - Binary selection

INSOMNet
Putting It All Together

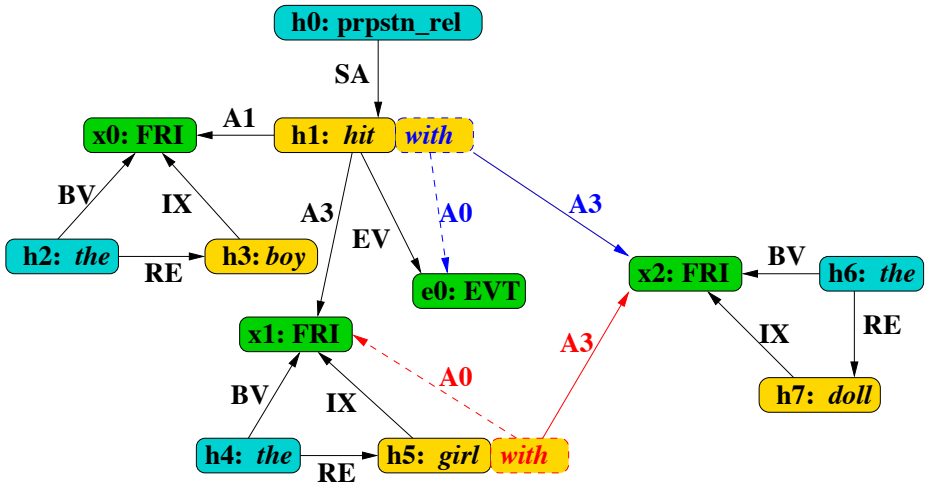


Sequence Processor reads in sentence incrementally

Semantic Frame Encoder/Decoder represents MRS frames

Frame Selector models graded frame selection

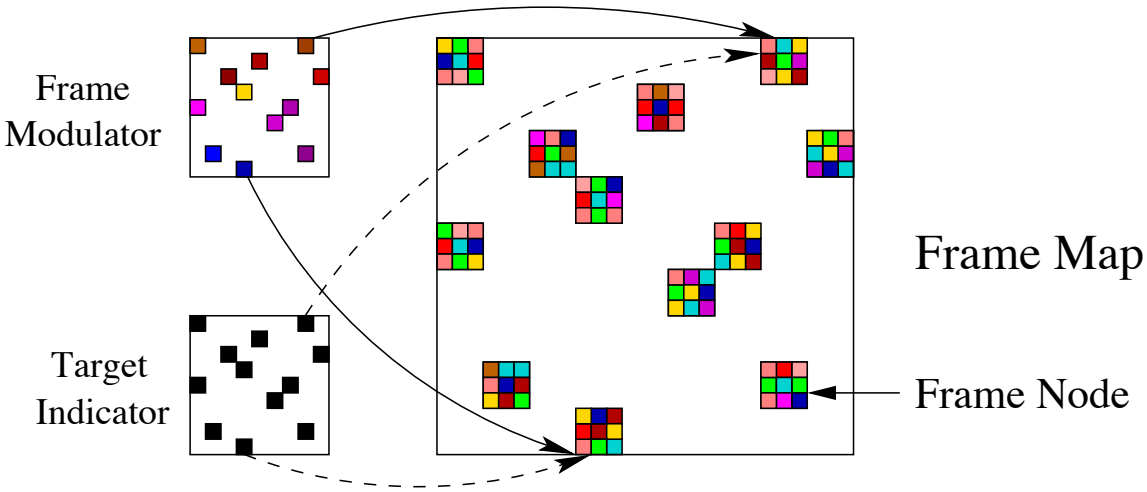
Encoding Semantic Characteristics
with RAAM (Pollack 1990)



	h0 prpstn_rel SA h1		h4 *the* BVDMRESC x1 h5	h2 *the* BVDMRESC x0 h3
h1 *hit* A0A1A3DMEV x0 x1 e0				h6 *the* BVDMRESC x2 h7
	h1 *with* A0A3DMEV e0 x2	h5 *with* A0A3DMEV x1 x2	e0 EVT DVASMOTN	
			x1 FRI DVGNPTPN	x2 FRI DVGNPTPN
h5 *girl* A3IX x1	h7 *doll* IX x2	h3 *boy* A3IX x0		x0 FRI DVGNPTPN

- ⋆ Build reduced descriptions of frames and store in handles
- ⋆ Encode recursively (green $\prec$ gold $\prec$ cyan)
- ⋆ Order is important
- ⋆ Represents semantic characteristics of frames
- ⋆ Handles function as *content-addressable pointers*

Semantic Self-Organization



Frame Modulator Map

- ⋆ provides graded Frame Node selection
- ⋆ trained through back propagation
- ⋆ target is Target Indicator Map

Target Indicator Map

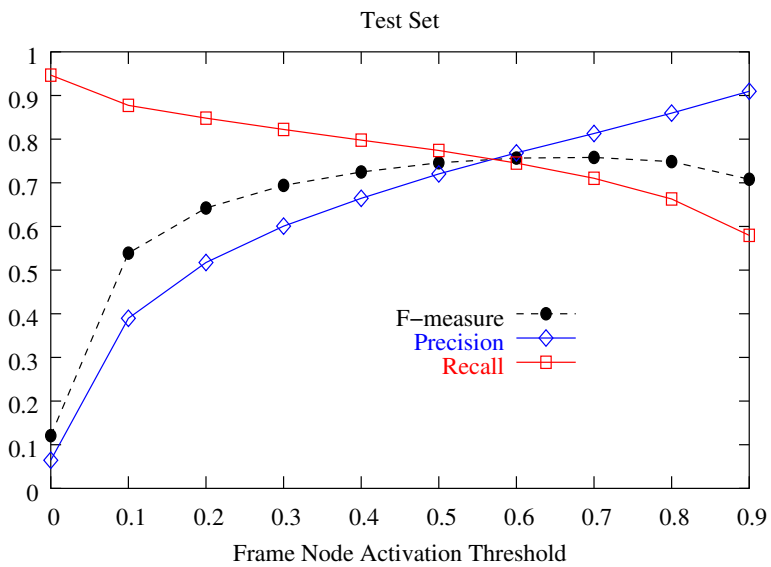
- ⋆ is only used during training
- ⋆ is self-organized using handles
- ⋆ selects Frame Nodes to encode MRS frames

LinGO Redwoods

S. Oepen, D. Flickinger, C. Manning, & K. Toutanova
Stanford University

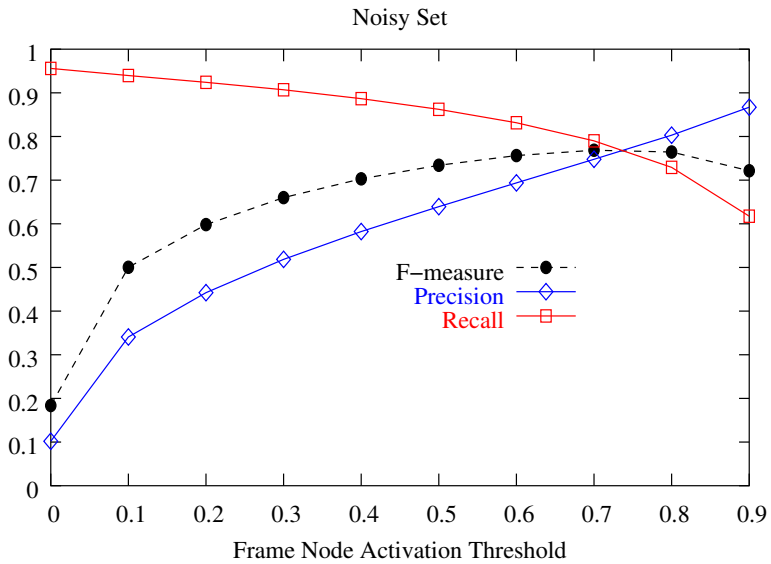
- ⋆ A Rich and Dynamic Treebank for Head-driven Phrase Structure Grammar (HPSG)
- ⋆ Deep annotation of syntactic and semantic information
- ⋆ More than 5000 sentences from the VerbMobil project (dialogues of face-to-face business travel arrangements)
- ⋆ Existing statistical model to rank parses from Treebank grammar provides basis for rough comparison
- ⋆ *"Actually next Friday does not look too bad."*
- ⋆ *"I have a dentist appointment on the sixteenth but I can call him and reschedule it."*

Performance Results



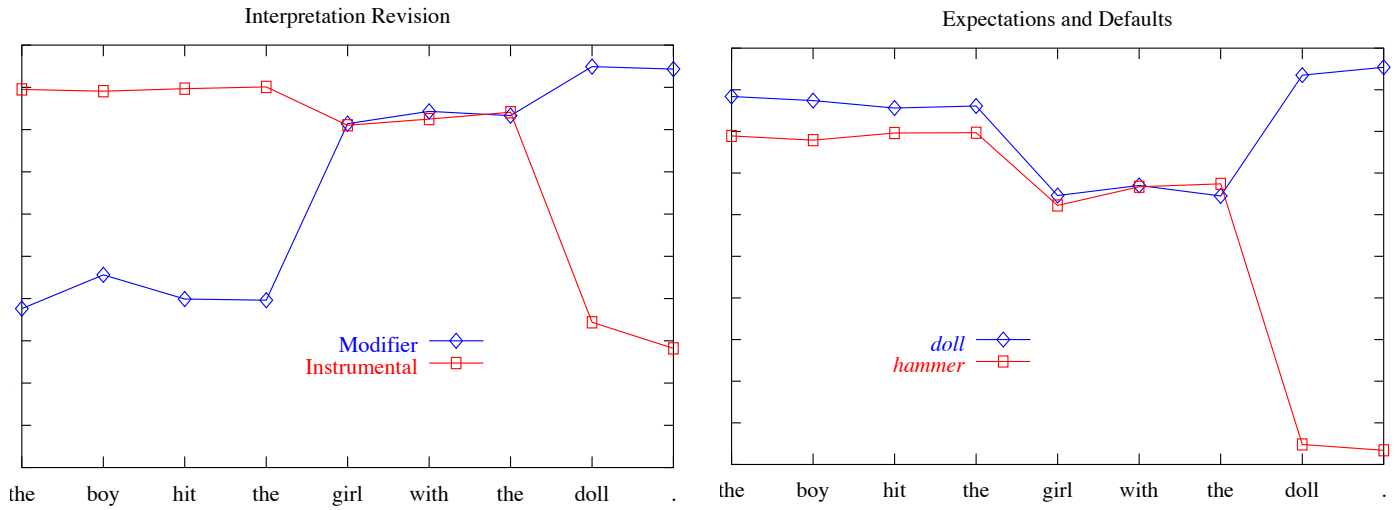
- ⋆ Ten-fold cross-validation on 4817 sentences
- ⋆ *Exact handle match criterion*: an item is counted only if
  - – it is decoded correctly
  - – the frame containing it is above threshold
  - – if item is a handle, the frame it points to is above threshold
- ⋆ Precision/Recall curves for exact handle match criterion (comprehension accuracy over frame components factored in)
- ⋆ F-Measure is ≈ 0.76 at intersection
- ⋆ **Compared to** CSLI log-linear model: $\approx 68\%$ accuracy on test set using *exact tree match criterion* (grammar provides trees which model ranks)

Robustness Results



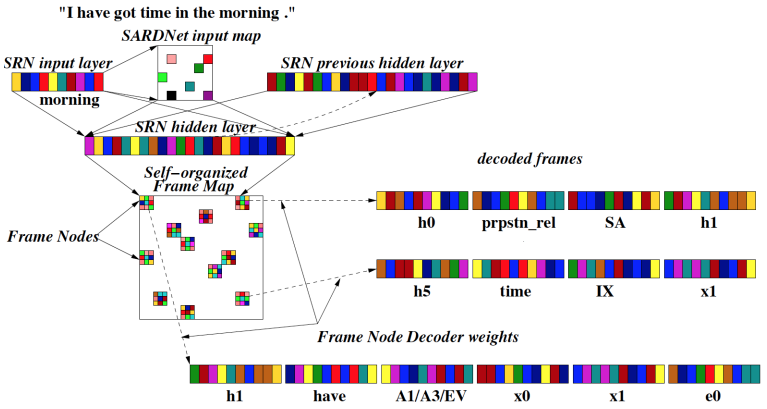
- ⋆ Pre-trained scale up network tested on 5068 noisy sentences
- ⋆ Average of five insertions per sentence
  - repairs: *# uh the fir I m free uh the first through the fourth.*
  - dysfluency: *I am free # from ah that # from nine , until oh six.*
  - ungrammatical input: *Here is some clues.*
- ⋆ Precision/Recall curves for *exact handle match criterion* (with comprehension accuracy factored in)
- ⋆ F-measure is ≈ 0.75 at intersection

Cognitive Plausibility



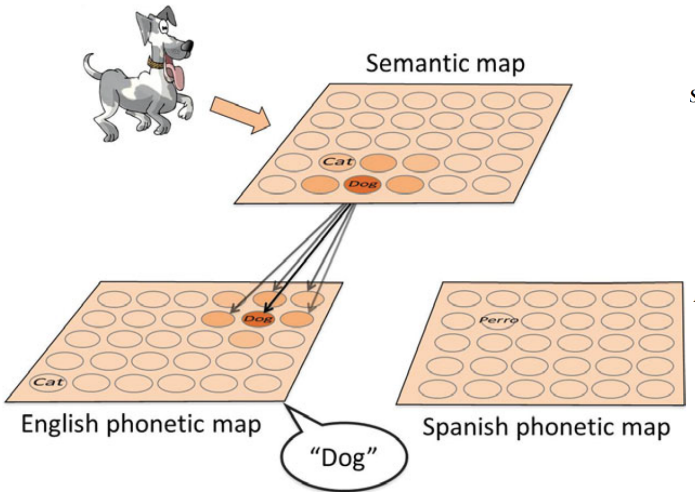
- ⋆ Trained and tested on McClelland & Kawamoto corpus
- ⋆ Plausible account of ambiguity: *with*
  - early preference for instrumental interpretation
  - modifier interpretation increases at *girl*
- ⋆ Defaults, expectation, and semantic priming
  - early expectation of *hammer* as instrument
  - decays away at *doll*
  - semantic flipping
- ⋆ Semantic interpretation actively revised during processing

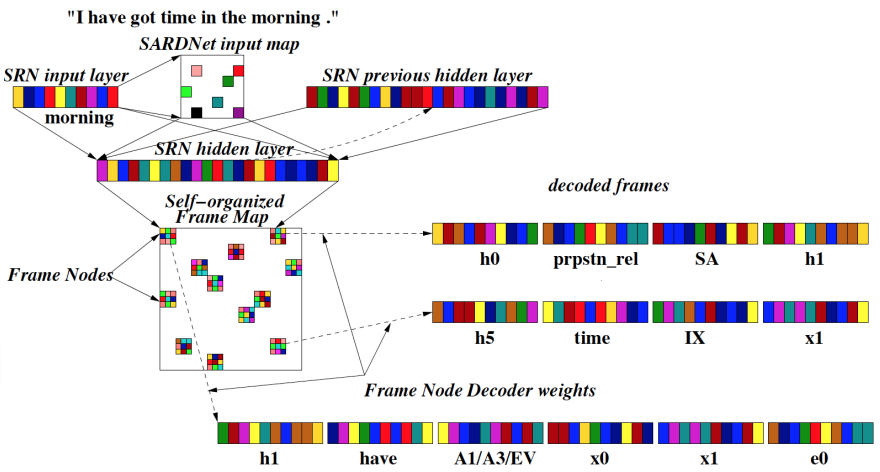
Sentence Semantics Conclusion



- INSOMNet = Semantic parsing with soft constraints
- Novel representation of sentence meaning
  - Dynamically constructed from constituents on a map
  - Graded activation of semantics
- A very different parsing approach
  - Scales up to real langauge
  - Robust against speech errors
  - Cognitively appealing performance

Conclusion





- Language semantics on a map
  - Individual locations for words
  - Distributed patterns for sentences
- Can account for robustness, cognitive effects, impairments
- May be possible to verify with imaging soon
  - fMRI for structure, TMS for function
- Artificial systems can be built
  - Robust language processing
  - A next step from Siri/Voice?