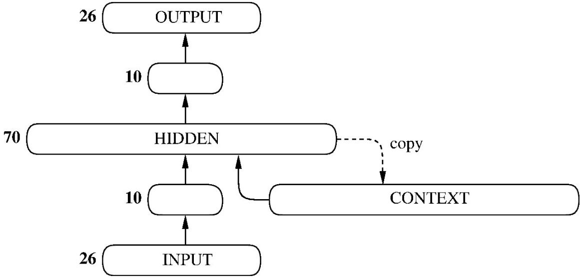
Traditional view of language

- Language knowledge largely consists of an explicit *grammar* that determines what sentences are part of a language
  - Isolated from other types of knowledge—pragmatic, semantic, lexical(?)
- Language learning involves identifying the single, correct grammar of the language
- Grammar induction is under-constrained by the linguistic input given lack of explicit negative evidence
  - Impossible under near-arbitrary positive-only presentation (Gold, 1967)
- Language learning requires strong *innate linguistic constraints* to narrow the range of possible grammars considered

Statistical view of language

- Language environment has rich *distributional regularities*
  - May not provide correction but is certainly not *adversarial* (cf. Gold, 1967)
- Language learning requires only that knowledge across speakers converges sufficiently to support effective communication
- No sharp division between linguistic vs. extra-linguistic knowledge
  - Effectiveness of learning depend both on the structure of the input and on *existing knowledge* (linguistic and extra-linguistic)
- Distributional information can provide *implicit* negative evidence
  - Example: *implicit prediction* of upcoming input
  - Sufficient for language learning when combined with *domain-general* biases

A connectionist approach to sentence processing (Elman, 1991)



S $\rightarrow$ NP VI . | NP VT NP .
NP $\rightarrow$ N | N RC
RC $\rightarrow$ who VI | who VT NP | who NP VT
N $\rightarrow$ boy | girl | cat | dog | Mary | John | boys | girls | cats | dogs
VI $\rightarrow$ barks | sings | walks | bites | eats | bark | sing | walk | bite | eat
VT $\rightarrow$ chases | feeds | walks | bites | eats | chase | feed | walk | bite | eat

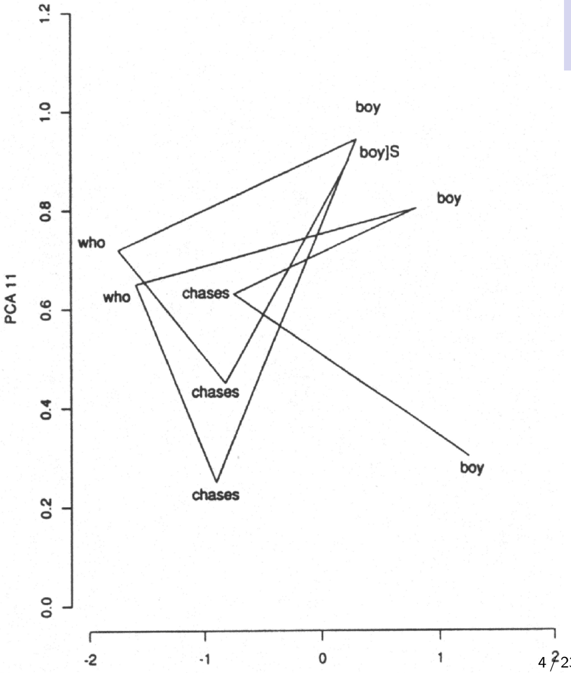
- Simple recurrent network trained to predict next word in English-like sentences
  - Context-free grammar, number agreement, variable verb argument structure, multiple levels of embedding
  - 75% of sentences had at least one relative clause; average length of 6 words.
  - e.g., **Girls who cat who lives chases walk dog who feeds girl who cats walk .**
- After 20 sweeps through 4 sets of 10,000 sentences, mean absolute error for new set of 10,000 sentences was **0.177** (cf. initial: 12.45; uniform: 1.92)

Principal Components Analysis

Boy chases boy who chases boy who chases boy .

Principal Components Analysis (PCA) of network's internal representations

- Largest amount of variance (PC-1) reflects **word class** (noun, verb, function word)
- Separate dimension of variation (PC-11) encodes **syntactic role** (agent/patient) for nouns and level of embedding for verbs



Sentence comprehension

Traditional perspective

- Linguistic knowledge as grammar, separate from semantic/pragmatic influences on performance (Chomsky, 1957)
- Psychological models with initial syntactic parse that is insensitive to lexical/semantic constraints (Ferreira & Clifton, 1986; Frazier, 1986)

Problem: Interdependence of syntax and semantics

The spy saw the policeman with a revolver
The spy saw the policeman with binoculars
The bird saw the birdwatcher with binoculars

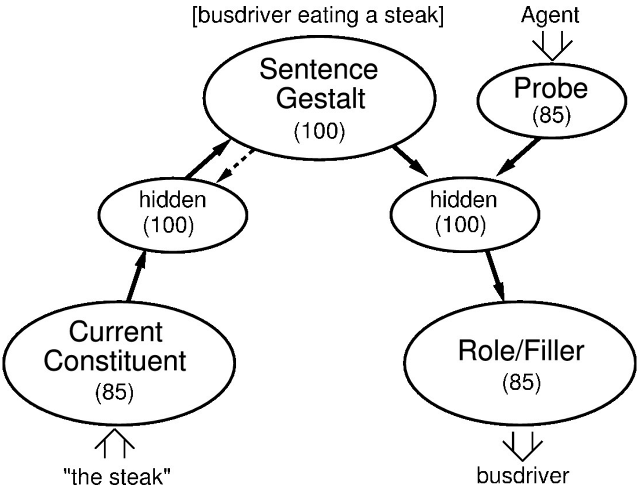
The pitcher threw the ball
The container held the apples/cola
The boy spread the jelly on the bread

Alternative: Constraint satisfaction

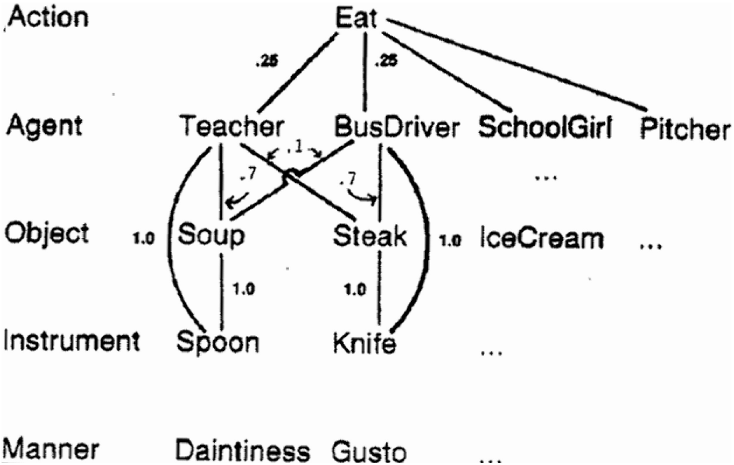
Sentence comprehension involves integrating multiple sources of information (both semantic and syntactic) to construct the most plausible interpretation of a sentence (MacDonald et al., 1994; Seidenberg, 1997; Tanenhaus & Trueswell, 1995)

Sentence Gestalt Model (St. John & McClelland, 1990)

- Trained to generate thematic role assignments of event described by single-clause sentence
- Sentence constituents (≈ phrases) presented one at a time
- After each constituent, network updates internal representation of sentence meaning ("Sentence Gestalt")
- Current Sentence Gestalt trained to generate *full set* of role/filler pairs (by successive "probes")
    - Must **predict** information based on partial input and learned experience, but must **revise** if incorrect



Event structures

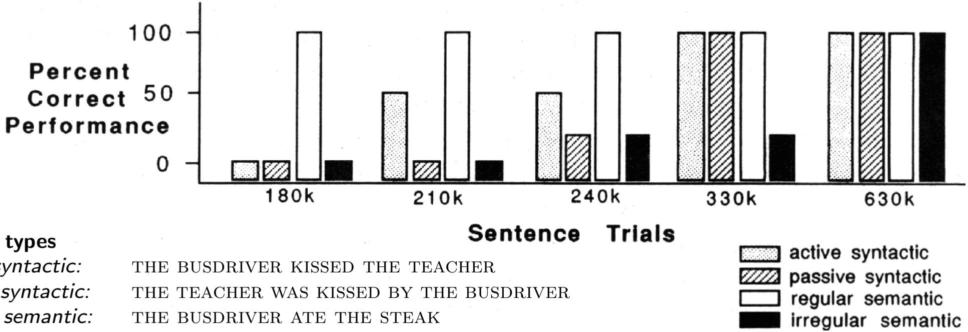


- 14 active frames, 4 passive frames, 9 thematic roles
- Total of 120 possible events (varying in likelihood)

Sentence generation

- Given a specific event, probabilistic choices of
  - Which thematic roles are explicitly mentioned
  - What word describes each constituent
  - Active/passive voice
- Example: busdriver eating steak with knife
  - THE-ADULT ATE THE-FOOD WITH-A-UTENSIL
  - THE-STEAK WAS-CONSUMED-BY THE-PERSON
  - SOMEONE ATE SOMETHING
- Total of 22,645 sentence-event pairs

Acquisition



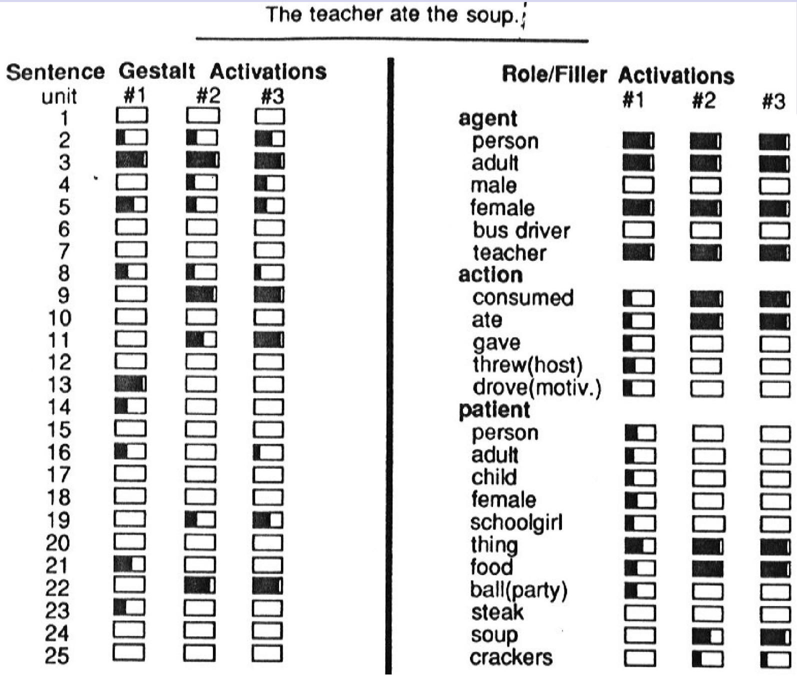
Sentence types

Active syntactic:	THE BUSDRIVER KISSED THE TEACHER
Passive syntactic:	THE TEACHER WAS KISSED BY THE BUSDRIVER
Regular semantic:	THE BUSDRIVER ATE THE STEAK
Irregular semantic:	THE BUSDRIVER ATE THE SOUP

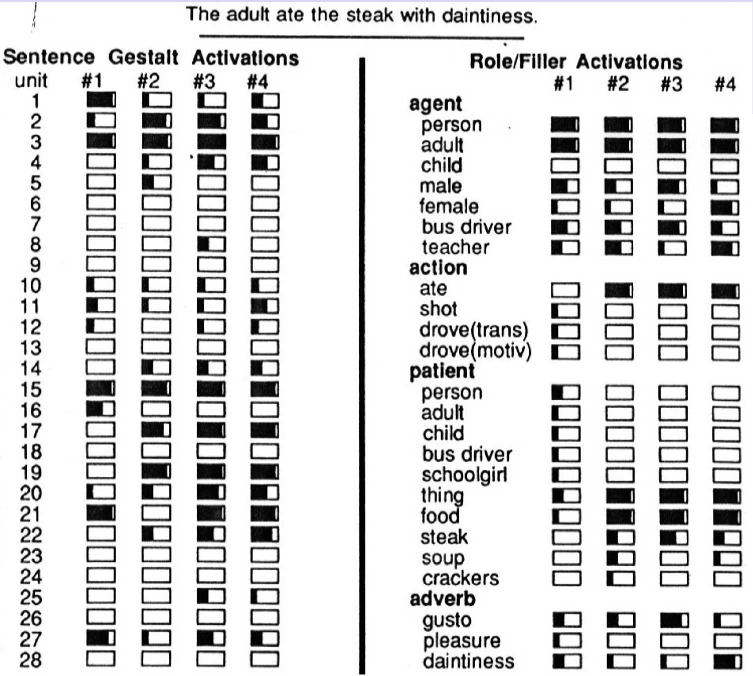
Results

- Active voice learned before passive voice
- Syntactic constraints learned before semantic constraints
- Final network tested on 55 randomly generated unambiguous sentences
    - Correct on 1699/1710 (99.4%) of role/filler assignments

Implied constituents

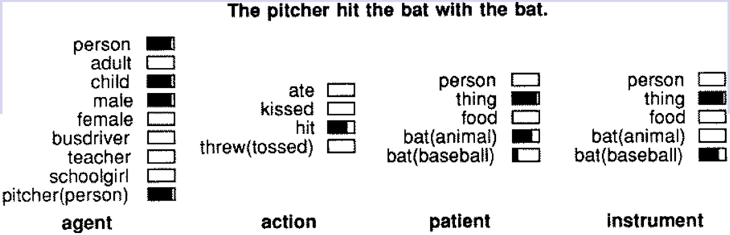


Online updating and backtracking

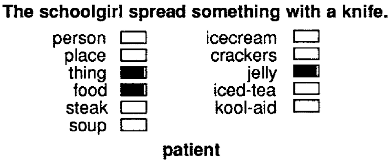


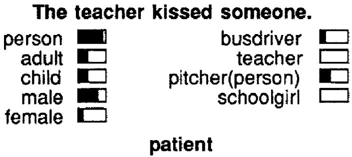
Semantic-syntactic interactions

Lexical ambiguity

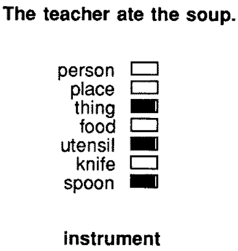


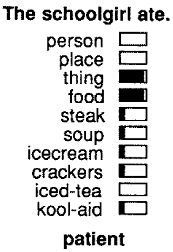
Concept instantiation



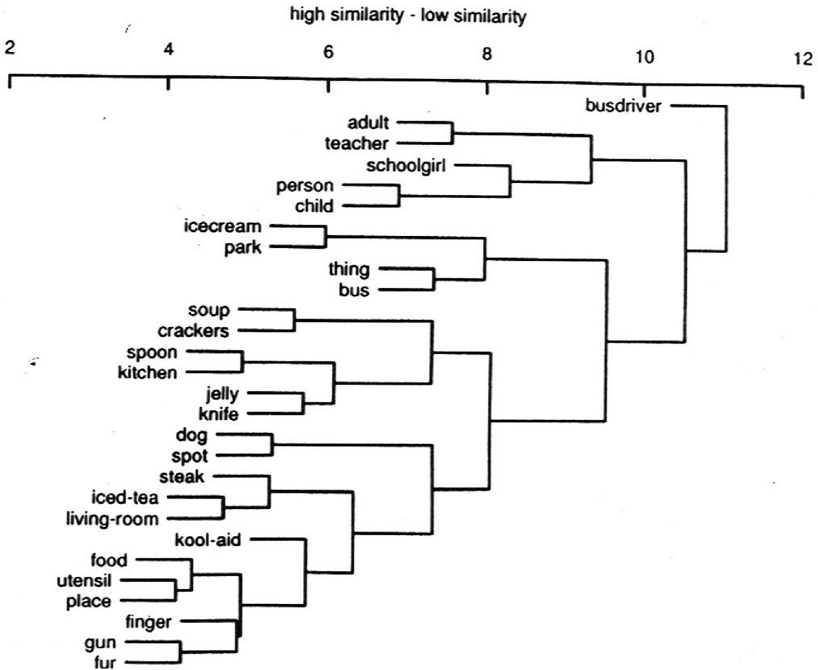


Implied constituents

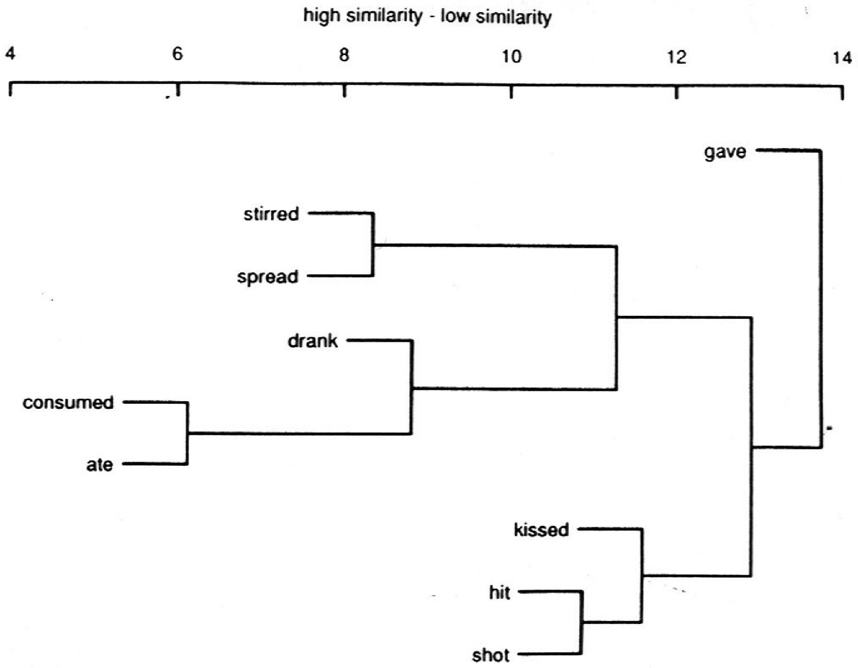




Noun similarities

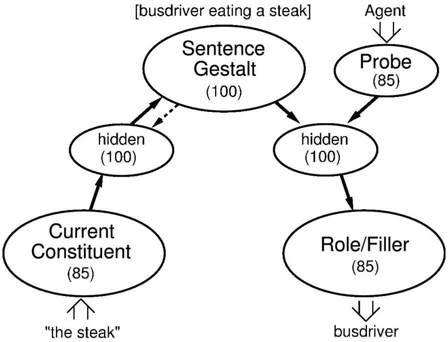


Verb similarities

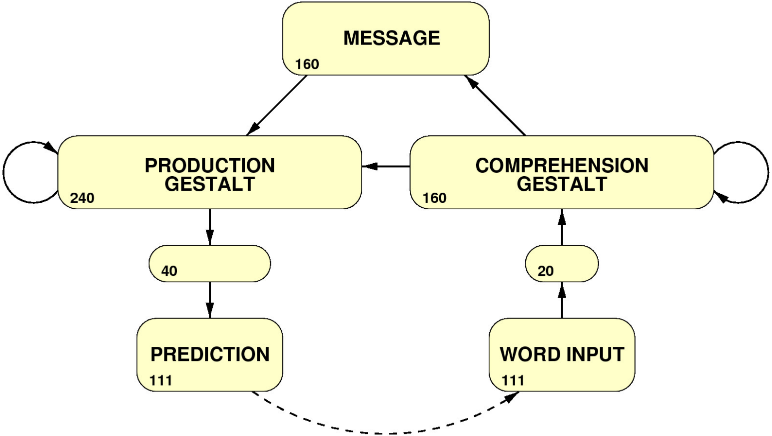


Summary: St. John and McClelland (1990)

- Syntactic and semantic constraints can be learned and brought to bear in an integrated fashion to perform online sentence comprehension
- Approach stands in sharp contrast to linguistic and psycholinguistic theories espousing a clear separation of grammar from the rest of cognition



Sentence comprehension and production (Rohde Ph.D. thesis)



- Extends approach of Sentence Gestalt model to multi-clause sentences
- Trained to generate learned "message" representation and to predict successive words in sentences when given varying degrees of prior context

Training language

- Multiple verb tenses
  - e.g., ran, was running, runs, is running, will run, will be running
- Passives
- Relative clauses (normal and reduced)
- Prepositional phrases
- Dative shift
  - e.g., gave flowers to the girl, gave the girl flowers
- Singular, plural, and mass nouns
- 12 noun stems, 12 verb stems, 6 adjectives, 6 adverbs

Examples

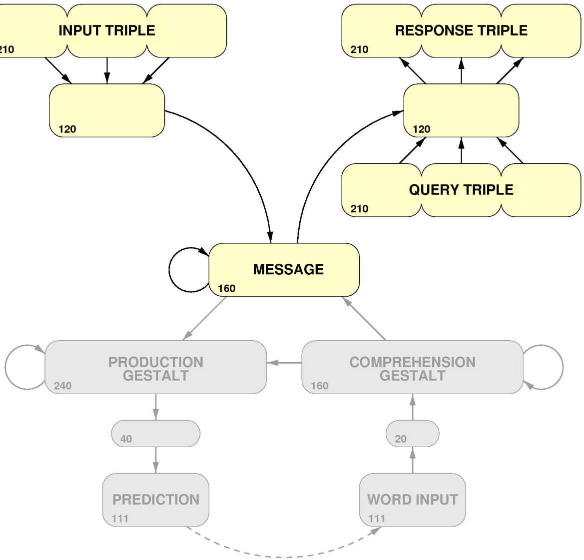
- The boy drove.
- An apple will be stolen by the dog.
- Mean cops give John the dog that was eating some food.
- John who is being chased by the fast cars is stealing an apple which was had with pleasure.

Encoding messages with triples

The boy who is being chased by the fast dogs stole some apples in the park .

steal action transitive instantaneous past	**boy** entity singular definite animate human	**apple** entity plural indefinite
chase action passive ongoing present which	**boy** entity singular definite animate human	**dog** entity plural definite animate
property	**dog** entity plural definite animate	**fast** quality
location	**steal** action transitive instantaneous past	**park** entity singular definite

Message encoder



Methods

- Triples presented in sequence
- For each triple, all presented triples queried three ways (given two elements, generate third)
- Trained on 2 million sentence meanings

Results

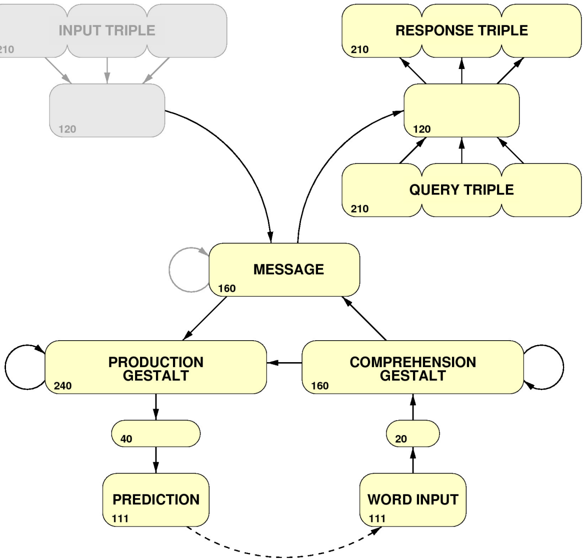
- Full language

Triples correct:	91.9%
Components correct:	97.2%
Units correct:	99.9%

- Reduced language ($\leq$10 words):

Triples correct:	99.9%

Training: Comprehension (and prediction)



Methods

- No context on half of the trials
- Context was *weak clamped* (25% strength) on other half
- Initial state of message layer clamped with varying strength

Results

- Correct query responses with comprehended message:

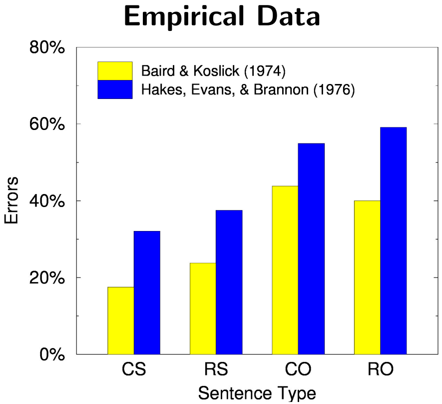
  Without context: 96.1%

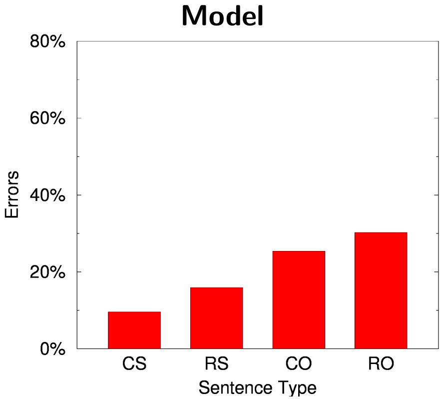
  With context: 97.9%

Testing: Comprehension of relative clauses

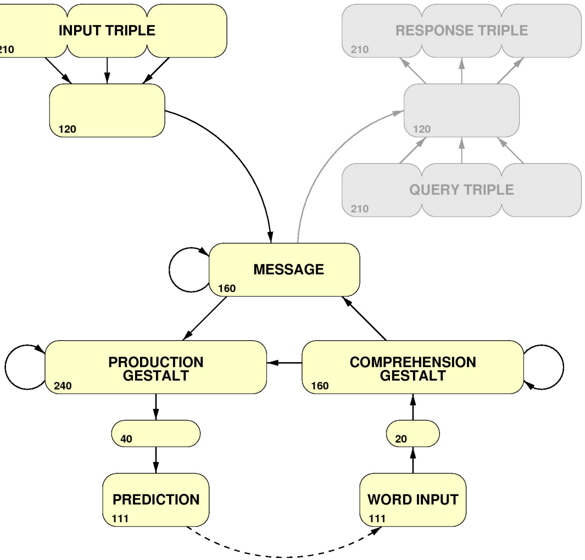
Single embedding: Center- vs. Right-branching; Subject- vs. Object-relative

- CS: A dog [who chased John] ate apples.
- RS: John chased a dog [who ate apples].
- CO: A dog [who John chased] ate apples.
- RO: John ate a dog [who the apples chased].





Testing: Production



Methods

- Message initialized to correct value and weak clamped (25% strength)
- Most actively predicted word selected for production
- No explicit training

Results

- 86.5% of sentences correctly produced.

Transformers (state-of-the-art natural language processing: e.g., ChatGPT)

- Sequence-to-sequence encoding
- Traditional approach learns single hidden representation that encodes entire input sequence (and generates entire output sequence
- Transformers learn separate context-sensitive encodings of input elements which are combined flexibly (using "attention") to generate output elements