Language Acquisition:

Computational Modeling

Afra Alishahi, Heiner Drenhaus

Fall 2010, Winter 2011
Computational Linguistics and Phonetics
Saarland University

What is Computational Modeling of Human Language Acquisition?

- Human language acquisition
  - Identify processes and mechanisms involved in learning language
  - Detect common behavioural patterns among children
- Computational modeling
  - Simulate a cognitive process via computational tools and techniques
  - Use the model to explain the observed human behaviour
- Computational modeling of human language acquisition
  - Develop computational simulations of the process of human language acquisition

Computational Modeling of Human Language Acquisition

- Using computational methods for modeling cognitive processes of language learning enables us to
  - study these processes through simulation
  - evaluate the plausibility of existing theories of language learning and understanding
  - explain the observed human behavior during the process of learning and using a natural language
  - predict behavioral patterns that have not yet been experimentally investigated

Various Aspects of Language Acquisition

- Word segmentation: extract words from the speech stream
- Phonology: acquire the sound system of the language, and the correct form of each word
- Word meaning: map each word form to the concept it represents in the outer world
- Morphology: learn the regularities governing the structure of each word form
- Syntax: combine words and construct well-formed sentences
- Semantics: interpret the (relational) meaning of a phrase or sentence
- Pragmatics and discourse: use context to augment the meaning

The Focus of this Course

- Word segmentation: extract words from the speech stream
- Phonology: acquire the sound system of the language, and the correct form of each word
- Word meaning: map each word form to the concept it represents in the outer world
- Morphology: learn the regularities governing the structure of each word form
- Syntax: combine words and construct well-formed sentences
- Semantics: interpret the (relational) meaning of a phrase/sentence
- Pragmatics and discourse: how context attributes to meaning

Part I

General Issues

Characteristics of Human Language Acquisition

- Children learn to speak a language fluently at a young age
- Their linguistic knowledge is robust in the face of noise and incomplete data
- Speakers of the same language agree on grammaticality
- Humans are also flexible and creative when using language
- They face limitations on processing resources
- They learn and process language incrementally

Main Questions

- Representation of the linguistic knowledge
  - How is the knowledge organized in mind and brain?
    - Separate areas for representing different types of knowledge?
  - What is innate, what is learnable?

- Acquisition of the linguistic knowledge
  - Are different types of knowledge acquired in order?
  - What are the processes involved in language learning?

Language Modularity

- Representation of the linguistic knowledge
  - How is the knowledge organized in mind and brain?
    - Separate areas for representing different types of knowledge?
  - What is innate, what is learnable?

- Acquisition of the linguistic knowledge
  - Are different types of knowledge acquired in order?
  - What are the processes involved in language learning?

Modularity of Mind

- What is the architecture of the brain?
- Highly modular architecture (e.g., Fodor'83)
  - Each task (including language) is performed by domain-specific, encapsulated and autonomous modules
  - Interaction between these modules is minimal
- Functionalist approach (e.g., Sperber'94, Pinker'97)
  - Modules are defined by the specific operations they perform on the information they receive
- Many variations in between (e.g., Coltheart'99, Barrett & Kurzban'06)

Modularity of Language

- How is language related to other cognitive abilities?
- Highly modular architecture
  - Language is handled by a highly specific “mental organ” or “language faculty”
  - Evidence from studies of the Specific Language Impairment (SLI): language is isolated from other cognitive processes
- Functional approach
  - Language is represented and processed using the same general-purpose skills which underly other cognitive tasks
  - Evidence from Visual World Paradigm: language and other modules (e.g. vision, gesture) interact at process level

What is a Module?

- Do distinct modules exist within the language processor?
  - E.g. word segmentation, lexical development, syntax
- How to define a module:
  - Representational autonomy: each module has its own representational framework, but learning mechanisms are similar
  - Procedural autonomy: different mechanisms are involved in the acquisition of each aspect, but representations are shared
- The modularity debate is highly interleaved with nativism, or language innateness

Language Learnability

- Representation of the linguistic knowledge
  - How is the knowledge organized in mind and brain?
    - Separate areas for representing different types of knowledge?
  - What is innate, what is learnable?

- Acquisition of the linguistic knowledge
  - Are different types of knowledge acquired in order?
  - What are the processes involved in language learning?

Learnability and Nativism

- The Innateness Hypothesis (IH):
  - Humans have innately specified knowledge in several areas
  - Humans' innate abilities of language are domain-specific
    - I.e., highly detailed linguistic knowledge

- Localization:
  - Processing language is localized to specific regions of brain

- Innateness is <u>not</u> the same as localization

Dual Approach to Studying Language

- Linguistics: focus on "competence"
  - Representational frameworks which precisely and parsimoniously formalize a natural language according to adult speakers
- Psycholinguistics: focus on "performance"
  - process of learning and using a language by children and adults
- The Competence Hypothesis
  - Weak competence: people recover representations that are isomorphic to those of linguistic theories
  - Strong competence: people directly use grammatical knowledge and principles of linguistic theories

How to Approach these Questions?

- Language modularity and learnability have been discussed for decades
- The debate must ultimately be settled by neurological evidence, but for now we have
  - indirect evidence from psycholinguistics on how language is learned as used
  - insight from computational simulation of the plausible mechanisms of language acquisition

Experimental Investigation

- Controlled experimental studies of language
  - One aspect or property of a task or stimuli is manipulated, and other factors are held constant (controlled)
  - The effect of the manipulated condition is investigated among a large group of subjects
- Advantages
  - Isolate different language-related factors in the stimuli
  - Examine significance of the impact of each factor on the process
- Limitations
  - Only the the input (and not the process) can be manipulated
  - Each subject has a different learning history

Computational Simulation

- Computational models require detailed specification of the input properties and the processing mechanism
- Methodological advantage:
  - Explicit assumptions: all bias or constraint on the characteristics of the input data and learning mechanism are specified
  - Controlled input: researcher has full control over the input that the model receives in its life time
  - Observable behaviour: impact of every factor in the input or the learning process can be directly studied in the output
  - Testable predictions: novel situations or combinations of data can be simulated

Computational Language Acquisition

- We use computational modeling of human language acquisition for
  - suggesting cognitively plausible formalisms for representing linguistic knowledge
  - developing algorithms that can acquire knowledge of language from exposure to linguistic data
  - explaining the observed patterns and predicting new ones in the experimental data

Marr's Levels of Modeling

- Theories provide a high-level characterization of a process
- Marr's (1982) 3 levels of describing cognitive processes
  - Computational: what knowledge is computed
  - Algorithmic: how computation takes place
  - Implementation: how algorithms are realized in brain

- A computational model must specify, and be evaluated based on the level it attempts to simulate a process

What if the Model is Flawed?

stated at computational level

Theory

↓

Model

built at algorithmic level, therefore details of processing have to be specified

Cognitive Plausibility

- Realistic input data
  - Make realistic assumptions about the actual properties of the data available to children, e.g. noise, no negative evidence
- Language-independent strategies
  - Do not rely on learning techniques that only work for some languages, e.g. exploiting fixed word order
- Memory and processing limitations
  - Avoid unrealistically computation-heavy algorithms, e.g. remembering every sentence or processing data iteratively
- Incrementality
  - Process every piece of data when received

What to Expect from a Model

- A computational model can, at best
  - show that certain types of knowledge can be learned from certain types of input
  - suggest that a particular mechanism/algorithm is plausible due to the behavioural patterns it yields
- Computational cognitive models should conform to psychological plausibility criteria
  - At computational level, a cognitive model must make realistic assumptions about the properties of input
  - At algorithmic and implementation level, a model should conform to incrementality and processing limitations

Modeling Frameworks

- Symbolic models
  - rule-based, computationally well-understood, transparent with respect to their linguistic basis
- Connectionist models
  - inspired by the structure of brain: distributed representations of the input, output, and linguistic knowledge
- Probabilistic models
  - transparent linguistic basis, combined with experience-based learning and inference mechanisms
- Hybrid models
  - a combination of the above approaches, e.g. a symbolic representation of linguistic knowledge paired with a probabilistic learning mechanism

Symbolic Modeling

- Explicit formalization of the representation and processing of language through a symbol processing system
  - Linguistic knowledge
    - A set of symbols and their propositional relation
  - Learning and processing mechanism
    - Processing and updating knowledge via general rules or schemas, and under certain constraints
    - Each rule is augmented by a list of exceptions, i.e. tokens for which the rule is not applicable

Symbolic Modeling - Example

- Context Free Grammar (CFGs)
  - A symbolic formalism for representing grammatical knowledge of language

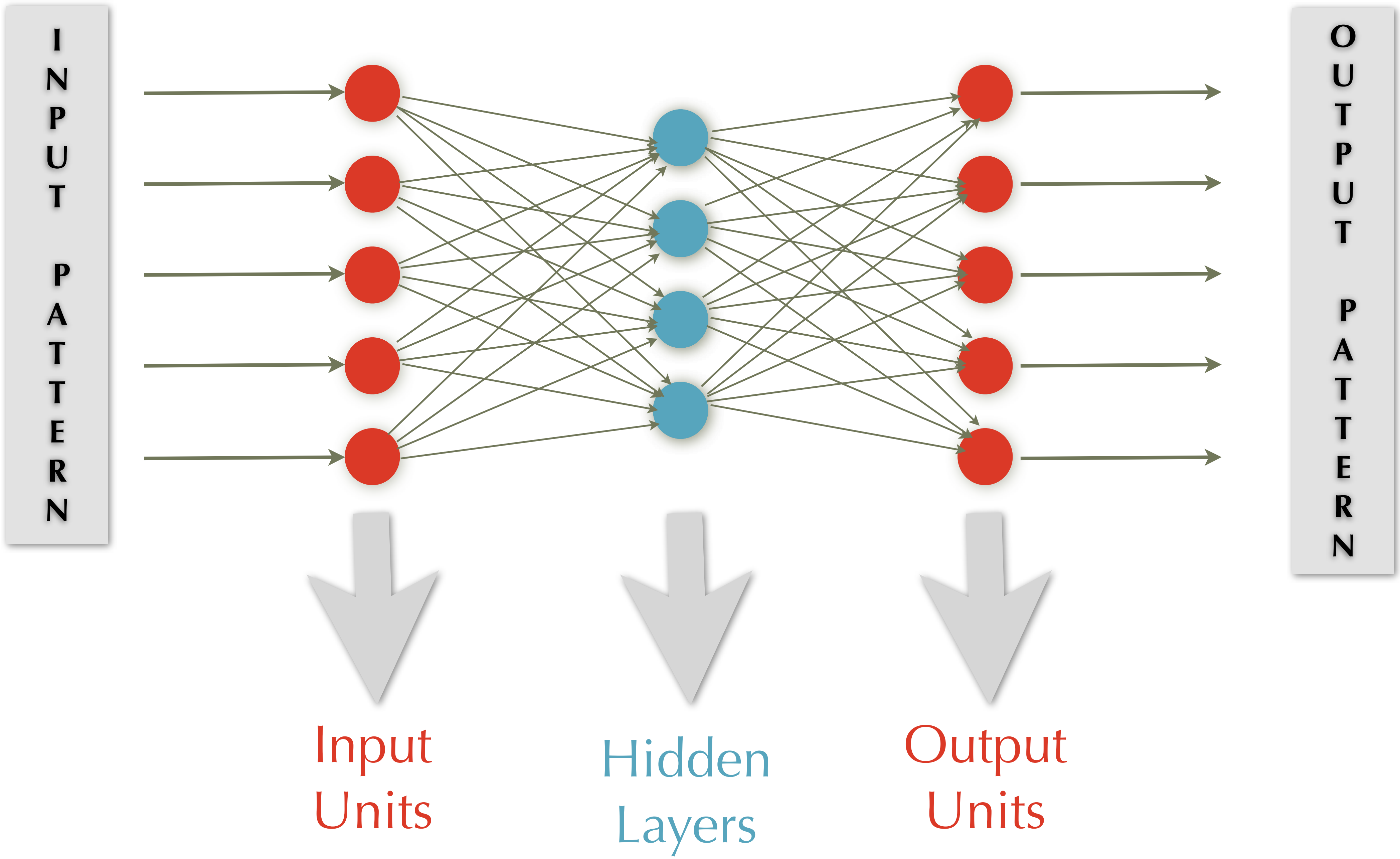
English Past Tense

```
Rule:          Vpast → Vroot + "ed"
Exceptions:    go → went, put → put, ...
```

Connectionist Modeling

- Inspired by simple neuronal processing in the brain
  - Linguistic knowledge
    - Distributed activation patterns over many neurons, and the strength of connections between them
  - Learning and processing mechanism
    - A neuron receives, processes and passes signals to other neurons
    - Connection weights between neurons change over time to improve the performance of the model in certain tasks
  - Cognitive processes
    - Large numbers of neurons perform basic computations in parallel

Connectionist Modeling - Example



Probabilistic Models

- Apply Probability Theory on previous language exposure
  - Linguistic knowledge
    - Weighted information units that reflect bias or confidence based on previous observations
  - Learning mechanism
    - Principled algorithms for weighting and combining evidence to form hypotheses that explain data best

- Bayesian modeling
  - Inference on observed data to infer the probability of a hypothesis

Bayesian Inference

- Bayes′ rule: break down complex probabilities into ones that are easier to compute

Likelihood of evidence e assuming hypothesis i

Prior probability of hypothesis i

Probability of a hypothesis i given some evidence e

$$P(i|e) = \frac{P(e|i)P(i)}{P(e)}$$

Prior probability of evidence e

- Find the hypothesis i that maximizes $P(i|e)$

Hybrid Models

- A combination of the techniques and formalisms from different frameworks
- Example:
  - a symbolic rule-based representation, where each rule is augmented with a probability value indicating its applicability
  - English past-tense formation rules:

```
Rule 1:    Vpast → Vroot + “ed”        probability: 0.7
Rule 2:    Vpast → Vroot               probability: 0.08
...        ...                         ...
```

Evaluation of Computational Models

- Cognitive models cannot be solely evaluated based on their accuracy in performing a task
    - The behavior of the model must be compared against observed human behavior
    - The errors made by humans must be replicated and explained
- Evaluation of cognitive models depends highly on experimental studies of language

Language Acquisition Models: Evaluation

- What humans know about language can only be estimated/evaluated through how they use it
  - Language processing and understanding
  - Language production
- Analysis of child production data yields valuable clues
  - Developmental patterns such as error and recovery
- Comprehension experiments reveal biases and preferences
  - knowledge sources that children exploit, and their biases towards linguistic cues

Language Production Data

- CHILDES database (MacWhinney, 1995)
  - An ever-growing collection of the recorded interactions (text, audio, video) between children and their parents

```
@Languages:     en
@Participants:  CHI Adam Target_Child, URS Ursula_Bellugi Investigator, MOT Mother, ...
@ID: en|brown|CHI|3;1.26|male|normal|middle_class|Target_Child||
@ID: en|brown|PAU|||||Brother||
@ID: en|brown|MOT|||||Mother||
..
@Date:          30-AUG-1963
@Time Duration: 10:30-11:30
*CHI:           one busses .
%mor:           det:num|one n|buss-PL .
%xgra:          1|2|QUANT 2|0|ROOT 3|2|PUNCT
*URS:           one .
%mor:           det:num|one .
%xgra:          1|0|ROOT 2|1|PUNCT
*CHI:           two busses .
%mor:           det:num|two n|buss-PL .
%xgra:          1|2|QUANT 2|0|ROOT 3|2|PUNCT
*CHI:           three busses .
%mor:           det:num|three n|buss-PL .
%xgra:          1|2|QUANT 2|0|ROOT 3|2|PUNCT
```

Experimental Methods

- Online methodologies
  - Reading time studies: measure relative processing difficulties
  - Eye-tracking studies: Monitor gaze as people hear a spoken utterance; anticipatory eye-movements reflect interpretation
  - Visual world paradigm: monitor subjects' eye movements to visual stimuli as they listen to an unfolding utterance
- Offline methodologies
  - Preferential looking studies: monitor infants' preferences of certain scene depictions based on linguistic stimuli
  - Act-out scenarios: describe an event and ask the child to act it out using a set of toys and objects
  - Elicitation tasks: persuade the child to describe an event or action

Reading Times

- Reading the whole sentence

`The man held at the station was innocent`

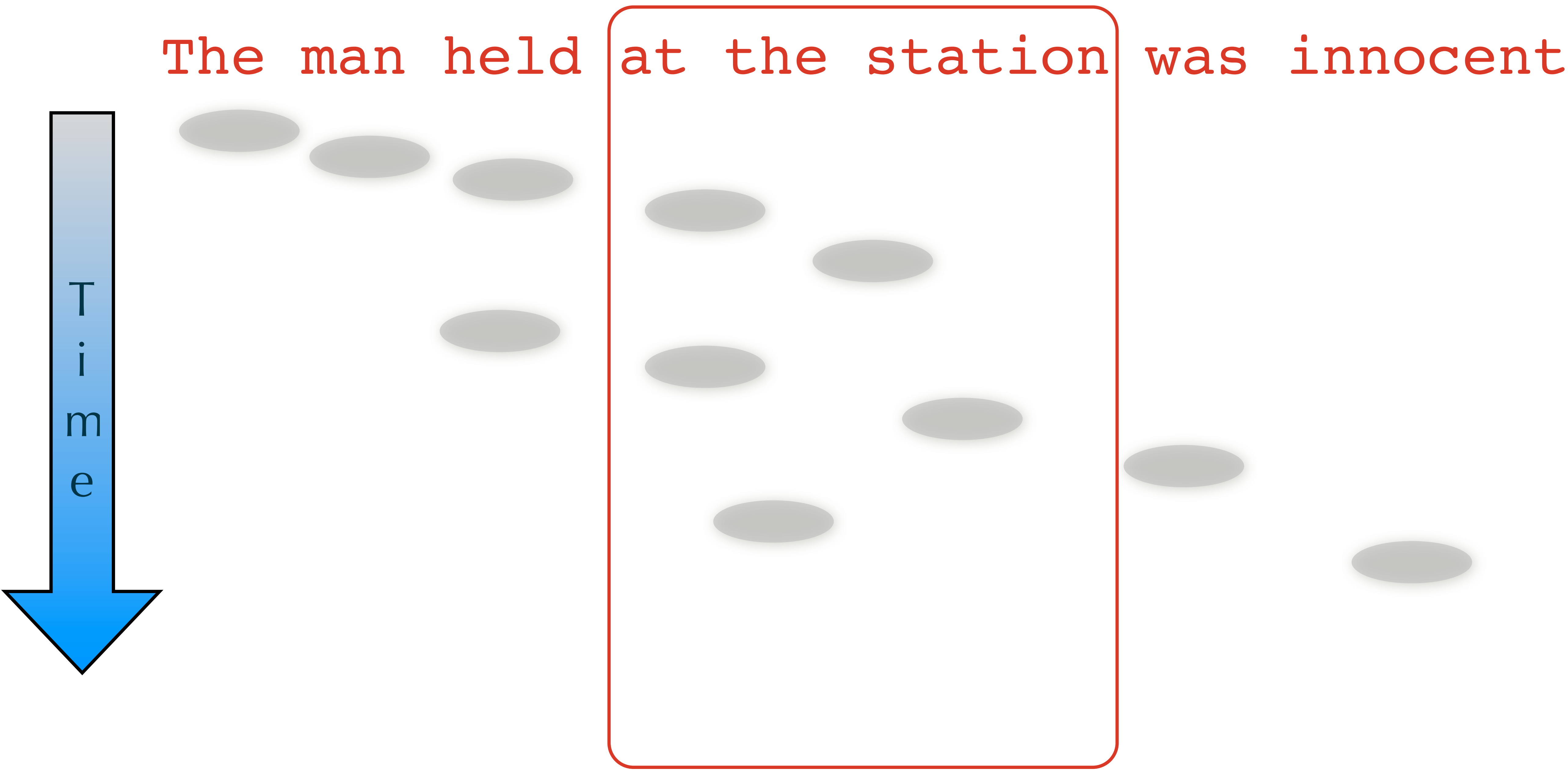
- Self-paced reading, central presentation

[illegible]

- Self-paced reading, moving window

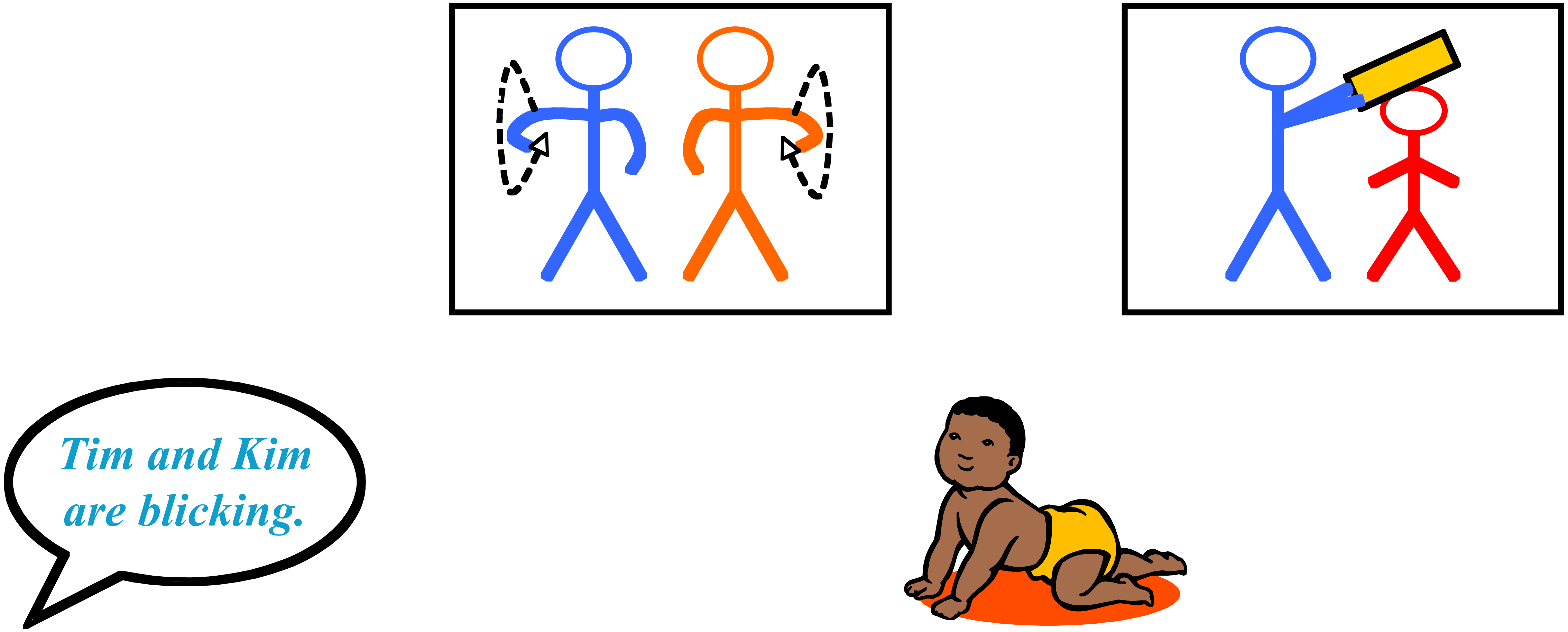
`The man held at the station was innocent`

Eye-tracking



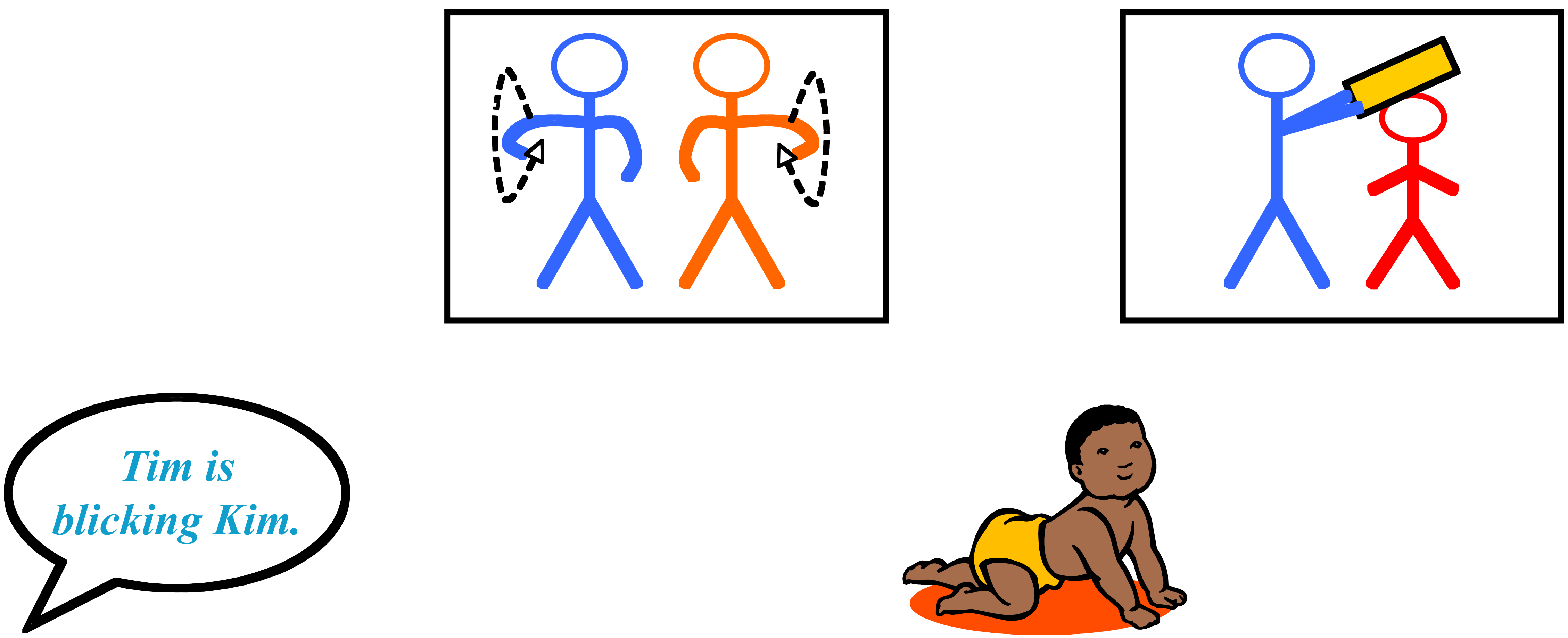
Preferential-looking Studies

- Monitor infants' preference of visual stimuli based on linguistic stimuli



Preferential-looking Studies

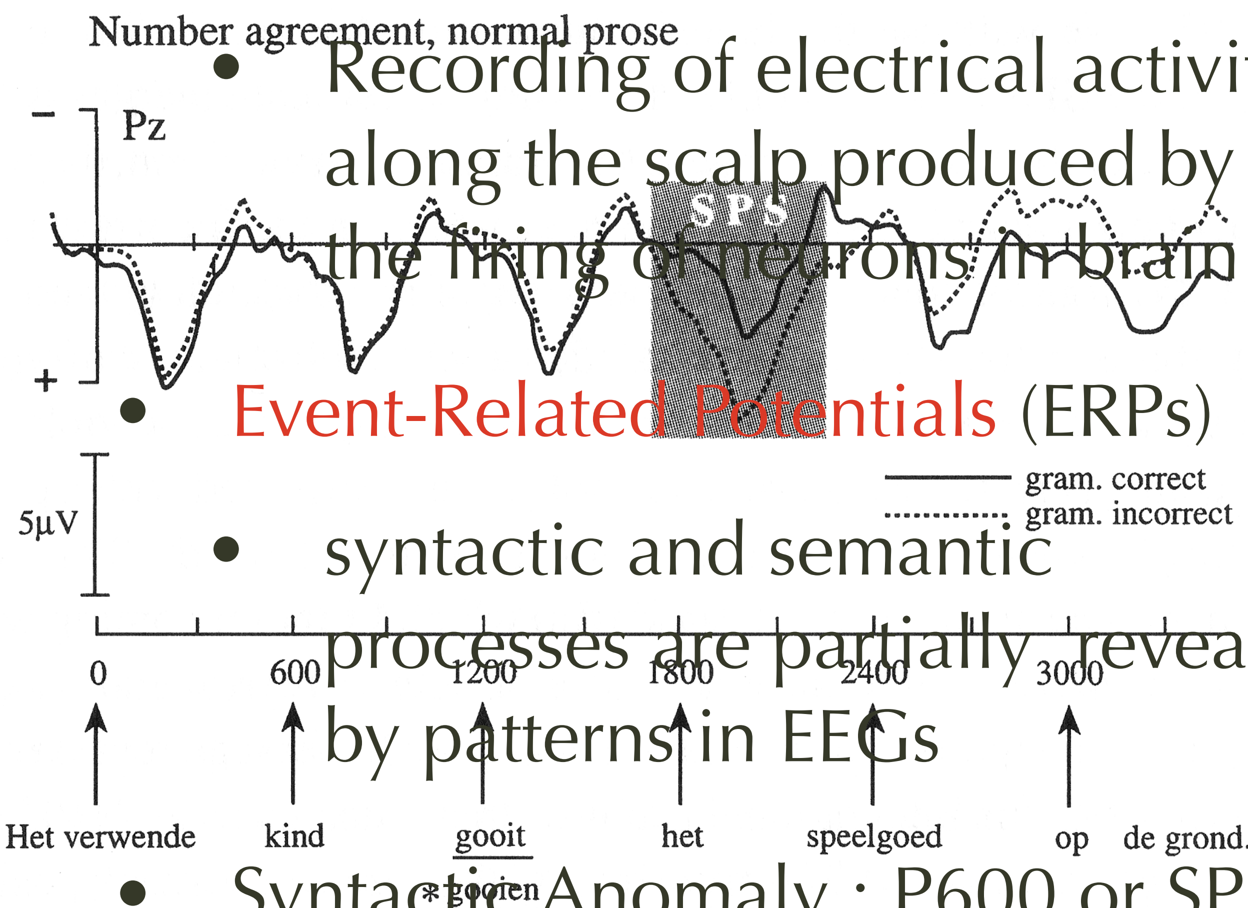
- Monitor infants' preference of visual stimuli based on linguistic stimuli

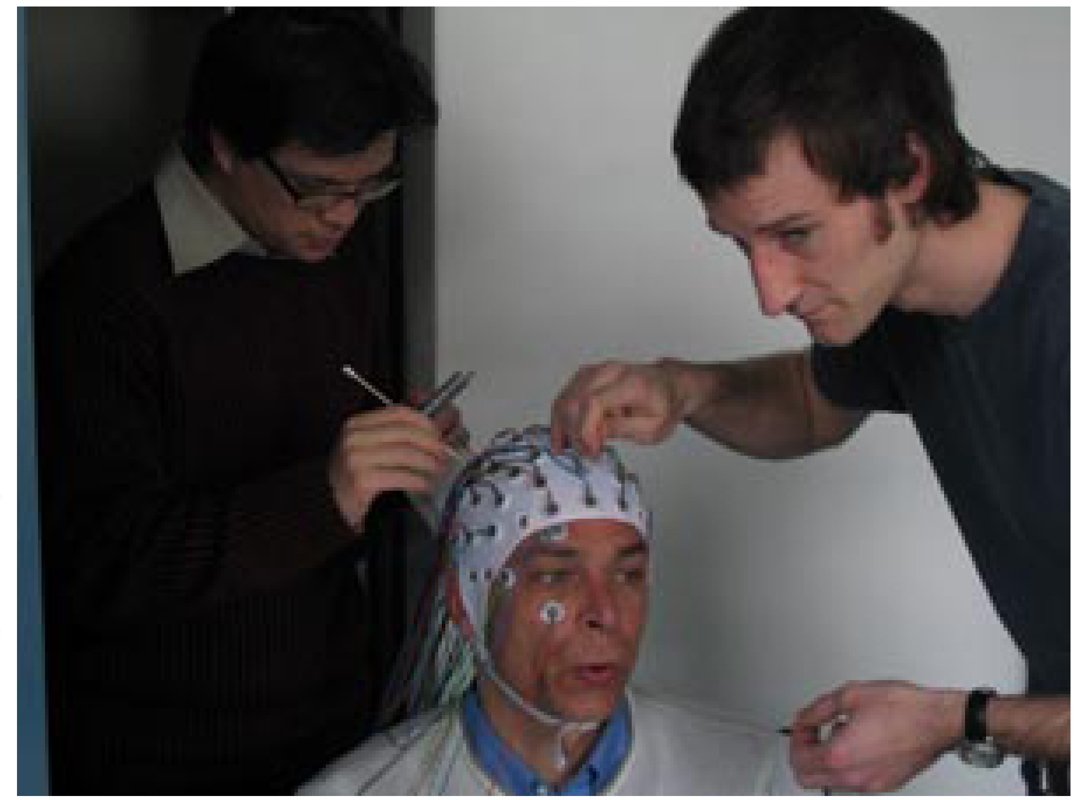


Neuroscientific Methods

Syntactic and semantic processes are partially revealed by activation patterns in brain

- Electroencephalography (EEG)
  - Recording of electrical activity along the scalp produced by the firing of neurons in brain
- Event-Related Potentials (ERPs)
  - syntactic and semantic processes are partially revealed by patterns in EEGs
- Syntactic Anomaly : P600 or SPS
- Semantic Anomaly: N400





"The spoilt child throw(s) the toy on the ground"