Cognition, Language & Communication'14

MSc Brain & Cognitive Science, UvA
track Cognitive Science

Jelle Zuidema
Institute for Logic, Language & Computation
zuidema@uva.nl

week 2: Models of sequences

Recap: Things to investigate

- How does language relate to human and non-human communication? What are its defining features?
- What is the relation between language and other cognitive domains? Is language involved in other uniquely human skills for reasoning, music, mathematics, consciousness, … ?

Recap: Unique "Design Features"?

- Displacement
- Compositionality
- Arbitrariness
- Cultural transmission
- Discreteness
- Stimulus freedom
- Duality of Patterning
- Open-endedness, Recursion

Recap: Unique “Design Features”?

- ~~Displacement~~
- ~~Compositionality~~
- ~~Arbitrariness~~
- ~~Cultural transmission~~
- ~~Discreteness~~
- ~~Stimulus freedom~~
- Duality of Patterning
- Open-endedness, Recursion

Recap: Human language:

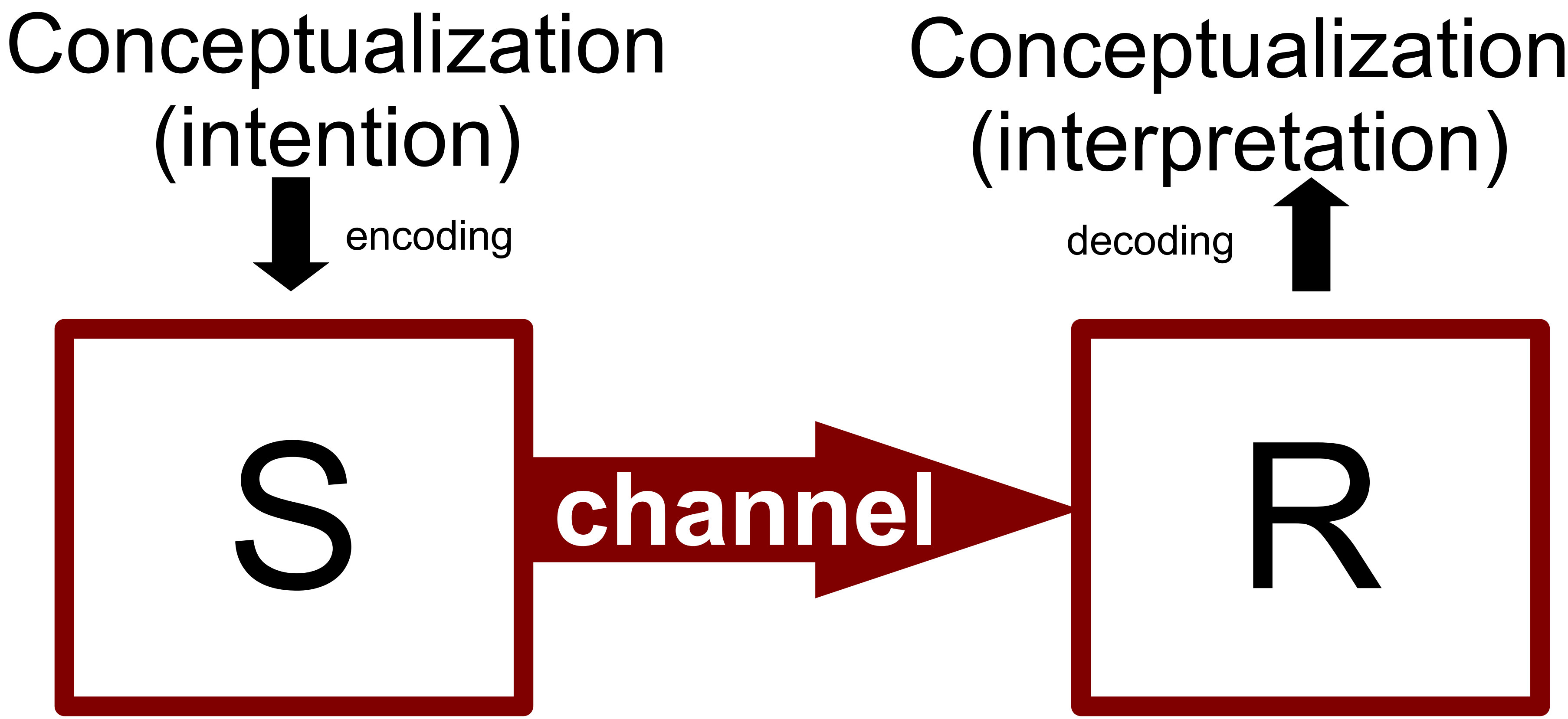
- Is an extremely complex and varied phenomenon;
- Orders of magnitude more complex than any animal communication system discovered so far;
- Requires extensive memory and sophisticated computations to be produced, interpreted and learned.

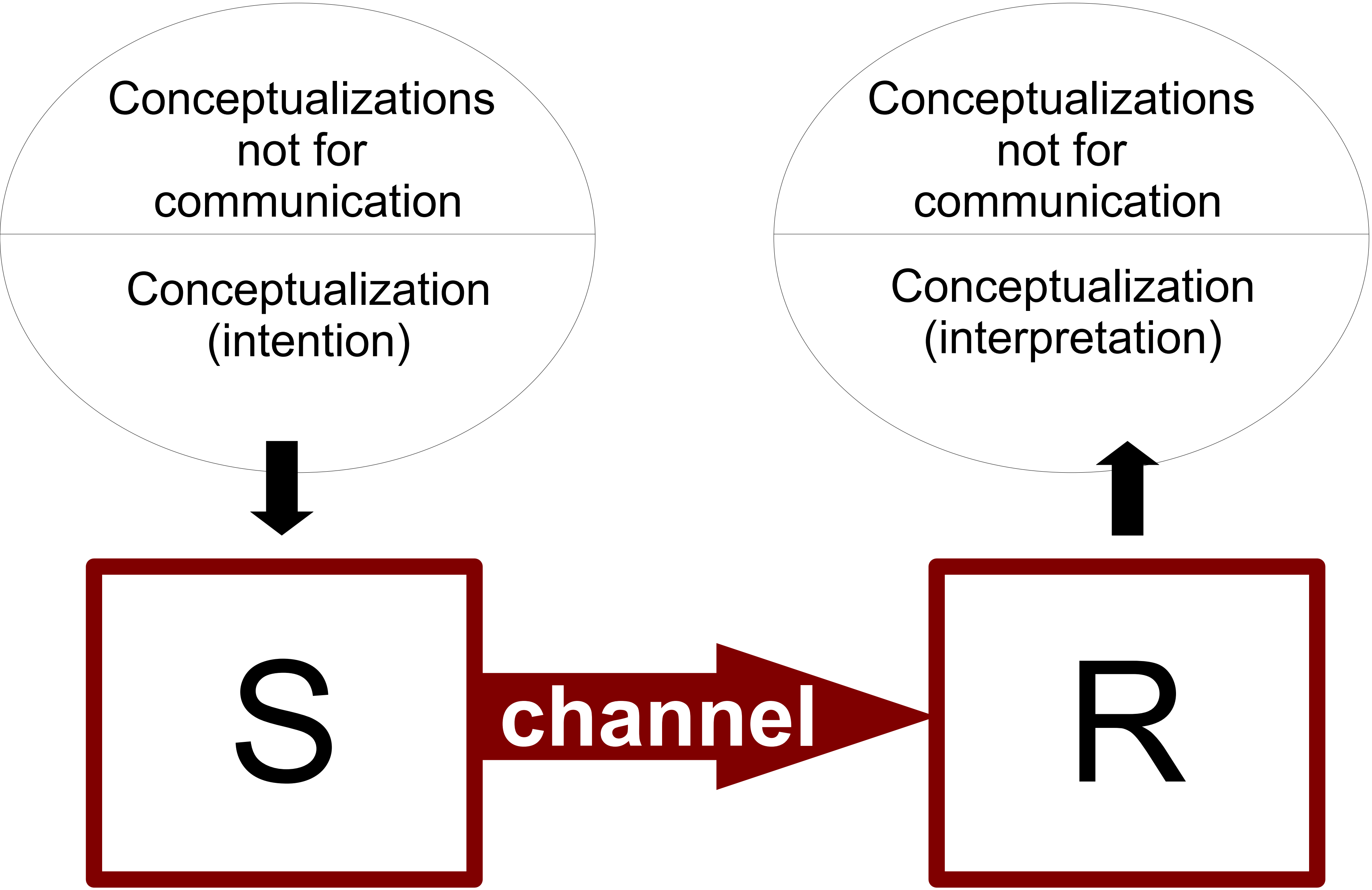
Hockett's design features do not address what human language & animal communication are *for*

Other differences between language and animal communication might lie in the function of language/communication

How can we think systematically about function?

Shannon's model

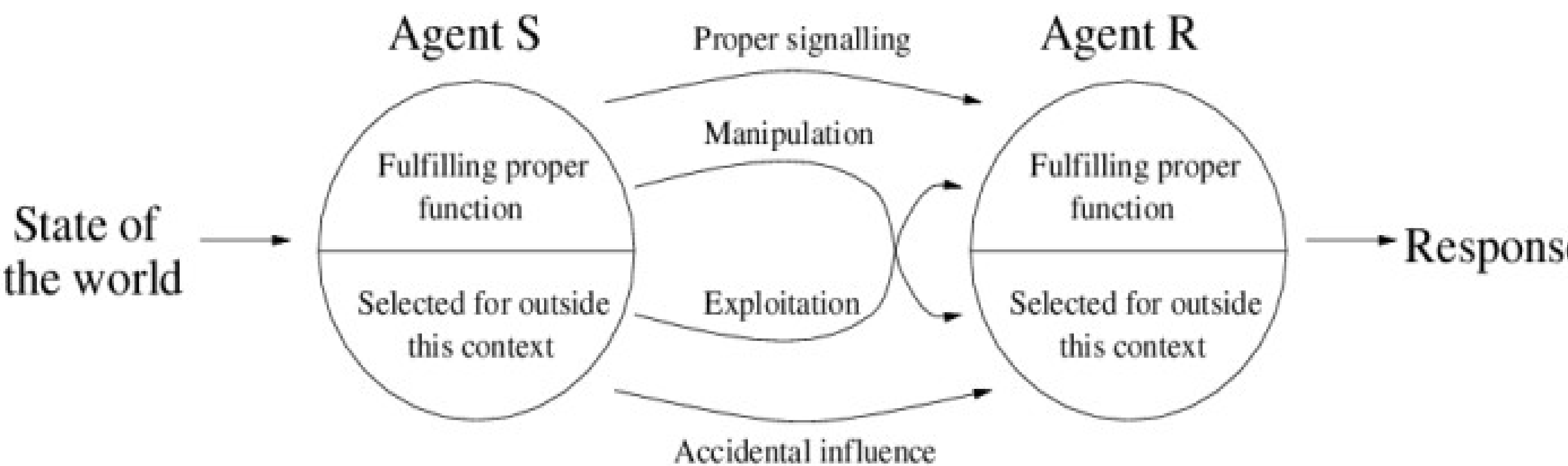




→ Non-human animals might have rich conceptual representations that are for some reason not accessible for communication (Jackendoff, 2002).

The animal behavior perspective

Millikan (ref Noble, 1998)



- Accidental influence: e.g., pig scares mouse
- Exploitation: e.g., cheetah catches injured gazelle
- Manipulation: e.g., broken wing display
- Proper signalling: e.g., bee dance

Maynard-Smith & Harper'03

- Cue: A feature of the world, animate or inanimate, that can be used by an animal as a guide for future action.
- Signal: an act or structure that alters the behaviour of another organism, which evolved because of that effect and which is effective because the receiver's response has also evolved.
- Ritualization: the evolutionary process whereby a cue maybe converted into a signal

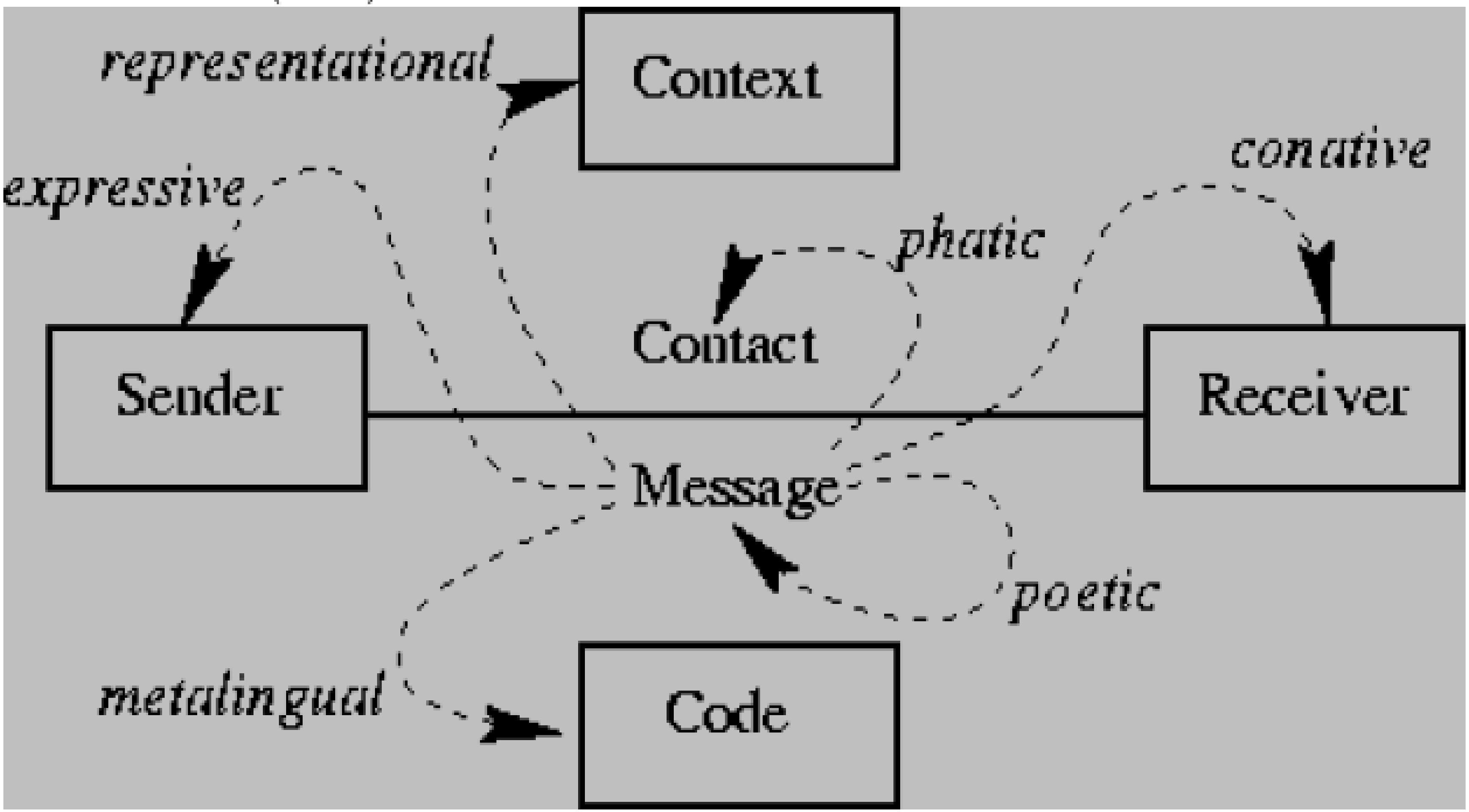
- Cue
- Ritualization
- Signal
- Sign
- Icon

Maynard-Smith & Harper'03

- The problem of reliability: what maintains the honesty of signals?
- Three possibilities:
  - Index: a signal that cannot be faked because its intensity is physically connected to the quality being signalled.
  - Common interests
  - Handicap principle
- Cost: loss of fitness resulting from making a signal, which includes:
  - efficacy cost: the cost needed to ensure that the information can be reliably perceived
  - strategic cost: cost needed, by the handicap principle (Zahavi, 1975), to ensure honesty

The linguistics perspective

Jakobson (1967)



- Expressive, e.g., *ouch!*
- Representational, e.g., room 2.02 is over here
- Phatic, e.g., *how are you?*
- Conative, e.g., imperatives
- Poetic, e.g., *absence of evidence*
- Metalingual, e.g. definitions

Intentionality

- Grice (1957): in *meaningful* communication the signaller has:
  - the intention the influence the recipient's behaviour
  - the intention for the recipient to recognise this intention

Dennett's levels of intentionality

(Dennett, 1983)

- zero-order: no mental states (such as beliefs and desires)
- first-order: sender has beliefs and desires, but no beliefs and desires about the mental states of others
- second-order: beliefs and desires about the mental states of others
- third-order: x wants y to believe that x believes he is all alone
- …

→ Perhaps nonhuman animal communication is limited to first-order intentionality (Fitch, 2010; Cheney & Seyfarth, 1997)

Crockford et al., 2012, Current Biology

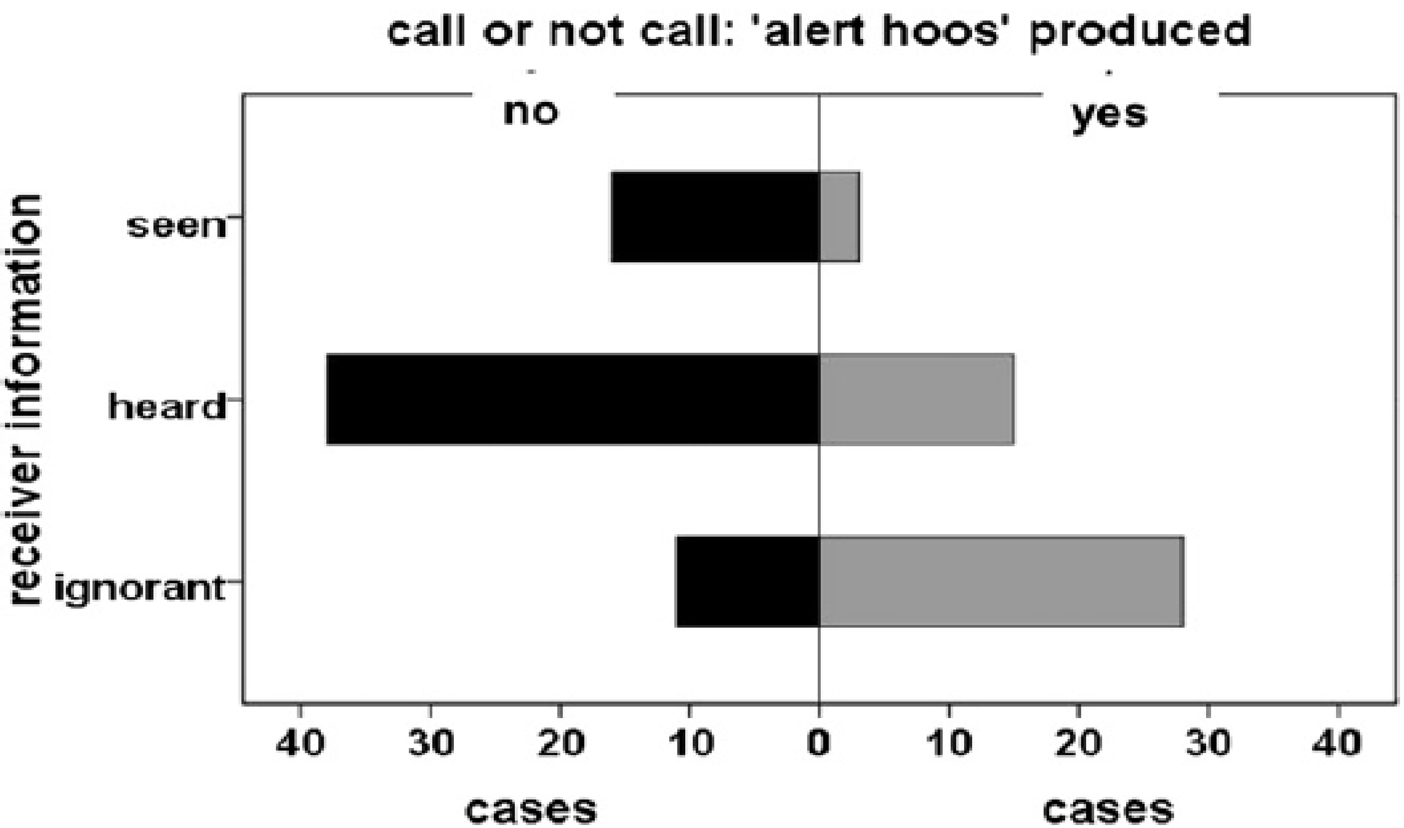


Figure 1. Influence of Receiver Information on Subjects' Likelihood to Emit Alert Hoos upon Seeing the Snake Model

Black indicates no alarm calls produced; gray indicates at least one alarm call produced. "Receiver information" indicates receiver ignorance or knowledge from the perspective of the subject, divided into the following three categories. "Seen" indicates knowledgeable receivers: the subject had seen all receivers see the snake model. "Heard" indicates partially knowledgeable receivers: the subject had heard an alarm call when all receivers were within 50 m of the snake model but could not have seen all receivers see the snake model. "Ignorant" indicates that the subject could not have seen all receivers see the snake and had not heard an alert hoo when all current receivers were within earshot (50 m) of the alert hoo.

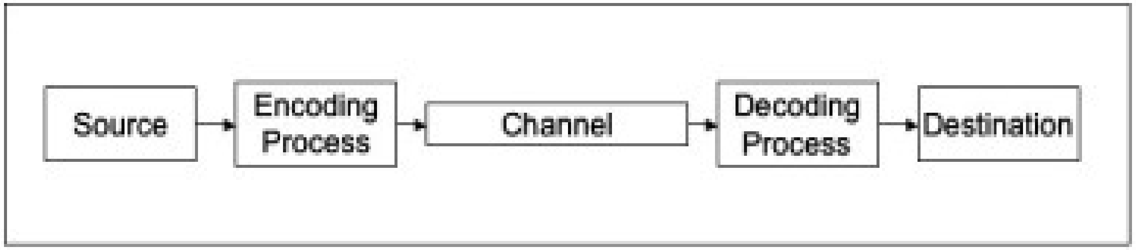
Claude Shannon: the engineering perspective

- 1916-2001
- MSc 1937: Boolean algebra in computers
- PhD 1940: Population genetics
- 1948 Information Theory
- Mechanical mouse, Rocket powered flying discs, “Ultimate Machine”

Shannon / information theory

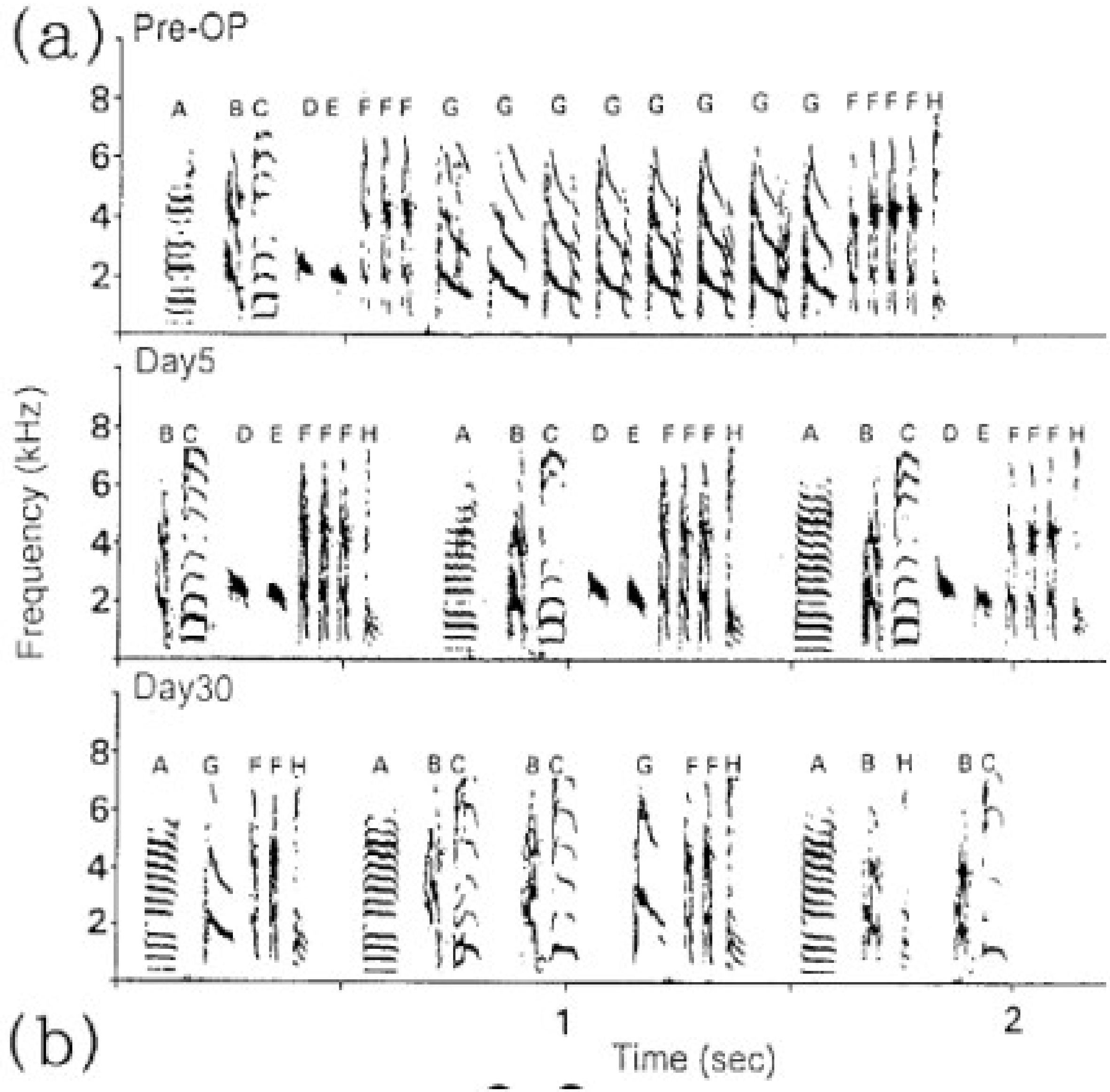
(Weaver, 1949)

- Three levels of analysis:
  - Technical level
  - Semantic level
  - Effectiveness level
- At the technical level, the content of communicative act is irrelevant; the source is viewed as a stochastic process;
- Shannon's concept of information: reduction in uncertainty about the source;
- (Note: a subjectivist interpretation of probabilities)



Markov models

- Shannon wants to consider different 'sources'; needs models that define probabilities over sequences: Markov models

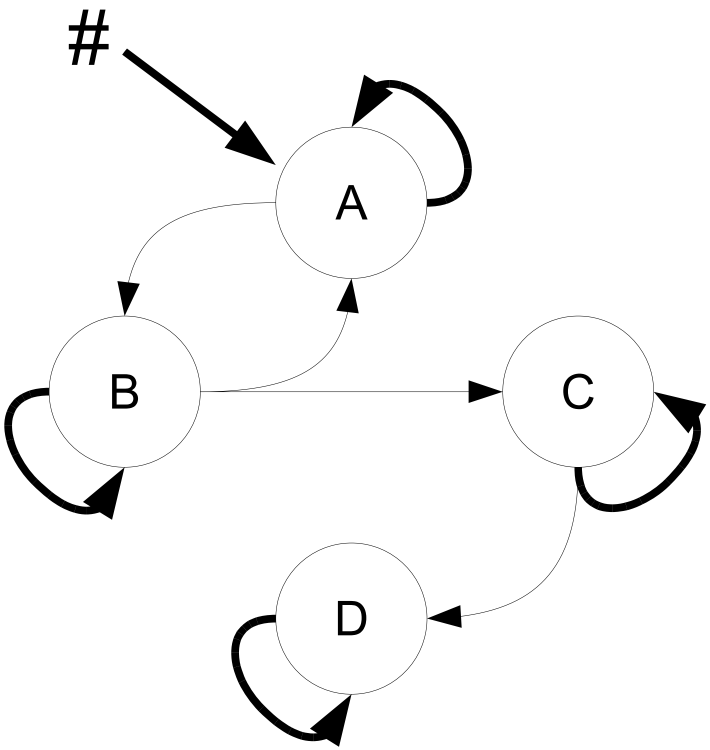

(a)
Pre-OP
A BC DE FFF G G G G G G G G FFFFH
Day5
BC D E FFFH A BC D E FFFH A B C D E FFFH
Day30
A G FFH A BC BC G FFH A B H BC
Frequency (kHz)
2
4
6
8
1
2
Time (sec)
(b)

Markov models

- Shannon wants to consider different 'sources'; needs models that define probabilities over sequences: Markov models
- Markov property: the probability of the next event is only dependent on the current state

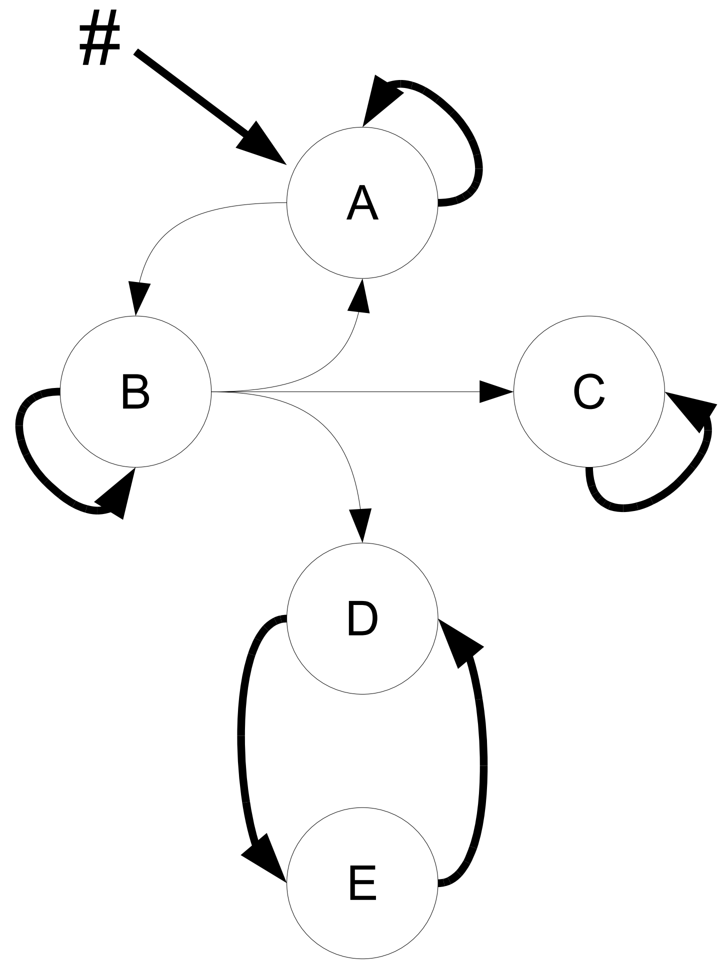
Transitional Probabilities

	A	B	C	D
#	1	0	0	0
A	0.8	0.2	0	0
B	0.1	0.8	0.1	0
C	0	0	0.8	0.2
D	0	0	0	1



D is a “sink” (point attractor)

Transitional Probabilities



This system has multiple attractors

C is a “sink” (point attractor)

D-E is a “limit cycle”

Shannon 1948

Approximations of English based on character transition probabilities:

0-order: XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD

1st-order: OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL

2nd-order: ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

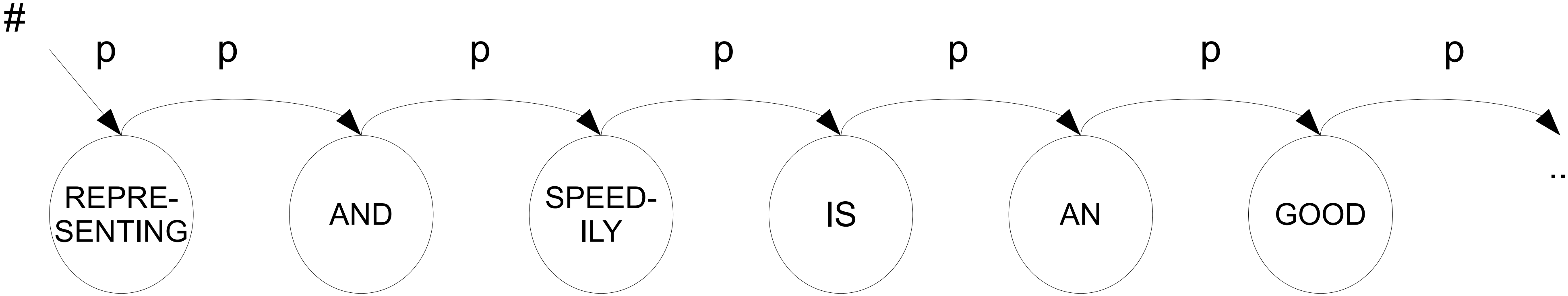
3d-order: IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PON-DENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

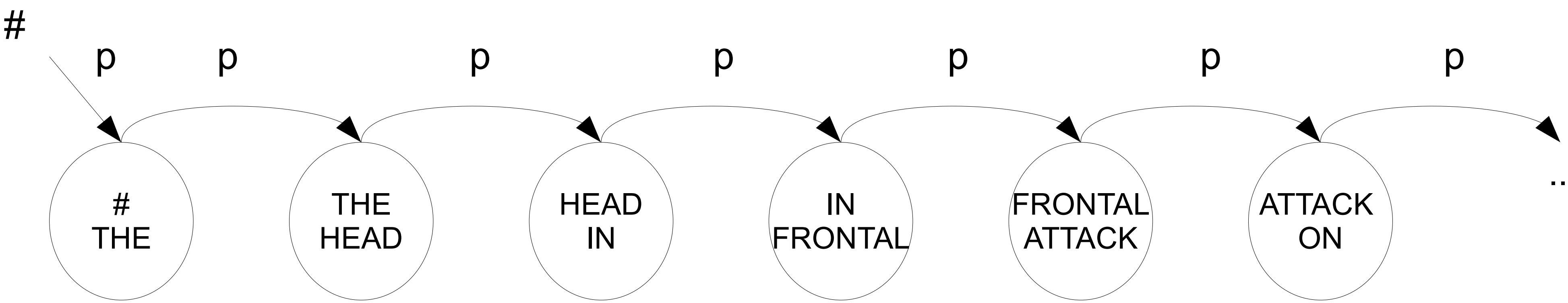
Shannon 1948

Approximations of English based on word transition probabilities:

1st-order: REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE

2nd-order: THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED



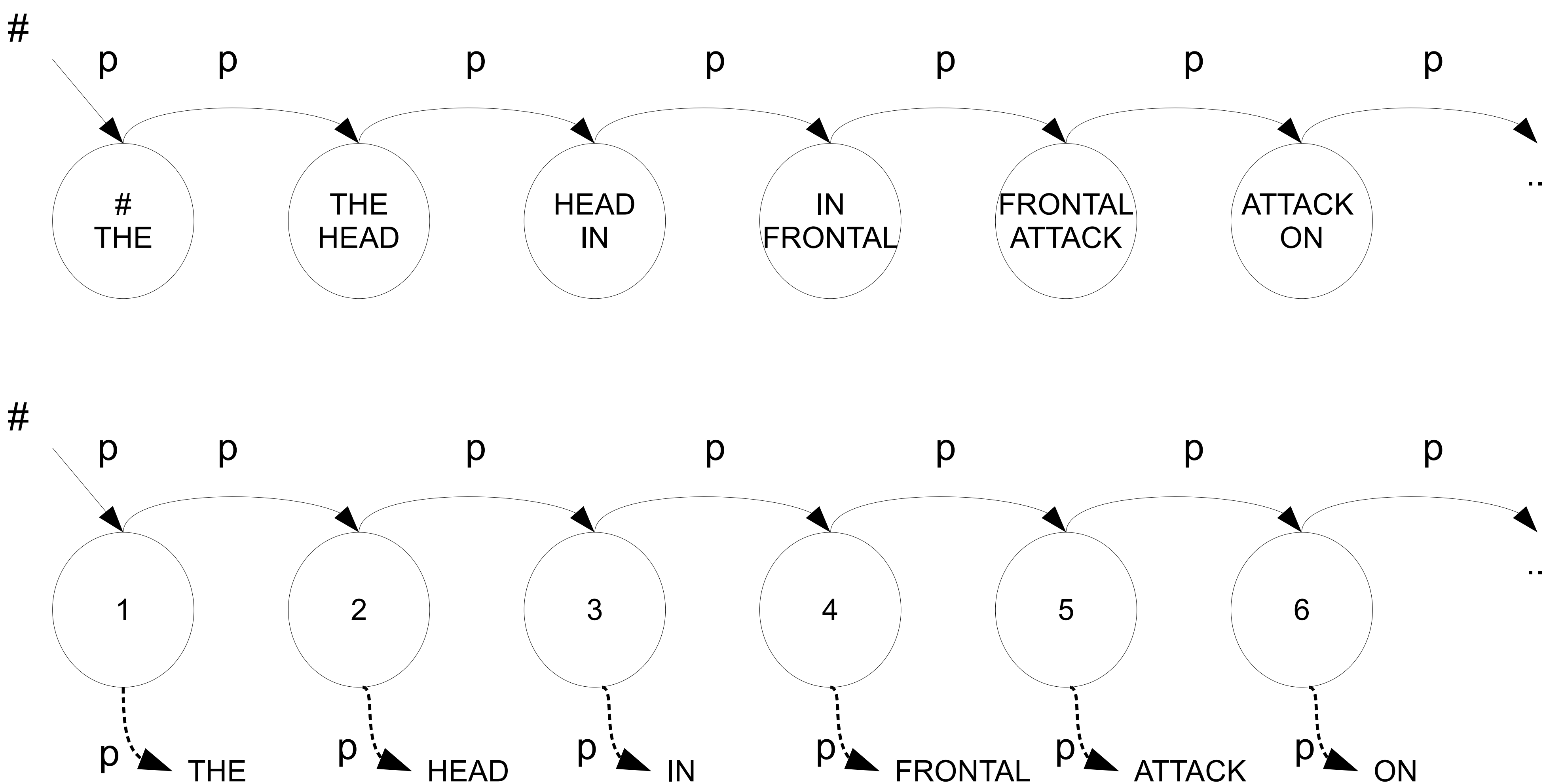


- Markov order 1: the probability of the next state depends only on the current state
- Markov order 0: the probability of the next state is independent of the current state
- Markov order n: the probability of the next state depends on the current state and the previous (n-1) states
- Equivalently: the previous (n-1) states are incorporated in the current state description!
- In the language domain, (n+1)-th order Markov models are also called ngrams!

Markov models

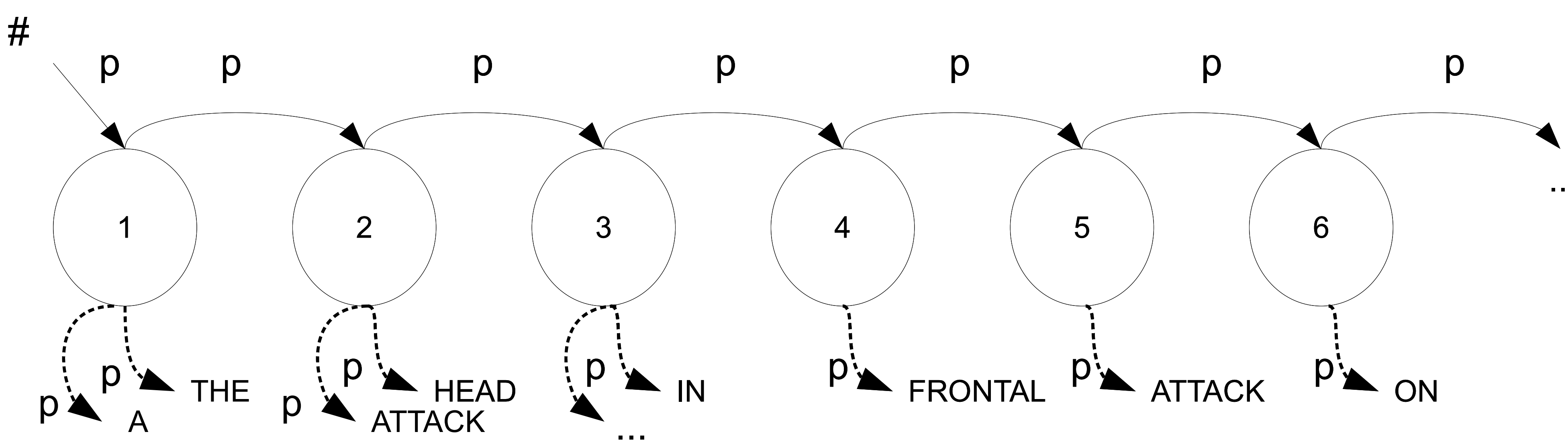
- Shannon wants to consider different 'sources'; needs models that define probabilities over sequences: Markov models
- Markov property: the probability of the next event is only dependent on the current state
- Terms:
  - (In)dependence of current state
  - Transitional probabilities, transition matrix
  - Sink / point attractor, Limit cycle
  - Markov order

Generalizing over states



Hidden Markov Model

- Finite number of hidden states
- “Transition probabilities” from state tot state
- Finite number of observable symbols
- “Emission probabilities” from hidden states to observable symbols



Computing with HMMs

- Forward algorithm:

  $P(\mathbf{o}, HMM)$

- Viterbi algorithm:

  $\text{argmax}\_\mathbf{h}\ P(\mathbf{o}|\mathbf{h}, HMM)$

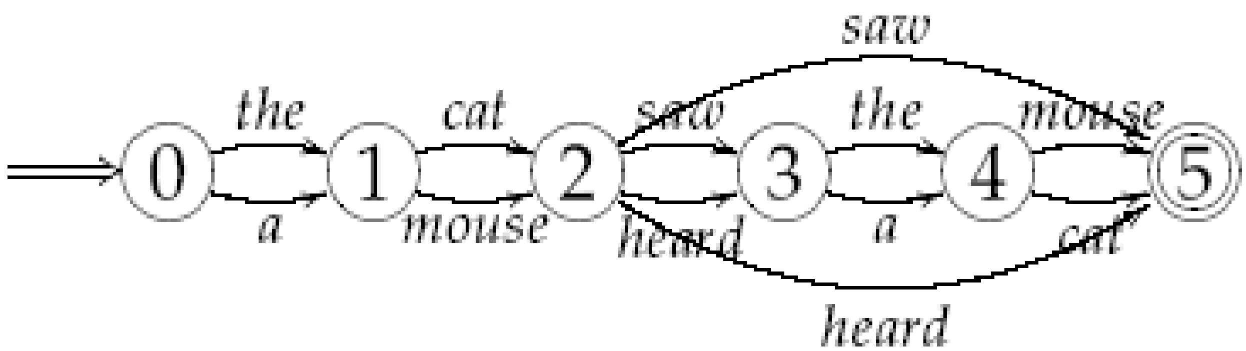
- Baum-Welch algorithm (Forward-Backward):

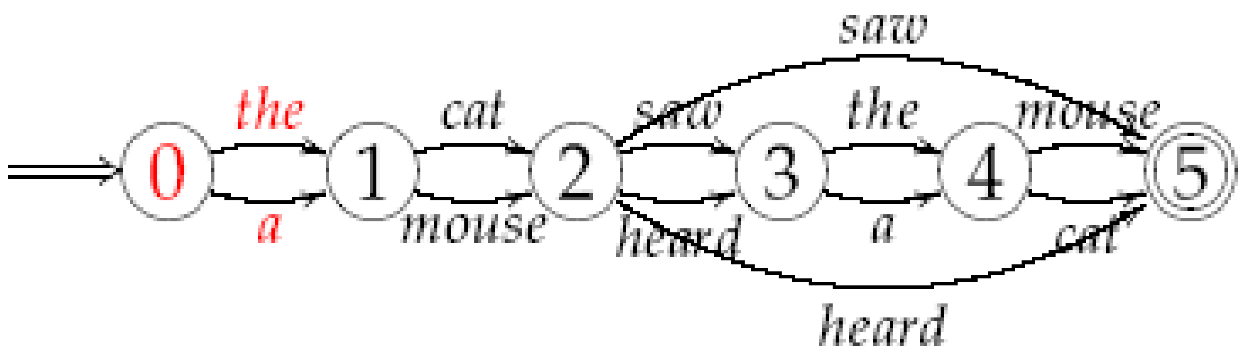
  $\text{argmax}\_HMM\ P(\mathbf{o}|HMM)$

Finite-state Automaton

- Finite number of hidden states
- Transitions between states
- Transitions labeled with observable symbols
- Ignoring the probabilities, FSA's are equivalent to HMMs.
- FSA's are also equivalent to “left-linear rewrite grammars”

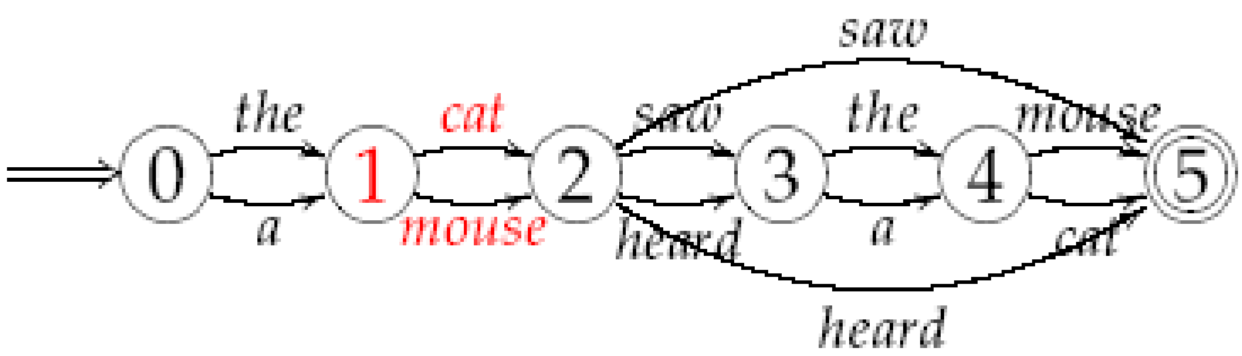
- the cat saw the mouse
- a mouse heard a cat
- the mouse heard
- a cat saw


saw
the
cat
saw
the
mouse
0
1
2
3
4
5
a
mouse
heard
a
cat
heard

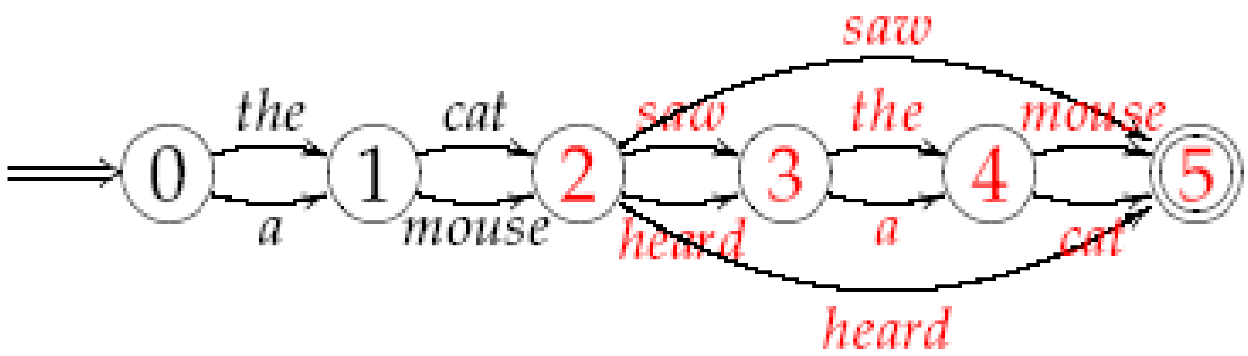


0→the 1

0→a 1



0→the 1
0→a 1
1→cat 2
1→mouse 2



0→the 1
0→a 1
1→cat 2
1→mouse 2
2→saw
2→heard
2→saw 3
2→heard 3
3→the 4
3→a 4
4→cat
4→mouse

Terms to know:

- finite-state automaton (FSA)
- hidden markov model (HMM)
- Forward algorithm:

$$P(\mathbf{o}, HMM)$$

- Viterbi algorithm:

$$\text{argmax}_{\mathbf{h}} P(\mathbf{o} | \mathbf{h}, HMM)$$

- Baum-Welch algorithm:

$$\text{argmax}_{HMM} P(\mathbf{o} | HMM)$$

FSA's are inadequare

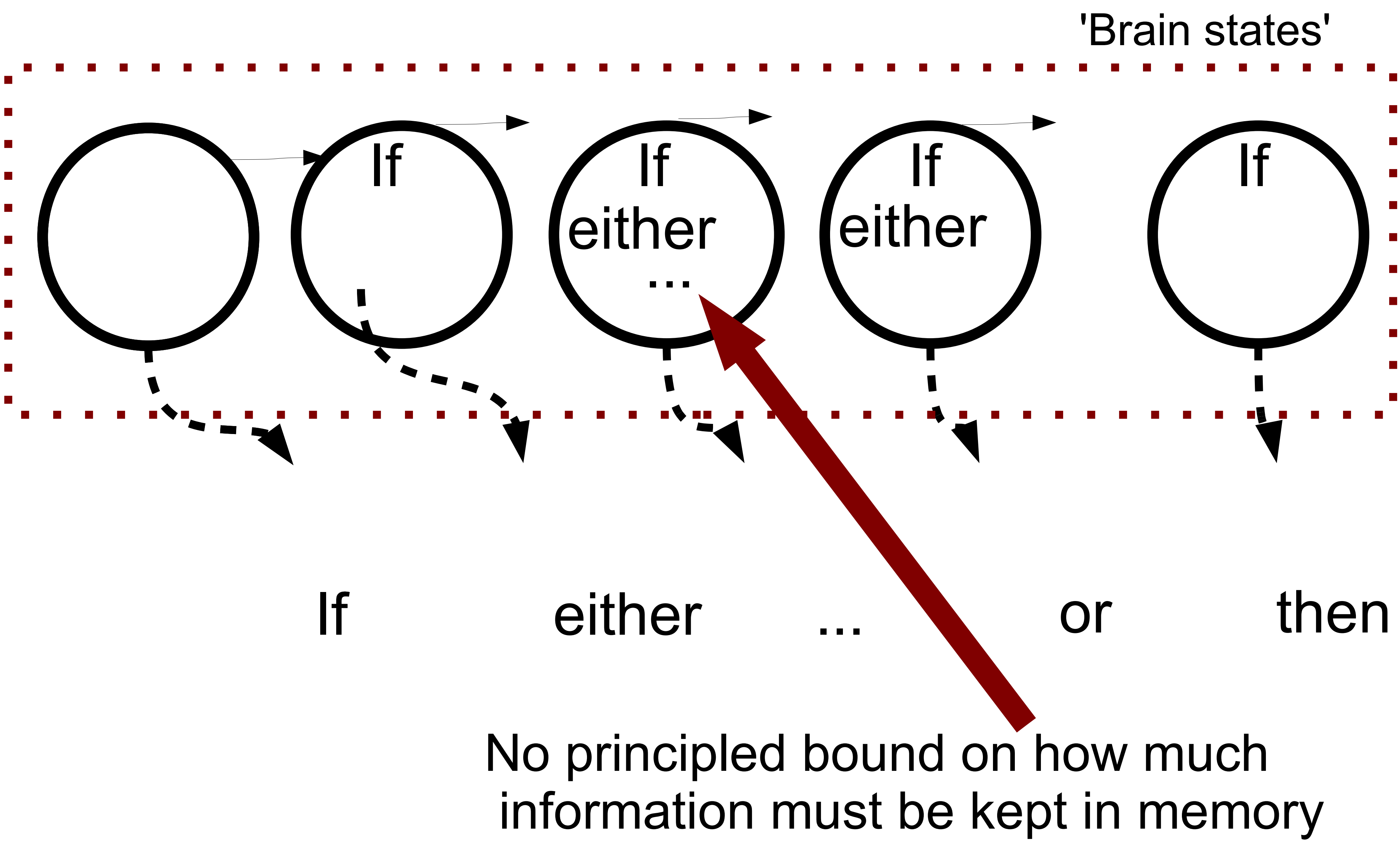
(Chomsky, 1957)

Let S1, S2, S3, S4, S5 be simple declarative sentences in English. Then also

(2) If S1, then S2.
(3) Either S3 or S4.
(4) The man who said that S5, is arriving today

are sentences of English.

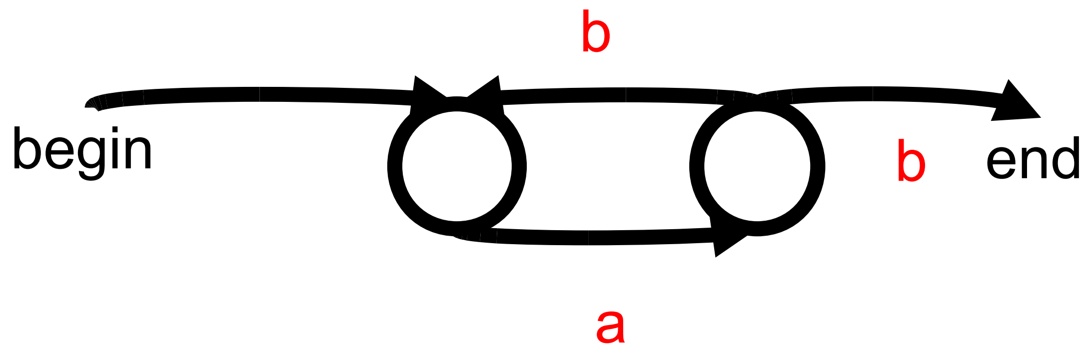
E.g., *if either you are with us or you are against us applies here, then there is nothing more to discuss*.


'Brain states'
If
If
either
...
If
either
If
If
either
...
or
then
No principled bound on how much
information must be kept in memory

Simplest example of a "finite-state language":

$(ab)^n$

E.g. ab, abab, ababab, abababab



Simplest example of a "context-free language":

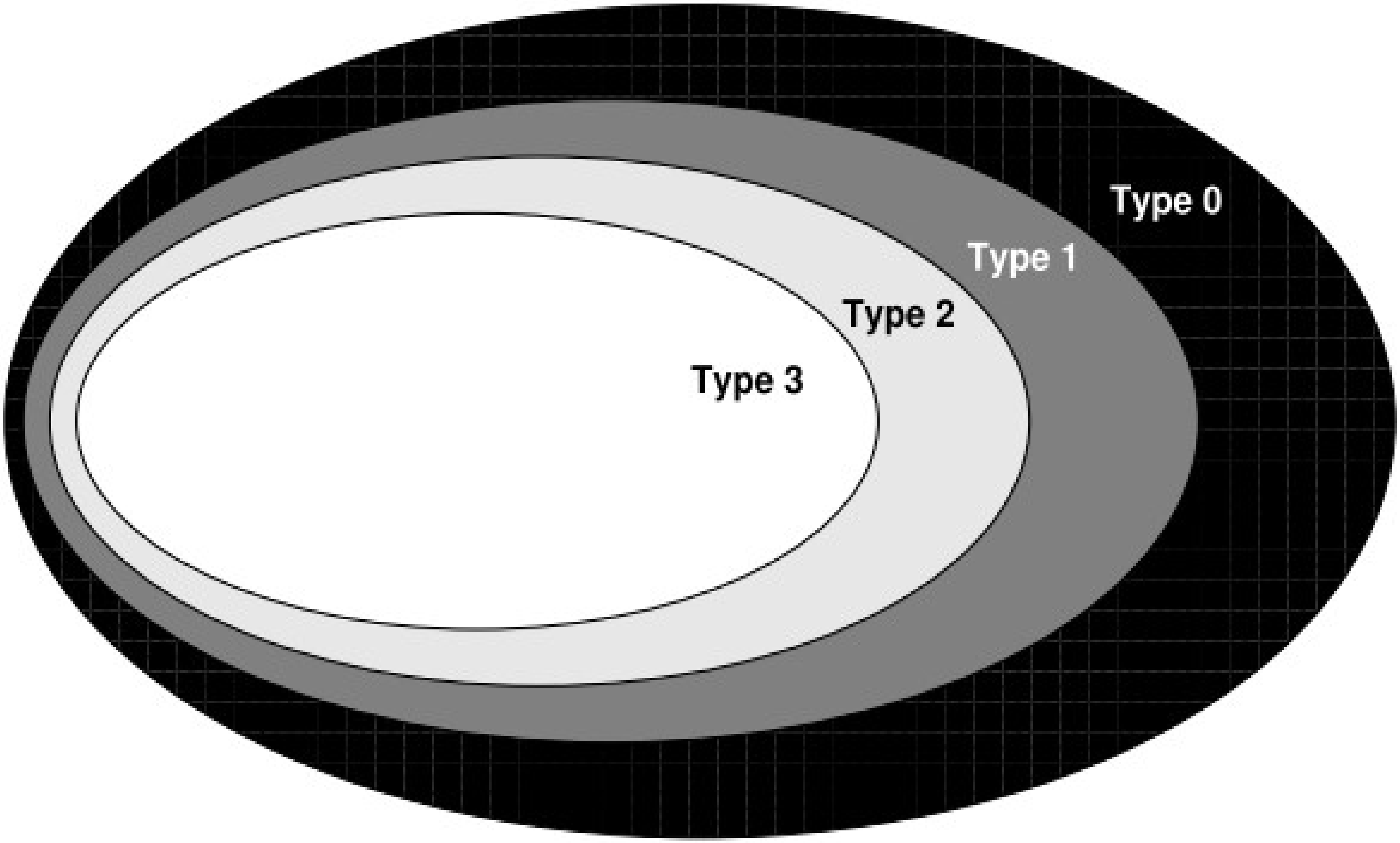
$a^n b^n$

E.g. ab, aabb, aaabbb, aaaabbbb, ...

Chomsky Hierarchy

3. Finite state grammars	$A \rightarrow a,\ A \rightarrow aB$	$(ab)^n,\ a^n b^m$
2. Context-free grammars	$A \rightarrow \gamma$	$a^n b^n$
1. Context-sensitive grammars	$\alpha A \beta \rightarrow \alpha \gamma \beta$	$a^n b^n c^n$
0. Unrestricted grammars	$\alpha \rightarrow \gamma$	$\{a^n b^m c^l \mid l = n * m\}$

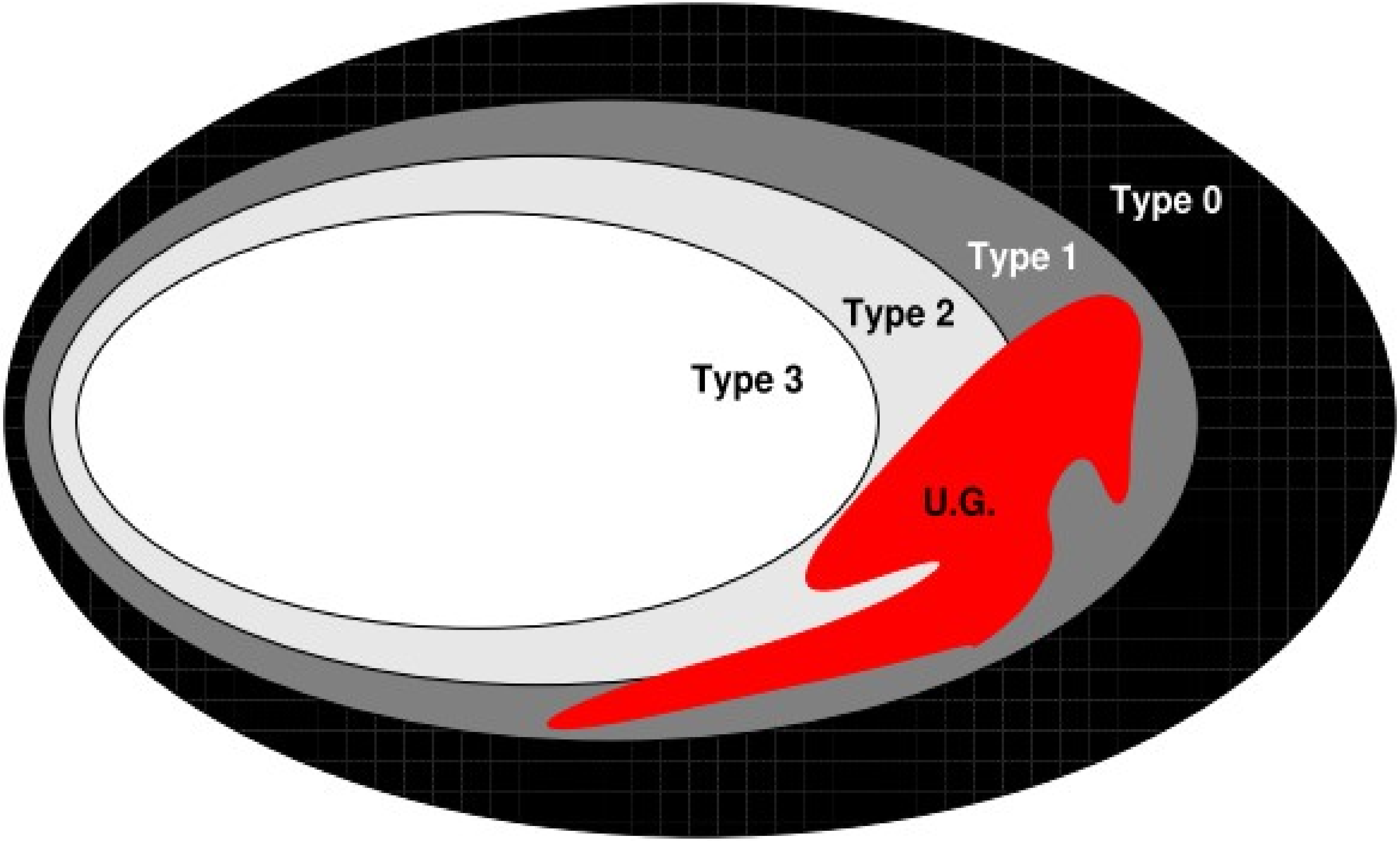
The Chomsky Hierarchy



(1) a. Gilligan claims that Blair deceived the public.
b. Gilligan claims that Campbell helped Blair deceive the public.
c. Gilligan claims that Kelly saw Campbell help Blair deceive the public. (tail recursion)

(2) a. Gilligan behaupte dass Kelly Campbell Blair das Publikum belügen helfen sah. (center embedding)
b. Gilligan beweert dat Kelly Campbell Blair het publiek zag helpen bedriegen. (crossing dependencies)

The Chomsky Hierarchy



Terms to know

- Rewrite grammars, rewrite operation
  - Production rules
  - Terminal alphabet / observable symbols
  - Nonterminal alphabet / hidden states
  - Start symbol
  - Derivation
  - Phrase-structure
- Contextfree grammars, contextfree constraint
- Push-down automaton