Infant artificial language learning and language acquisition

Gómez and Gerken (2000)

Presented by Sophia Wiedmann

Why use artificial languages for testing language acquisition?

1. Natural languages are hard to control
   – Simplicity of artificial languages allow experiments to isolate factors of interest

Why use artificial languages for testing language acquisition?

1. Natural languages are hard to control
   – Simplicity of artificial languages allow experiments to isolate factors of interest
2. Prior learning in utero
   – Newborns are already ‘biased’ towards their mother’s language
   – Perceive rhythm, melody, and intonation in utero

Experiment 1: Word Segmentation

Saffran, Aslin, and Newport (1996): Statistical Learning by eight-month-old-infants

Within-word syllables have higher *transitional probabilities* than between-word syllables:

“Pretty baby”, “Pretty baby”

→by | ba has higher transitional probability than ba | ty

Experiment 1: Word Segmentation

Saffran, Aslin, and Newport (1996): Statistical Learning by eight-month-old-infants

Within-word syllables have higher *transitional probabilities* than between-word syllables:

“Pretty baby”, “Pretty baby”

→ by | ba has higher transitional probability than ba | ty

Research question: Can infants use transitional probabilities to segment words in continuous speech?

Experiment 1: Word Segmentation

Saffran, Aslin, and Newport (1996): Statistical Learning by eight-month-old-infants

Train: Familiarized for 2 minutes with nonsense tri-syllabic words

🗣 bidakupadotigolabutupiro

Experiment 1: Word Segmentation

Saffran, Aslin, and Newport (1996): Statistical Learning by eight-month-old-infants

Train: Familiarized for 2 minutes with nonsense tri-syllabic words

🗣 bidakupadotigolabututupiro

Test: Can they discriminate new words composed of different syllables pairings?

tupiro vs tilado

bidaku vs dapiku

“Nonwords” have transition probability of 0

Experiment 1: Word Segmentation

Saffran, Aslin, and Newport (1996): Statistical Learning by eight-month-old-infants

Train: Familiarized for 2 minutes with nonsense tri-syllabic words
🗣 bidakupadotigolabutupiro

Test: Can they discriminate new words composed of different syllable pairings?
tupiro vs tilado
bidaku vs dapiku

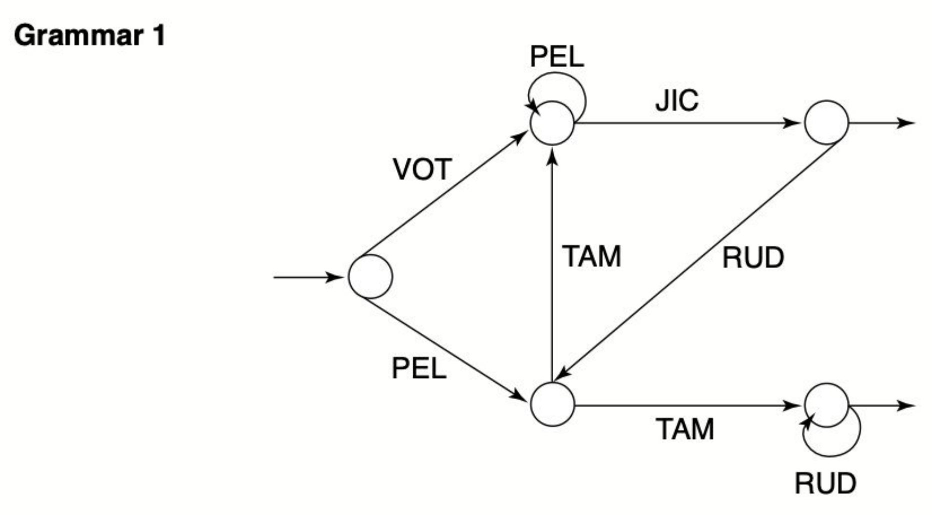
“Nonwords” have transition probability of 0
Infants showed differential attention to familiar vs. unfamiliar words → demonstrate sensitivity to transitional probabilities

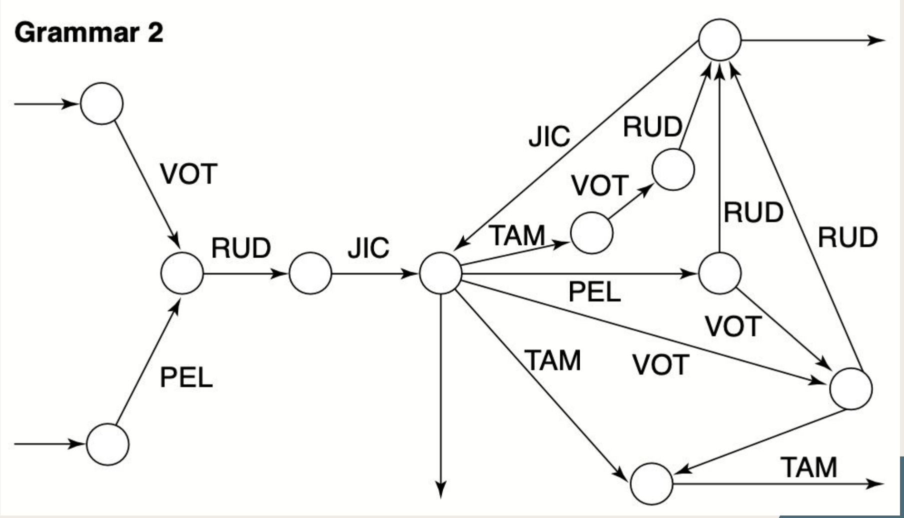
Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

How do learners acquire legal ordering of words in sentences?

Babies exposed to strings formed by 1 of 2 grammars:

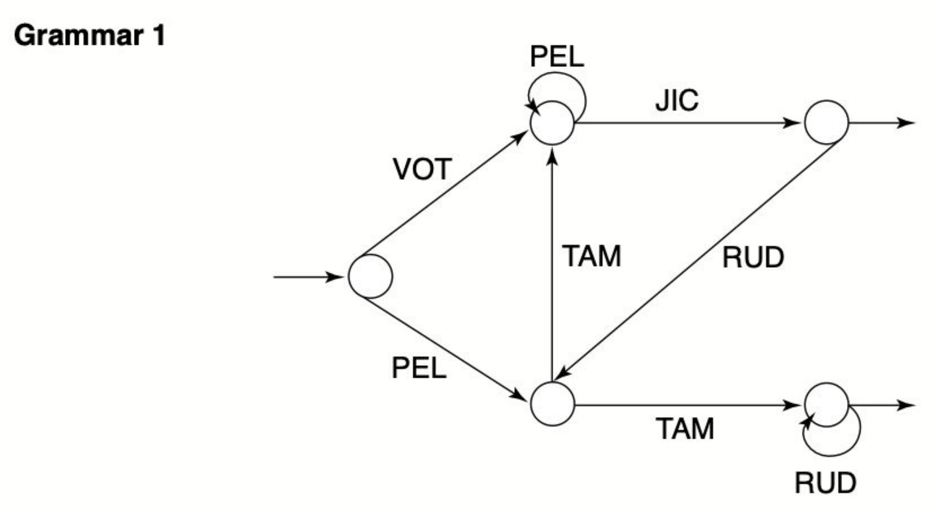


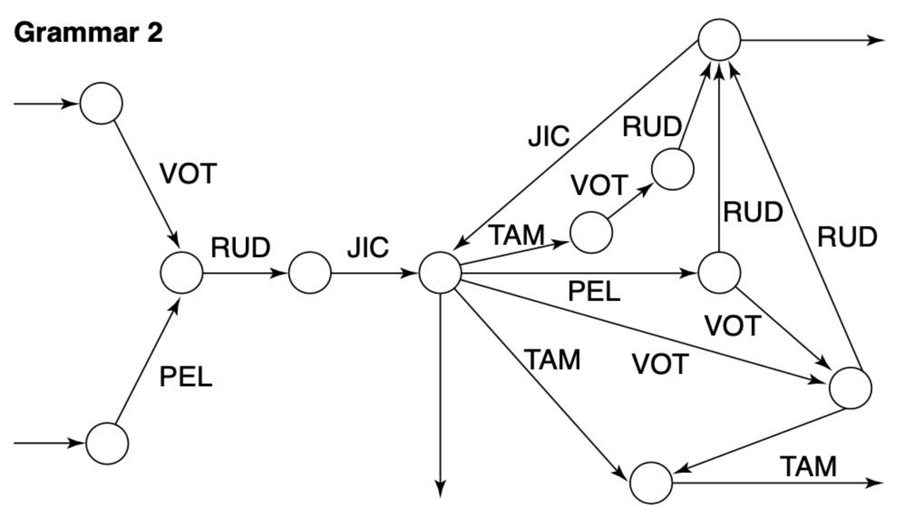


Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Consist of same vocabulary: VOT, PEL, JIC, TAM, RUD

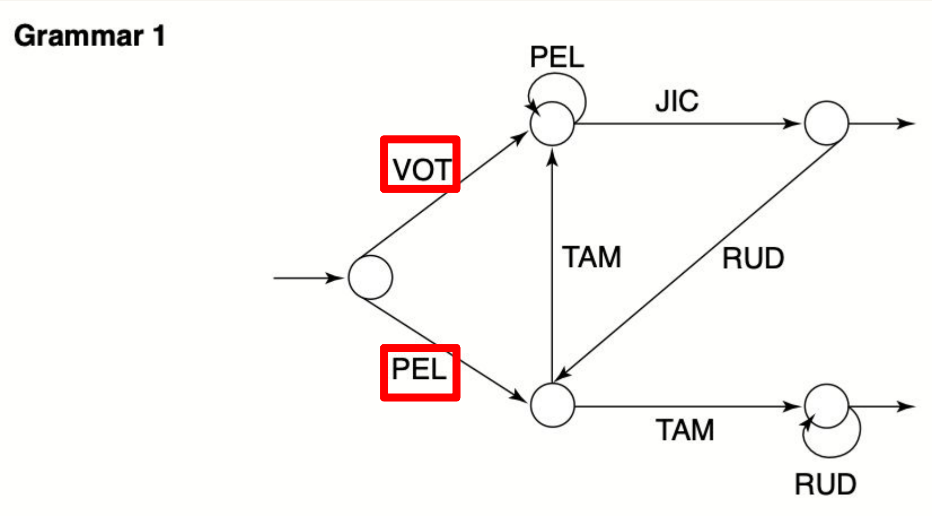


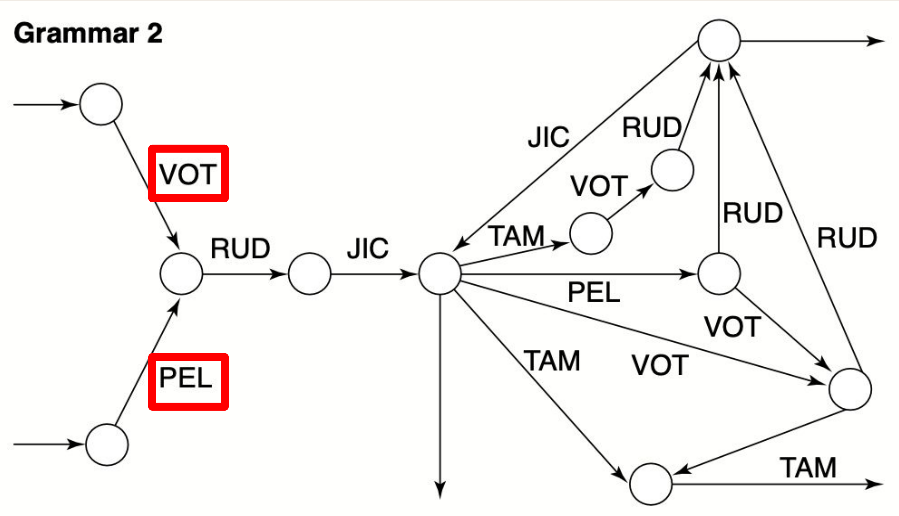


Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Start with same words

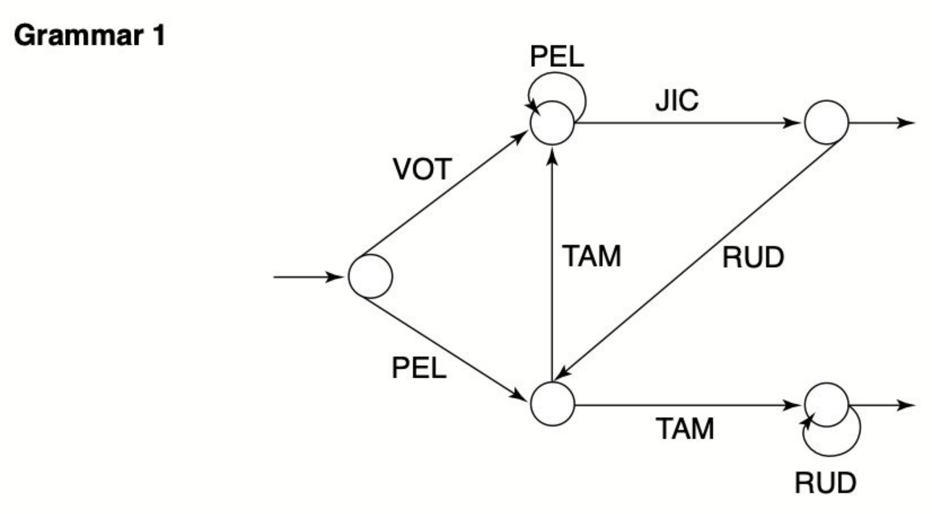


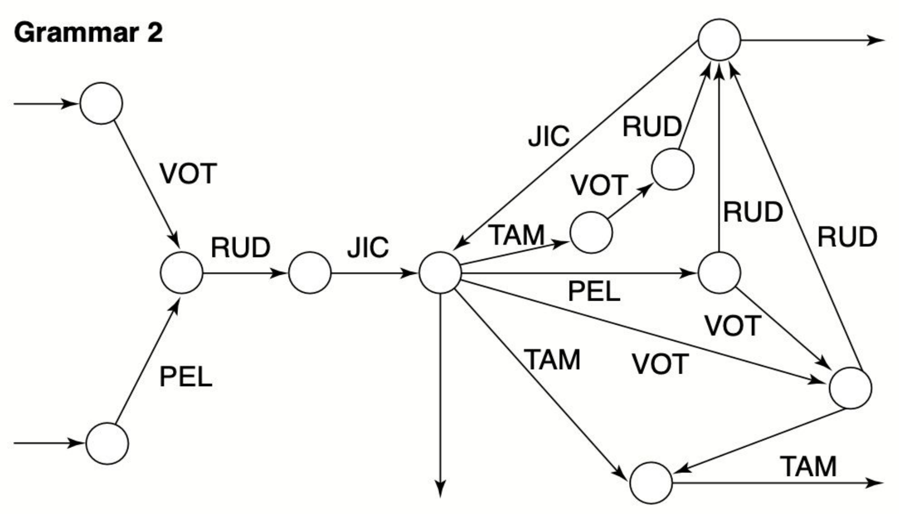


Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Only differ in word order





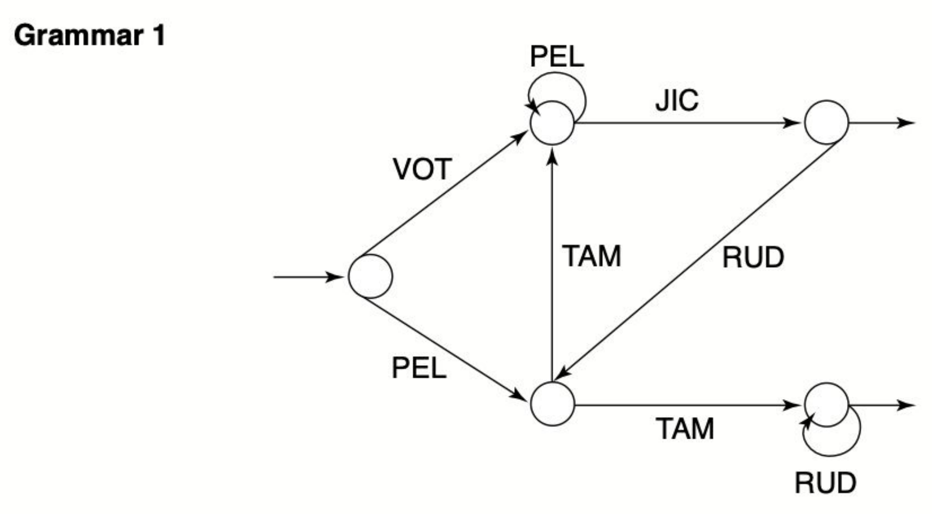
VOT-PEL-JIC-RUD-TAM-RUD

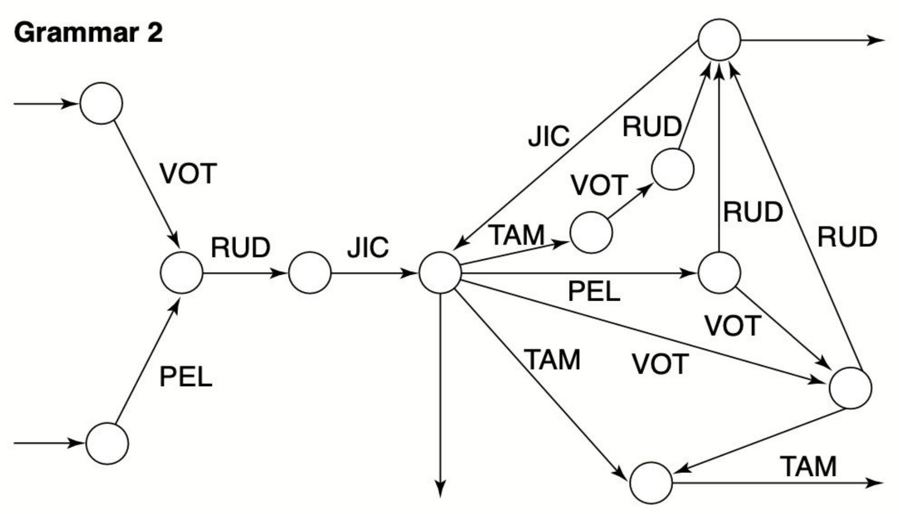
VOT-RUD-JIC-TAM-VOT-RUD

Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Train: Listen to strings in either grammar for 50-127 seconds

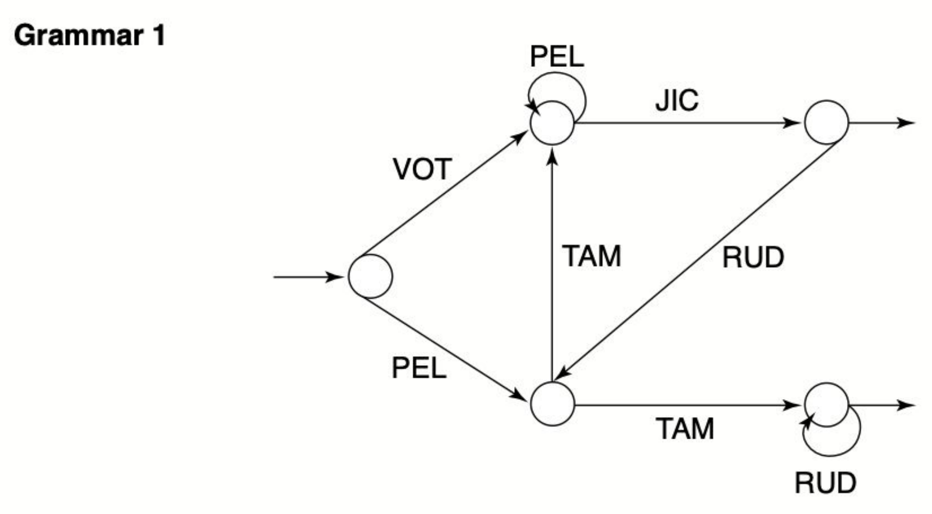


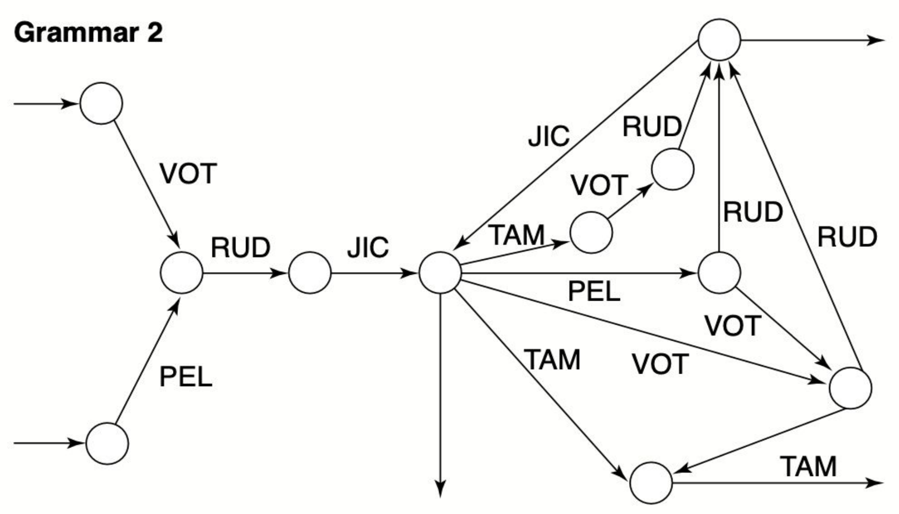


Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Test: Listen to strings in both grammars and see if they can differentiate



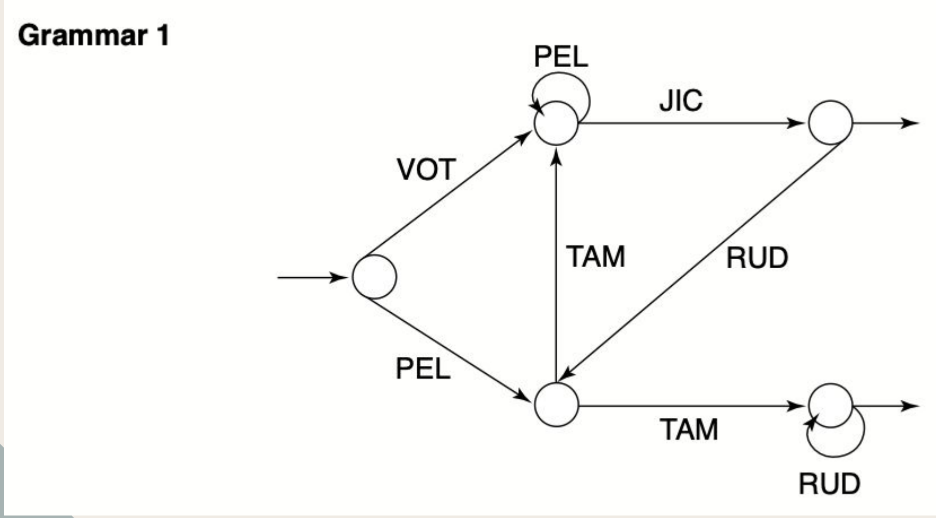


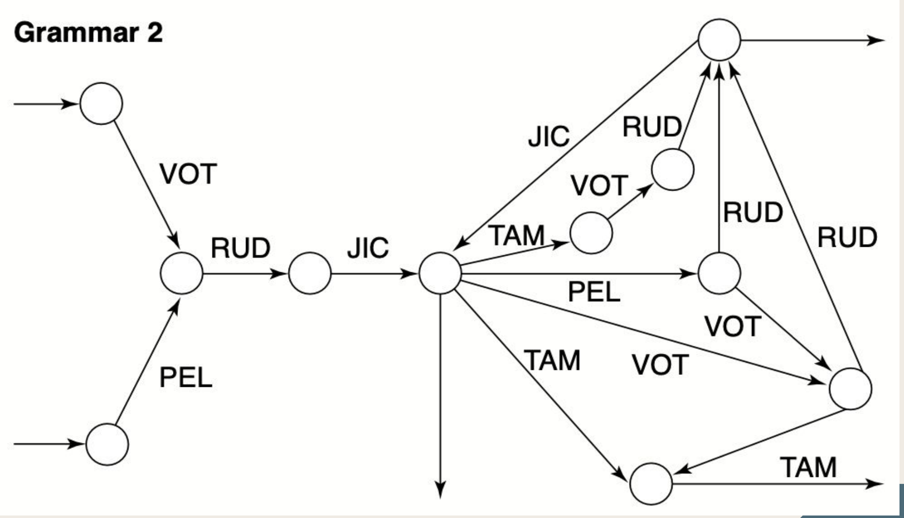
Important: Train strings were never encountered in test strings

Experiment 2: Words in Sequence

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Results: Babies listened longer to strings from their trained grammar

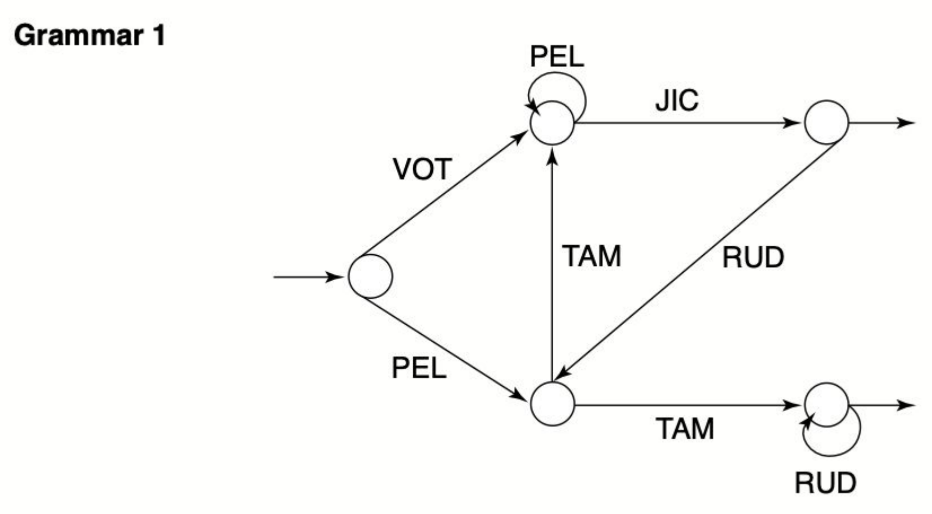


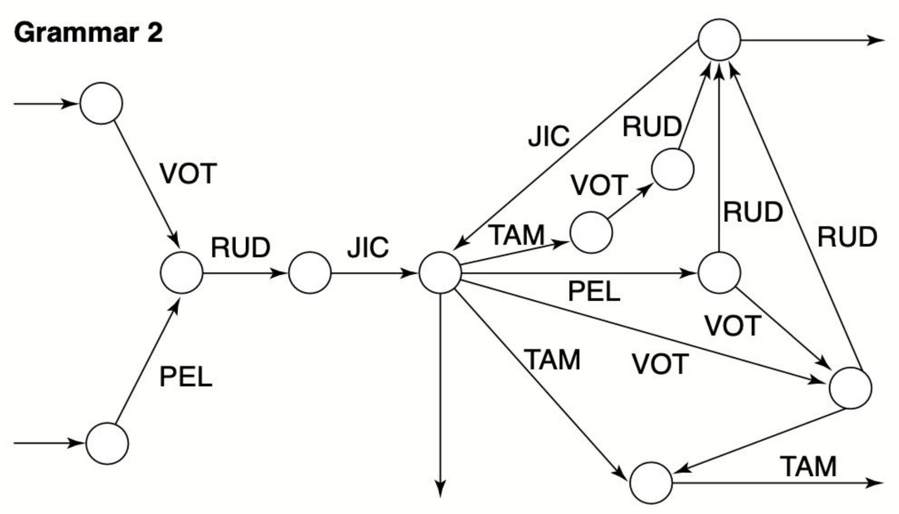


Experiment 2.2: Words in Abstract Patterns

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Same idea, but: train set vocabulary differs from test set vocabulary



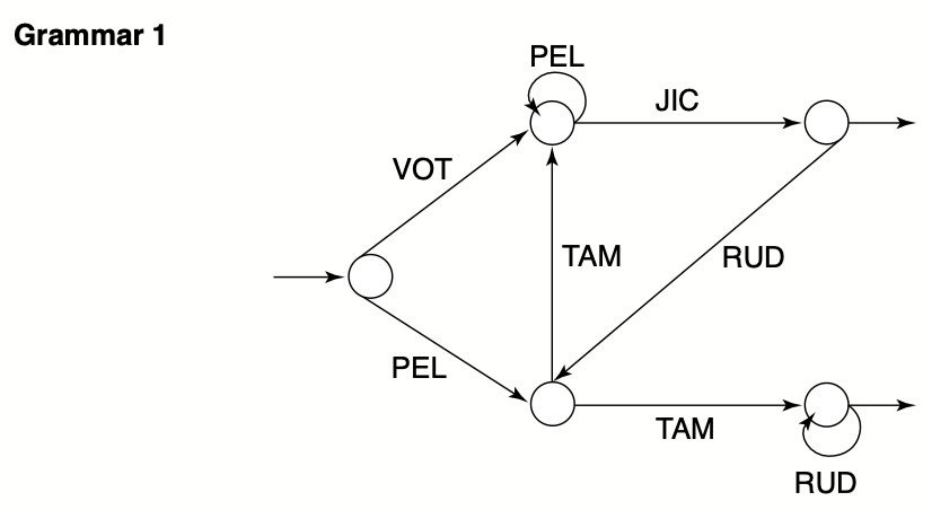


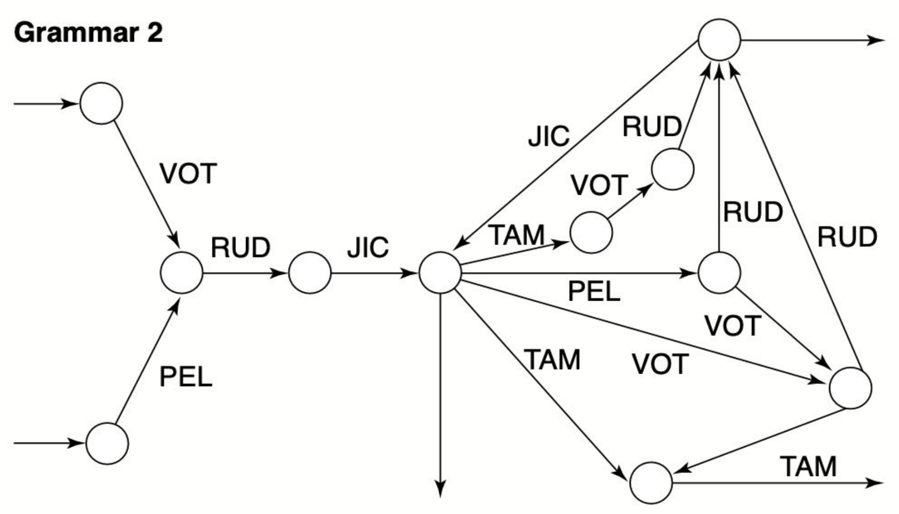
Train: PEL-TAM-PEL-PEL-JIC
Test: FIM-SOG-FIM-FIM-TUP

Experiment 2.2: Words in Abstract Patterns

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Same idea, but: train set vocabulary differs from test set vocabulary



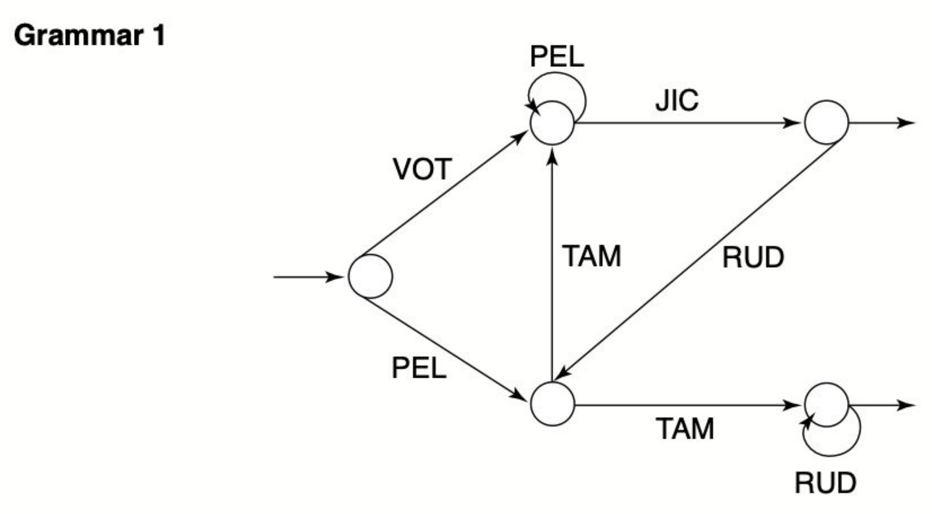


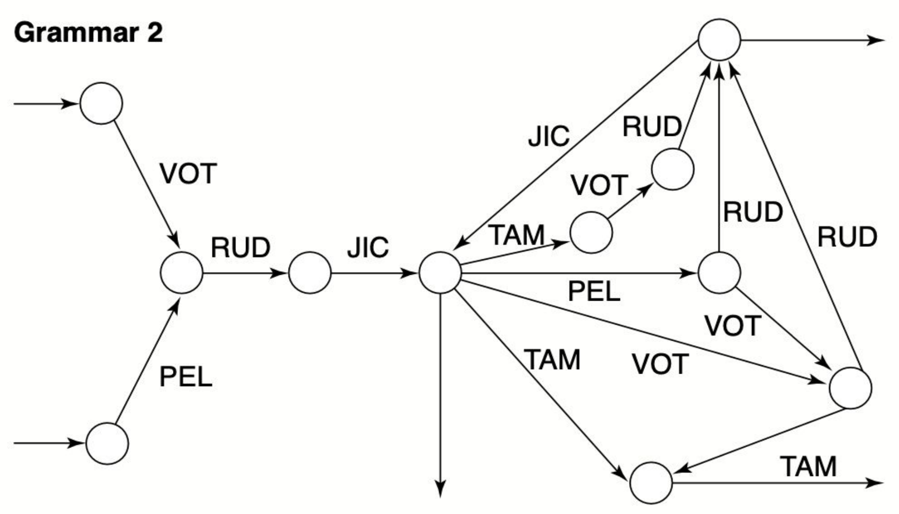
Learners could not simply memorize word pairs: vocabulary was different and again, train strings were never presented in test strings

Experiment 2.2: Words in Abstract Patterns

Gómez and Gerken (1999): Artificial grammar learning by one-year-olds leads to specific and abstract knowledge

Results: Babies discriminated between grammars despite change in vocabulary



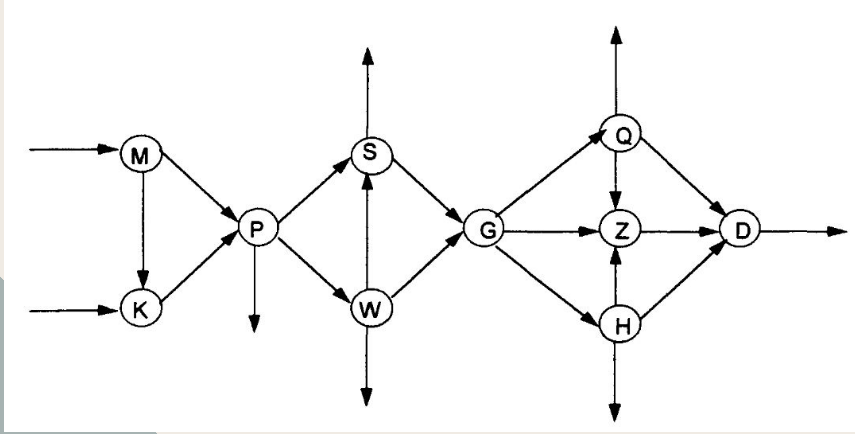


Experiment 3: Limitations of pattern-based representations

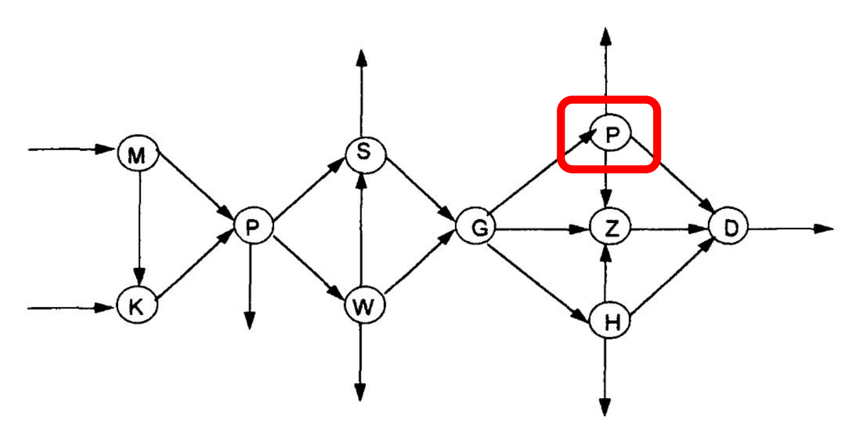
Gómez et al. (2000): The basis of transfer in artificial grammar learning.

Expose adult learners to 1 of 2 grammars and test if they can recognize ungrammatical forms

No repeating/identical elements



One identical element in ⅓ of sentences



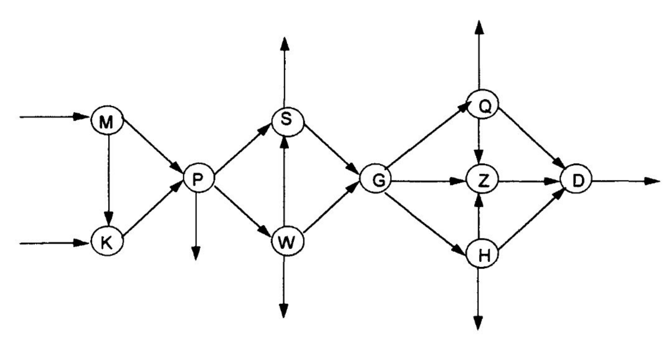
Experiment 3: Limitations of pattern-based representations

Gómez et al. (2000): The basis of transfer in artificial grammar learning.

Train: Listen to strings formed by one grammar

Test: Listen to strings which are grammatical and ungrammatical, according to their trained grammar

Non-repeating condition

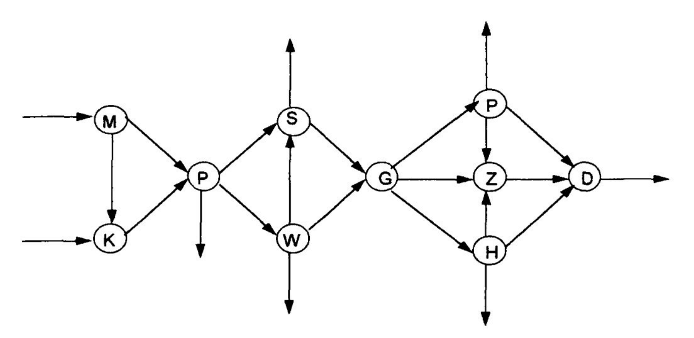


MK**PD**GQ

MKPWGQ

One illegal transition per item

Repeating condition



MP**SWS**GH

MPWSGH

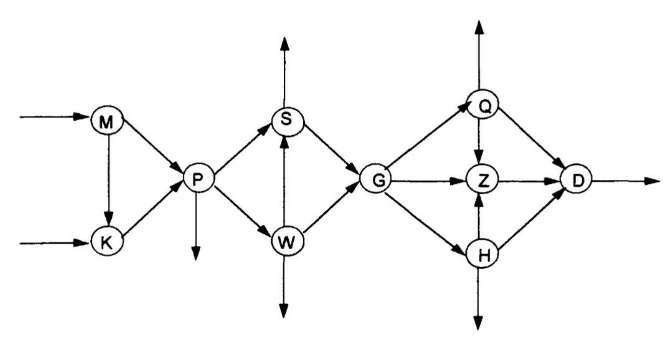
One illegal transition + one illegal repetition per item

Experiment 3: Limitations of pattern-based representations

Gómez et al. (2000): The basis of transfer in artificial grammar learning.

Results: Learners able to discriminate ungrammatical forms, regardless of which grammar they were trained on.

Non-repeating condition

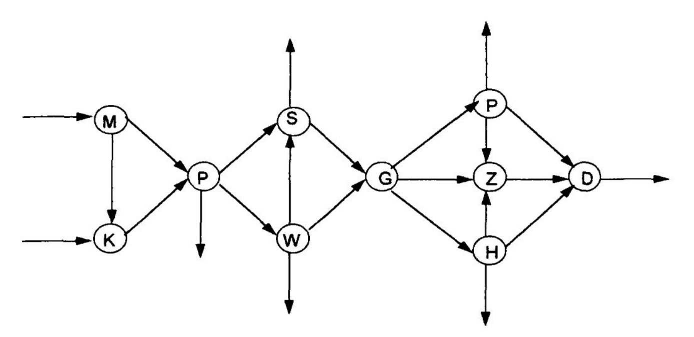


MK**PD**GQ

MKPWGQ

One illegal transition per item

Repeating condition



MP**SWS**GH

MPWSGH

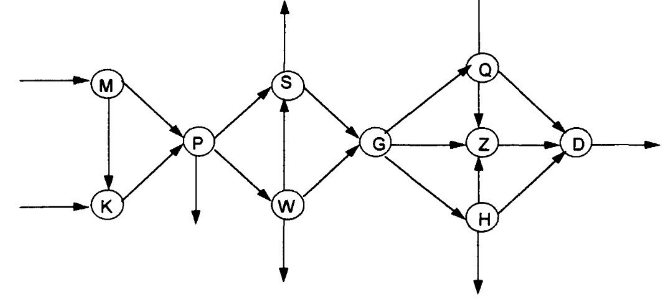
One illegal transition + one illegal repetition per item

Experiment 3: Limitations of pattern-based representations

Gómez et al. (2000): The basis of transfer in artificial grammar learning.

Results: However, when using different test than train-vocabulary, learners could *only generalize in the repeating condition.*

Non-repeating condition train set

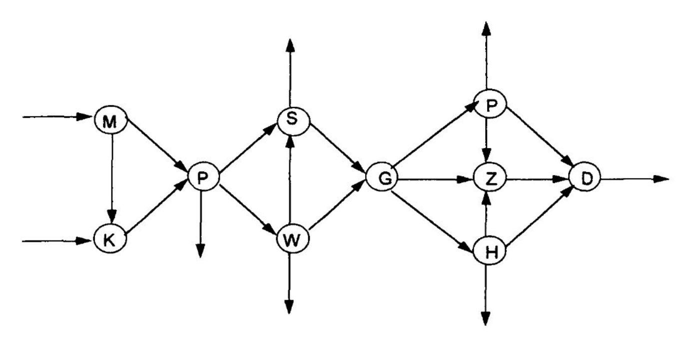


MK**PD**GZD

MKPWGZD

One illegal transition per item

Repeating condition train set



MP**SWS**GH

MPWSGH

One illegal transition + one illegal repetition per item

Pattern vs. Category-Based Abstraction

🤔 How do language learners abstract grammatical rules?

Pattern-based: perceptually bound
– ABA word pattern - humans can obviously recognize identical words because they sound the same
– Rising or falling frequency patterns

Category-based: over abstract categories
– humans can further abstract into broader categories:
– can access rules associated with categories
“Here is a wug. Now there are two of them. There are two ___”
“A bear washed himself” vs. “A bear washed him”
“The” precedes a noun but never a verb

Pattern vs. Category-Based Abstraction

Smith (1969): Learning co-occurrence restrictions: rule learning or rote learning?

	N_1	N_2	N_3
M_1	x	x	?
M_2	?	x	x
M_3	x	?	x

	Q_1	Q_2	Q_3
P_1	x	?	x
P_2	x	x	?
P_3	?	x	x

Train: Learners given some but not all possible M→N pairs and P→Q pairs

Note: the M,N,P,Q categories consists of 12 letters: h j l n p q r s w x y z
Example: M = {h, z, q}, N = {j, x, l}, P = {n, w, y}, Q = {s, r, p}
hx ql ws np

Pattern vs. Category-Based Abstraction

Smith (1969): Learning co-occurrence restrictions: rule learning or rote learning?

	N_1	N_2	N_3
M_1	x	x	$\boxed{?}$
M_2	$\boxed{?}$	x	x
M_3	x	$\boxed{?}$	x

	Q_1	Q_2	Q_3
P_1	x	$\boxed{?}$	x
P_2	x	x	?
P_3	$\boxed{?}$	x	x

Train: Learners given some but not all possible M→N pairs and P→Q pairs

Test: Can learners generalize to new pairs?

Example: M = {h, z, q}, N = {j, x, l}, P = {n, w, y}, Q = {s, r, p}

zj? nr? hp? zs?

Pattern vs. Category-Based Abstraction

Smith (1969): Learning co-occurrence restrictions: rule learning or rote learning?

	N_1	N_2	N_3
M_1	x	x	?
M_2	?	x	x
M_3	x	?	x

	Q_1	Q_2	Q_3
P_1	x	?	x
P_2	x	x	?
P_3	?	x	x

Train: Learners given some but not all possible M→N pairs and P→Q pairs

Test: Can learners generalize to new pairs?

No. Learners may accept M1→Q3. Not enough information to generalize dependencies between categories.

Pattern vs. Category-Based Abstraction

Smith (1969): Learning co-occurrence restrictions: rule learning or rote learning?

	N_1	N_2	N_3
M_1	x	x	?
M_2	?	x	x
M_3	x	?	x

	Q_1	Q_2	Q_3
P_1	x	?	x
P_2	x	x	?
P_3	?	x	x

💡Problem is fixed when category members are marked with salient perceptual cues:

N members: **[kais]**ermil**[rish]**

Q members: **[wan]**ersum**[glot]**

→ Can use category markings to make inferences about new pairs

Pattern vs. Category-Based Abstraction

🤔 **How do language learners abstract grammatical rules?**
by gaining sufficient evidence from systematically related cues

Pattern-based: perceptually bound
– ABA word pattern - humans can obviously recognize identical words because they sound the same
– Rising or falling frequency patterns

Category-based: over abstract categories
– humans can further abstract into broader categories (**by using salient perceptually-bound cues)**
– can access rules associated with categories

“Here is a wug. Now there are two of them. There are two ___” **“s” indicates plurality**

“A bear washed himself” vs. “A bear washed him” **“self” indicates reflexivity**

“The” precedes a noun but never a verb

Pattern vs. Category-Based Abstraction

🤔 How do language learners abstract grammatical rules?

By gaining sufficient evidence from systematically related cues

More examples from English:

– Verbs have inflectional endings (“-ed”, “-ing”)

– Different stress patterns:

“The”, “a” exhibit reduced stress

“Per**mit**” is a verb vs “**Per**mit” is a noun

→ 🤔 How *much* evidence is needed to generalize?

📚Review

Infants can ..

– parse word constituents based on syllabic relationships

– learn word-order constraints

– abstract familiar patterns to unseen vocabulary by learning ‘abstract-level’ categories

In using *artificial languages* we can demonstrate learning of syntactic form without initial connection to semantic form

Final question…

How representative are artificial grammars/languages of natural languages anyway?

→ we should drive feature manipulation based on knowledge of natural language acquisition