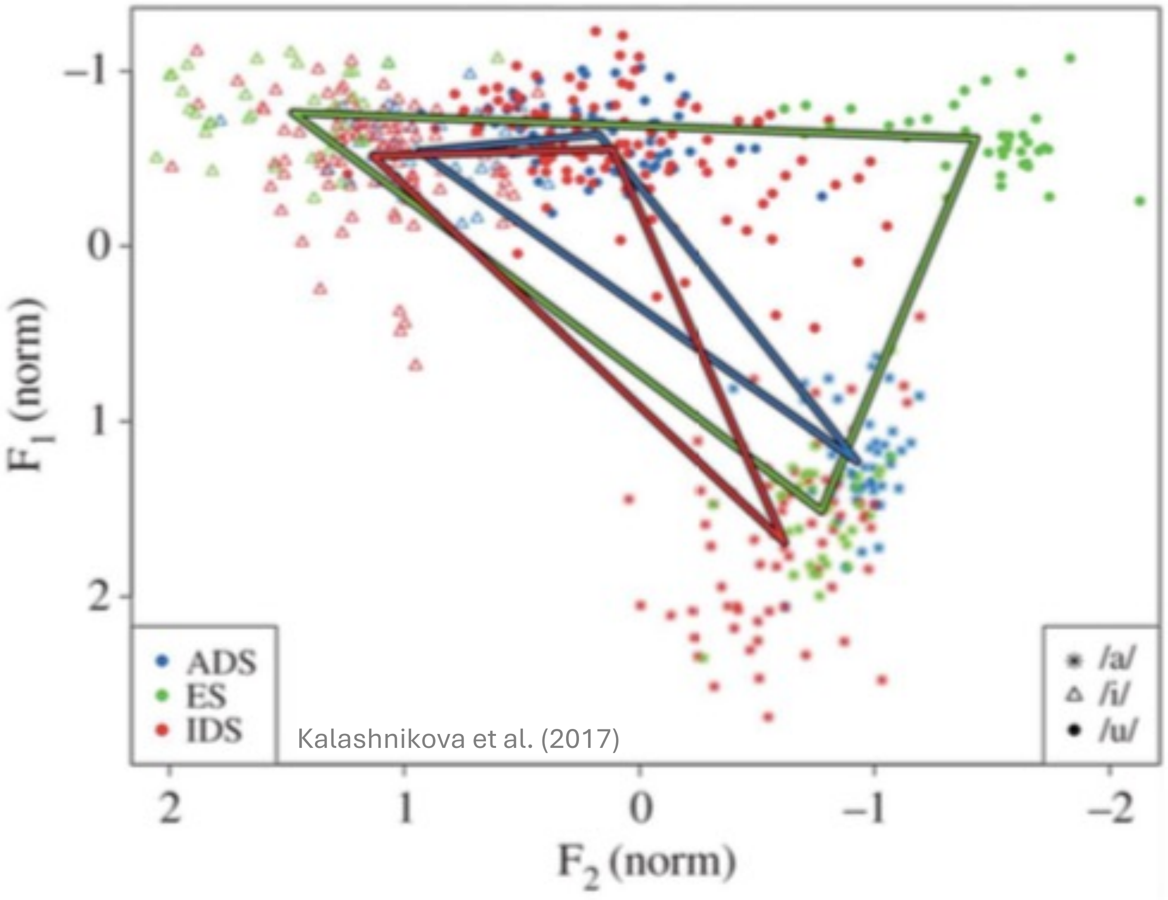
Statistical information in (early) language learning

Nicole Altvater-Mackensen

University of Mannheim,
Chair of Psycholinguistics



Some contextualization...

Perception & Processing

Mechanisms and biases in (speech) perception

Neurocognitive development

Language acquisition

How do children manage to learn their native language(s) without explicit instruction and with only limited cognitive ressources?

Social cognition & individual differences

Social context and structure of the input

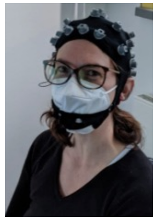
Interplay with general cognitive develpment

Language comprehension before the first word(s): **speech perception and lexical development**



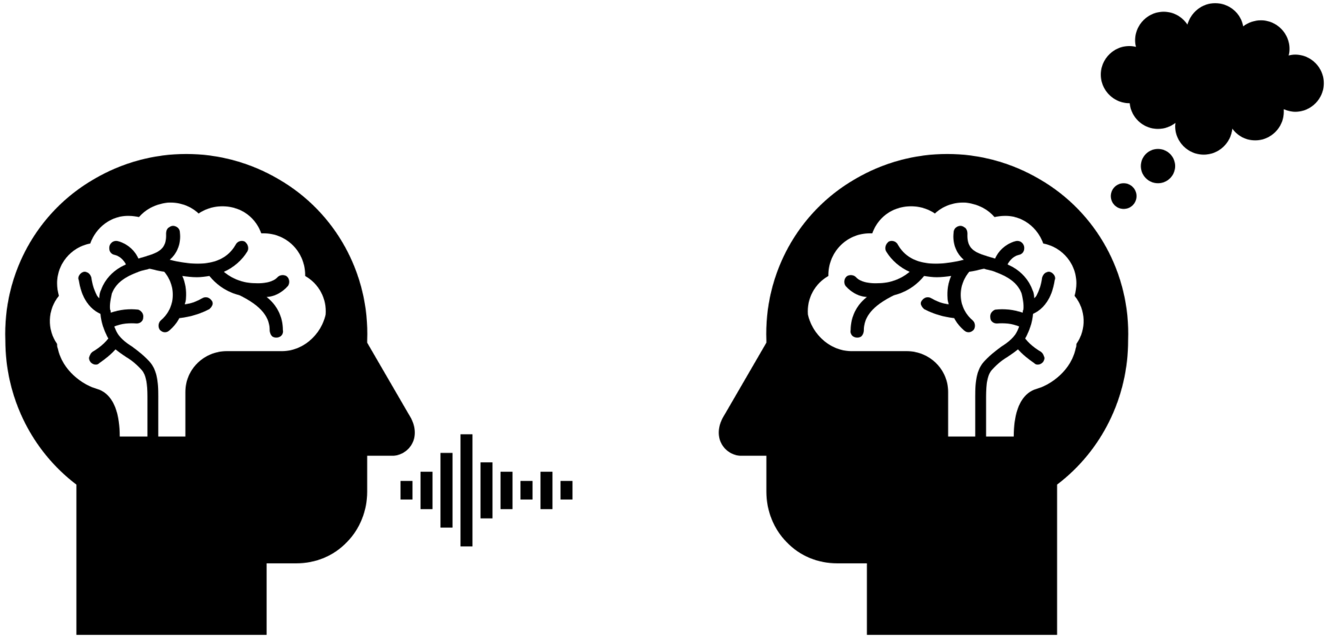




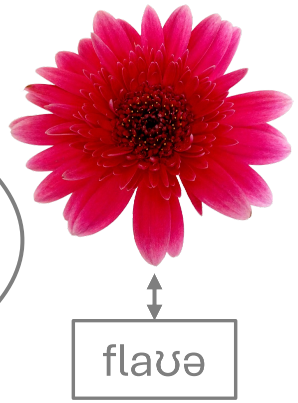
Language as a system

- Multimodal social behavior → communicative function
- Limited inventory of signs that are recombined to encode meaning → symbolic function



What needs to be learned?

LEXICON
Storage of form-meaning pairs



/g/ -/k/
dog – dock
**ơ[lg – ơ[gl*
**lgad – glad*

Phonology

Sound system

they play
*he play**__s__***

Morphology

Grammatical markers on words

Mummy likes flowers.
**Mummy flowers likes.*

Syntax

Rules on word combinations

Semantics

Assigning meaning

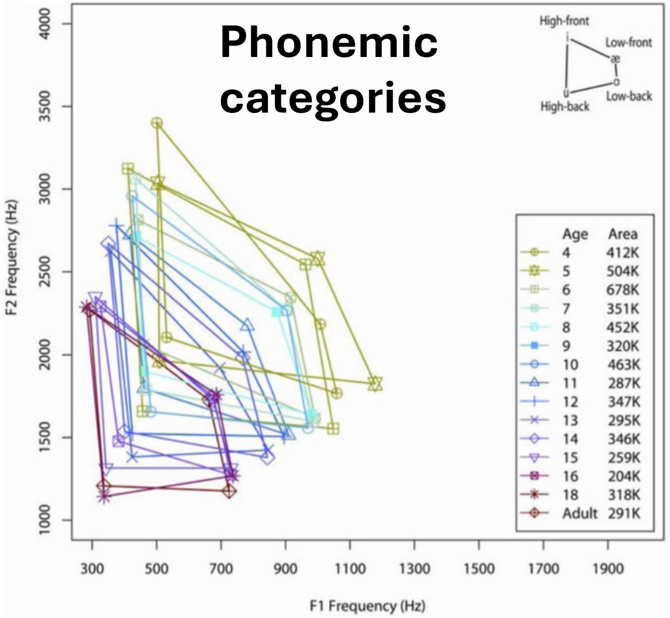
When may I ask a question?

Pragmatics

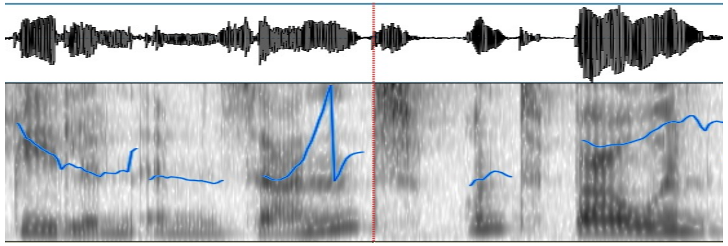
Conversational rules

What is the problem?

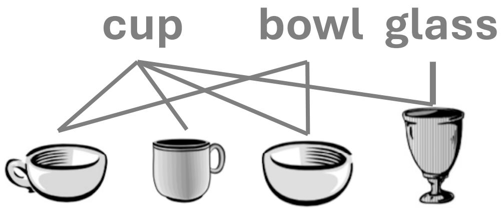
The input is noisy and does not present clear-cut categories. Children have to deduce abstract representations from a broad range of overlapping concrete exemplars.



Word segmentation



Word meaning



Morphological regularities

Hund *dog* – Hunde *dogs*
Puppe *doll* – Puppen *dolls*
Kater *male cat* – Kater *male cats*
Tochter *daugther* – Töchter *daugthers*

Word order

Le petit prince (DAN) the little prince
En infant terrible (DAN) a terrible child
.....

Ressources to solve the learning problem

Innate biases

Predispositions that help children to attend to relevant input, e.g. preference for socially relevant stimuli (faces, language, infant-directed communication, pitch countours, ...)

Scaffolding

Adult behavior that supports language learning (infant-directed speech, imitation, turn-taking, ...)

Cognition

General cognitive mechanisms supporting learning (categorical perception, associative learning, fast mapping, statstical learning, ...)

Principles

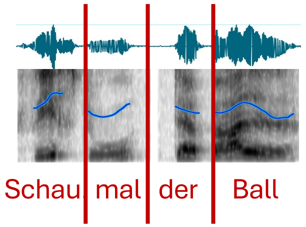
Domain-specific linguistic principles that facilitate learning (mutual exclusivity, Gricean maximes, universal grammar?)

Statistical learning

Machine learning

→ finding a predictive function that maps input/output pairs based on a (limited) data set

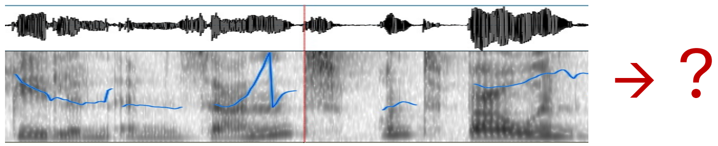
→ often supervised learning with training on an annotated set and testing prediction on an untrained set



Human learning (domain-general)

→ detect statistical regularities in the world to categorize and learn about the environment

→ mostly unsupervised, implicit learning (or reinforcement learning) with opaque input from which categories have to be derived



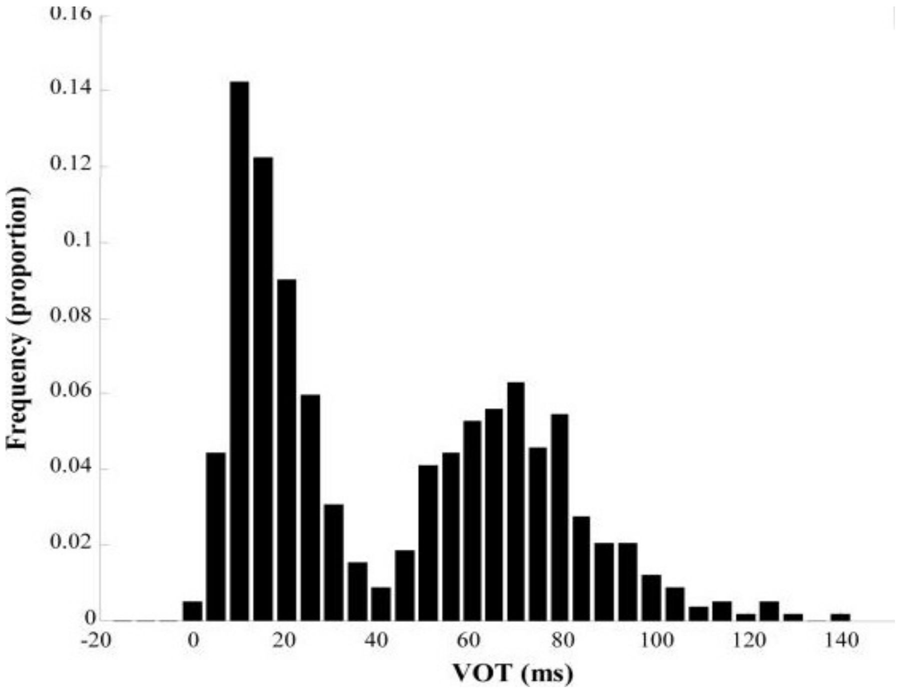
Statistical regularities in language

Linguistic forms are associated with specific frequency and co-occurrence information that may help to infer/learn linguistic categories.

Distributional information

Example: phonemic categories

Even though tokens of phonemes differ, they cluster around frequency peaks in specific acoustic dimensions. This distributional information can be exploited for category formation.



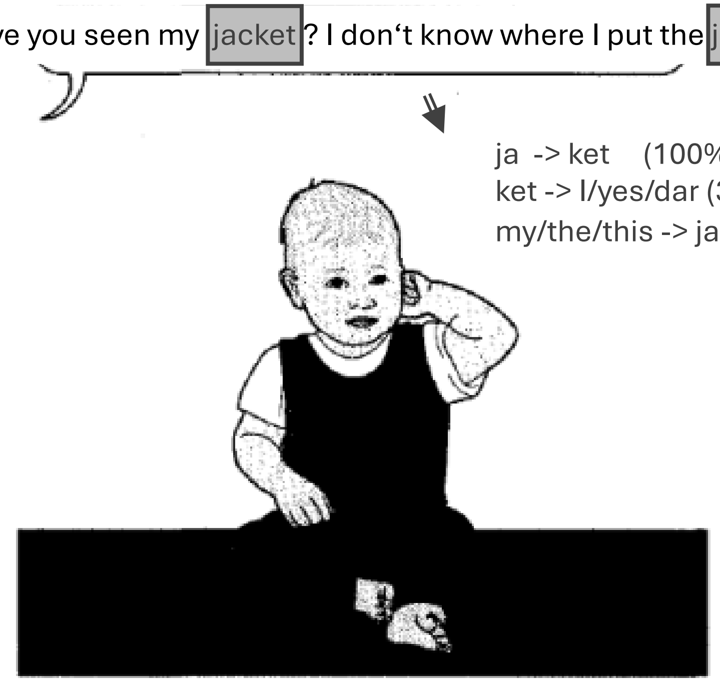
Statistical regularities in language

Linguistic forms are associated with specific frequency and co-occurrence information that may help to infer/learn linguistic categories.

Transitional probabilities

Example: speech segmentation

Syllables within a word reliably co-occur, while syllables that span word boundaries have smaller transitional probabilities. This statistical information can help to find word boundaries (next to, e.g., prosodic boundary information).



Statistical regularities in language

Linguistic forms are associated with specific frequency and co-occurrence information that may help to infer/learn linguistic categories.

Co-occurrence information

Example: word learning

In interactions with young children, words tend to co-occur with their referents (e.g., during play or picture book reading). Across multiple situations this allows disambiguation of potential referents.

The **dog** fetched the **ball**

The **dog** tugs at the **blanket**

The **child** throws the **ball**

The **child** sleeps under the **blanket**

Statistical regularities in language

Linguistic forms are associated with specific frequency and co-occurrence information that may help to infer/learn linguistic categories.

Non-adjacent dependencies

Example: Tense marking

Linguistic elements are not randomly ordered and so a specific element allows to make predictions about upcoming elements. This dependency information can be used to learn morpho-syntactic rules.

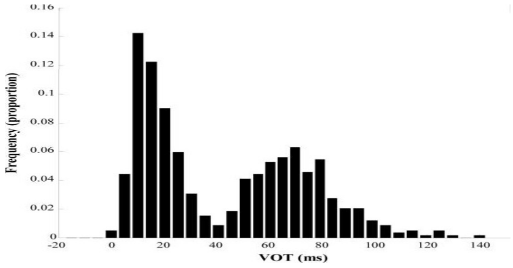
Past tense:

He **has** push**ed**
She **has** pull**ed**
She **has** mov**ed**
He **has** call**ed**
....

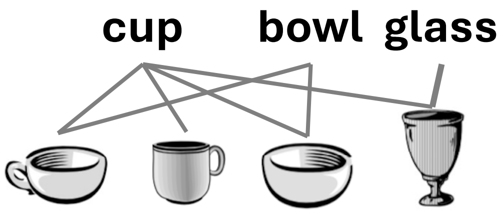
-> *S/he h**as** X-**ed***

Evidence for different types of statistical learning in language acquisition

Distributional information
(e.g., phonological categories)



Co-occurrence information
(e.g., word learning)



Transitional probabilities
(e.g., speech segmentation)

The -> gui -> tar -> is ....
ball looks
cookie sounds
... .

(Non-) adjacent dependencies
(e.g., word order, grammatical markers)

[[Determiner + [Noun+PL]] ...

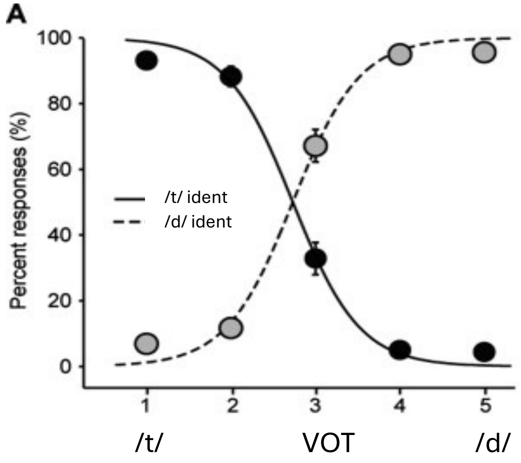
... has .... -ed / ... is ...-ing

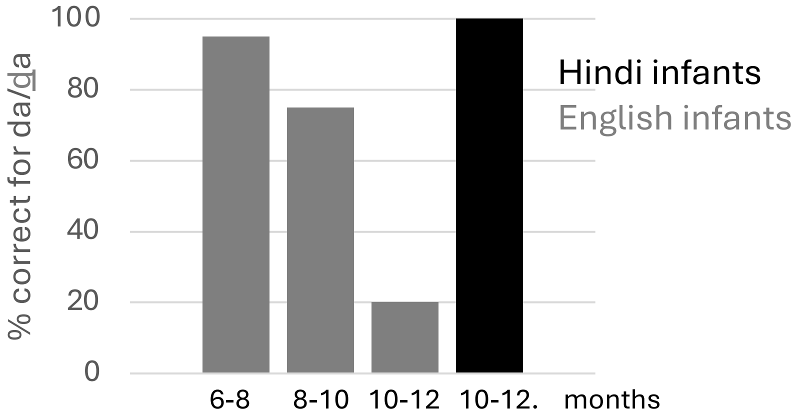
Learning phonological categories

Infants (and other primates) perceive speech sounds categorically. And they are sensitive to a wide variety of speech sound contrasts from birth.

Within the first year of life, they attune their perception to the native language, i.e., they improve their perception of native contrasts while they loose sensitivity to (most) non-native contrasts.

→ Distributional statistics as a key to detect relevant acoustic distinctions.





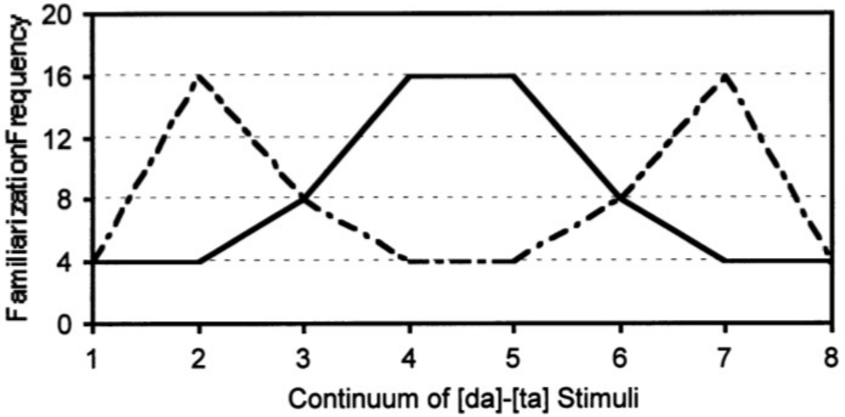
e.g. Kuhl et al., 20066; Werker & Tees, 1998

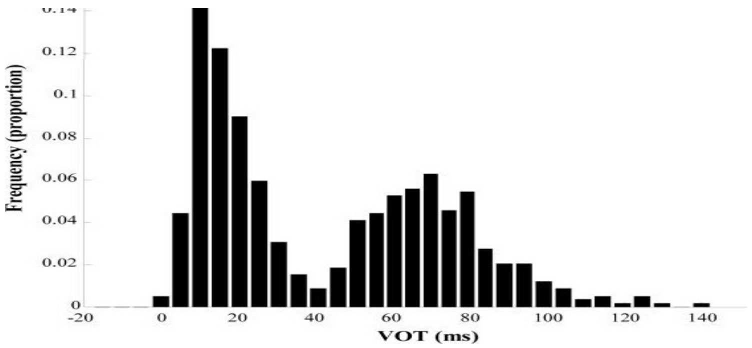
Distributional phonetic information

Preferential listening experiment

24 American-English 6-month-olds &
24 American-English 8-month-olds

2.5 minutes familiarization with six blocks of 24 syllables from the da-ta continuum with either ***unimodal*** or ***bimodal*** distribution with respect to VOT; colorful flowers as visual stimulus



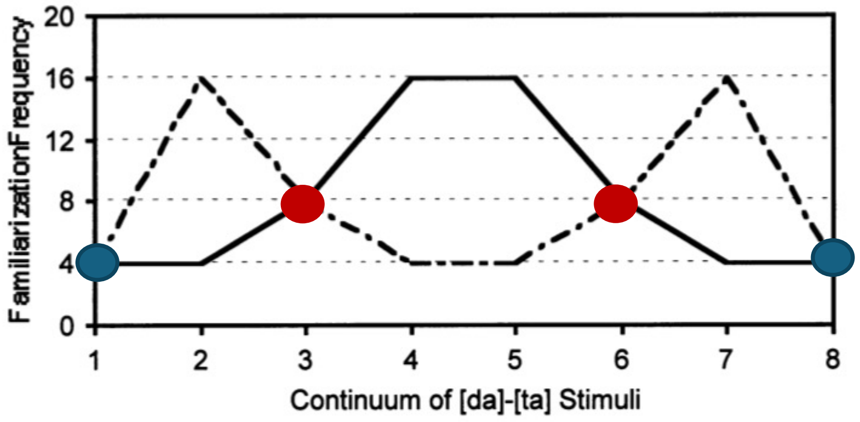


Maye et al. (2002)

Distributional information

Preferential listening experiment

Test phase presenting 4 ***alternating*** and 4 ***non-alternating*** trials, half presenting **tokens 1 and 8** (endpoints) and **tokens 3 and 6** (category boundaries) alongside visual checkerboard pattern



→ Infants only discriminate da/ta after listening to bimodal distribution

→ Distributional information in the input modulates sound discrimination, i.e. statistical information in the input provides a mean to learn phonemic categories (even with only limited short-term exposure).

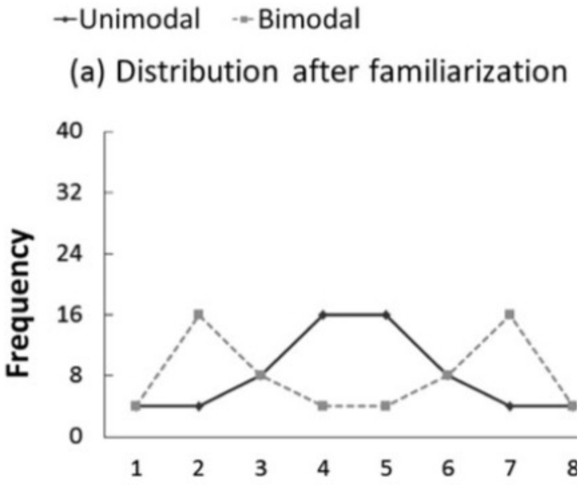
Distributional learning as domain-general mechanism…

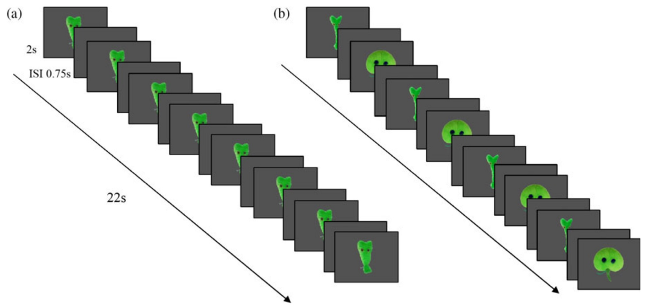
DISCRIMINATION OF OBJECTS

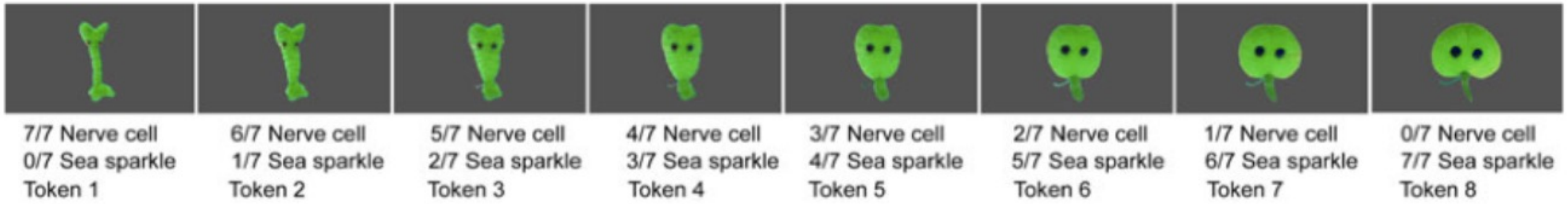
Method: eye tracking
4 alternating (token 1 and 8)/
4 non-alternating (token 3 or 6) test trials

Age group: Dutch 9-to 11-month-olds

Result: discrimination after bimodal,
but not unimodal familiarization







Junge et al. (2018)

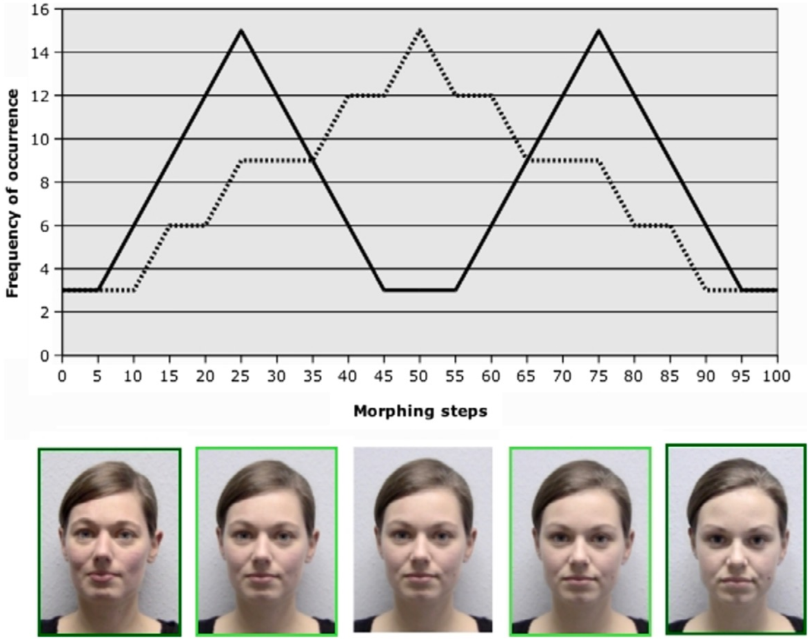
Distributional learning as domain-general mechanism…

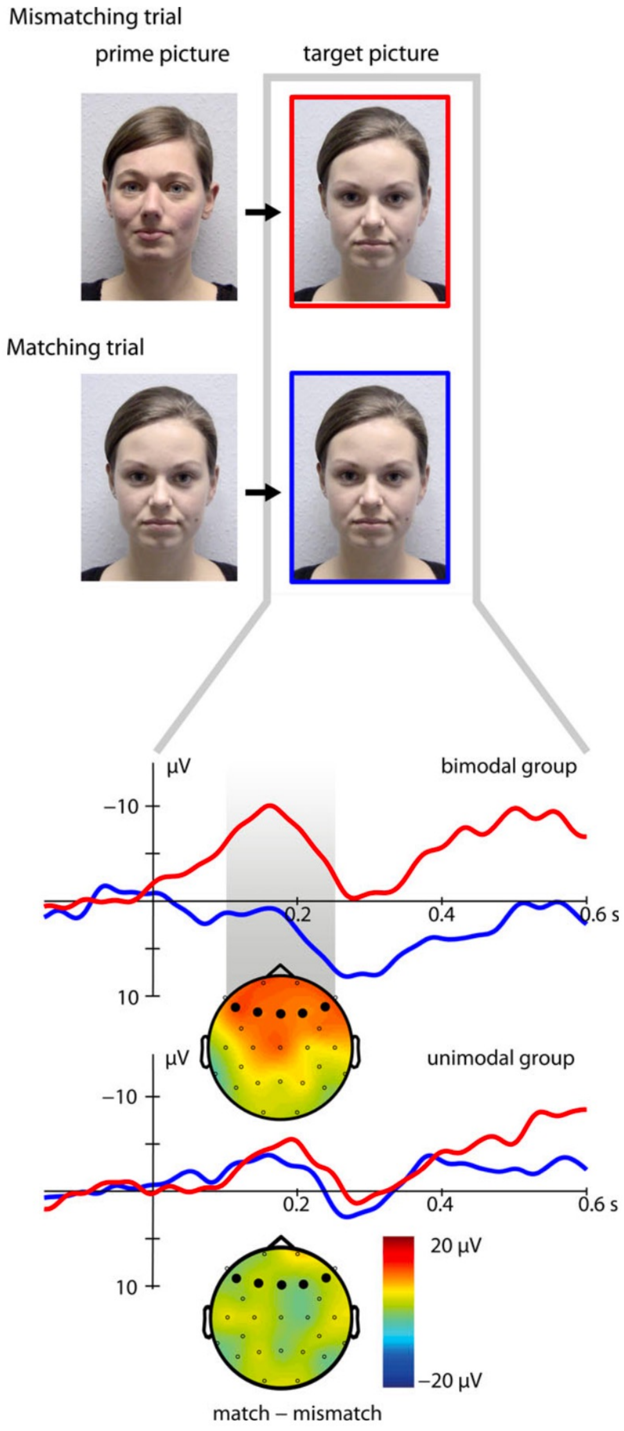
DISCRIMINATION OF FACES IN INFANTS

Method: ERPs, repetition suppresion paradigm

Age group: 6.5-month-olds

Result: discrimination after bimodal, but not unimodal familiarization





Altvater-Mackensen et al. (2017)

… with limits

DISCRIMINATION OF FACES AND PHONEMES IN ELDERLY

Method: similar procedure in familiarization but forced choice task for test,
stimuli were syllables (as in Maye et al., 2002) and faces (as in Altvater-Mackensen et al., 2017)

Age group: young (18-24 years) and old (60-84 years) healthy adults

Result: no evidence of learning in either modality without an explicit task

→ Distributional information might be easier to pick up for infants than adults (maybe because learning is less automatic and more top-down regulated in adults).

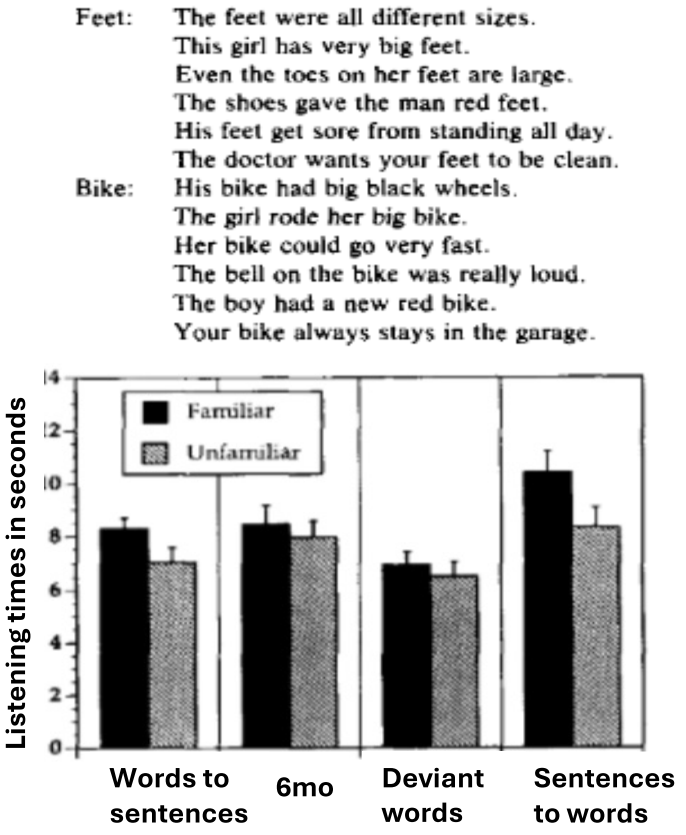
Schevenels et al. (2023)

Detecting words in the speech stream

Around 7 to 9 months of age, infants are able to extract (and memorize) words from a continuous speech stream

They exploit different cues to word boundaries, such as typical stress patterns, (il)legal consonant clusters and already familiar words.

→ Transitional probability on syllable combinations as a further key to detect common syllable sequences = words.



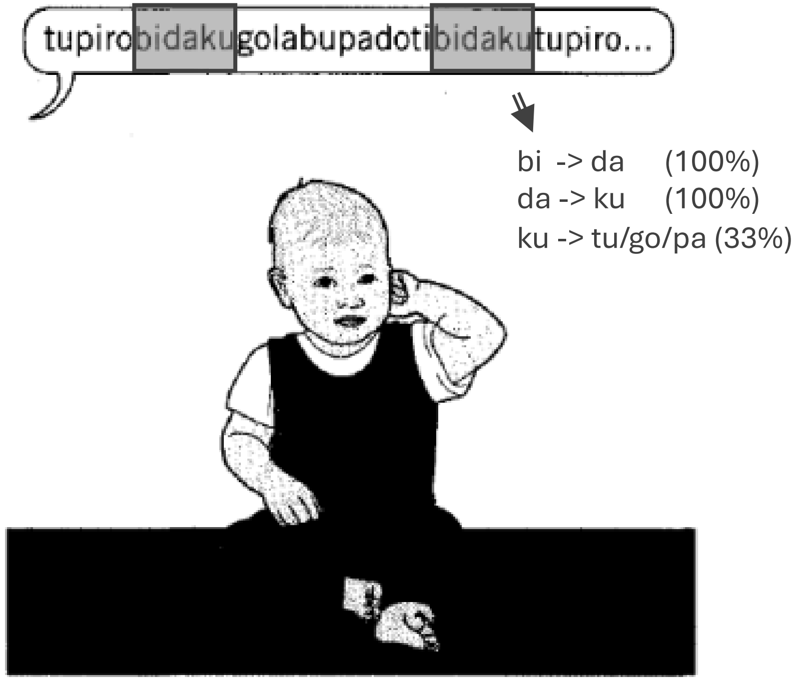
e.g. Juscyzk & Aslin (1996); Altvater-Mackensen & Mani (2013)

Transitional probabilities

Artificial language learning experiment using HTPP

24 American-English 8-month-old

2 minutes of familiarization to continous speech stream (fade in and out, no pauses, no prosody): concatenation of four three-syllable nonsense words *bidaku, padoti, golabu, tupiro*

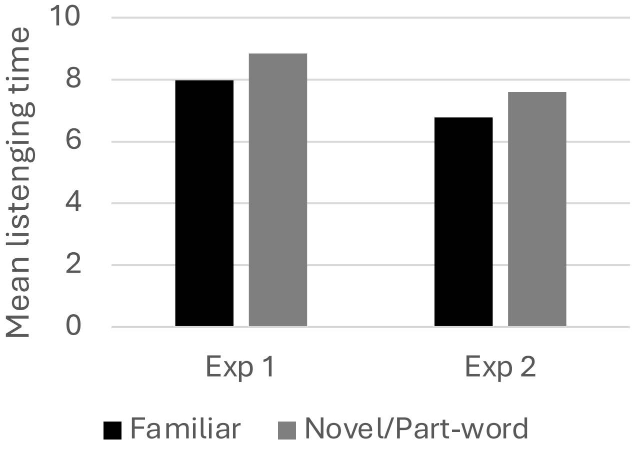


Saffran et al. (1996)

Transitional probabilities

Exp 1: test phase with two trials each consisting of
familiarized words (*bidaku, padoti, golabu, tupiro)*
novel syllable orders *(dabiku, dopati, lagobu, pituro)*

Exp 2: test phase with two test trials each with
familiarized words (*bidaku, padoti, golabu, tupiro)*
part-words *(daku-pa, doti-go, labu-tu, piro-bi)*



→ Longer looking to novel/part-words than familiarized words, indicating that infants can discriminate novel word forms from exposed word forms
→ Infants use transitional probability information to extract words from the speech stream.

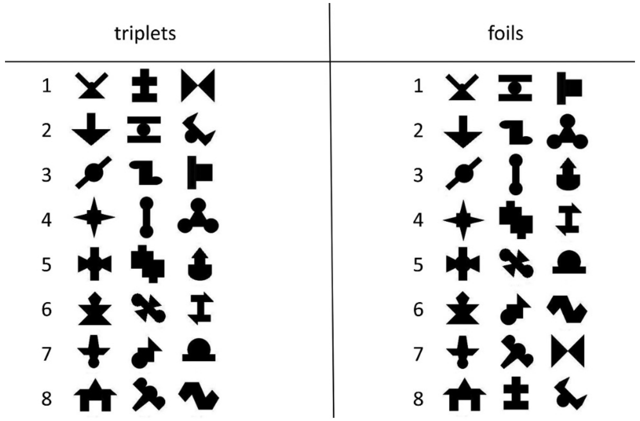
Transitional probabilities as domain-general (and age-robust) mechanism…

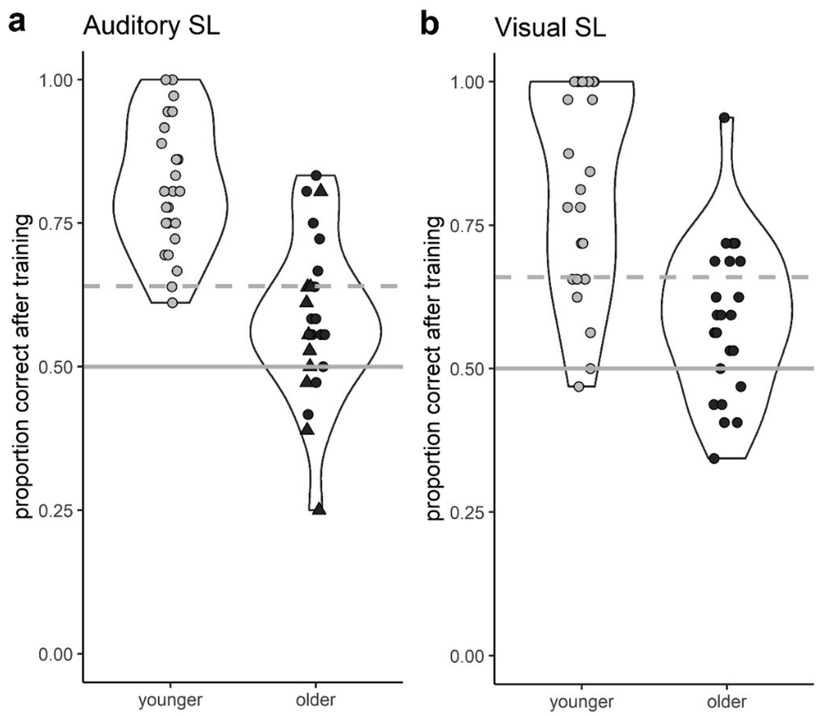
TRIPLET LEARNING IN AUDITORY AND VISUAL MODALITY

Method: familiariziation similar to Saffran et al. (1999), but behavioral forced choice task for test

Age group: young (18-24 years) and old (60-84 years) healthy adults

Result: discrimination in both modalities, but stronger effects for auditory than visual domain and less pronounced effects in elderly





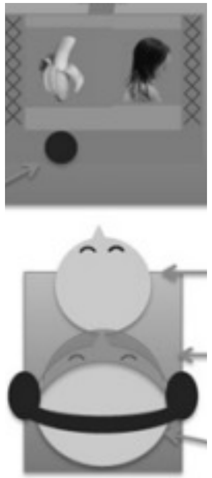
Fiser & Aslin (2001); Schevenels et al. (2023)

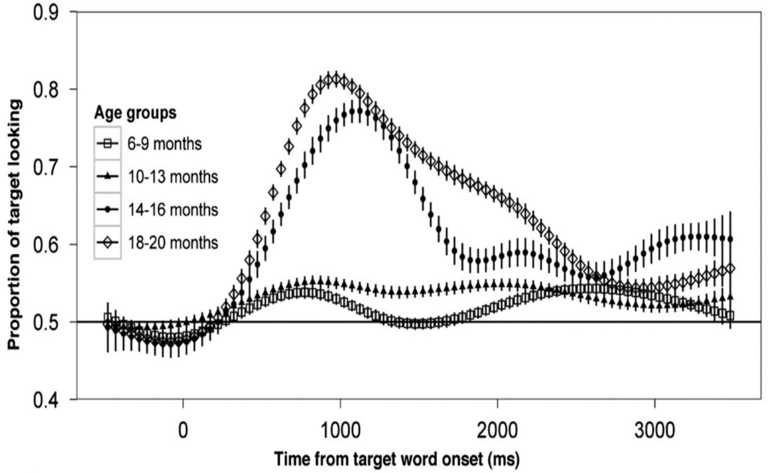
Learning form-meaning pairs

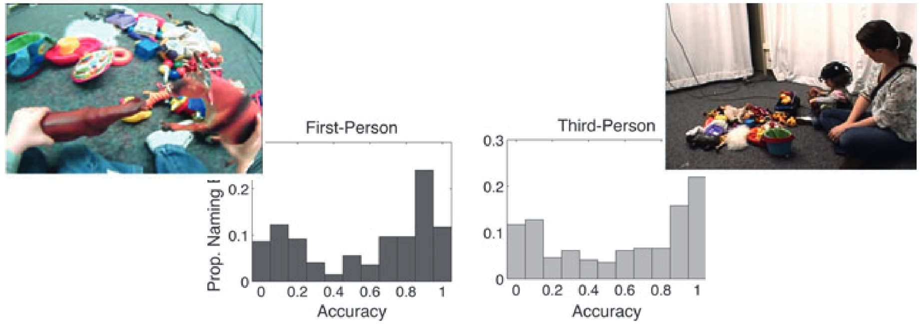
By 6- to 9-month of age infants show first recognition of highly frequent words and accumulate a sizable vocabulary within the following months.

Remarkable as word-meaning mappings are largely arbitrary and naming events are mostly ambiguous (cf. *Quine's problem*), even though the child's view might often be less crowded than the environment suggests.

→ Cross-situational co-occurrence of word and referents as an important source for lexical development.







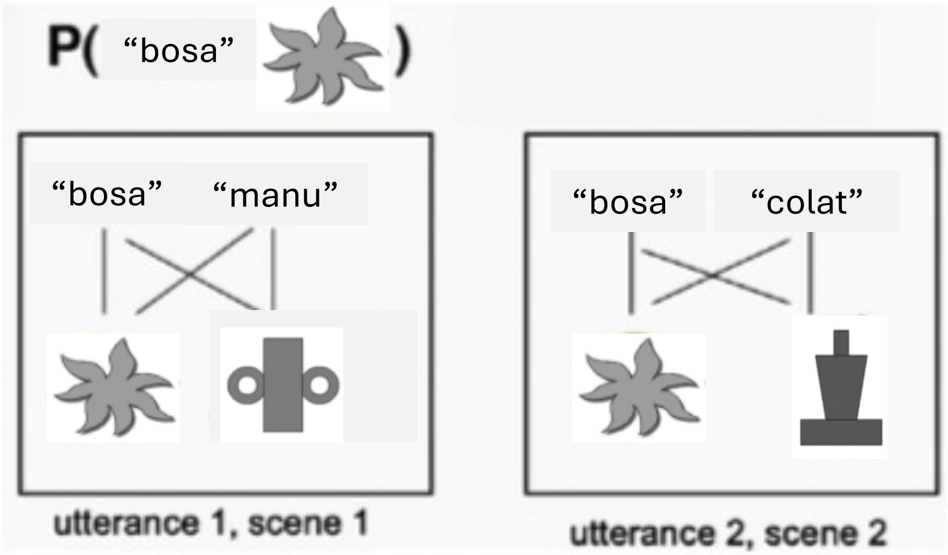
e.g., Yurovsky et al. (2013); Bergelson & Swingley (2012)

Tracking of co-occurrence information

Word learning task

28 American-English 12-month-olds &
27 American-English 14-month-olds
(other study successfully tested adults)

Familiarization with 30 slides presenting two shapes on each slide and labelling one of them (drawn from a set of six word-object pairs)

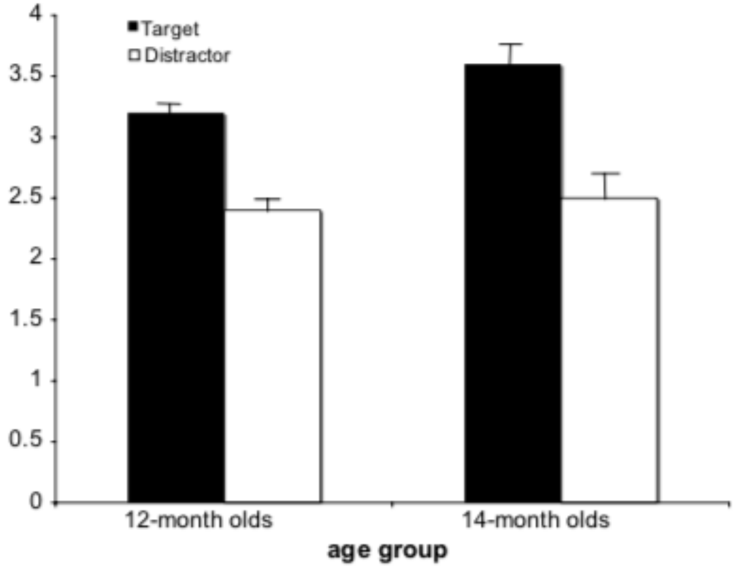


Smith & Yu (2008)

Tracking of co-occurrence information

Word learning task

12 test trials presenting each object-word combination twice, using preferential looking procedure with proportion of target looking as dependent variable



→ Longer looking to labelled object
→ Infants can use cross-situational co-occurrence information to track (and learn) word-object pairings.

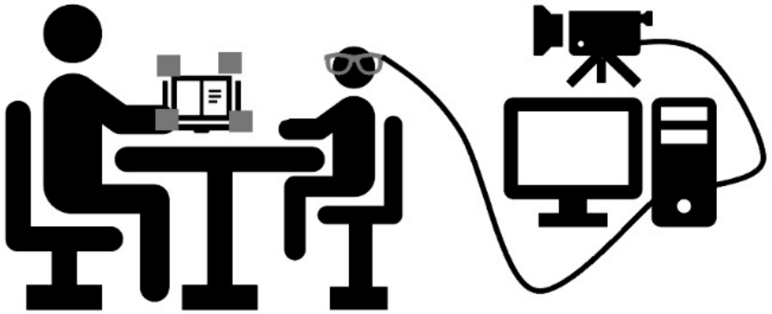
Tracking of co-occurrences in more natural learning situations ...

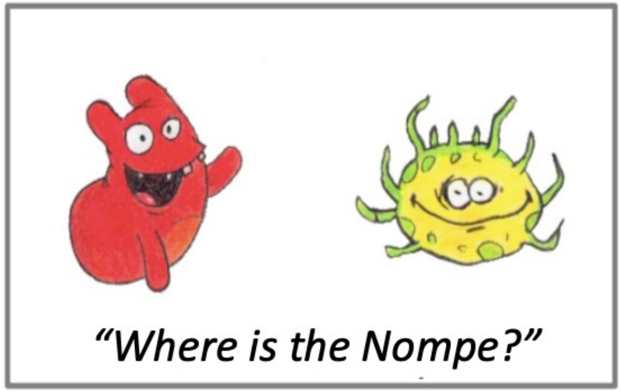
WORD LEARNING FROM PICTURE BOOKS IN TODDLERS

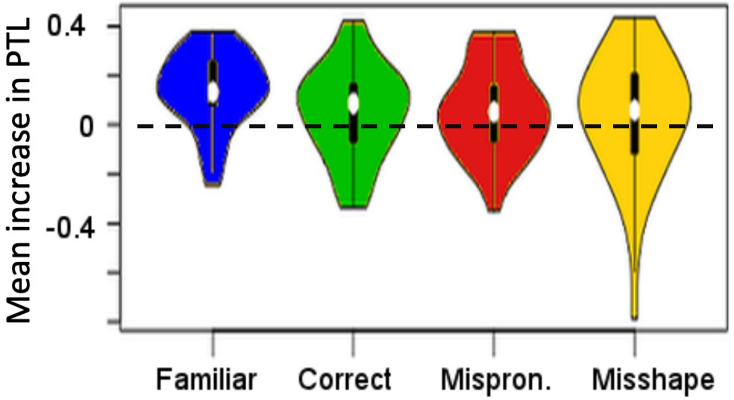
Method: natural exposure in book reading; preferential looking task

Age group: 2- to 4-year-olds

Result: successful learning of novel word-object pair (name of microbe) from crowded display; better learning for toddlers that showed higher attention during reading



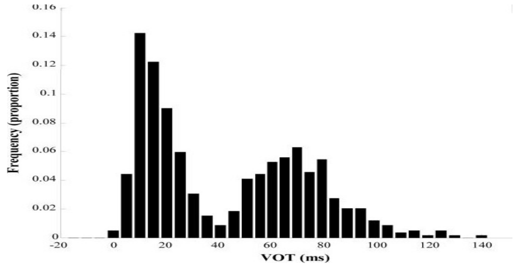




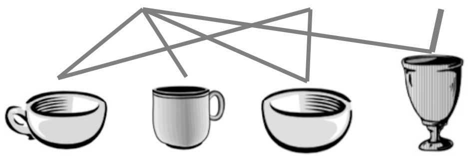
Altvater-Mackensen (2021)

Evidence for different types of statistical learning in language learning

Distributional information
(e.g., phonological categories)



Co-occurrence information
(e.g., word learning)



Transitional probabilities
(e.g., speech segmentation)

The -> gui -> tar -> is ....
ball looks
cookie sounds
... .

?

How can infants keep track of statistical regularities?
(in particular if they do not yet know what is the relevant unit of representation)

Why do we see differences in performance depending on age?

Conceptualizations of learning

⇄ Simple heuristics

- Take every syllable to be a word
- Associate what is co-occuring
- Detect repetitions
- Focus on salient parts of input
- Allow only one meaning for each word (and only one word for each meaning)
- …

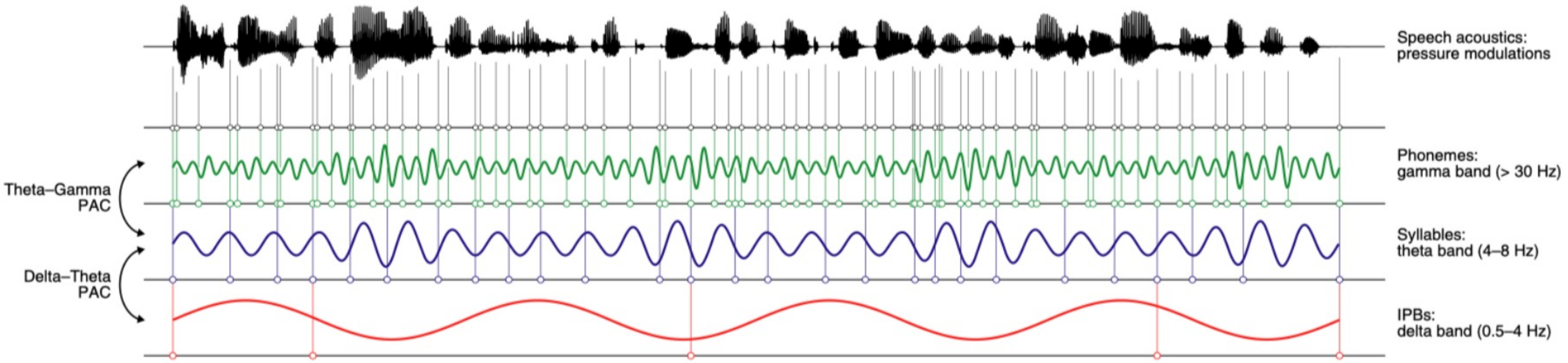
Hypotheses testing

- Take first occurrence and check it against the next occurrence. Update if different, strengthen if similar.
- Assume a rule/regularity and tolerate deviation as long as exceptions are not exceeding a certain treshold (cf. Yang‘s *tolerance principle*)
- …

Attention and brain oscillation?

Attention is dynamic and synchronizes with sensory information (cf. *DAT theory*).

This might be reflected in brain oscillations that align with regularities in the input at different levels, which might in turn help to track these regularities (cf. *neural entrainment*).

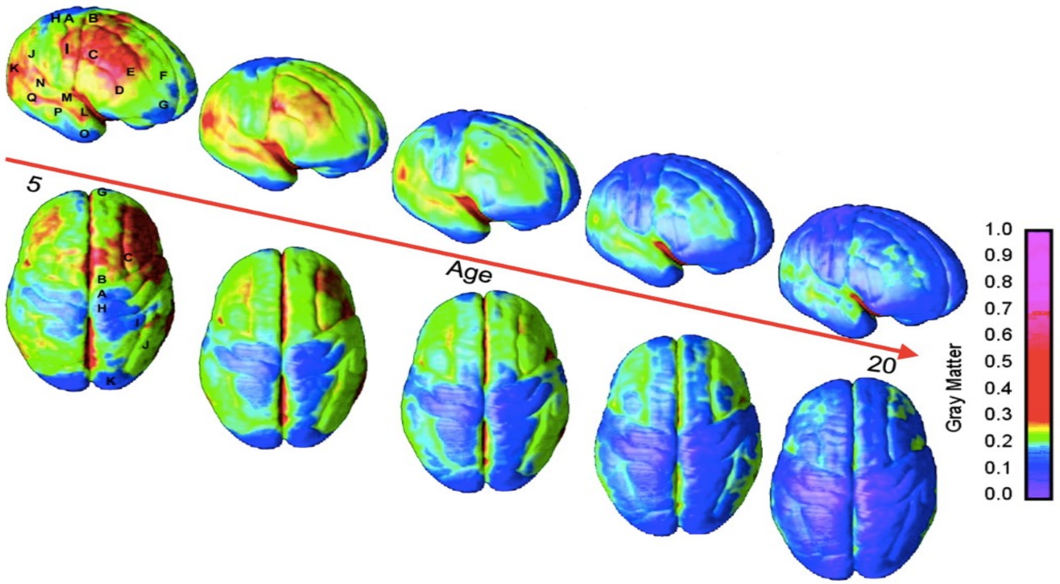


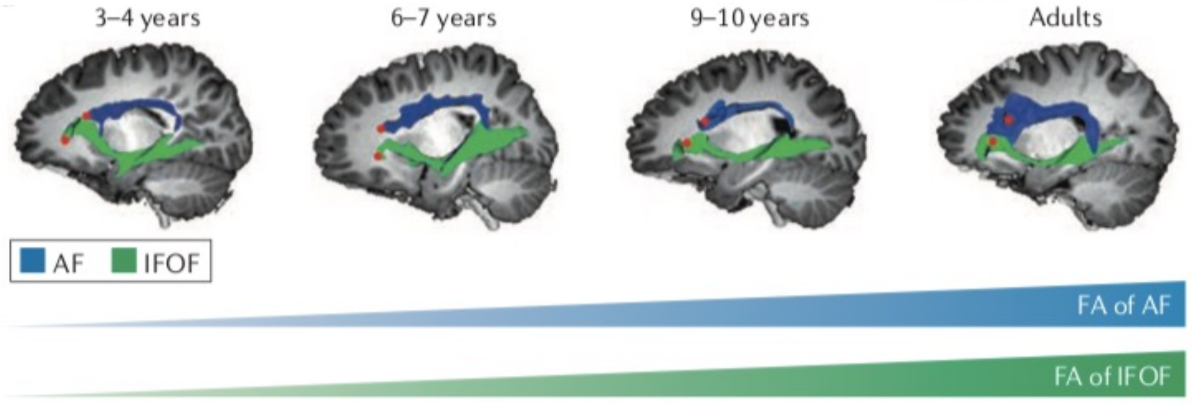
Jones (1976)

Processing and brain maturation?

Infants seem to be better implicit statistical learners than adults who often need a task.

This might reflect differential impact of bottom-up vs. top-down information in processing, which might in turn be related to the maturation of prefrontal cortex as well as fiber tracts connecting lower- to higher-level processing areas of the brain.





Skeide & Friederici (2016); Gogtay et al. (2004)

Some open questions/problems…

- Work with complex/more natural stimuli is very limited (e.g. transitional probabilities often reduced to learning of doublets/triplets and a limited inventory)

→ **ecological validity**

- It is largley unclear how statistical cues are weighed/integrated with other cues (some work, e.g. showing that prosody wins out over stats)

→ **comprehensive models**

Recommendation for further reading:
Conway (2020). *Neuroscience and Biobehavioral Reviews, 112, 279–299.*
Frost et al. (2019). *Psychological Bulletin, 145(12),* 1128-1153.