Lecture 4: Language acquisition 2 - words and grammar

Cognition and Communication, Monday, Sep. 20th 2021

Kristian Tylén

Next week

- No regular teaching in Cognition and Communication (Monday and Tuesday)

- Wednesday – Friday:

  Python/PsychoPy Workshop
  - Class 1: 9:15 – 12:00
  - Class 2: 13:15 – 16:00

- More info pending …

Agenda

- Generativist accounts
  - Chomsky and poverty of the stimulus
  - Productivity and recursion
  - Critical/sensitive periods

- Usage based accounts
  - Statistical learning
  - Fast mapping

- The nature of children's linguistic representations
  - The building blocks of language
  - Acquiring the first grammar
  - Productivity

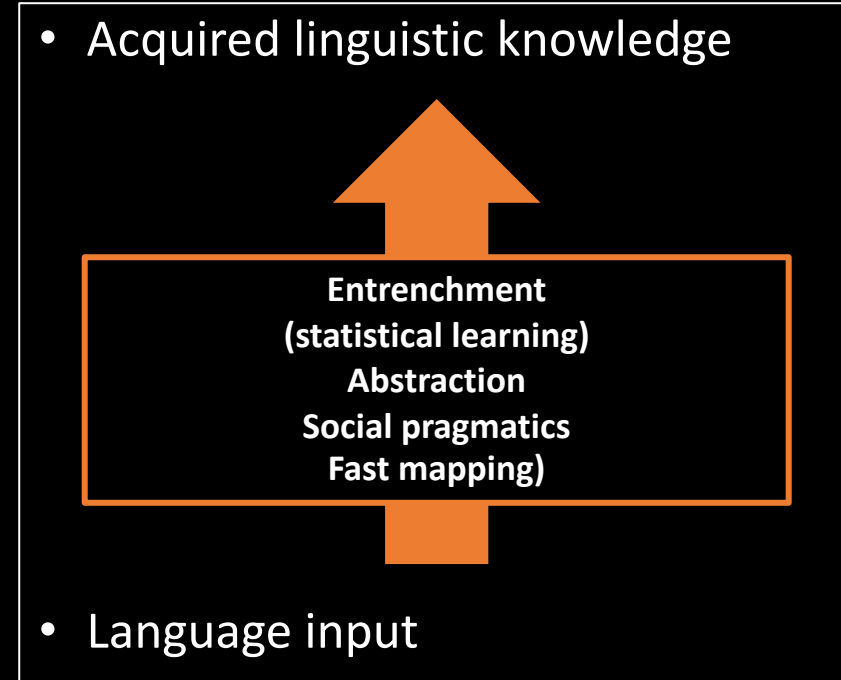- Resources for the study of language acquisition

Human language might be based on…

- Domain specific prerequisites for language (e.g. Chomsky & Pinker):
  - Innate linguistic "knowledge"
  - Universal grammar, including recursion (the ability to embed a structure within another structure)
  - LAD: a 'language acquisition device'

- Innate linguistic knowledge

  Universal Grammar
  Language Acquisition Device
  principles and parameters

- Specifics of native language
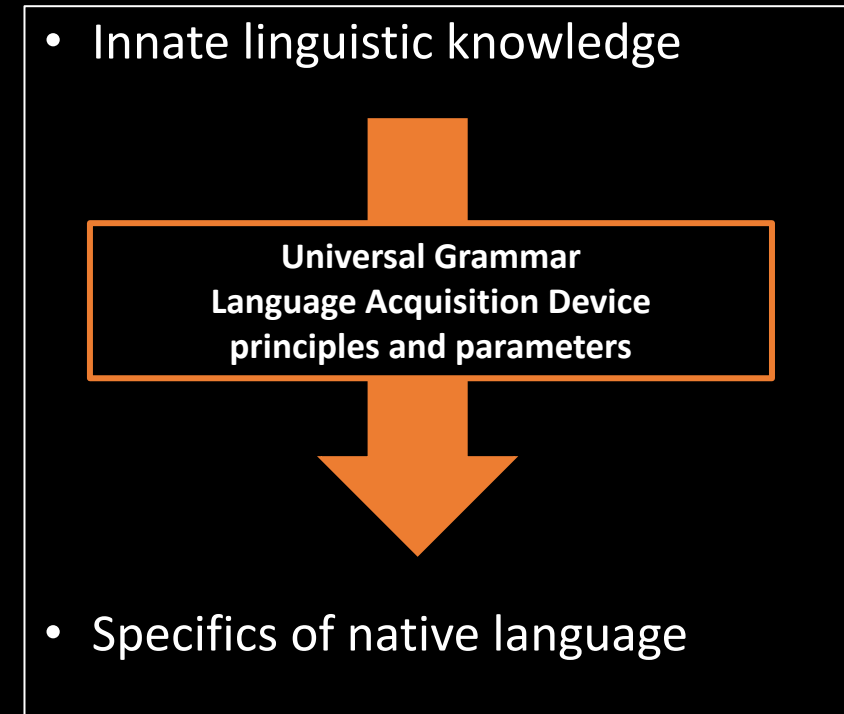
Human language might be based on…

- Domain general prerequisites for language:
  - Species unique motivation to collaborate and share information
  - Special mind-reading abilities (ToM)
  - Special learning mechanisms (e.g. imitation, sensitivity to social contingency and ostensive cues)

- Acquired linguistic knowledge

  **Entrenchment
  (statistical learning)
  Abstraction
  Social pragmatics
  Fast mapping)**

- Language input

Language acquisition and the "poverty of the stimulus" argument

- Noam Chomsky (1980): Children effortlessly learn the grammar of their native language

- However the speech that children are exposed to is often …
  - Unstructured and partially ungrammatical
  - Consistent with numerous possible grammars
  - I.e. the linguistic stimuli that the children are exposed to are too "poor" for them to acquire all the rules of their mother tongue

- Example:
  1. While he was dancing, the Ninja Turtle ate pizza.
  2. He ate pizza while the Ninja Turtle was dancing.
  - In English, the pronoun "he" can refer to the Ninja Turtle in 1 but not 2

- When children nonetheless acquire the complex target grammar, it must be due to innate "universal grammar"/language acquisition device

- "Evidently, development of language in the individual must involve three factors: genetic endowment, which sets limits on the attainable languages, thereby making language acquisition possible; external data, converted to the experience that selects one or another language within a narrow range; principles not specific to the Faculty of Language." (Chomsky, 2007)

From the CHILDES Sarah corpus:

*MOT: clown .
*MOT: all the babies (.) how they grow and grow and grow .
*MOT: who's that ?
*MOT: who ?
*MOT: oh (.) bow_wow@o (.) yeah .
*MOT: oh (.) see the babies .
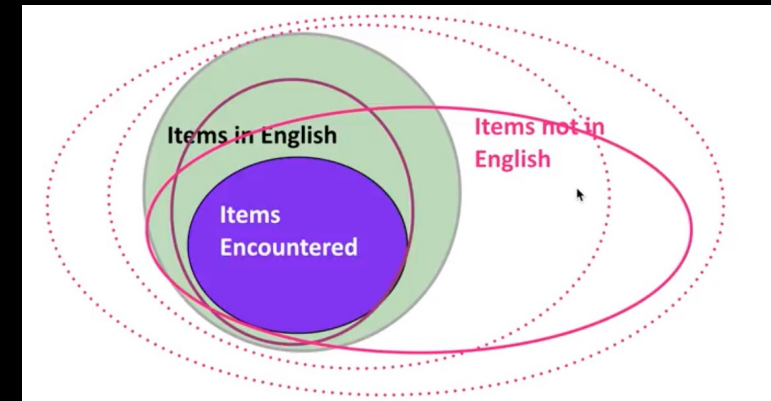
Linguistic productivity

- Linguistic productivity: the ability to combine a finite set of words into an infinite number of new sentences

- Recursion:
  - The ability to embed structures within structure
  - Suggested to be what makes human thought and language unique and powerful
  - … and in the core of *Universal Grammar*
  - *"Jean knows that Charlie said that Sue suspects that Bill thinks that Beth believes that…"*
  - *"The man the woman kissed loved Harley"*
  - *"The rat the cat the man the woman kissed loved chased escaped" (?)*
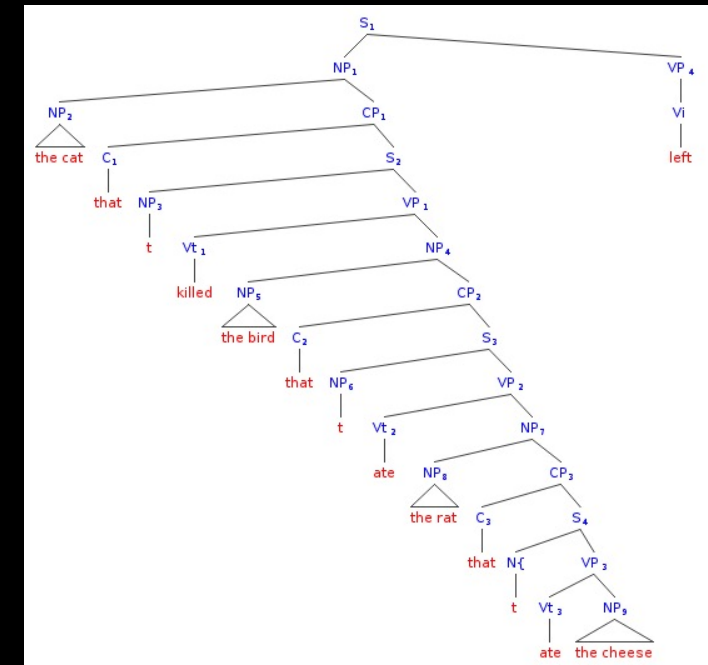
A critical period for language acquisition?

- Eric Lenneberg (1967)
  - There is a developmental window within which language has to mature (due to plasticity of the language module of the brain)
  - After this period it is very difficult/impossible to learn language
  - Based primarily on anecdotal case studies (children deprived from linguistic stimulation)

Genie Wiley

- Several *sensitive* periods?
  - Some evidence that infants sensitivity for speech sounds change around approx 1 year
  - But no consistant evidence for vocabulary/syntax
  - 2nd language acquisition: difficulties acquiring perfect pronounciation as an adult
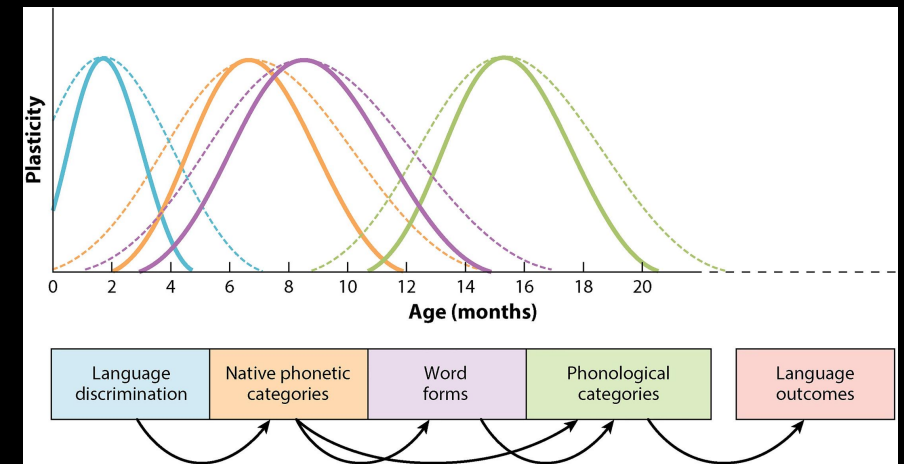

Werker and Hensch 2015

The "usage-based" approach

- Language is learned from experience (through domain general learning mechanisms)

- Abstract linguistic knowledge/categories are "extracted" from experience with language use

- Young children do generally not construe sentences they have not heard before, i.e. true productivity - develops late

- Lieven *et al*. 1997:
  - Children 2–3 years of age use virtually all of their verbs in one and only one form and sentence context
  - 92% of the earliest multi-word utterances were related to one of their first 25 lexically based patterns
  - These were different for different children

How do children acquire language from experience?

- If language is not innate, how do children come to recognize individual words and gramnmatical structure in a continuous speech stream?

    - Itisverydifficulttobeaninfantwhohastoguesswhereonewordstopsandthenextstarts

    - Itisverydifficulttobeaninfantwhohastoguesswhereonewordstopsandthenextstarts
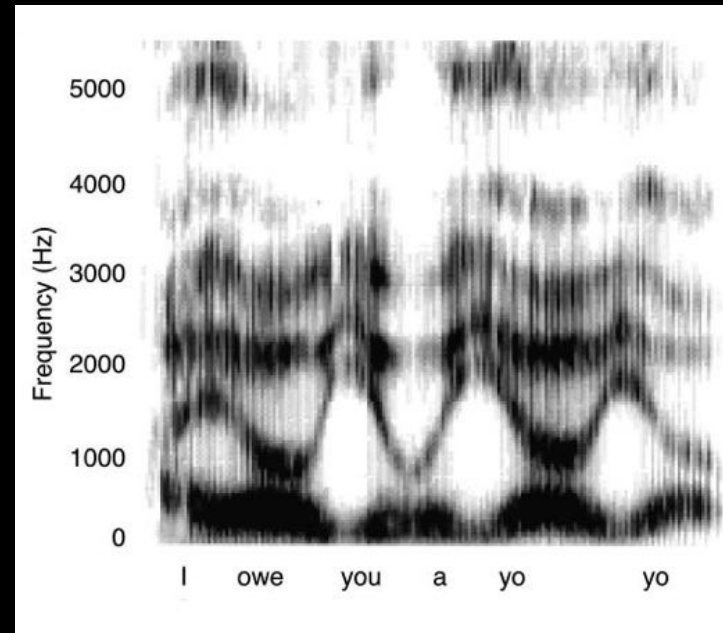
Statistical learning 🔊

- Saffran et al (1996): Statistical learning by 8-month-old infants

- Bidagugolabupadotitupirogolabutupirobidagupadotigolabupadotitupirogolabutupirobidagutupiro padotibidagugolabu

- Consist of the "words": bigadu, golabu, padoti, tupiro

- Bidagugolabupadotitupirogolabutupirobidagupadotigolabupadotitupirogolabutupirobidagutupiro padotibidagugolabu

- Contrasted to "non-words": dugola, bupado, titupi, robiga

- Infants were found to be surprised and thus sustained attention longer for non-words

- Sensitivity to "transitional probabilities":

*Sound sample from Trecca et al 2019
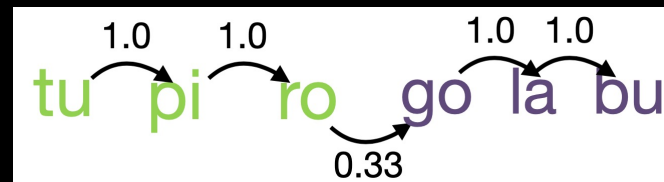
Fast mapping

- A inferential mechanism that supports word learning (Bloom and Markson 1998)

- Spoon/whisk-experiment:
  - Two objects are placed in front of the child – a spoon and a whisk. The adult ask for "the fendle".
  - The child knows the word "spoon" and therefore infer that the new word probably map onto the less known object

- The "modi"-experiment (Tomasello 2003):
  - An adult and a child are playing with a couple of objects that they slide down a ramp.
  - In condition 1, the adult sends off two objects and then says "now modi" and sends off the third object.
  - In condition 2, they again play various games with the three toys. Then the adult says "now modi" and sends off an object on the ramp.
  - What does the child learn? What is the meaning of the word "modi" in the two situations?

Constructing a language
Michael Tomasello (2003)

- In linguistics, traditionally "the sentence" has been the object of study

- However, the "full sentence" is mainly a property of written discourse:
  - In everyday oral discourse there is surprisingly few full, well-formed, grammatical sentences

- Michael Tomasello: the 'utterance' as our minimal object of study
  - An expressive act (vocal, gestural, multimodal) that express a communicative intention
  - Focus moves from sentence structure to the social/functional pragmatics of communication as constitutive

Michael Tomasello

CONSTRUCTING A LANGUAGE

A Usage-Based Theory of Language Acquisition
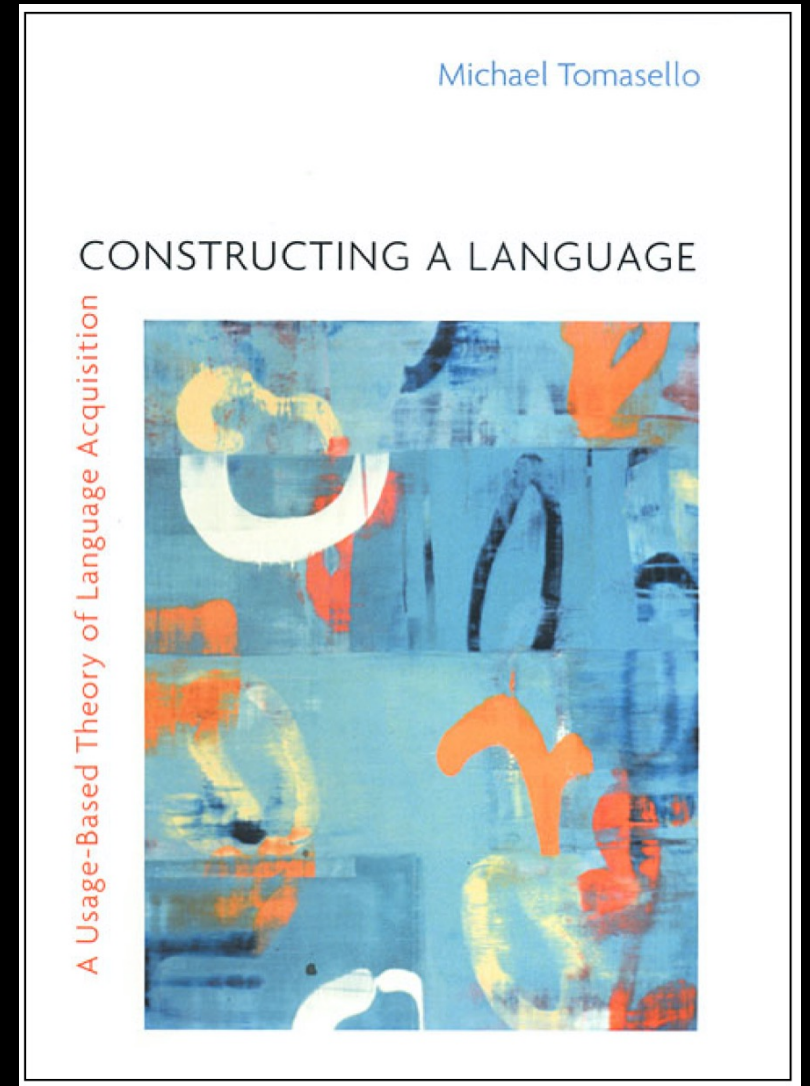
Constructing a language
## Michael Tomasello (2003)

- Children are not born with adult categories such as noun, verb, adverb etc

- Children's linguistic representations are – at first – different than adults'

- Even if they form correct sentences, their underlying representations are not necessarily the same

- ## My blue ball is gone!

Adult: pos.pr. adj noun   aux adv

Child: chunck 1   chuck 2

- Adult like linguistic categories emerge slowly with experience as children generalize over their experiences

Michael Tomasello

CONSTRUCTING A LANGUAGE

A Usage-Based Theory of Language Acquisition
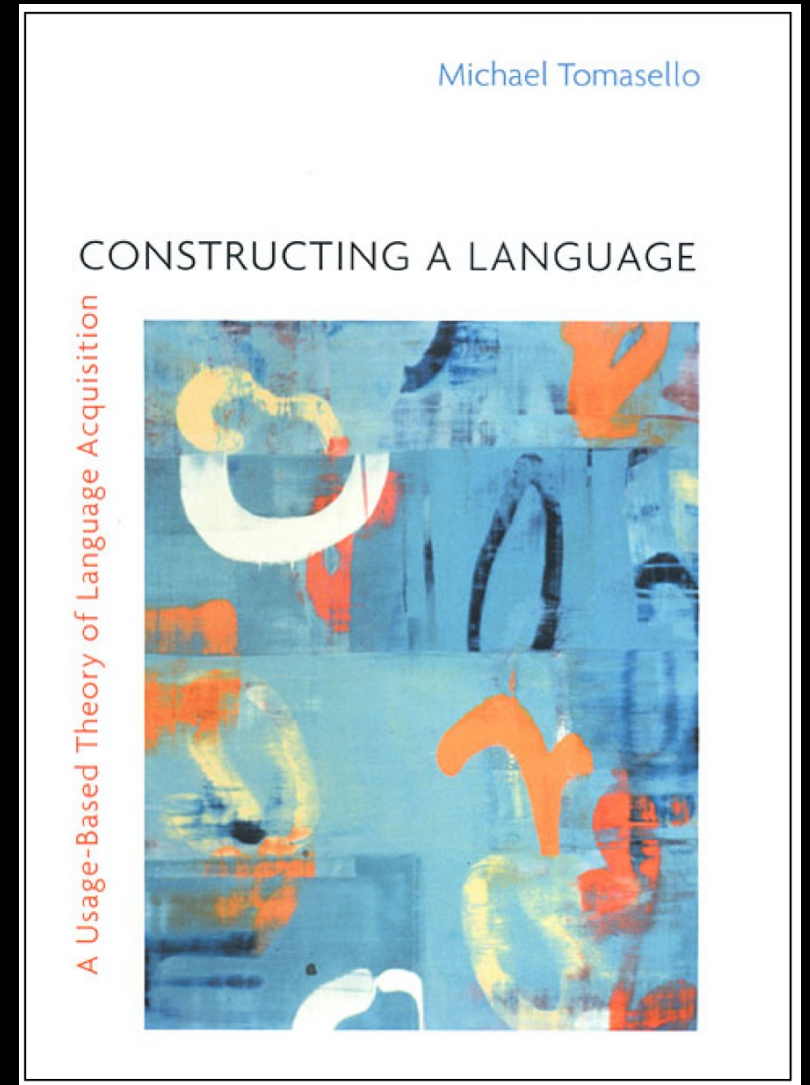
Childrens emerging linguistic knowledge

- Children might use a grammatical form – e.g. past tense – in one context, with one verb, but not generalize to other verbs

- Does the child have a generalized concept of nouns and verbs?

- The "WUG" experiment

The 'Tamming' experiment
Tomasello & Brooks, 1998

- Infants' syntactical productivity/creativity:
  - Intransitive vs. transitive verbs ("the car drives down the street"/"Paul drives the car down the street")

- 2 – 3 year-old children saw a novel activity and were told that "The sock is tamming" (a teddy bear did something to the sock which made it 'tam')

- In subsequent situation with new items (a dog and a car) the adult is asking: What is the doggie doing? (the dog makes the car 'tam').

- Only very few children in this experiment produced a transitive form suggesting limited knowledge of how verbs behave

Acquiring the first grammar

- Holophrases

  - Unanalyzed 'chunks'/'frozen phrases': are used as whole words/utterances

  - Learned 'by root' rather than 'by rule'.

  - Examples:

    One-word takes the meaning of a full utterance:
    - "Ball?" (where is the ball?),
    - "Water!" (I am thirsty),
    - "Gone!" (the object is gone)
    - Or short phrases: "its-mine!", "over-there", "my-turn!".

CHILDES, Sarah, 2.3 years

CHI	a eye she .
CHI	a yyy .
CHI	a bear .
CHI	yyy .
CHI	a bear xxx .
CHI	a .
CHI	xxx .
CHI	come here .
CHI	up there .
CHI	up there .
CHI	xxx .
CHI	up there .
CHI	bike .
CHI	ride bike .
CHI	xxx .
CHI	eat there .

Acquiring the first grammar

- Pivot schemas:

  - Children's first multi-word-constructions

  - A phrase that has a fixed part and an open slot that can take multiple alternatives:

    CHILDES, Sarah, 3 years

    - "where's X?",
    - "I want X",
    - "more X",
    - "Its an X", etc.

CHI	I want go doctor's .	1964–05–28
CHI	I want fix my xxx .	1964–05–28
CHI	I want I want xxx .	1964–05–28
CHI	want my finger .	1964–05–28
CHI	where bandaid ?	1964–05–28
CHI	I want do it .	1964–05–28

Acquiring the first grammar

- Verb island constructions (syntactical constructions centered around a verb):

  - E.g. 'to hit' has an open slot for the agent-'hitter' and the patient-'hittee', and the instrument

  - Example verb-island construction:
    - "_____hits _____ with a _____", or
    - "_____ gives _____ a _____"
    - "Cut___"

- A kind of constrained productivity?

CHILDES, Sarah, 3.5 years

CHI	I want Kellogg xxx can't play with it .
CHI	yyy ?
CHI	let's see he can play with it .
CHI	Chantilly .
CHI	he go hon +...
CHI	that that Mommy .
CHI	and get that other toys .
CHI	I going get that that that .
CHI	–um he's playing with the toys .

Linguistic productivity

- Only quite late in development, infants start to abstract and generalize across constructions (linguistic productivity)

- Before the age of 4, children do rarely use verbs outside the specific constructions in which they have encountered them (i.e. productively)

- The infant rarely generalizes linguistic knowledge from one verb to another: the fact that they use the past tense form for one verb does not mean that they will use any other verb in the past tense

TalkBank

The TalkBank System

TalkBank is a project organized by Brian MacWhinney at Carnegie Mellon University with the support and cooperation of hundreds of contributors and dozens of collaborators. The goal of TalkBank is to foster fundamental research in the study of human communication with an emphasis on spoken communication. Currently, TalkBank provides repositories in 14 research areas, as represented by the links on this page. Data in TalkBank have been contributed by hundreds of researchers working in over 34 languages internationally who are committed to principles of open data-sharing. These data are used by thousands of researchers resulting in many thousands of published articles. Data in TalkBank use a consistent XML-compatible representation called CHAT which facilitates automatic analysis and searching, using open-source and free programs we have developed.

System	Programs	Manuals
Ground Rules	CLAN	CHAT - CLAN - MOR
Hints on Downloading	MOR grammars	Tutorial Screencasts
Contributing	XML creator and XML Schema	SLP's Guide to CLAN and 中文
IRB Principles	Other Software	

Conversation Banks	Child Language Banks	Multilingualism Banks
CABank	CHILDES	Second Language Tutors
SamtaleBank	PhonBank	BilingBank
ClassBank	HomeBank	SLABank

Clinical Banks	Clinical Banks	Other
DementiaBank	AphasiaBank	Database Versioning
RHDBank	ASDBank	TalkBank DB - Search the Databases
TBIBank	FluencyBank	

Recording	Resources	Symposia
Digital Video	CLARIN Knowledge Center	CMU June 5-8
Digital Audio	GoogleGroups	
Picture Stimuli		

22

The `childes-db` project aims to make CHILDES transcripts more accessible by reducing the amount of preprocessing (e.g., CLAN or specific preprocessing libraries) and by making the individual tokens, utterances, transcripts, and corpora available in a tidy, tabular format that is accessible across programming languages. We release new versions of this dataset periodically to facilitate reproducibility. We also provide an R package (childesr) and a Python package (childespy) which allow users to access this database without having to write complex SQL queries.

Citation policy

If you use `childes-db` to access CHILDES in your research, please note the database version you used (i.e., `2018.1`) and cite:

1. The `childes-db` paper in *Behavior Research Methods*:

   *Sanchez, A., *Meylan, S.C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). "childes-db: a flexible and reproducible interface to the Child Language Data Exchange System." *Behavior Research Methods* 51 (4), 1928–1941.

   * indicates co-first authorship.
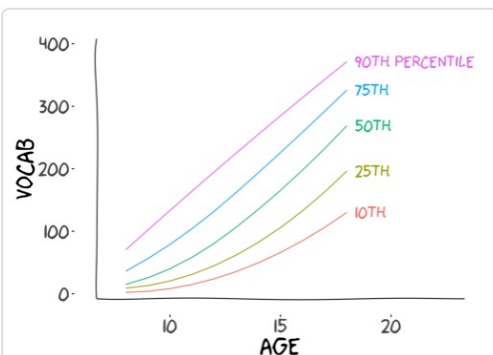2. CHILDES itself – both the database and the corpora you use – following the Talkbank policy.
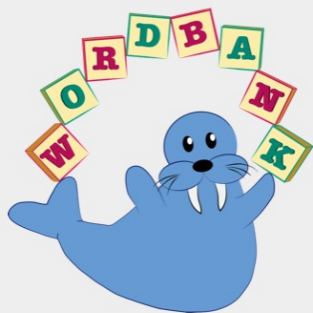
Wordbank

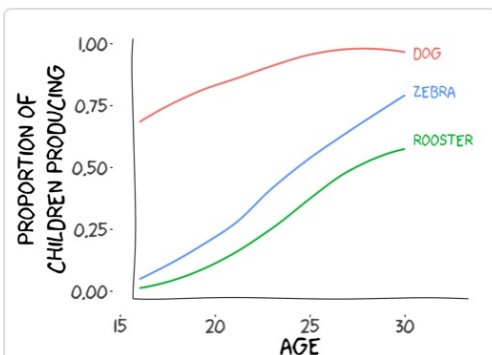An open database of children's vocabulary development

Wordbank contains data from 75,144 children and 82,983 CDI administrations, across 29 languages and 56 instruments:
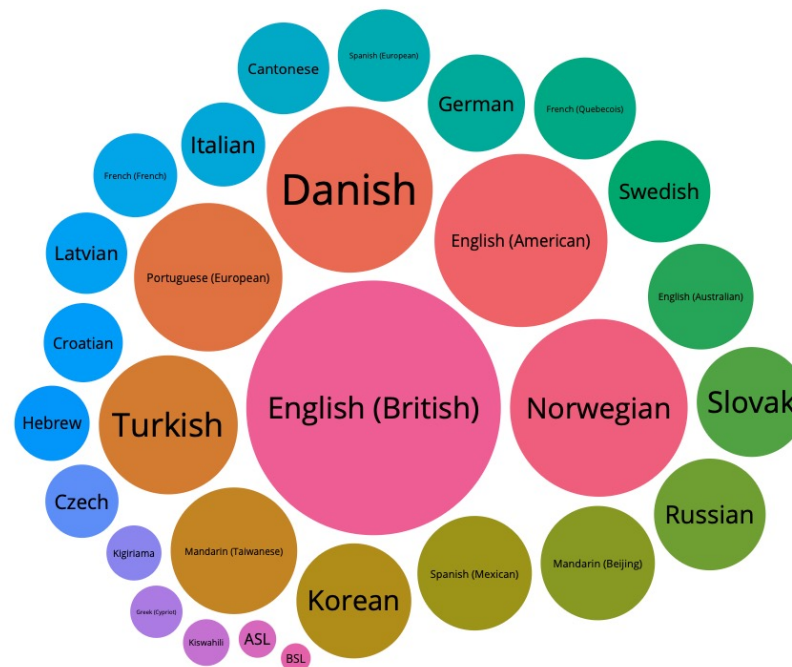
Vocabulary Norms

Explore vocabulary size growth curves for various languages and demographic groups.

Item Trajectories

Explore trajectories of individual words, word categories, and grammar items.

Take Home

- Following Tomasello, children acquire their linguistic knowledge through language *use*

- They are sensitive to statistical regularities, which supports segregation of the speech signal into words

- And they rely on "fast mapping" inferences to connect words and meanings

- Evidence suggest that children do not initially possess adult abstract linguistic categories and representations such as noun and verb (Tomasello 2003)

- Rather, infants seem to acquire a repertoire of concrete linguistic constructions and only slowly abstract more general structural properties

- Only eventually they generalize between items to form adult-like linguistic categories and use those productively

- Tomasello: the generative underlying algorithm that enables us to combine a finite set of words to an infinite set of sentences is this an *emergent* product of extended language use – not an *innate* capacity