Luigi Rizzi **TGLA 2.**

The experimental study of the initial cognitive state

0. An instinctive tendency to speak.

(1) "As Horne Took, one of the founders of the noble science of philology, observes, language is an art, like brewing or baking. But writing would have been a much more appropriate simile.
It certainly is not a true instinct, as every language has to be learnt.
It differs, however, widely from all ordinary arts, for **man has an instinctive tendency to speak**, as we see in the babble of our young children, while no child has an instinctive tendency to brew, bake, or write… "

Charles Darwin, The Descent of Man (1871)

(2) "There is a tendency among social scientists to regard language as a wholly learned and cultural phenomenon…
"[but] … man may be equipped with highly specialized biological propensities that favor and, indeed, shape the development of speech in the child … roots of language may be as deeply grounded in our natural constitution as, for instance, our predisposition to use our hands."
… "clarity of the problem of the **biological foundation of language** is of utmost importance in formulating both questions and hypotheses regarding the function, mechanism, and history of language."

E. Lenneberg 1964, see also Lenneberg 1967 Biological Foundations of Language

(3) "From the point of view that I adopt here, the fundamental empirical problem of linguistics is to explain how a person can acquire knowledge of language."

N. Chomsky 1973 Conditions on Transformations

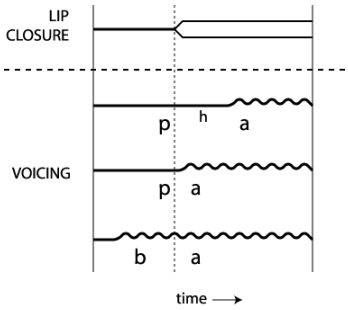
1. Categorical perception in adults and babies.

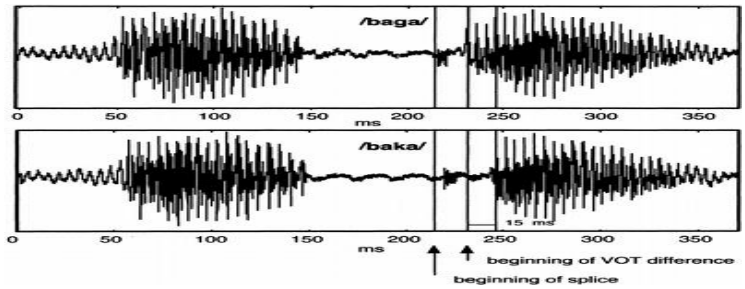
(1) A. Liberman, M. Studdert-Kennedy et al. (1967) had observed that speech perception is categorical in adults: a continuously varying stimulus between [pa] and [ba] is perceived categorically as either [pa] or [ba] with a sudden phase transition around a natural perceptual boundary.

The relevant physical parameter is Voice Onset Time, VOT: voicing starts before or around the point of release of air in voiced stop consonants, and slightly later in voiceless stop consonants. The VOT can be manipulated experimentally, so that physically intermediate stimuli can be created ranging from a "perfect" [ba] and a "perfect" [pa].

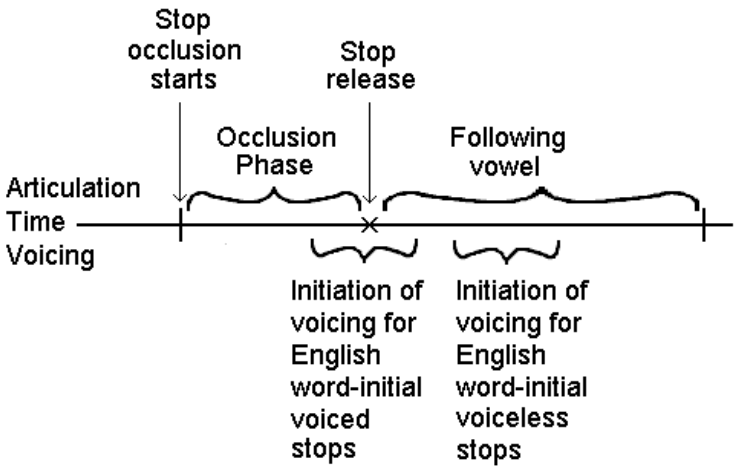
But then speakers, asked to judge physically intermediate stimuli, do not perceive something intermediate: up to a certain VOT they perceive [ba] and then they start perceiving [pa]. So, a continuous physical stimulus is perceived as belonging to one or another discrete category

(2) Voice Onset Time (VOT):



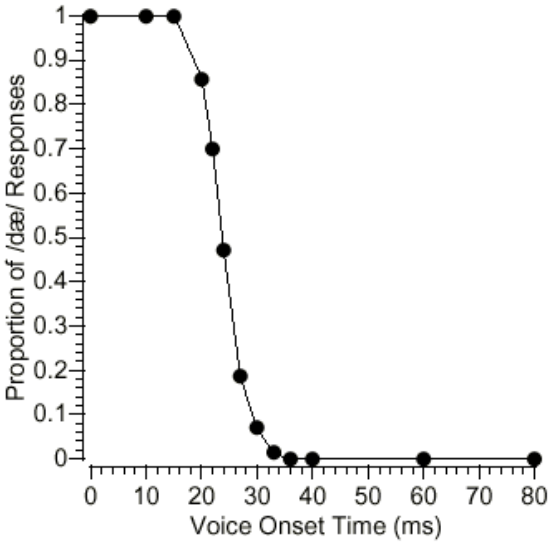


/g/ vs /k/: if the voice onset is in the interval between -20 and +20 ms, the segment is perceived as voiced:



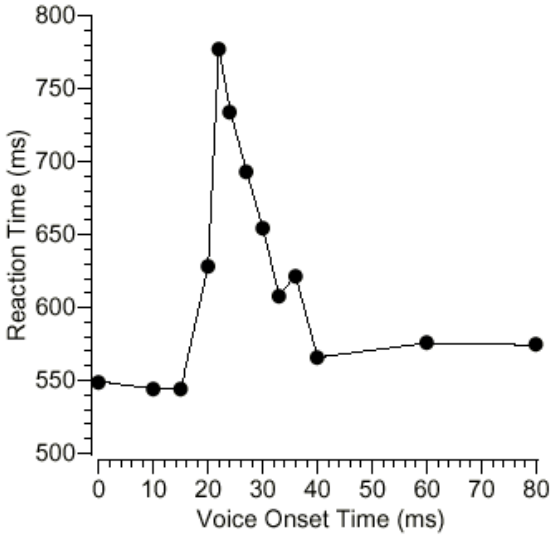
(3) Categorical perception:

A



Responses for discrimination of /da/, as opposed to /ta/

B



Timing of response: the decision is more difficult around the perceptual boundary

Figure A shows the sudden transition in the discrimination of a segment as voiced or voiceless at a perceptual boundary. Figure B indicates the reaction time: around the perceptual boundary the discrimination is more difficult, and the reaction time is higher.

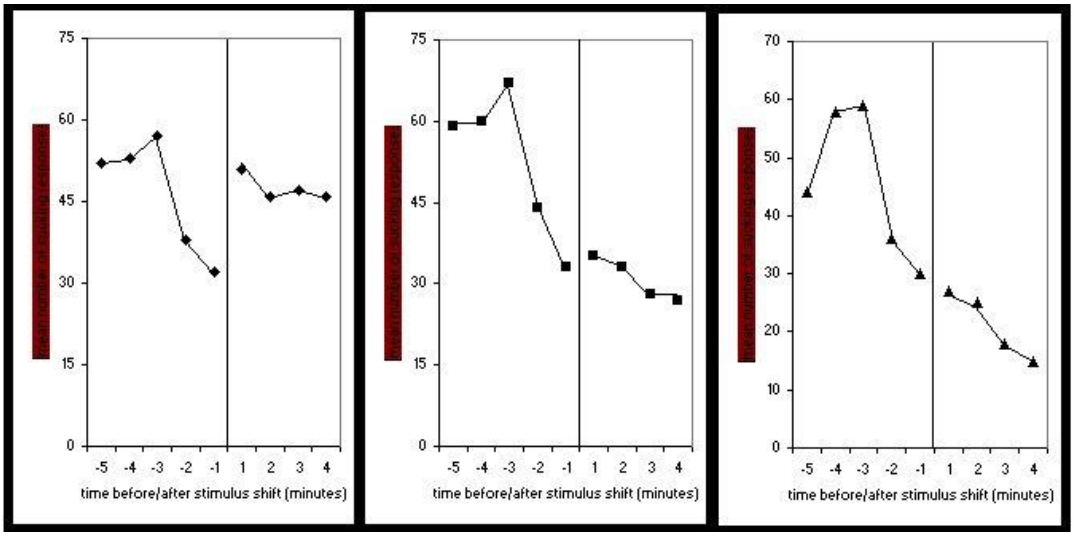
(4) What about babies? Is categorical perception an innate property, or is it acquired through experience? The study of the discrimination capacities of the baby is conducted through the main behavioural response s/he can offer: the sucking rate ("succion non-nutritive", "non-nutritive suction", "high amplitude sucking procedure", etc.).

Habituation phase: the baby hears a stimulus (a sound,...) which is repeatedly presented as a function of the sucking rate. For some time sucking increases, then the baby gets habituated to it, and the rate decreases.

Experimental phase: two groups of babies: The experimental group hears a new stimulus (a different sound,...); the control group continues to hear the same stimulus. If the sucking rates diverge, this means that the babies are able to discriminate between the two stimuli.

(5) Eimas et al. (1971) : categorical perception of [pa] / [ba] at 4 months.

I. pa / ba: stimuli crossing a perceptual category boundary	II. Stimuli within the same category boundary	III. Same stimulus



I and II present experimental stimuli that are equally different in physical terms from the habituation phase to the experimental phase (same difference in VOT), but the perceptual boundary is crossed in I but not in II. The pattern in II is statistically indistinguishable from the pattern in III, in which the same stimulus is presented in the habituation and experimental phase.

(6) Is categorical perception specific to our species? In the sixties it was considered uniquely human, part of the cluster of properties which make "speech special". Later, it turned out that other species (monkeys and other higher mammals, birds,...) have it (even though they require a much longer training than humans). Presumably it is a feature of the auditory system of all higher animals, and an ability which has a long evolutionary history, and which was "recruited" for language, with other preexisting abilities, when our species acquired the fundamental combinatorial capacities that make human language so special.

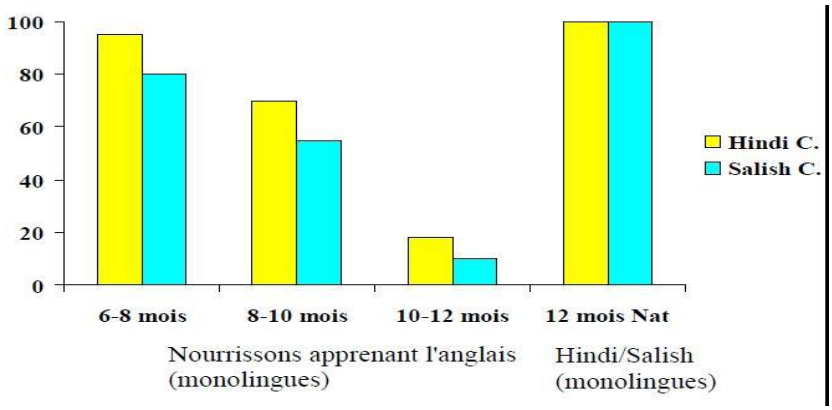
2. Learning by forgetting.

(9) "L'apprentissage d'une langue se traduit en fait par une perte partielle, par la sélection de certains contrastes et l'oubli d'autres, qui ne sont pas pertinents… Les tout petits bébés discriminent tous les contrastes auxquels ils sont confrontés, tandis que vers huit ou dix mois, ils montrent la même aptitude, mais seulement pour les contrastes de leur propre langue. A douze mois, ils se conduisent comme des adultes et ne sont plus sensibles aux contrastes étrangers" (Mehler & Dupoux 1990, 236).

(11) Werker & Tees 1984 : /ta/ vs /ʈa/ dental vs retroflex stops are discriminated by infants at 6-8 and 8-10 months (they used a head turning procedure). Afterwards, they are discriminated only if the target language uses the features as distinctive (Hindi, but not English).

Glottalized vs uvular velar stops are discriminated by infants till 8-10 months ; afterwards, they are discriminated only if the target language uses the features (Salish, but not English).

(12) Table from Werker & Tees (1984)



NB: the first three columns indicate the progressive loss of the capacity to discriminate in babies exposed to English. The fourth column indicated the persistence of the capacity in babies exposed to languages that use that particular phonetic distinction.

(13) The loss of contrasts around the first year of life holds for contrasts that are similar to those used by the acquired language. The ability to perceive distant contrasts (eg., contrasts in Zulu clicks for learners of English) is not affected. Close contrasts are affected: Tsushima, et. Al 1994: /l/ - /r/ is perceived by 6-8 months old Japanese learners, but lost at 12 months.

(14) Learning by forgetting is not a loss in the system of acoustic perception: it is a functional reorganisation of the mental representation of linguistic sounds, a more central cognitive development which makes it possible for the child to disregard distinctions which are useless for distinguishing words in his language. Werker (1995): this loss facilitates lexical learning, as it reduces the features that the child must pay attention to, so it is not accidental that this development coincides with the beginning of the process of building up a large lexicon.

(15) Learning by forgetting takes place in a specific time window. Infants generally stop discriminating non-native contrasts around 12 months of age. But Werker & Hensch (2015) have shown that preterm infants born at 28 weeks (3 months before term) gestation stop discriminating non-native contrasts at 15 months. So the loss of the sensitivity to non-native contrasts is a function of the time from conception, not a function of the time after birth.

(16) Kuhl (1993) : in perception, infants converge to the vowel system of the target language around 6 months.

(18) The functional reorganization is not irreversible during childhood: children exposed to a second language very early on (say at 2, 3, etc.) acquire the distinctions in the second language in a way which is indistinguishable from monolinguals and bilinguals from start. But there clearly are limits: adult L2ers speak with an accent and don't easily perceive certain distinctions. Pallier et al. (1997) have shown that Spanish speakers exposed to Catalan around 6 tend not to master the distinction /e/ - /ɛ/, which is not used in Spanish. It may be already too late for the functional reorganization to take place in the second language.

(19) Selective vs. instructive theories of learning:
"Looking back into the history of biology, it appears that wherever a phenomenon resembles learning, an instructive theory was first proposed to account for the underlying mechanisms. In every case this was later replaced by a selective theory"
…[in immunology, an animal] cannot be stimulated to make specific antibodies, unless it has already made antibodies of this specificity before the antigen arrives"
… [selective mechanisms are mechanisms] through which products which are already present in the system prior to the arrival of the signal are selected and amplified".
(Jerne 1985; Piattelli-Palmarini 1989).

3. Discriminating different languages.

(20) Vouloumanos & Werker (2007): babies at birth prefer listening to speech than artificial non-speech stimuli of comparable complexity. So, there seems to be an in-born bias to pay attention to linguistic sounds. But much more specific preferences had been observed earlier.

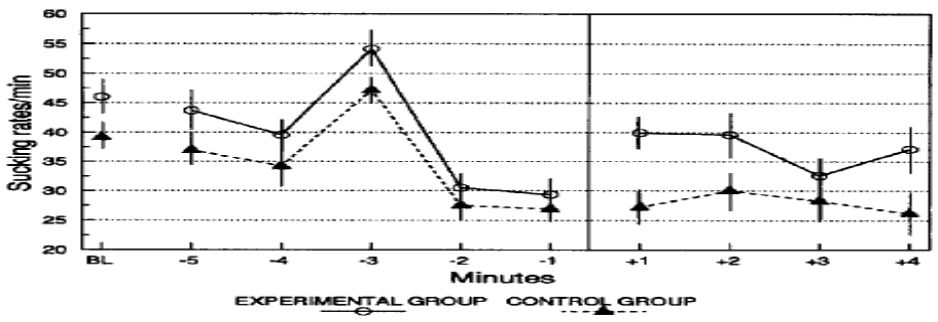
(21) Mehler et al. (1978): the baby manifests a preference for the mother's voice a few days after birth, but what is crucial is the dynamic aspect of the mother's speech, the particular sequence of events: if the tape is played backward, no preference is shown.

(21') Mehler et al (1988): preference for the mother's language at 4 days (a bilingual subject, not the mother, speaks in French and Russian; the transition is detected by the baby). Again, no preference if the tape is played backward (they are not sensitive to coarse cues like mean energy).

(22) Is the foetus habituated to his mother's voice, or language? The auditory system starts being functional from the 25th week and is well-developed around the 35$^{th.}$ (term at 40th week). The foetus then perceives sounds (high frequency sounds are filtered out, but low frequency sounds reach the foetus through the amniotic liquid).The habituation to the mother's voice, or to specific linguistic texts (a particular poem, even read by another person) is revealed by a deceleration of the heart pace, showing the detection of an interesting stimulus.

(23) Later the same authors showed that the newborn baby is able to discriminate different languages distinct from the mother's language. For instance, a baby of 4-5 days exposed to French can discriminate between English and Japanese, or English and Italian (Mehler & Christophe 1995), which he has never heard before. So, the pre-birth or very early experience is not necessary for the language discrimination capacity.

(23') Nazzi et al. (1998) discrimination capacities of 4 day old babies exposed to French:



Control group: English – English; experimental group: English - Japanese

(24) How do infants discriminate languages well before any lexical knowledge? The specific properties of individual segments are not necessary: if frequencies higher than 400 Hz (normal range of audible sounds is 20 Hz – 20.000 Hz) are filtered out, individual segments are not recognizable, but babies still discriminate. Mehler and his collaborators concluded hat the crucial property could be either intonation or rhythm.

(25) Ramus, Nespor & Mehler (1999) dissociated the two factors: they took sentences in English and Japanese, and transformed

- all the segments into “a”, with intonation (variation of F0 frequency) retained, and
- all the consonantic segments into “s” and the vocalic segments into “a”, with intonation eliminated:

I. The next local elections will take place during the winter

II. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa (with intonational contour)

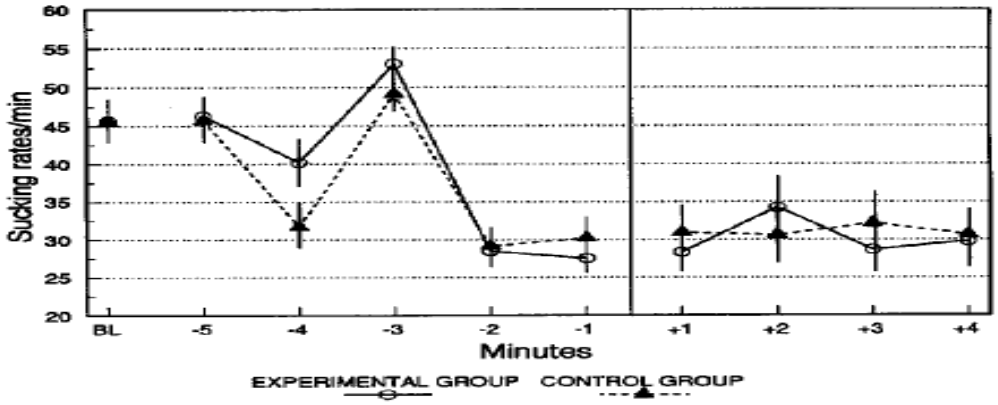
III. sasassssasasasasassassasasssasassasasasasasssasassas (without intonation contour)

So that “go” becomes “sa”, “cat” becomes sas”, “next” becomes “sasss”, “smart” becomes “ssass”. Adults and children were unable to discriminate English and Japanese on the basis of intonation only, and were able to discriminate on the basis of **rhythm** only.

(26) Rhythm refers to the periodicity with which certain events recur. Linguistic rhythm may correspond to the periodicity of consonant-vowel alternations. So, structural linguists had distinguished between “syllable-timed” and “Stress-timed” languages (Pike 1948). According to one hypothesis, they involve isochrony of syllables and stress intervals, respectively. A third type is mora timed languages, based on isochrony of the mora, a subsyllabic unit (light syllables (CV) consist of one mora, heavy syllables (CVV, CVC) of two). So, we have at least three rhythmic classes:
- Stress-timed languages (English, Dutch, Polish,…)
- Syllable-timed languages (Italian, Spanish,…)
- Mora-timed languages (Japanese,…)

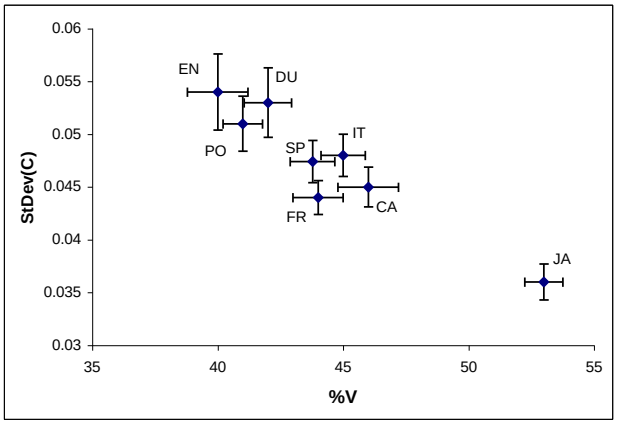
(28) Both adults and newborn babies are able to discriminate on this basis between two languages belonging to different rhythmic groups, but not within the group: newborn babies acquiring French can discriminate between English and Italian, but not between Dutch and English (Nazzi, Bertoncini & Mehler 1998).

Nazzi et al 1998: lack of discrimination by French 4 day old between English and Dutch



Control group: English – English; experimental group: English – Dutch.

(30): Results from the study by Ramus, Nespor and Mehler (1999).
Distribution of languages over the percent vowel per utterance versus the Standard deviation of the consonants. Error bars represent ± one standard error.



(31) Clearly a language with syllable types (C)(C)(C)V(C) will have a lower proportion of vowels and a higher variability of intervowel consonantal cluster than a language allowing only (C)V. These criteria permit the quantitative identification of classes such as

i. English, Dutch, Polish,…
ii. Spanish, French, Italian, Catalan,…
iii. Japanese, Tamil, Hawaiian...

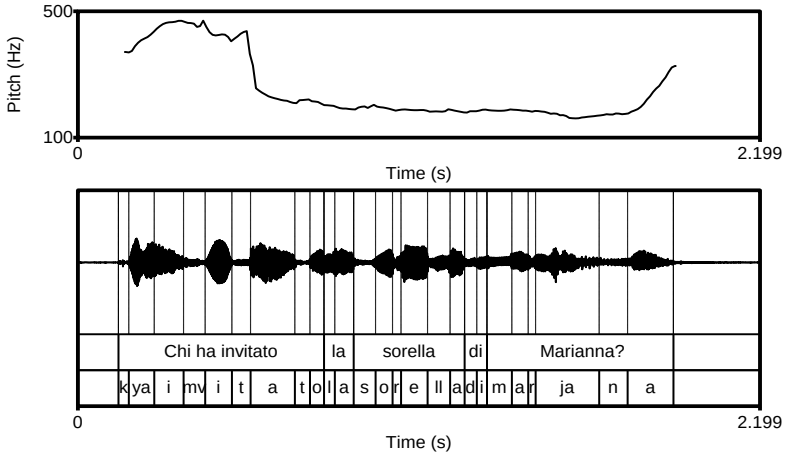
(32) Mora-based languages tend to have smaller syllabic repertoires (2-3 in Japanese and Hawaiian) than syllable-timed (8-9 in Spanish and Italian) and stress-timed (16-19 in English and Dutch); they also tend to have longer words, whereas in stress-timed languages, which allow for many different monosyllables, most of the basic frequent vocabulary is monosyllabic. (Zipf's law: high frequency words tend to be shorter than low-frequency words). So, one may conjecture that the very early sensitivity to rhythm may guide the acquisition of syllable types and of the basic word length, thus helping the acquisition of the phonology and lexicon of the language: the rhythmic class provides information for the early fixation of parameters on syllable structure, and on the average word length to be expected.

Eg., monosyllabic words are 42% of the lexicon a Dutch child has at 3, but only 4% of the lexicon of the Italian child. Conversely, trisyllabic or multisyllabic words are 35% of the early Italian lexicon, but only 13% of the early Dutch lexicon.

6. Segmentation

(1) Normally, word boundaries are indicated in writing, but not in the speech signal

Points of reduction of phonetic emission don't correspond to pauses between words, but to obstruents, typically word internal:



In many cases the phonetic signal can be segmented in different ways, giving rise to ambiguities:

(2) I scream - Ice cream

The adult speaker can segment the input string through his/her lexical knowledge: certain syllable sequences match his lexical representations and are then segmented and identified as words; in some cases two distinct representations can be associated to the same input string, and ambiguity arises.

(2) But for the child a paradox arises: the child must build a mental lexicon, so initially he cannot rely on any previous lexical knowledge to segment the input string; then, how can the initial segmentation take place?

(3) Different strategies are used, none of which may be sufficient, but which may converge to make word segmentation possible:

- words presented in isolation by adults (but this happens relatively rarely: maybe it works for the baby's name, which is acquired already at 5-6 months, but most other words are typically not presented in isolation).

- Statistical regularities: two syllables (or larger chunks) within a word have a higher probability of occurring in sequence than two syllables (or larger chunks) occurring across words (gi-raffe vs are-smart). The baby is a good statistician (Saffran, Aslin & Newport (1996): already at 8 months the baby figures out transition probabilities with a 2 minutes

exposure of an invented language) so he/she may exploit these different transitional probabilities to segment the input signal.

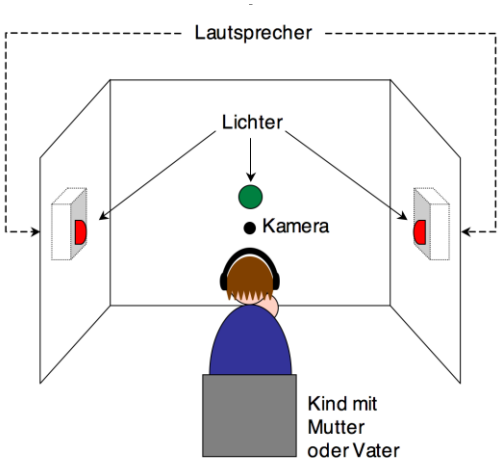
- Phonological bootstrapping: exploit various cues in the phonological representation (suprasegmental cues, phonotactics based on language specific constraints on syllable structure, etc. associated to the speech signal to segment the words).

(4) Phonological bootstrapping: if word boundaries are normally not represented in the speech signal, boundaries between clauses and other phrases are normally expressed by certain cues (pauses, resetting of the intonational contour expressed by the fundamental frequency, which declines and raises again at the beginning of a new unit). These cues allow the child to narrow down the problem by chunking the input string into smaller units, each containing only a few words.

(6)a While Mary was reading the book fell from the table
b While Mary was reading // the book fell from the table
c While Mary was reading the book // …

(7) There are acoustic cues signalling clausal boundaries:
a. pauses
b. syllable lengthening
c. the intonational contour (F0) tends to decline at the end of the clause

(8) Infants are sensitive to these cues. For instance, Hirsh-Pasek and Golinkoff (1987), using the head-turning paradigm, showed that 7-10 months old infants learning English prefer listening to stories with pauses inserted at clause boundaries (hence consistent with the other cues), rather than inserted within clauses (and inconsistent with other cues). Other studies with the sucking paradigm showed a much earlier sensitivity to such cues.
Head turning paradigm:



(9)a A man came to Mary's house # he was accompanied by a woman and a child # the child was crying…
b A man came # to Mary's house he was accompanied by # a woman and a child the child # was crying…

(10) Gerken, L.-A., Jusczyk, P.W., & Mandel, D. (1994). 9 months olds' sensitivity to phonological versus syntactic phrases. *Cognition, 51,* 237-265. Infants preferred and listened longer to structures

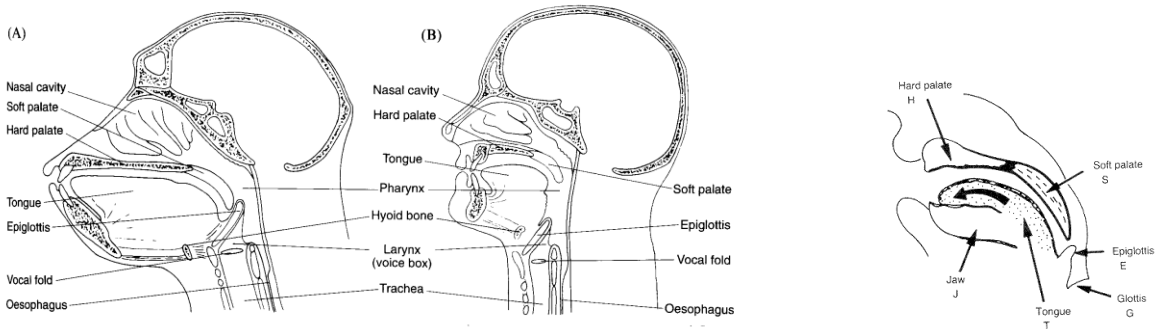
with pauses inserted between the subject and the verb, as in (11)a, w.r.t. structures with pauses inserted between the verb and the object:

(11)a The boy # saw the girl
b The boy saw # the girl

7. Production

(35) The oral tract of babies resembles the tract of apes, with a higher larynx and a reduced internal space. The larynx starts to descend at 4-6 months and the maturation of the system is completed only at 3 years. (non-descended larynx permits simultaneous breathing and swallowing)

(36) Vocal tract in ape and man



Vocal tract in ape in man in infant

(38) The fact that linguistically relevant sounds are not produced in the first months (while sophisticated perception of linguistic sounds is already operative) is certainly related to the immature vocal tract, but may be a more general fact: Petitto et al. (1991) have shown that manual babbling in deaf infants starts about at the same time as vocal babbling.

(39) From birth: cries and other vegetative sounds; from 4-5 months some modulated vocalic-like sounds, but no other sign of language-like production.

(40) 6 months: beginning of babbling, production of syllabic structures with no associated meaning. Babbling is typically a sequence of CV syllables; can be
- canonical (repetition of the same syllable: babababa) or
- variegated (badada, badabada).

(41) Babbling often starts abruptly, and in the initial weeks it seems to conform to universal rules:
- Stop consonants (p, b, t, d) and nasals (m) are more frequent than fricatives (f, v, s,…) and liquids (l, r);
- Low vowels (a) are more frequent than high vowels (i, u)
- Syllables with no coda are more frequent than closed syllables

This "universal" babbling starts automatically at this age, independently of the acoustic stimulus: it is also found in deaf children with analogous characteristics (but the point is controversial).

(42) 8-10 months: there is a progressive fine-tuning to the properties of the language the infant is exposed to, so babbling shows signs of language specificity. Boysson-Bardies: 70% of babbling

produced at 8 months can be discriminated by adult speakers of French, Arabic, Cantonese; presumably, the adult speakers are sensitive to language-specific segmental and prosodic properties which are assimilated in the babbling. This phase is sometimes called "conversational babbling": the infant reproduces conversational patterns typical of adult conversation.

(43) Boysson-Bardies et al (1991) : the frequency of phonemes in babbling at 10 months matches the relative frequency of the ambient language : labials more frequent in French than in English, dentals more frequent in Japanese than in French, and these preferences are matched in babbling.

(44) So, the age of 8-10 months shows converging signs of acquisition of the sound pattern of the particular language:

- in perception through the forgetting of phonetic contrasts not selected by the target language, and
- in production with the emergence of language specific properties of babbling.

(45) Babbling is modality independent : Petitto & collaborators (1991, 1995) have shown that deaf children exposed to sign language begin to babble manually around the same age (6 months), with the same characteristics:

- "syllabic" structure, signs which do not have an independent meaning, but are the constituents of meaningful signs,
- use of a subset of the basic configurations used in the language,
- canonical or variegated,
- with fine-tuning to the repertoire of the community around 10 months as speaking children.

These authors claim that manual babbling in this community is distinct from the class of spontaneous manual gestures that every child makes according to several parameters: fewer forms, movements and hand orientations in ordinary manual gestures, and different frequency (lower frequency in babbling than in ordinary gestures).

(46) Goldin-Meadow (1998, 2011): deaf children not exposed to a conventional sign language invent their own system of signs to communicate with the family (home sign systems). Such systems are remarkably uniform across cultures (Goldin-Meadow studied the similarities between American and Chinese home sign systems).

(47) So, language language can be considered an amodal capacity: it is somehow linked to speech and the oral modality (otherwise human communities would presumably arbitrarily choose different modalities for externalisation), but it is very easily adaptable to another modality when the auditory articulatory channel is precluded. In any case, the ease of the adaptability to another modality suggests that the connection to speech and the oral modality is not a deep factor of language.

(48) 10-12 months: the first recognizable words are produced by the child and the one word stage begins. Babbling and one word stage overlap for 4-5 months.

(49) As for the brain organization, the neural circuitry involved in language is still organizing itself in babies. Nevertheless, brain imaging studies show that there is an innate tendency to process linguistic properties in the left hemisphere. For instance, Dehaene-Lambertz et al (2013) show, through functional near-infrared spectroscopy, that even premature babies born after 28 weeks show a left-lateralize response to linguistic minimal pairs (/ba/ - /ga/), and a right-lateralized response to non-linguistic contrasts (male-female voice).

8. Determining the meaning of nouns

(13) Once words are segmented, they must be mapped to a meaning. Assigning a meaning to a noun is relatively easier than assigning a meaning to a verb on the basis of the information provided by the extralinguistic context of an utterance. Still it is a difficult problem: a given extralinguistic situation may be described in innumerable different way; even an isolated noun may refer to an object or to one of its parts, or to broader aspects of the situation (Quine's "gavagai" problem); abstract nouns like sincerity, freedom, but also yesterday, this morning, etc. do not correspond to any obvious entity salient in the extra-linguistic context of discourse.

(14) One important help may be provided by the innate capacity to identify the focus of attention of the speaker, ie by observing the direction of the speaker's gaze (Baldwin 1991). In the absence of this kind of cue, offering a situation of shared focus of attention, the association between word and referent does not take place, at least not in very young children (13 months: Woodward, Markman et al. 1994).

(15) And then the child probably follows certain strategies or biases which narrow down the class of possible meanings of a given word form (Markman 1994):

- Whole object: a novel label is likely to refer to a whole object, not to its parts.

- Mutual exclusivity: each object will have one and only one label (this interacts with Whole object: if a new label is presented for an object which already has a name, the hypothesis is considered that the new label may refer to a part of the object, or to a larger category): there is a strong bias against synonymy.

9. Determining the meaning of verbs

(16) Verbs appear later than nouns in the initial productions. In the first lexicon of 50 words, verbs are a tiny minority (estimates vary considerably: Bates has 3% in English, Boysson-Bardies 13% in French, etc., while nouns are the core of the early lexicon).

(18) This may well be related to the difficulty of figuring out verb meanings: Gillette et al 1999: adults guessed intended nouns in silent videos 45% of the times (when the bell rings, guess what N they are pronouncing) , but verbs only 15% of the time (when the bell rings, guess what V they are pronouncing).

(19) Syntactic bootstrapping (Gleitman 1990, etc.): verb meaning may be determined in part by looking at the syntactic environment. Through the principles determining the syntax-semantics interface, the syntactic environment of a verb is very informative on the verb's argument structure, hence it narrows down the possible verb meanings (which can then be more easily determined by the context).

(20)a John gorped that Mary would come

b Bill sibbed

c Tom gobbed Mary a book

d Peter stog from Milan to Naples

(21) Naigles (1990): the syntactic form of a sentence narrows down the possible meanings. So, given a transitive sentence with an invented verb like (i), children at 2 looked preferentially at a picture displaying a causative action (ie, the duck pushing the bunny) then pictures in which both characters performed a monoargumental action (both characters rotating their right arm), and the other way around for (ii). So, a transitive clausal frame preferentially selects a causative (biargumental) action.

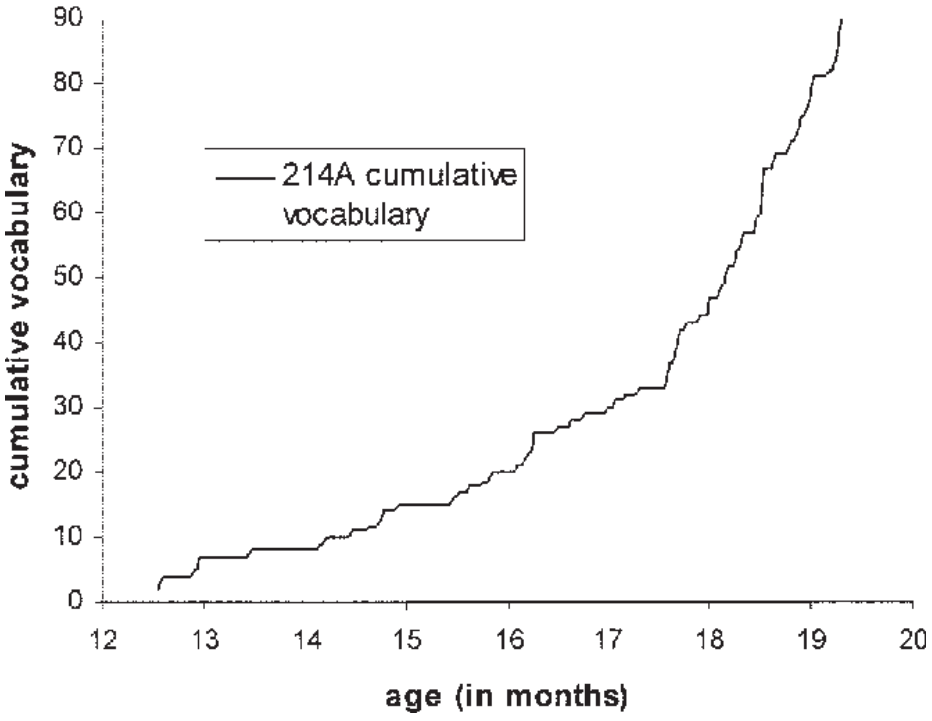
(i) The duck is gorping the bunny

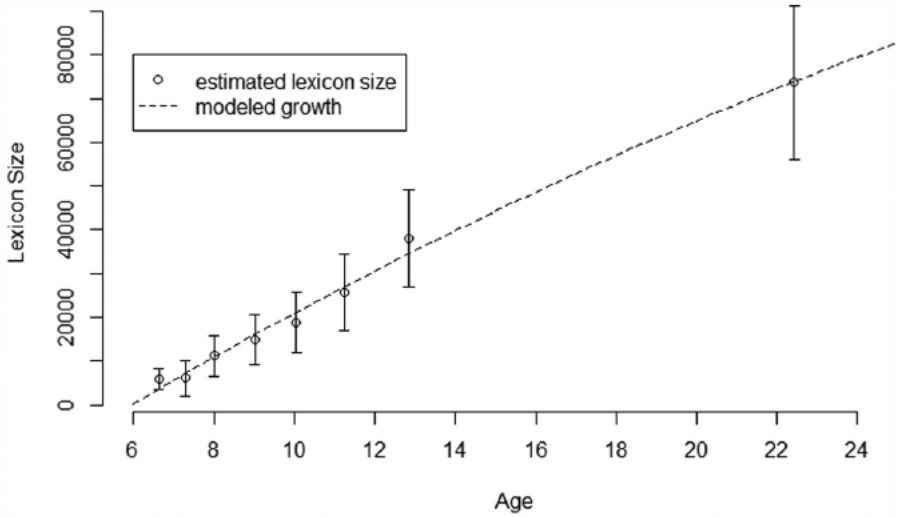
(ii) The duck and the bunny are gorping

(21) Semantic bootstrapping of syntax (Pinker 1984), etc.): the child is innately endowed with certain principles of the form-meaning interface expressing the Canonical Structural Realisation of certain semantic entities, such as:

- a word for a thing is a N
- a word for an action is a V
- a word for a quality/attribute is an A

(24) Lexical acquisition is initially slow: first recognizable words around 10-12 months, first vocabulary of 50 words in production around 18 months. Then at some point between 18 and 24 months, lexical acquisition becomes much faster, and the child acquires between 5 and 9 new words per day for several years.




estimated lexicon size
modeled growth
Lexicon Size
0
20000
40000
60000
80000
6
8
10
12
14
16
18
20
22
24
Age