

Cours 2013:

Le bébé statisticien

Stanislas Dehaene

Chaire de Psychologie Cognitive Expérimentale

Cours n°6

L'apprentissage de règles linguistiques

En quoi consiste la connaissance d'une langue?

La connaissance de la langue dépasse, de toute évidence, celle d'une simple matrice de probabilités de transitions entre mots.

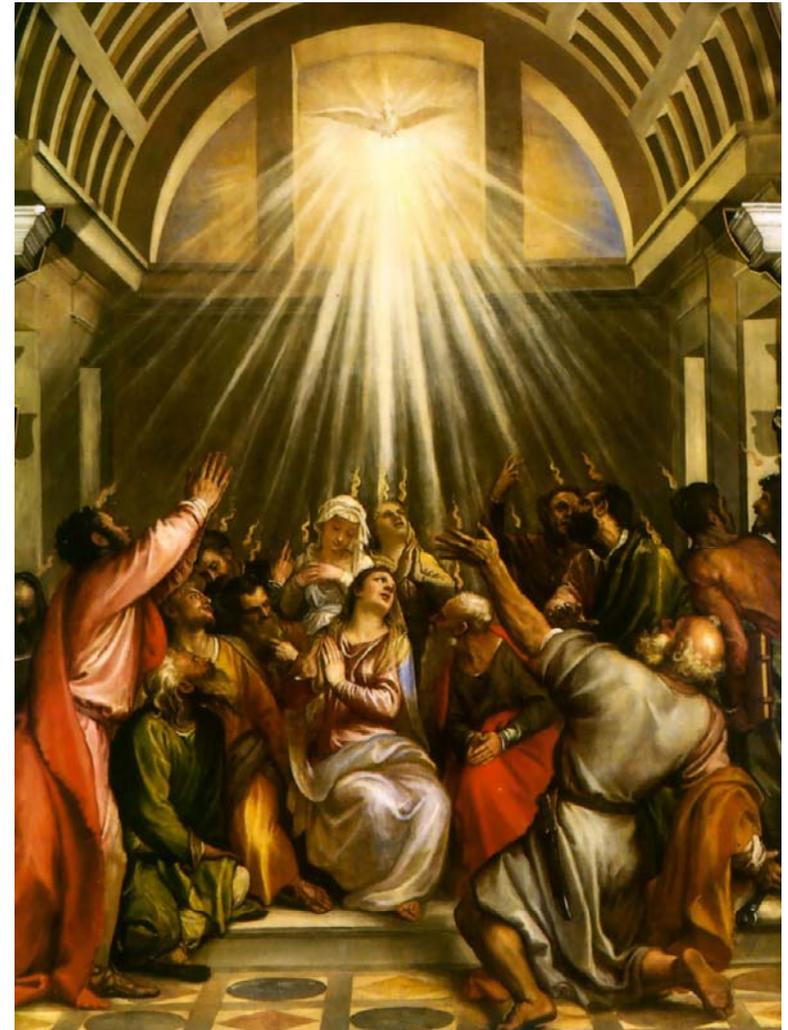
1. Nous devons, au minimum, considérer des règles plus abstraites qui font intervenir des **catégories** de mots (noms, adjectifs, verbes).

Chomsky (1957) "Colorless green ideas sleep furiously" versus "Ideas colorless sleep furiously green"

2. Même les probabilités de transition entre catégories de mots restent insuffisantes.

Toutes les langues possèdent une structure en constituants (en « arbres »), qui implique que les règles syntaxiques se traduisent par des dépendances à distance variable et arbitraire.

Exemple: is sing-ing; is (happily) singing; is (very happily) singing... etc



L'apprentissage des principes mêmes de la syntaxe peut-il s'expliquer par un réseau bayésien hiérarchique?

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*.

Selon Chomsky, les principes de la syntaxe ne peuvent pas être inférés par apprentissage statistique. Par exemple, comment l'enfant pourrait-il découvrir que la formation des questions fait appel à une structure arborescente et non pas linéaire?

The boy is smiling. Is the boy smiling?

The boy who is smiling is happy.

* (1) Is the boy who smiling is happy?

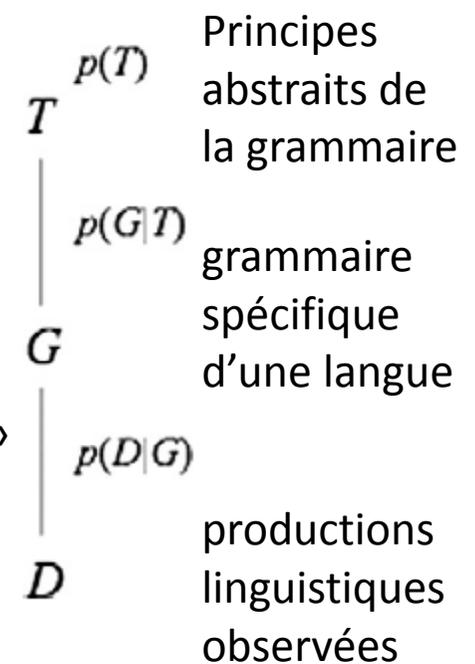
(2) Is the boy who is smiling happy?

Les enfants n'entendent presque jamais d'exemples de type (2).

"it is quite possible for a person to go through life without having heard any of the relevant examples that would choose between the two principles." (Chomsky, 1971). Principe de « pauvreté du stimulus » et inférence de l'existence d'une « grammaire universelle ».

Perfors et al. (2011) attaquent ce point de vue en montrant qu'il est possible d'utiliser les phrases entendues pour sélectionner les principes abstraits de la grammaire universelle.

Ils montrent quantitativement que, à partir des entrées que les enfants reçoivent (base Childes), l'hypothèse d'une grammaire non contextuelle (arborescente) est 10^{43} fois plus probable que celle d'une grammaire linéaire!



Comment séparer les mots à contenu des mots de fonction?

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115(3), 444-457.

On distingue classiquement les mots de classe fermée ou mots grammaticaux (*pronoms, auxiliaires, prépositions...*) des mots de classe ouverte ou à contenu (*noms, verbes*).

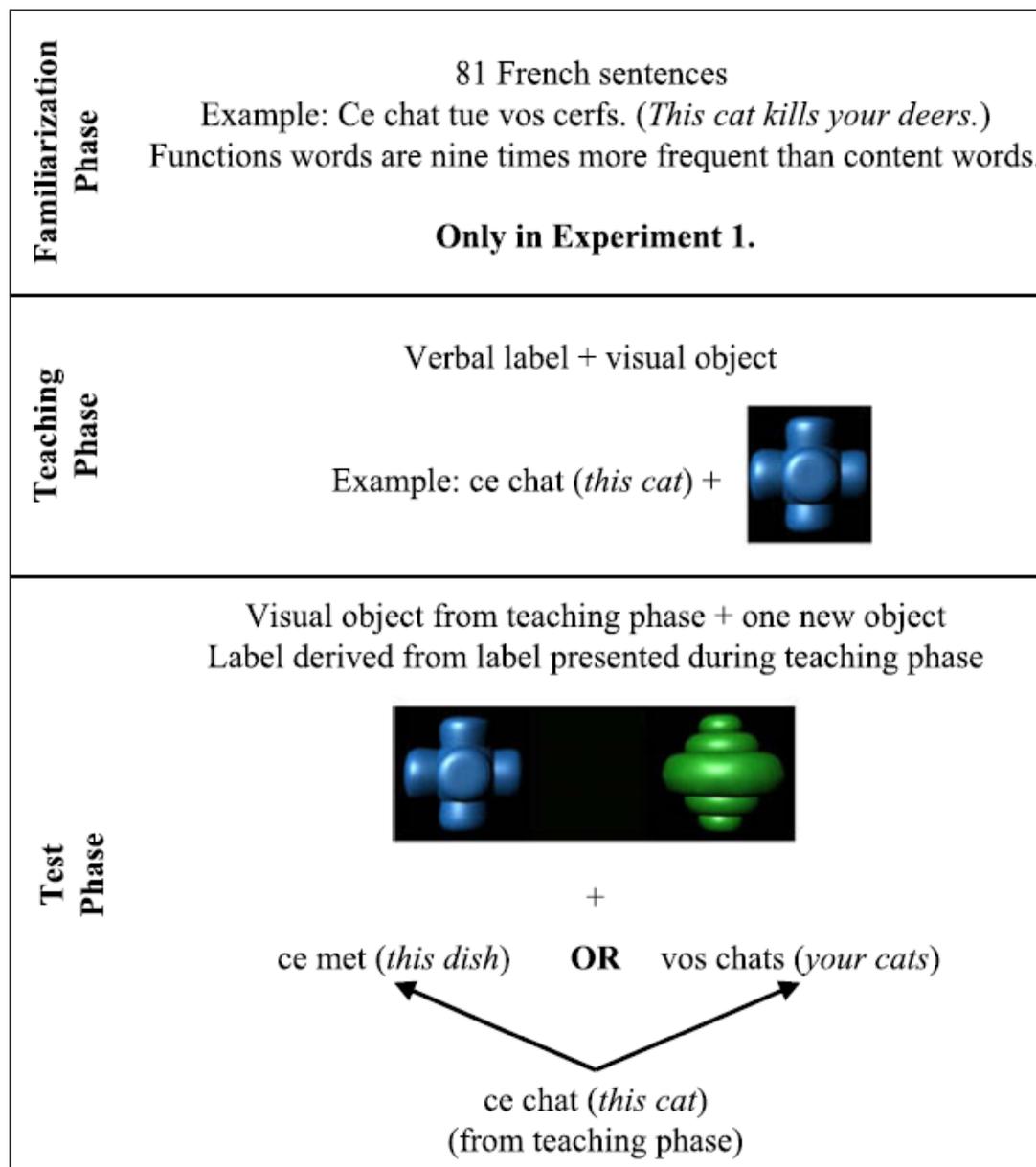
Dans la langue naturelle, les deux types de mots sont enchevêtrés.

Comment l'enfant fait-il cette distinction, et attache-t-il un sens uniquement aux mots à contenu?

Hochmann et al. (2010) font l'hypothèse que la fréquence joue un rôle majeur.

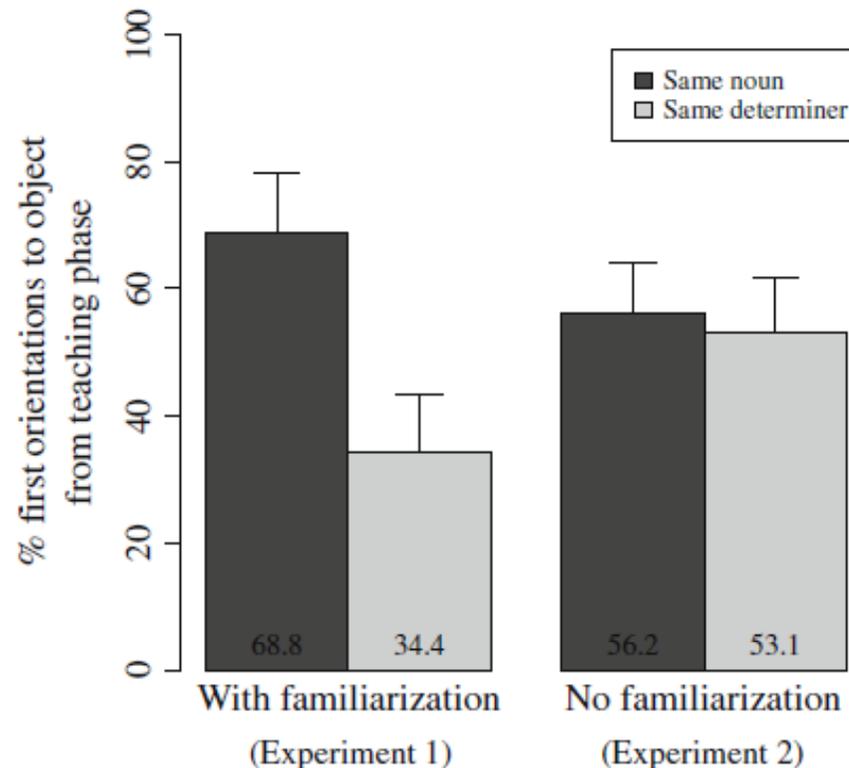
Ils exposent 28 enfants italiens de 17 mois à des structures du type « ce chat », accompagnées de l'objet en question.

L'enfant associe-t-il l'objet à « ce » ou à « chat »?



Comment séparer les mots à contenu des mots de fonction?

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115(3), 444-457.



Résultats: dans la phase de test, les enfants se tournent vers l'objet nouveau lorsqu'ils entendent « ce met », mais vers l'objet ancien lorsqu'ils entendent « vos chats ».

Cet effet disparaît lorsqu'on omet la phase de familiarisation.

Deux interprétations possibles:

1. Les enfants font des statistiques sur la fréquence d'association de chaque mot avec l'objet correspondant. Dans la phase de familiarisation, les déterminants sont présentés sans objet. « Au-delà d'une certaine fréquence, les enfants

'abandonnent' et concluent que ces formes n'ont pas de sens intrinsèque et sont peut-être des mots grammaticaux. »

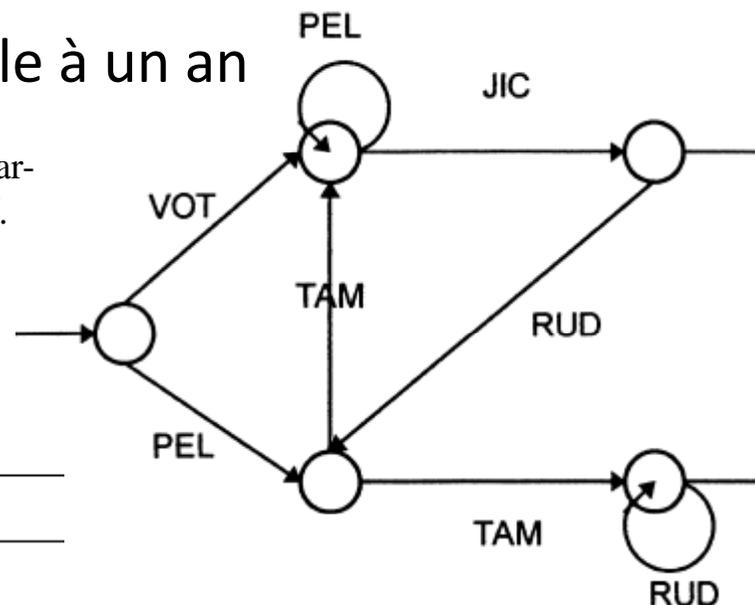
2. Les enfants appliquent la grammaire de leur langue maternelle à ces nouvelles entrées. La période de familiarisation leur permet de segmenter ces entrées en deux mots, et d'en conclure qu'il s'agit probablement d'un déterminant suivi d'un nom.

Dans tous les cas, ils font preuve d'un traitement statistique et 'proto-syntaxique' rapide.

L'apprentissage d'une grammaire artificielle à un an

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.

Des enfants d'un an sont d'abord exposés à 10 « phrases » issues d'une grammaire à états finis.



Acquisition strings used in Experiments 1 and 2

Set A

VOT PEL JIC
 PEL TAM PEL JIC
 PEL TAM JIC RUD TAM RUD
 PEL TAM JIC RUD TAM JIC
 VOT PEL PEL JIC RUD TAM

Set B

PEL TAM RUD RUD
 VOT JIC RUD TAM JIC
 VOT JIC RUD TAM RUD
 VOT PEL JIC RUD TAM
 PEL TAM PEL PEL PEL JIC

Test strings used in Experiment 1

Grammatical

VOT JIC RUD TAM
 VOT PEL PEL JIC
 PEL TAM JIC RUD TAM
 VOT PEL JIC RUD TAM JIC
 PEL TAM PEL JIC RUD TAM

 PEL TAM RUD
 PEL TAM PEL PEL JIC
 VOT PEL PEL PEL JIC
 VOT JIC RUD TAM RUD RUD

Ungrammatical

TAM JIC RUD VOT
 JIC PEL PEL VOT
 TAM TAM JIC RUD PEL
 JIC PEL JIC RUD TAM VOT
 TAM TAM PEL JIC RUD PEL

 RUD TAM PEL
 JIC TAM PEL PEL PEL
 JIC PEL PEL PEL VOT
 RUD JIC RUD TAM RUD VOT
 RUD PEL JIC RUD TAM VOT

Ensuite on teste le temps qu'ils passent à écouter des phrases partiellement nouvelles, qui sont compatibles ou incompatibles avec la grammaire précédente (les premiers et derniers mots sont échangés).

Les enfants regardent plus longtemps pour entendre les chaînes grammaticales.

... problème: les probabilités de transition sont également violées!

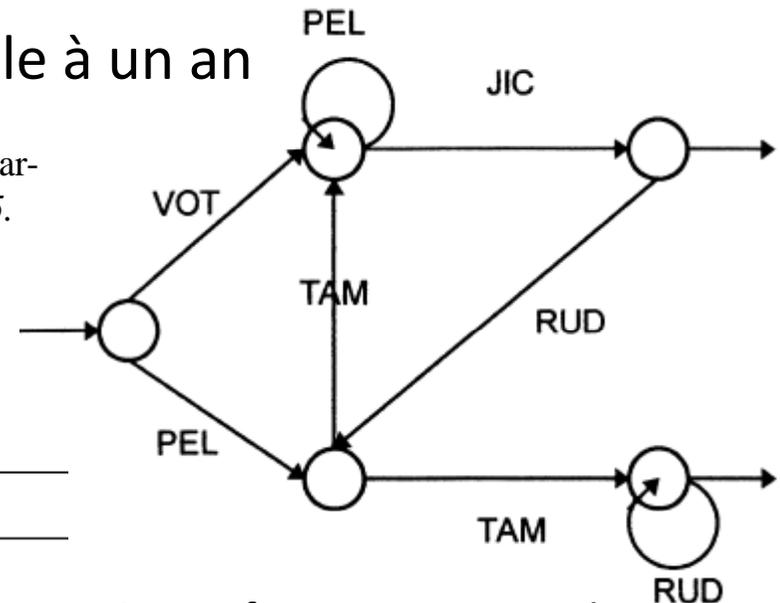
L'apprentissage d'une grammaire artificielle à un an

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.

L'expérience 2 utilise des phrases non-grammaticales dans lesquelles l'ordre interne est altéré.

Acquisition strings used in Experiments 1 and 2

Set A	Set B
VOT PEL JIC	PEL TAM RUD RUD
PEL TAM PEL JIC	VOT JIC RUD TAM JIC
PEL TAM JIC RUD TAM RUD	VOT JIC RUD TAM RUD
PEL TAM JIC RUD TAM JIC	VOT PEL JIC RUD TAM
VOT PEL PEL JIC RUD TAM	PEL TAM PEL PEL PEL JIC
Grammatical	Ungrammatical
VOT JIC RUD TAM	VOT*TAM PEL*RUD*JIC
VOT PEL PEL JIC	VOT*RUD*PEL JIC
PEL TAM JIC RUD TAM	PEL*RUD*JIC*PEL TAM
VOT PEL JIC RUD TAM JIC	VOT*TAM PEL JIC*PEL*RUD
PEL TAM PEL JIC RUD TAM	PEL*RUD*JIC*JIC*TAM*TAM
PEL TAM RUD	PEL*VOT*TAM
PEL TAM PEL PEL JIC	PEL*RUD*JIC*PEL TAM
VOT PEL PEL PEL JIC	VOT*RUD TAM*TAM
VOT JIC RUD TAM RUD RUD	PEL*VOT*RUD*PEL JIC*JIC
VOT PEL JIC RUD TAM RUD	VOT*RUD*PEL TAM PEL JIC



Les enfants continuent de regarder plus longtemps pour entendre les chaînes grammaticales.

→ l'enfant n'est pas uniquement sensible à la première et à la dernière syllabe.

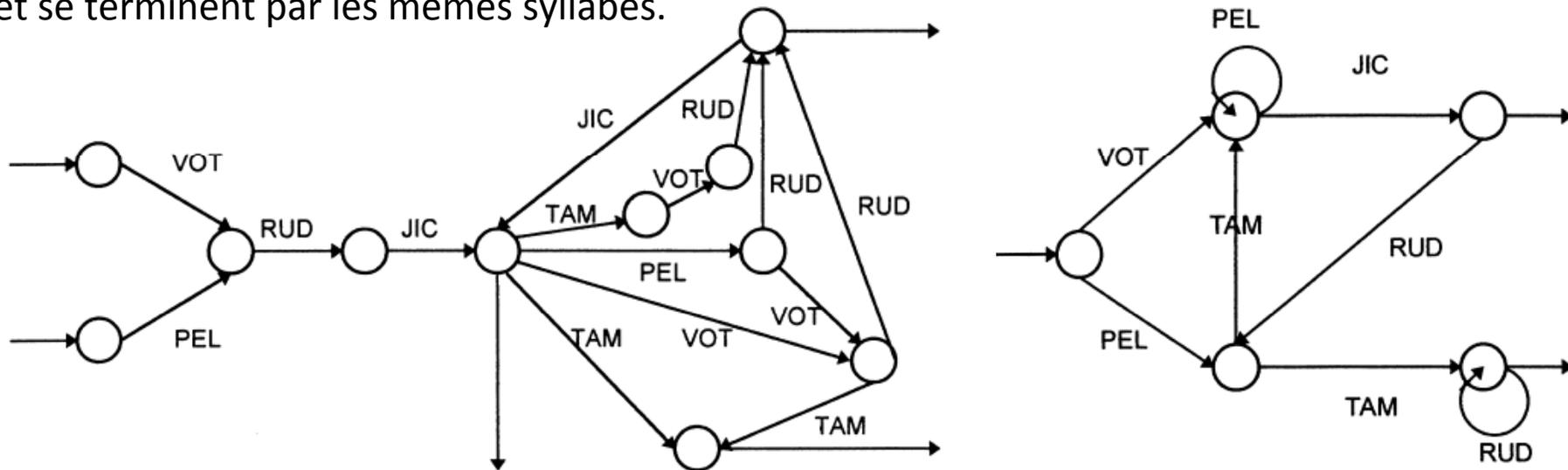
... mais les probabilités de transition sont toujours très différentes!

Par ailleurs, les phrases de test sont moins régulières que les phrases grammaticales

L'apprentissage d'une grammaire artificielle à un an

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.

L'expérience 3 utilise deux grammaires incompatibles entre elles – mais qui commencent et se terminent par les mêmes syllabes.



Les résultats sont toujours positifs... mais les probabilités de transition sont toujours très différentes.

Training grammar	Transitional probabilities		
	Training	Grammatical test	Ungrammatical test
G1 – Sets AB	0.44 (0.10 – 0.80)	0.40 (0.10 – 0.80)	0
G1 – Sets CD	0.38 (0.20 – 0.60)	0.39 (0.20 – 0.60)	0
G2 – Sets AB	0.61 (0.17 – 0.83)	0.60 (0.17 – 0.83)	0
G2 – Sets CD	0.61 (0.17 – 0.80)	0.60 (0.17 – 0.80)	0

L'apprentissage d'une grammaire artificielle à un an

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.

L'expérience 4 utilise les mêmes grammaires – mais teste la généralisation à des « mots » nouveaux, en remplaçant VOT par JED, PEL par FIM, JIC par TUP, RUD par DAK et TAM par SOG.

Dans cette expérience clé, les enfants continuent de regarder plus longtemps pour écouter les chaînes qui respectent, à un niveau abstrait, la grammaire qu'ils ont apprise au départ. Cette « grammaire » pourrait ne consister qu'en certains patrons de répétition.

Grammar 1

Set A

VOT PEL JIC
 VOT PEL PEL PEL JIC
 PEL TAM JIC RUD TAM JIC
 VOT PEL JIC RUD TAM RUD
 PEL TAM PEL JIC RUD TAM

Set B

VOT JIC RUD TAM
 PEL TAM RUD RUD
 PEL TAM PEL PEL JIC
 VOT PEL JIC RUD TAM
 PEL TAM JIC RUD TAM RUD

Grammar 2

Set A

PEL RUD JIC
 VOT RUD JIC VOT TAM
 PEL RUD JIC VOT RUD JIC
 PEL RUD JIC PEL VOT TAM
 VOT RUD JIC TAM VOT RUD

Set B

VOT RUD JIC TAM
 PEL RUD JIC PEL RUD
 PEL RUD JIC VOT RUD
 VOT RUD JIC VOT RUD JIC
 VOT RUD JIC PEL VOT RUD

Effectivement, le transfert ne se fonde que sur la présence ou non de répétitions: Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Mem Cognit*, 28(2), 253-263.

L'apprentissage de règles abstraites

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80.

16 bébés de 7 mois entendent des séquences de trois syllabes, tirées soit d'une grammaire « ABB », soit d'une grammaire « ABA ». Pendant 2 minutes, chaque enfant entend 3 répétition de 16 suites (par ex. « ga ti ti » ou « ga ti ga »).

Dans la phase de test, on présente exclusivement des séquences nouvelles, mais qui respectent le patron d'origine ou pas (par ex. « wo fe fe » ou « wo fe wo »).

Les enfants écoutent nettement plus longtemps les séquences qui s'écartent de la grammaire de départ.

Exp.	Mean listening time (s) (SE)		Repeated measures analysis of variance
	Consistent sentences	Inconsistent sentences	
1	6.3 (0.65)	9.0 (0.54)	$F(14) = 25.7, P < 0.001$
2	5.6 (0.47)	7.35 (0.68)	$F(14) = 25.6, P < 0.005$
3	6.4 (0.38)	8.5 (0.5)	$F(14) = 40.3, P < 0.001$

Ces résultats, très significatifs, sont répliqués avec des syllabes qui ne partagent aucun trait phonétique (exp. 2).

Dans l'expérience 3, les résultats sont étendus à AAB versus ABB.

Ces résultats ne peuvent s'expliquer par un apprentissage associatif ou de probabilités de transition. Les enfants seraient capables d'extraire des « **règles algébriques** ».

Quelques suites de l'article de Marcus et al. (1999)

Johnson, S. P., K. J. Fernandas, et al. (2009). "Abstract Rule Learning for Visual Sequences in 8- and 11-Month-Olds." *Infancy* 14(1): 2-18.

Entre 8 et 11 mois, le degré de précision de la règle change:

Les enfants les plus jeunes codent la répétition, mais pas sa position dans la séquence.

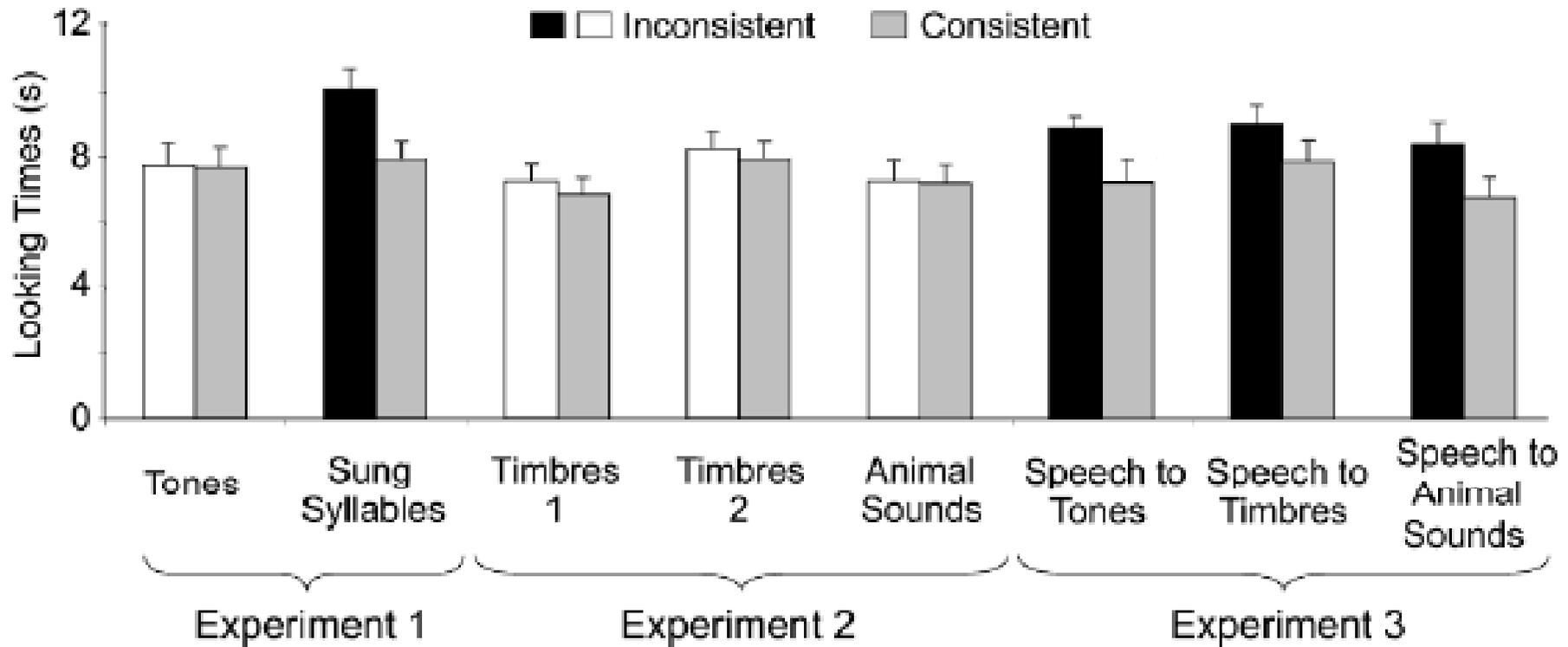
Les enfants de 11 mois codent la position de la répétition, mais pas la présence de répétitions non-adjacentes.

Condition	Outcomes	
	8-Month-Olds	11-Month-Olds
ABB vs. AAB (late vs. early repetition)	Fail	Succeed
ABB vs. ABA (late vs. nonadjacent repetition)	Succeed	Not tested
AAB vs. ABA (early vs. nonadjacent repetition)	Fail	Succeed
AAB vs. ABB (early vs. late repetition)	Not tested	Succeed
ABA vs. ABB (nonadjacent vs. late repetition)	Fail	Fail
ABA vs. AAB (nonadjacent vs. early repetition)	Not tested	Not tested

Quelques suites de l'article de Marcus et al. (1999)

Marcus, G. F., K. J. Fernandes, et al. (2007). "Infant rule learning facilitated by speech." *Psychol Sci* 18(5): 387-91.

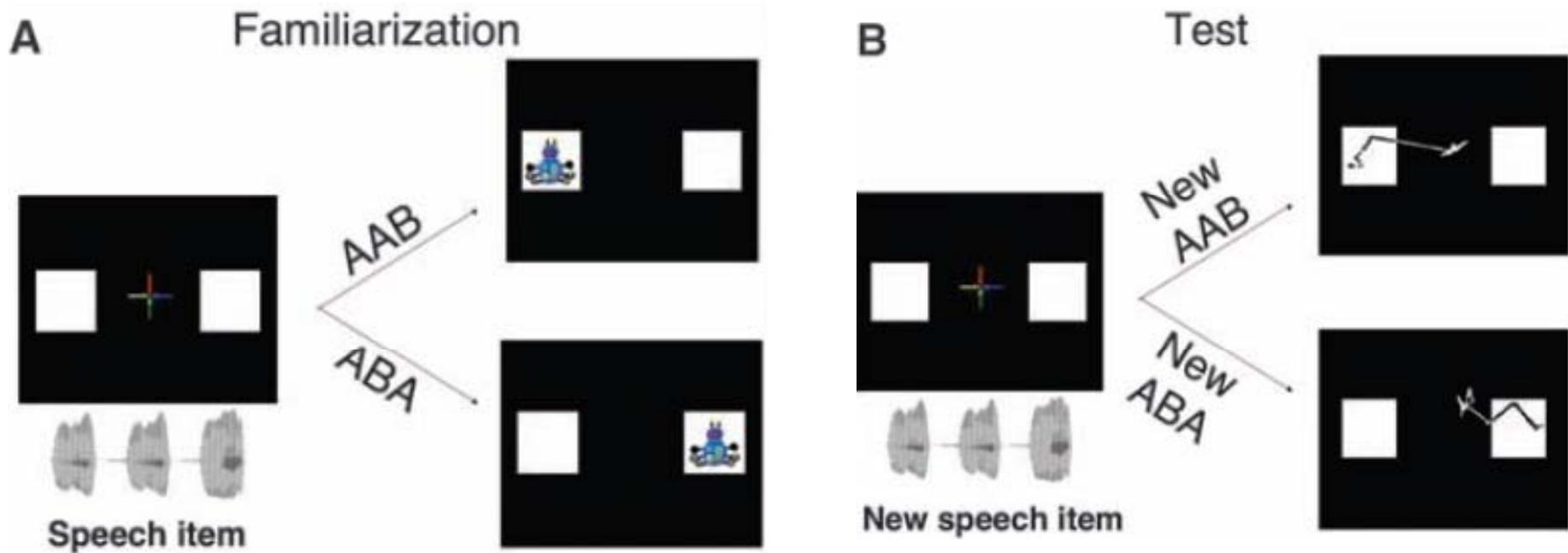
A sept mois, les enfants n'apprennent pas facilement des règles AAB, ABA ou ABB sur des tons purs ou des timbres – mais ils peuvent généraliser du langage à la musique.



Chez les enfants bilingues, l'apprentissage des règles serait facilité

Kovacs, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, 325(5940), 611-612.

Dans une période de familiarisation, les enfants apprennent à anticiper la présence d'un jouet à gauche après une séquence « AAB » et à droite après une séquence « ABA ».



On teste ensuite leur capacité d'anticiper du regard, en réponse à de nouvelles séquences « AAB » ou « ABA ».

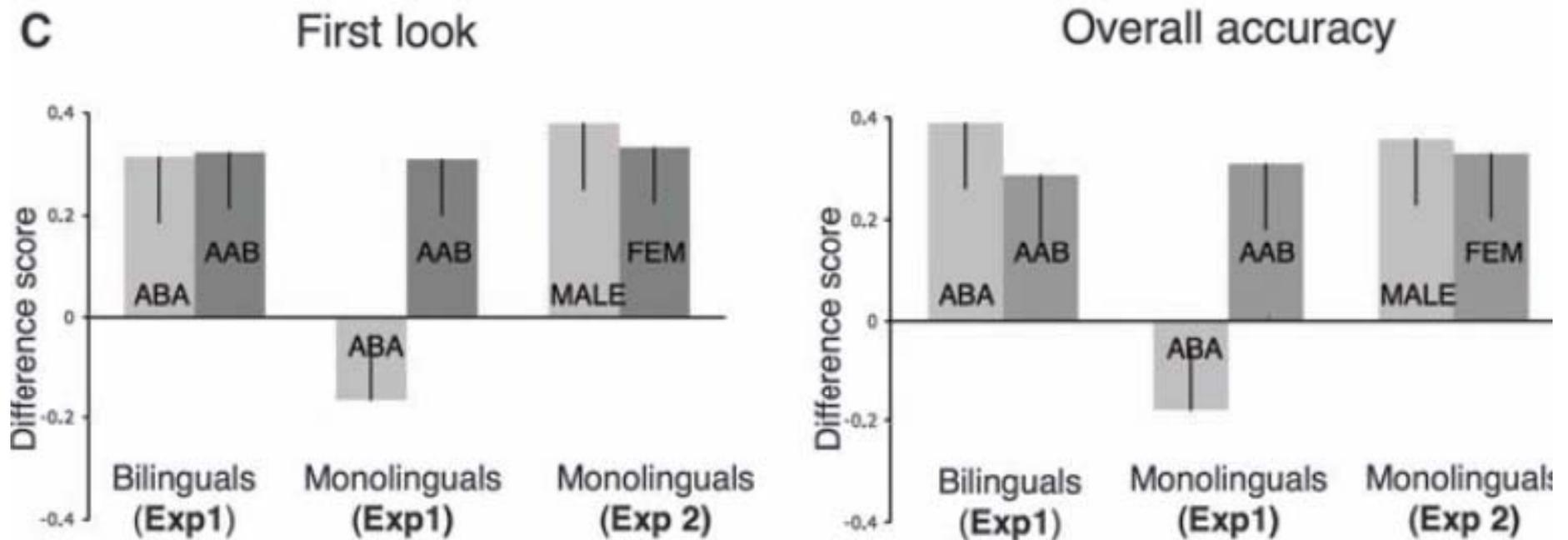
Participants: 44 enfants de 12 mois, 22 monolingues et 22 bilingues parfaits (italien/slovène)

Chez les enfants bilingues, l'apprentissage des règles serait facilité

Kovacs, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science*, 325(5940), 611-612.

Hypothèse: les enfants bilingues parviennent plus facilement à se représenter deux règles simultanément.

Résultat: les bilingues se tournent du bon côté pour les séquences « ABA » et « AAB ». Les monolingues ne détectent que la séquence « AAB » (ou même ont un biais pour se tourner toujours du même côté).



Un groupe de contrôle d'enfants monolingues, entraînés avec une information redondante sur la voix, parvient à réaliser la tâche.

D'autres données suggèrent une plus grande facilité de réapprentissage et de contrôle exécutif chez les enfants bilingues

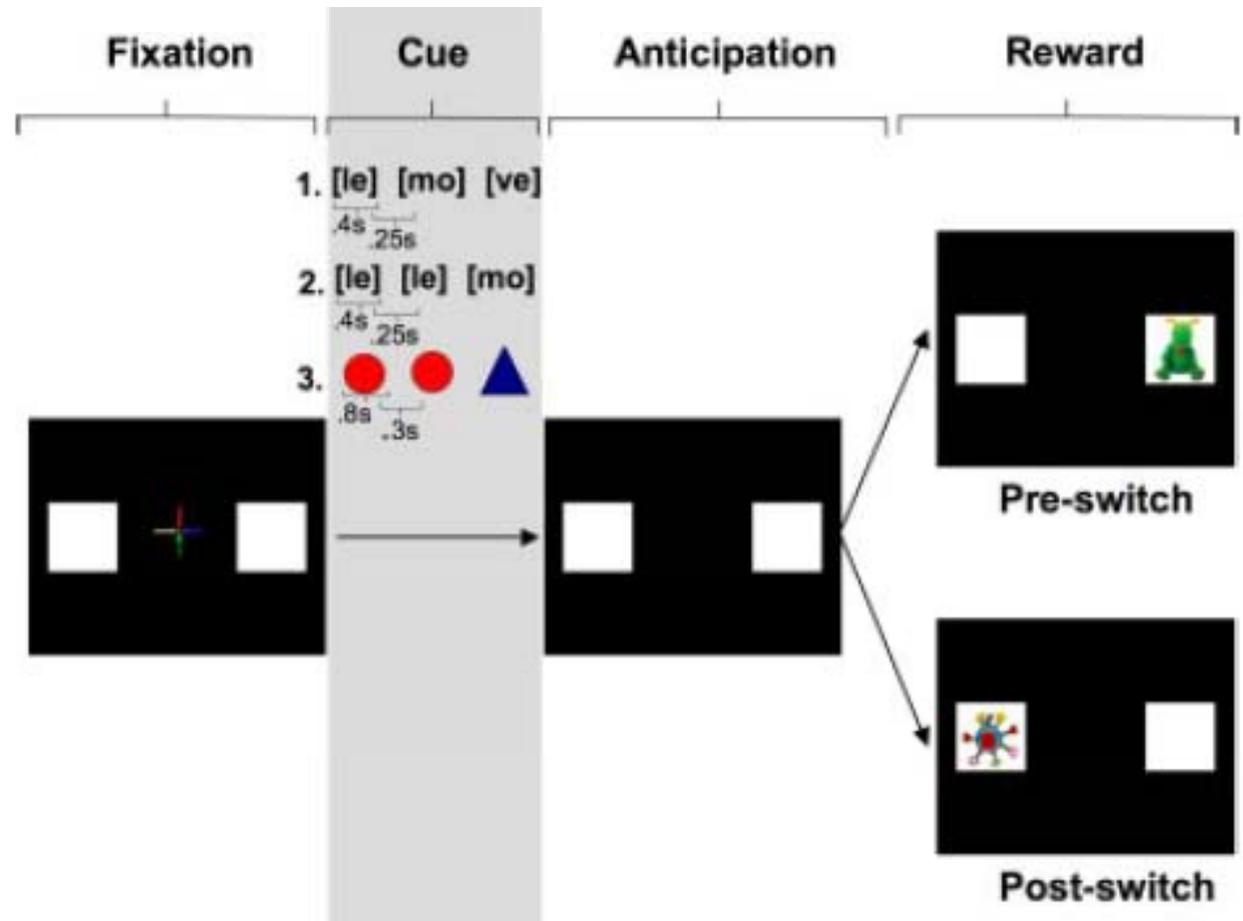
Kovacs, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proc Natl Acad Sci U S A*, 106(16), 6556-6560.

Test de la flexibilité de l'apprentissage:

Des enfants de 7 mois (monolingues ou bilingues) apprennent que, lorsqu'ils entendent des mots, ils peuvent anticiper un objet à droite.

Au bout de 9 essais, on inverse le côté, tout en changeant d'indice.

L'indice peut être un ensemble de mots particuliers (exp. 1), une structure (AAB ou ABB; exp. 2), ou une séquence visuelle (exp. 3).



D'autres données suggèrent une plus grande facilité de réapprentissage et de contrôle exécutif chez les enfants bilingues

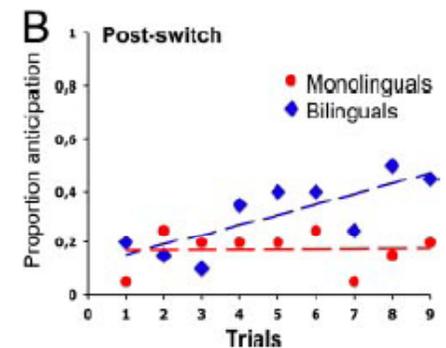
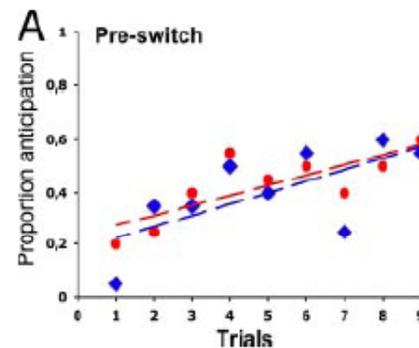
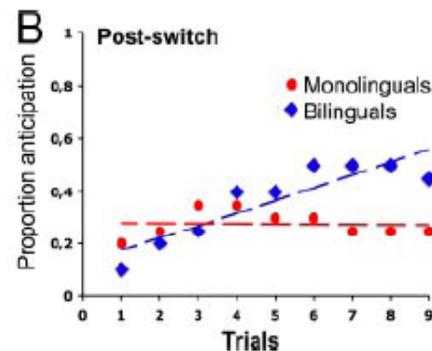
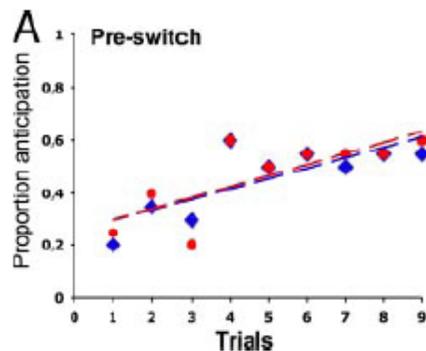
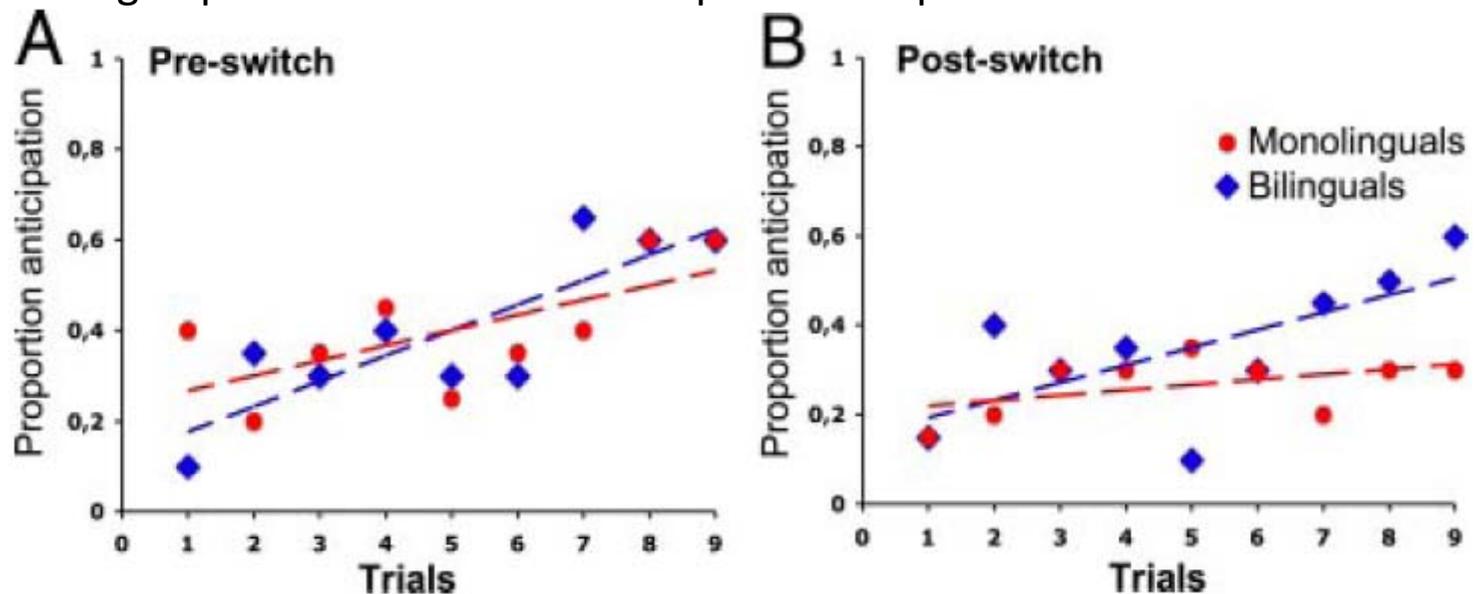
Kovacs, A. M., & Mehler, J. (2009). Cognitive gains in 7-month-old bilingual infants. *Proc Natl Acad Sci U S A*, 106(16), 6556-6560.

Résultats. Les deux groupes apprennent à anticiper...

mais seul le groupe bilingue parvient à inhiber ce comportement pour se tourner du bon côté.

Les données sont répliquables.

L'enfant bilingue semble présenter un apprentissage plus rapide et plus flexible.



L'apprentissage de dépendances à distance

Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604-607.

Des adultes français sont exposés à des séquences formées d'une concaténation de groupes de trois syllabes selon le patron « AXC », c'est-à-dire:

la première syllabe (A) prédit la troisième (C), et la seconde syllabe X peut prendre plusieurs valeurs (3 possibilités).

Exemple: puraki puliki pufoki

Trois régularités de type AXC sont présentées dans une séquence semi-aléatoire:

PURAKIBELIGATAFODUPUFOKITALIDUBERAGA

Si les auditeurs détectent cette régularité, ils peuvent grouper les syllabes par trois.

Mais la probabilité de transition seule ne suffit pas (probabilités 0.3, 0.3, 0.5).

Dans l'expérience 1, après l'exposition à la séquence continue, on présente des paires de trisyllabes. L'un est un « mot » de la séquence, l'autre une partie de mot (*part-word*). On demande de juger lequel de ces items appartient au langage de départ.

Les sujets préfèrent les mots aux « *part-words* », ce qui suggèrent qu'ils ont utilisé la règle pour segmenter la séquence.

Expérience 1.

« parties de mot » contre « mot »

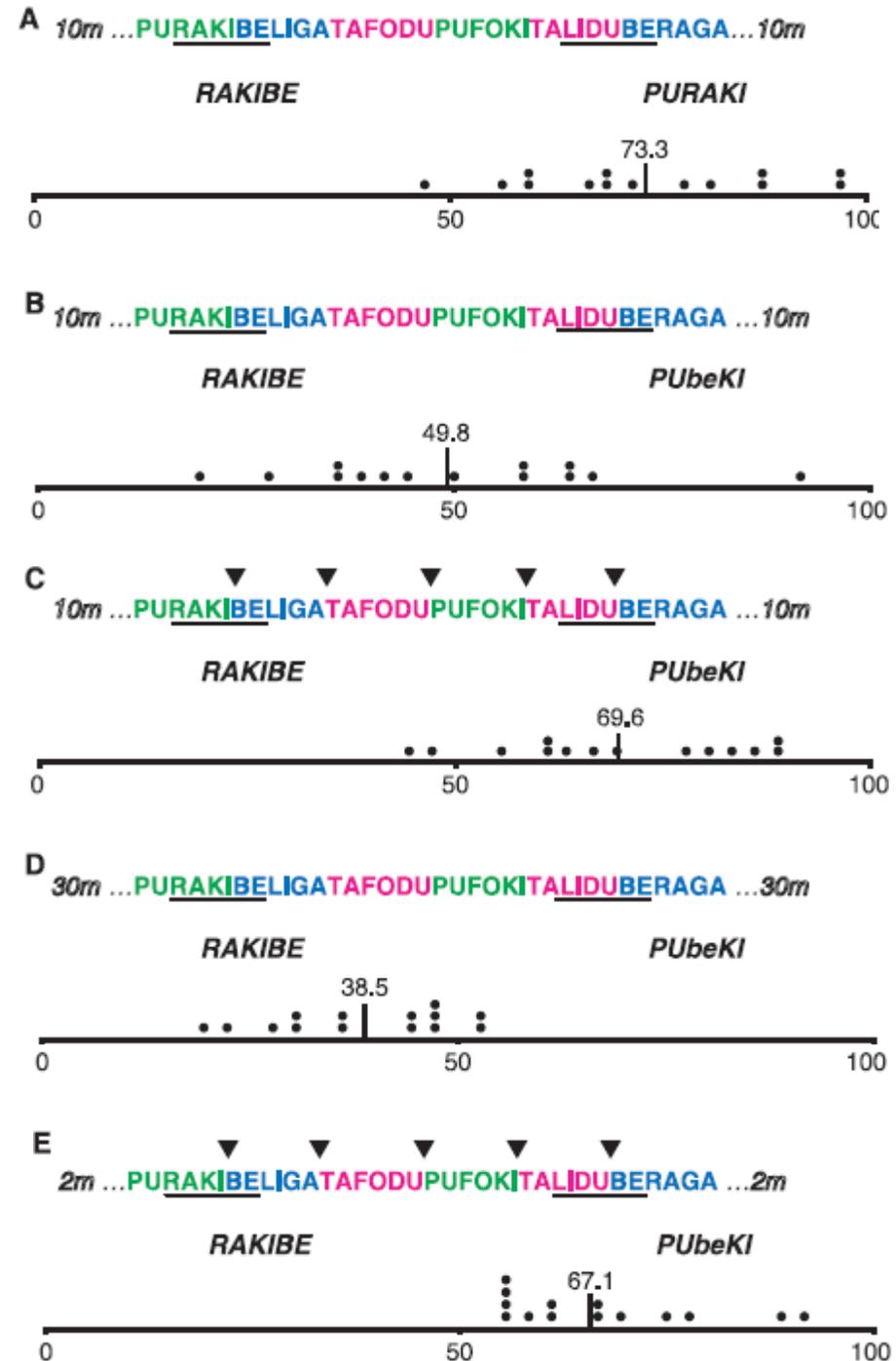
Expérience 2. Les sujets ont-ils appris simplement trois mots (puraki puliki et pufoki), ou bien une authentique règle? Test avec « parties de mot » contre « mot-règle »: pas de préférence!

Hypothèse: deux processus distincts: la **segmentation** est une processus statistique; l'**extraction d'une règle** demande d'inspecter des représentations discrètes des mots en mémoire de travail.

Expérience 3. introduction de pauses subliminales. Une préférence apparaît pour le « mot-règle ». Les probabilités de transition ne peuvent pas l'expliquer.

Expérience 4. Un sur-apprentissage de l'expérience 2 ne fait qu'augmenter la préférence pour les « parties de mots ».

Expérience 5. Inversement, même deux minutes d'exposition à l'expérience 4 suffisent à inférer la règle.



L'apprentissage de dépendances à distance

Pena, M., Bonatti, L. L., Nespors, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604-607.

Conclusions:

Deux mécanismes très différents seraient à l'œuvre dans l'apprentissage de séquences:

- un mécanisme d'apprentissage statistique, sensible aux probabilités de transition (de n à $n+1$, éventuellement de n à $n+2$), capable de proposer une segmentation probable (transitions de basse probabilité), qui alimente l'apprentissage du lexique.
- un mécanisme d'apprentissage de règles abstraites, tout-ou-rien, qui généralise à des instances nouvelles, et fournit des règles algébriques: ABB, AxC...

Double dissociation:

- la segmentation par apprentissage statistique s'améliore avec la quantité d'évidence
- l'apprentissage de règles n'a pas besoin de beaucoup d'évidence, mais opère sur des représentations segmentées (des « hypothèses » conscientes?)

Des règles ou des statistiques?

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Neuroscience. Does grammar start where statistics stop? *Science*, 298(5593), 553-554.

Seidenberg et al. critiquent Pena et al., en notant que de nombreuses autres règles peuvent être proposées – et certaines pourraient expliquer les résultats (?).

RULES OR STATISTICS IN LANGUAGE LEARNING?			
A Familiarization stream			
... PURAKIBELIGATAFODUPUFOKITALIDUBERAGATARADUPULIKIBEFOGA ...			
B Generalizations			
			Probability
1.	Words have PU_KI, BE_GA, TA_DU structure (Peña <i>et al.</i> 's rule, AXC)		1.0
2.	Initial syllables begin with a stop consonant		1.0
3.	Final syllables begin with a stop consonant		1.0
4.	Continuant consonants occur word medially		1.0
5.	PU predicts RA		0.33
6.	PU is not followed by BE		1.0
7.	KI predicts BE		0.5
8.	PURA predicts KI		1.0
9.	RAKI predicts BE		0.5
10.	PUBE is not followed by KI		1.0
C Experiment 1 forced choice			
		Choice supported by	Percentage chosen
Word	PURAKI	1 (rule), 2, 3, 4, 5, 8	73.3
Part-word	RAKIBE	3, 7, 9	26.7
D Experiment 2 forced choice			
		Choice supported by	Percentage chosen
Rule-word	PUBEKI	1 (rule), 2, 3, 4, 8	49.8
Part-word	RAKIBE	3, 6, 7, 9, 10	49.2

Et chez l'enfant?

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13, 431–436.

Gomez, R., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy*, 7(2), 183-206.

A 18 mois (mais pas à 15 mois), les enfants distinguent des phrases grammaticales et non-grammaticales comme « is running » et « can running », ou « is quickly running » et « can quickly running » (Santelmann & Jusczyk, 1998).

Ce résultat est répliqué par Gómez (2002) chez l'enfant de 18 mois:

L'enfant entend des stimuli issus de grammaires AxB et CxD (par exemple « pel-wadim-jic », « pel-dicey-jic »..., et de même « vot-wadim-rud »...), puis des violations du type AxD et CxB (« pel-wadim-**rud**», « vot-wadim-**jic** »).

Les enfants écoutent plus longtemps les stimuli qui violent la grammaire initiale.

De façon très intéressante, les enfants n'apprennent la règle que si x est suffisamment variable (24 possibilités, mais pas 2 ou 12). La diminution de la probabilité de transition force peut-être les enfants à chercher une régularité plus abstraite.

Les enfants de 12 mois n'apprennent pas de façon significative. Ceux de 15 mois (avec une variabilité maximale, 24 x possibles) montrent une préférence pour l'ancienne grammaire.

Conclusions: Dans le seconde année de vie, l'enfant est capable de découvrir des règles abstraites sur des items non-adjacents (mais, pas de preuve de récursivité)

La découverte d'un « cadre » de type AxB pourrait également permettre l'induction que tous les « x » appartiennent à la même catégorie. Par exemple « is x-ing » → x est un verbe. « the x-s » → x est un nom (Mintz, 2003, 2004).

L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

L'apprentissage des nombres pose un défi particulier:

-Les enfants apprennent facilement à réciter les nombres « un deux trois quatre... »

(Fuson, 1988)

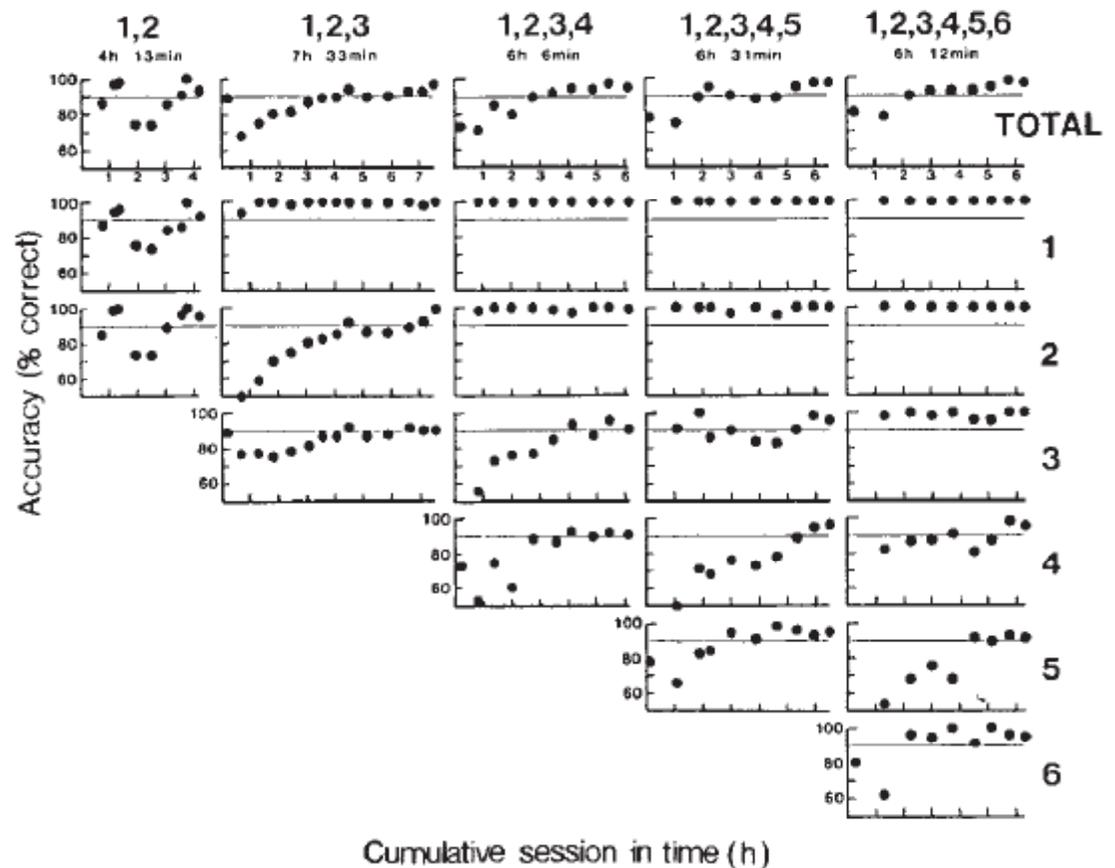
- Mais ils n'en connaissent pas nécessairement le sens (Wynn, 1990, 1992).

-Pendant plusieurs mois, ils apprennent, un à un, le sens des mots un, deux, trois

- Vers trois ans et demi, ils font un saut conceptuel remarquable: ils comprennent que chaque nom de nombre correspond à une quantité bien précise (*cardinal principle, CP*)

- Ce comportement est très différent de celui du chimpanzé Ai, qui apprend les chiffres un par un, très lentement.

(Toutefois, le perroquet Alex semblerait se comporter différemment; Pepperberg & Carey, 2012)



L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

Carey (2009) propose une théorie du « bootstrapping »: l'enfant observe les relations entre les quantités et les mots et, sur la base des premiers nombres, induit soudainement le sens des suivants. Il établit une correspondance entre « ajouter un objet » et « passer au mot suivant ».

Mais cette théorie reste vague. Comment se produit l'induction? Comment sélectionner parmi une infinité de modèles possibles? Le modèle de Carey ne présuppose-t-il pas, d'une manière ou d'une autre, la connaissance des entiers naturels?

Idée de Piantadosi et al. (2012): sélection bayésienne dans un espace de formules du lambda-calcul, un « langage de la pensée » doté de primitives et de lois de composition.

Piantadosi et al. modélisent la sélection d'un programme capable de prendre un ensemble en entrée et d'énoncer son cardinal en sortie.

Le système est doté d'un ensemble de primitives (*core knowledge*)

- Subitisation: singleton?, doubleton?, tripleton?
- Manipulation des ensembles: select, intersection, union, set-difference
- Logique: and, or, not, if
- Routine du comptage: next, prev, equal-word?
- Récursion: (L S) évalue une nouvelle fois la fonction L sur l'ensemble S

L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

Exemple de programme:

Produit « one » si l'ensemble est un singleton
« two » sinon

```
 $\lambda S \cdot (\text{if}(\text{singleton? } S)$   
  "one"  
   $(\text{next } (L (\text{select } S))))).$ 
```

Exemples de programmes avec ou sans signification

One-knower

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  undef)
```

Two-knower

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  (if (doubleton? S)  
    "two"  
    undef))
```

Three-knower

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  (if (doubleton? S)  
    "two"  
    (if (tripleton? S)  
      "three"  
      undef)))
```

CP-knower

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  (next (L (set-difference S  
            (select S))))))
```

Singular-Plural

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  "two")
```

Mod-5

```
 $\lambda S . (if (or (singleton? S)$   
  (equal-word? (L (set-difference S  
                  (select S))  
                "five"))  
  "one"  
  (next (L (set-difference S  
            (select S))))))
```

2-not-1-knower

```
 $\lambda S . (if (doubleton? S)$   
  "two"  
  undef)
```

2N-knower

```
 $\lambda S . (if (singleton? S)$   
  "one"  
  (next (next (L (set-difference S (select S))))))
```

L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

La règle d'apprentissage bayésien réalise un compromis entre deux facteurs:

-La complexité du programme

-La probabilité a priori d'une lambda-expression est égale à la probabilité qu'elle soit générée par une grammaire non-contextuelle aléatoire.

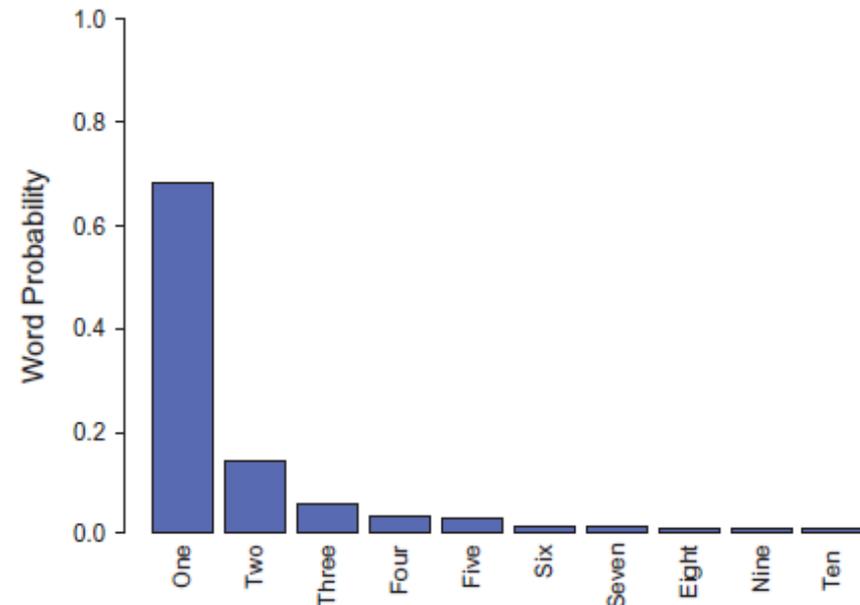
- Les règles récursives sont moins probables que les autres (paramètre γ)

-La capacité d'utiliser les nombres à bon escient

- A chaque essai, la simulation reçoit un ensemble d'objets (par ex {chat₁, chat₂, chat₃}) et le programme tourne pour générer un nombre (ou la réponse « undefined »).

- Les ensembles sont présentés avec une fréquence qui décroît avec leur cardinal

- Il y a également une petite probabilité de générer une réponse numérique au hasard

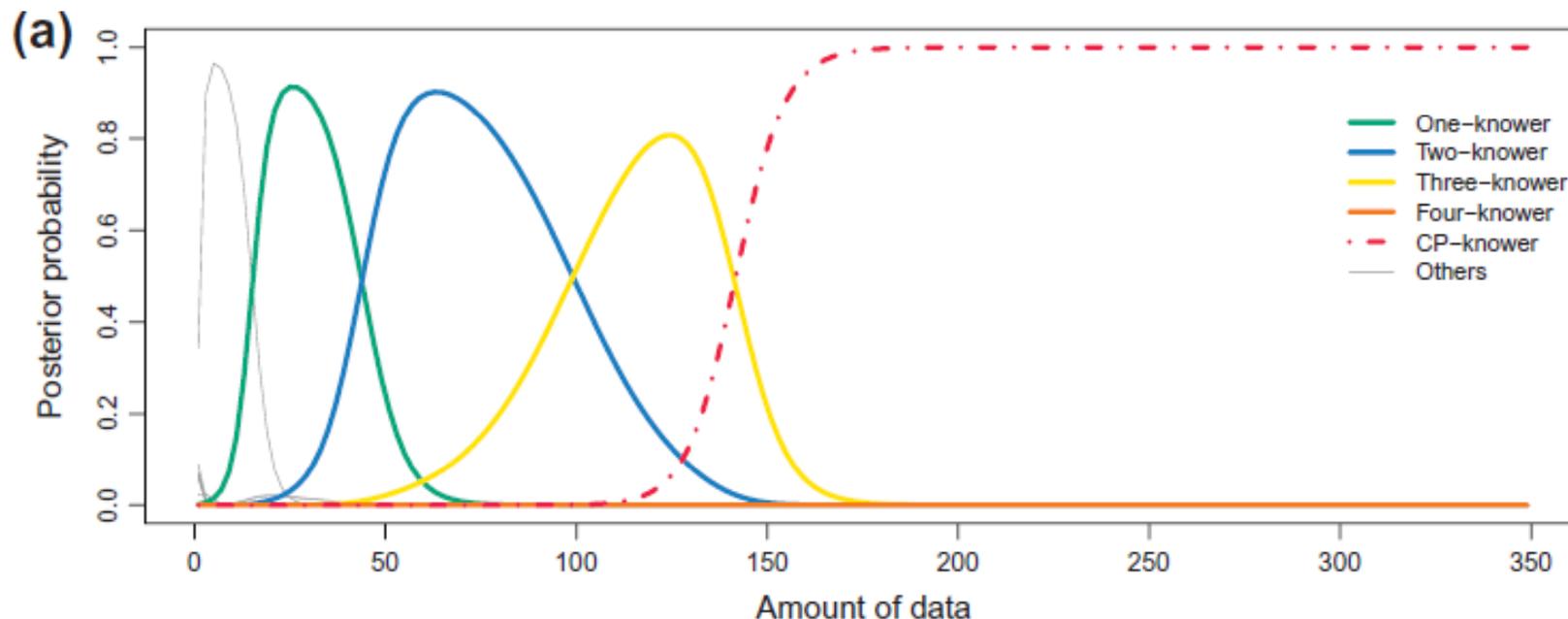


L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

Résultat des simulations:

En fonction du nombre d'essais d'entraînement, le programme passe par une série de « stades »: one-knower, two-knower, three-knower, CP-knower



Ces étapes émergent au moment où l'accumulation d'évidence parvient à compenser la probabilité a priori plus faible des lambda-expressions les plus complexes.

La probabilité que d'autres règles non-plausibles dominent est très faible.

Le point de transition vers une connaissance complète du « principe du cardinal » dépend du degré de pénalisation des règles récursives (paramètre γ).

Le modèle peut également apprendre d'autres règles (singulier/pluriel; comptage modulo 5)

L'apprentissage du vocabulaire numérique et du comptage.

Piantadosi, S. T., Tenenbaum, J. B., & Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition*, 123(2), 199-217.

Le modèle présente plusieurs propriétés nouvelles et intéressantes

-Il implémente l'idée du « bootstrapping » dans un modèle qui n'est clairement pas câblé pour l'apprentissage des nombres

-il présente des stades et des révolutions mentales, alors même que tout l'appareil de raisonnement logique est présent dès le départ.

-Même après apprentissage, il ne possède pas de connaissance explicite du « principe du successeur ». Cette connaissance n'est qu'implicite dans le déroulement du programme. Le « méta-raisonnement » qui la rend explicite reste à modéliser.

- Le modèle prédit que la transition vers la connaissance du « principe du cardinal » dépend

1. de l'existence d'une séquence de nombres (fonction *next*)

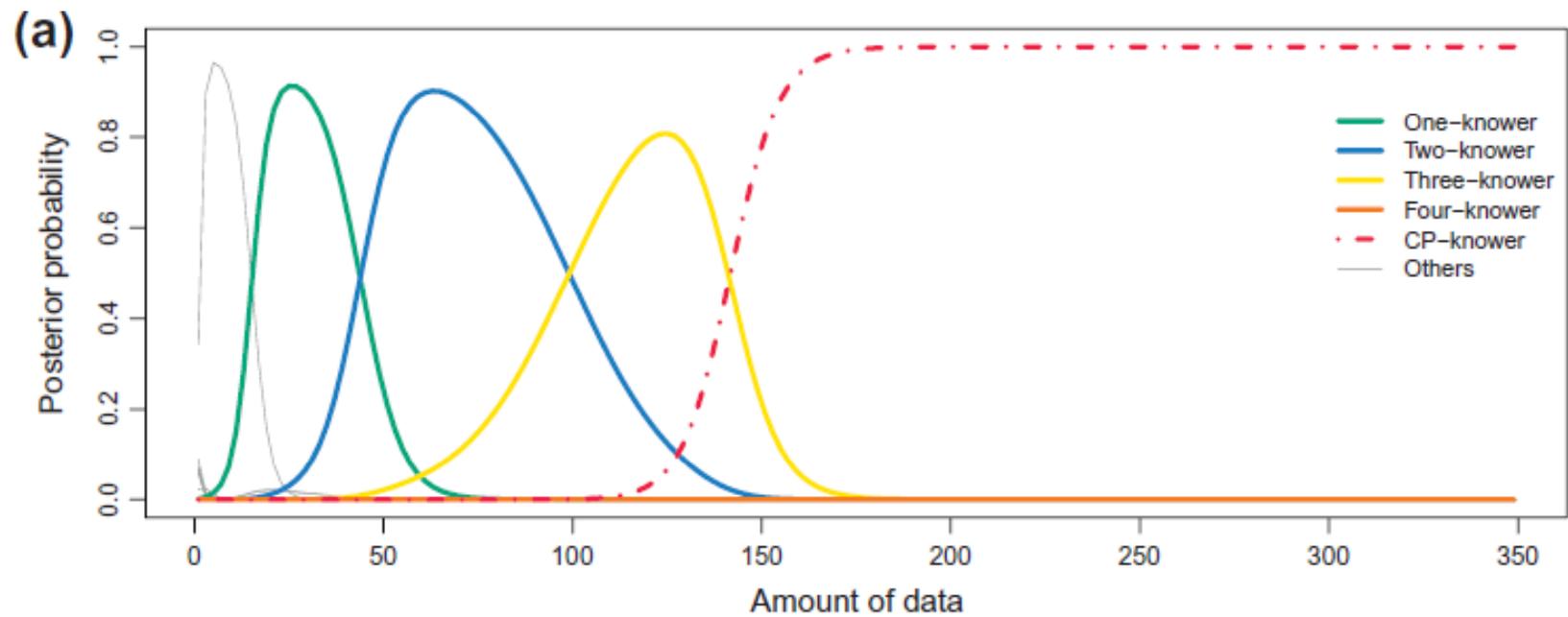
→ chez les mundurucus, cette connaissance n'émerge pas en l'absence d'éducation au comptage

2. de la capacité de récursion et de *compositionalité*, qui rassemble dans un même programme des fonctions modulaires très différentes

→ unique à l'espèce humaine?

- Le modèle pourrait être étendu à l'apprentissage des noms de nombres eux-mêmes (grammaire qui engendre des termes comme « cent cinquante et un »)

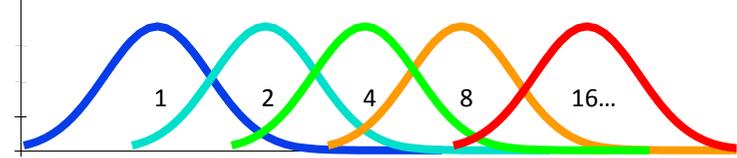
Pourquoi l'apprentissage des nombres est-il si lent?



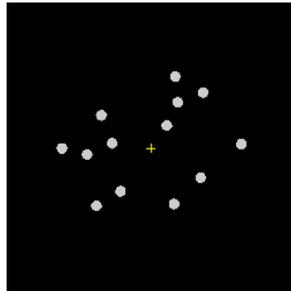
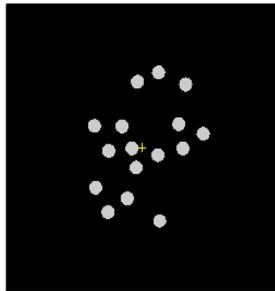
Pourquoi l'apprentissage des nombres est-il si lent?

Est-ce seulement dû à la fréquence des mots?
Ou bien à la précision du codage des quantités?

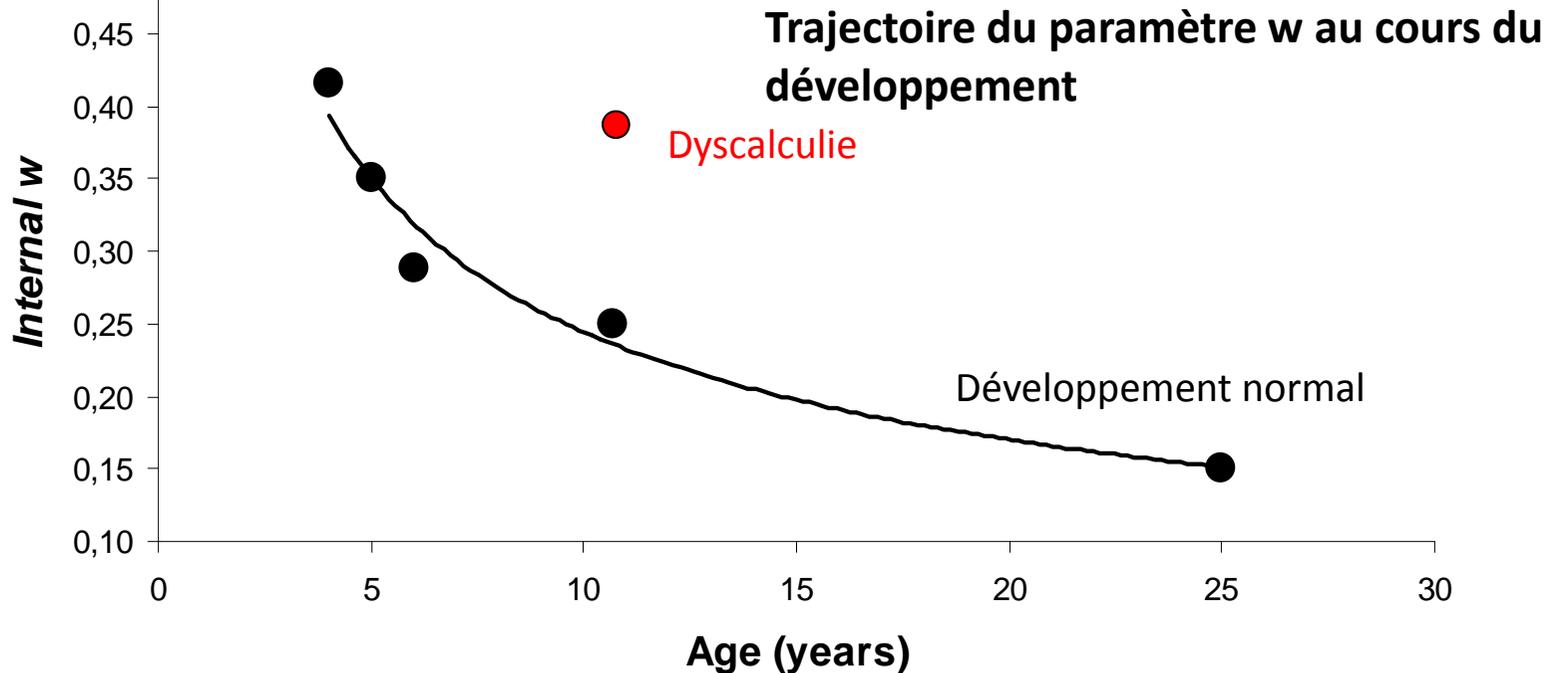
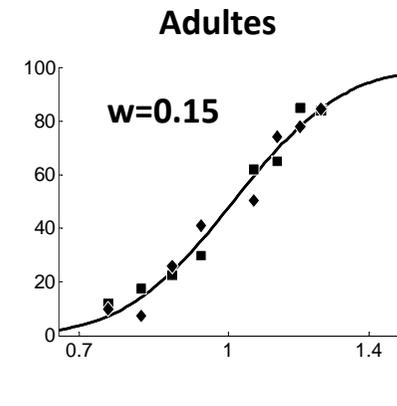
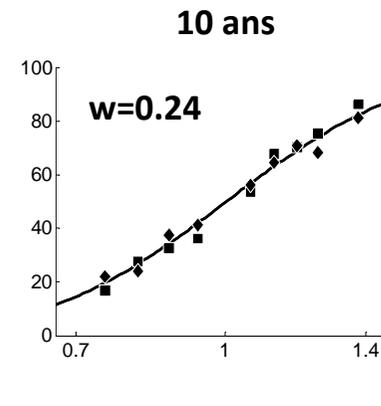
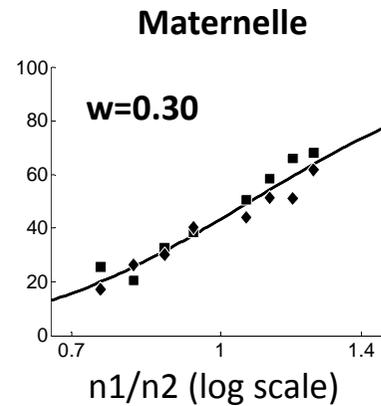
Modèle log-Gaussien du codage des nombres



Un seul paramètre libre: la largeur de la gaussienne



Quel nombre est le plus grand?



Conclusion: le bébé statisticien

L'apprentissage rapide du langage et des nombres n'implique pas d'adopter une perspective ultra-innéiste.

Les enfants s'avèrent capables

- d'extraire des règles fondées sur la répétition de mots (à partir de 7 mois)
- de détecter des dépendances à longue distance entre mots (entre 15 et 18 mois)
- d'inférer l'existence de catégories de mots (entre 1 et 2 ans) et de règles qui lient ces mots (voir l'exposé d'Anne Christophe dans le séminaire 2012)

De plus, la modélisation théorique montre qu'un algorithme bayésien hiérarchique est capable

- d'apprendre à découvrir les règles récursives qui régissent un domaine (les nombres)
- et même d'inférer qu'un domaine est régi par des arbres plutôt que des chaînes, des graphes, ou des cartes euclidiennes (cf Kemp & Tenenbaum, PNAS, 2008)

Les modèles bayésiens sont innéistes, mais en partie seulement:

- ils supposent l'existence d'une machinerie complexe d'inférence probabiliste
- et surtout d'espaces d'hypothèses qui, dans l'espèce humaine, sont peut-être plus vastes que chez tout autre espèce animale, et incluent les fonctions récursives.



« Apprendre, c'est éliminer » (Jean-Pierre Changeux)



Bon de Boullogne l'aîné (1649-1717) : *Jupiter confié aux Nymphes.* (Musée de Meaux, don Changeux)