

Cours 2013:

Le bébé statisticien

Stanislas Dehaene

Chaire de Psychologie Cognitive Expérimentale

Cours n°5

Vers une théorie bayésienne du lexique mental

L'apprentissage du sens des mots



L'apprentissage statistique extrait des mots et des objets

Dès la première année de vie, les enfants sont sensibles aux statistiques de leur environnement:

- sur le plan auditif, ils utilisent les probabilités de transition pour extraire des « mots » qui reviennent fréquemment.
- sur le plan visuel, ils compilent également des statistiques de cooccurrence des formes, à la fois dans l'espace et dans le temps.

Bayesian learning of visual chunks by human observers

Gergő Orbán^{*†}, József Fiser[†], Richard N. Aslin[‡], and Máté Lengyel^{*§¶}

Statistical learning of new visual feature combinations by infants

József Fiser^{*} and Richard N. Aslin

Les enfants de 9 mois regardent plus longtemps une combinaison qui revient plus fréquemment (exp. 1) ou qui possède une plus grande valeur prédictive (expériences 2 et 3)

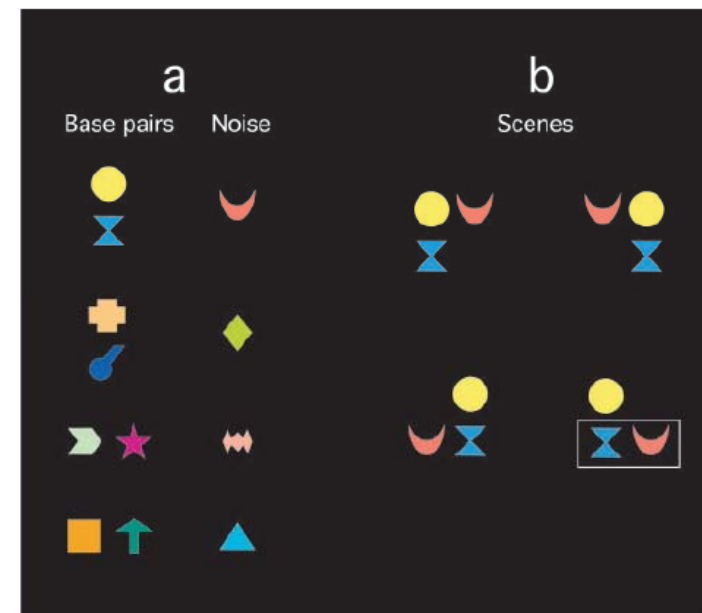


Fig. 1. Stimulus elements and scenes used in the experiments. (a) The twelve shapes were grouped into four base pairs and four noise elements, with each noise element appearing with only one base pair. (b) The four possible scenes created by one base pair and its noise element. In Experiment 1, all four scenes were presented during habituation; in Experiments 2 and 3, again, all four scenes were presented for low-frequency base pairs, but only two scenes (shown in the right column) were presented for high-frequency base pairs, which appeared twice as often as individual scenes with low-frequency base pairs. Because this doubling of appearance frequency was equally split between the two scenes containing the high-frequency base pair, the non-base pair (marked by the white rectangle in the lower right scene) appeared the same number of times as a low-frequency base pair.

L'apprentissage statistique crée des entrées lexicales que l'enfant peut attacher à des sens

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci*, 18(3), 254-260.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology*, 63(2), 93-106.

Des enfants de 17 mois sont exposés à des séquences dans lesquelles certaines transitions reviennent fréquemment, formant deux « mots » bisyllabiques.

manopigatimaymanodobu manotimaypigadobu pigadob
utimaydobu timaymano timaypigadobu pigatimayman...

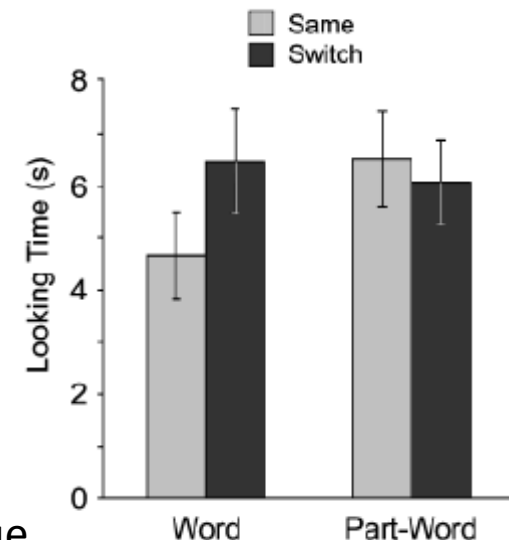
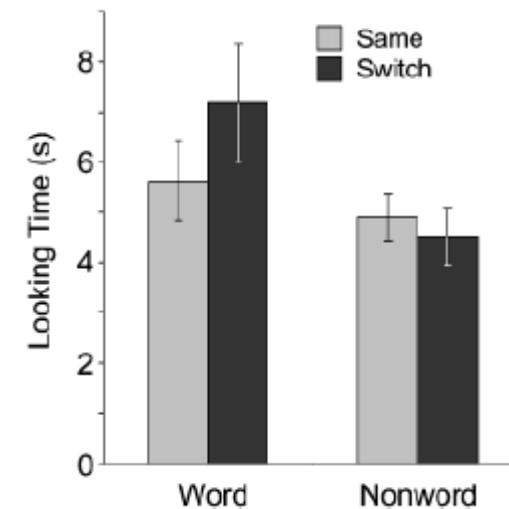
Par la suite, on associe ces deux mots à des objets.

D'autres mots bisyllabiques nouveaux (*part-words*) sont également associés à des objets.

Enfin on présente des essais de test où ces associations sont respectées ou violées.

Les enfants regardent plus longtemps lorsque les associations sont violées – mais uniquement pour les mots qui pouvaient être extraits de la séquence précédente.

Ces résultats sont partiellement répliqués avec la parole continue.



Cinq phénomènes clés dans l'apprentissage du sens des mots

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychol Rev*, 114(2), 245-272.

1. Le sens d'un mot peut être appris à partir de **très peu d'exemples** (un ou quelques-uns)
2. Les **exemples positifs** suffisent (l'enfant n'a pas besoin de contre-exemples)
3. Chaque entité, action, propriété ou relation peut être étiquetée par de nombreux mots. Il ne s'agit donc pas d'apprendre à partitionner le monde en catégories mutuellement exclusives, mais d'acquérir un **ensemble de concepts qui se recouvrent**.
4. Les inférences sur le sens d'un mot sont **graduelles**, avec des degrés de confiance variables.
5. Ces inférences sont considérablement influencées par le **contexte d'apprentissage**, particulièrement l'attention, les connaissances et les intentions du locuteurs.
 - exemple 1 (Grassman et al., 2009): des enfants de deux ans voient un adulte regarder un objet **nouveau** tout en prononçant un mot nouveau avec beaucoup d'excitation. Ils attachent ce mot à cet objet.

Ce n'est pas le cas si l'objet a déjà été vu par l'adulte auparavant. Dans ce cas, les enfants recherchent un autre référent possible pour le mot.
 - exemple 2 (Tomasello & Barton, 1994): un adulte dit « Let's glip the doll », puis fait une première action, dit « oops », fait une seconde action, dit « there ». L'enfant infère que le verbe renvoie à la seconde action.

Une illustration de la rapidité de l'apprentissage d'un mot

Tenenbaum, *Science*, 2011

Les objets rouges sont des « tufa ».



Un modèle bayésien de l'apprentissage du lexique

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychol Rev*, 114(2), 245-272.

Supposons que l'enfant observe n instances du nouveau mot C . Ses données sont:

$$X = x^{(1)}, \dots, x^{(n)}$$

où les x^i sont des entités dans un espace U (par exemple des objets particuliers).

Le modèle suppose que l'enfant dispose d'un espace $H = \{h\}$ d'hypothèses concernant les concepts auxquels le mot C peut renvoyer.

Chaque hypothèse pointe vers un sous-ensemble d'objets auquel le mot C peut renvoyer (par exemple: « tous les êtres vivants », « tous les chiens », « tous les dalmatiens », etc).

Sur la base de ses observations, l'enfant met à jour la probabilité que h soit vraie:

$$p(h | X) = \frac{p(X | h)p(h)}{p(X)}$$

La spécification de l'a priori $p(h)$ et de la vraisemblance $P(X|h)$ sont des aspects essentiels de la modélisation.

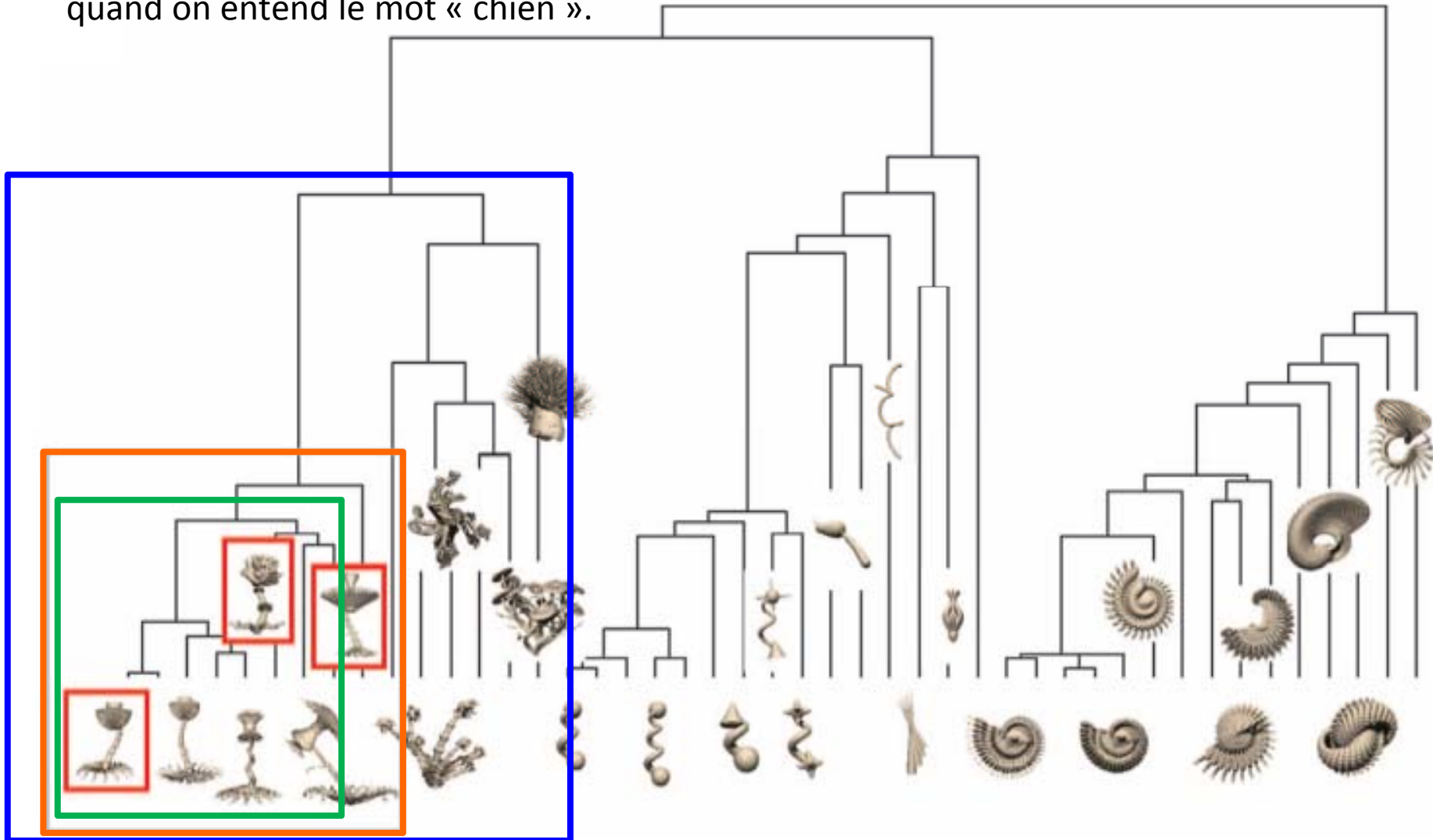
On suppose que la vraisemblance diminue avec la taille de la catégorie:

$$p(X | h) = \left[\frac{1}{\text{size}(h)} \right]^n$$

Illustration pour l'apprentissage du mot « tufa »:

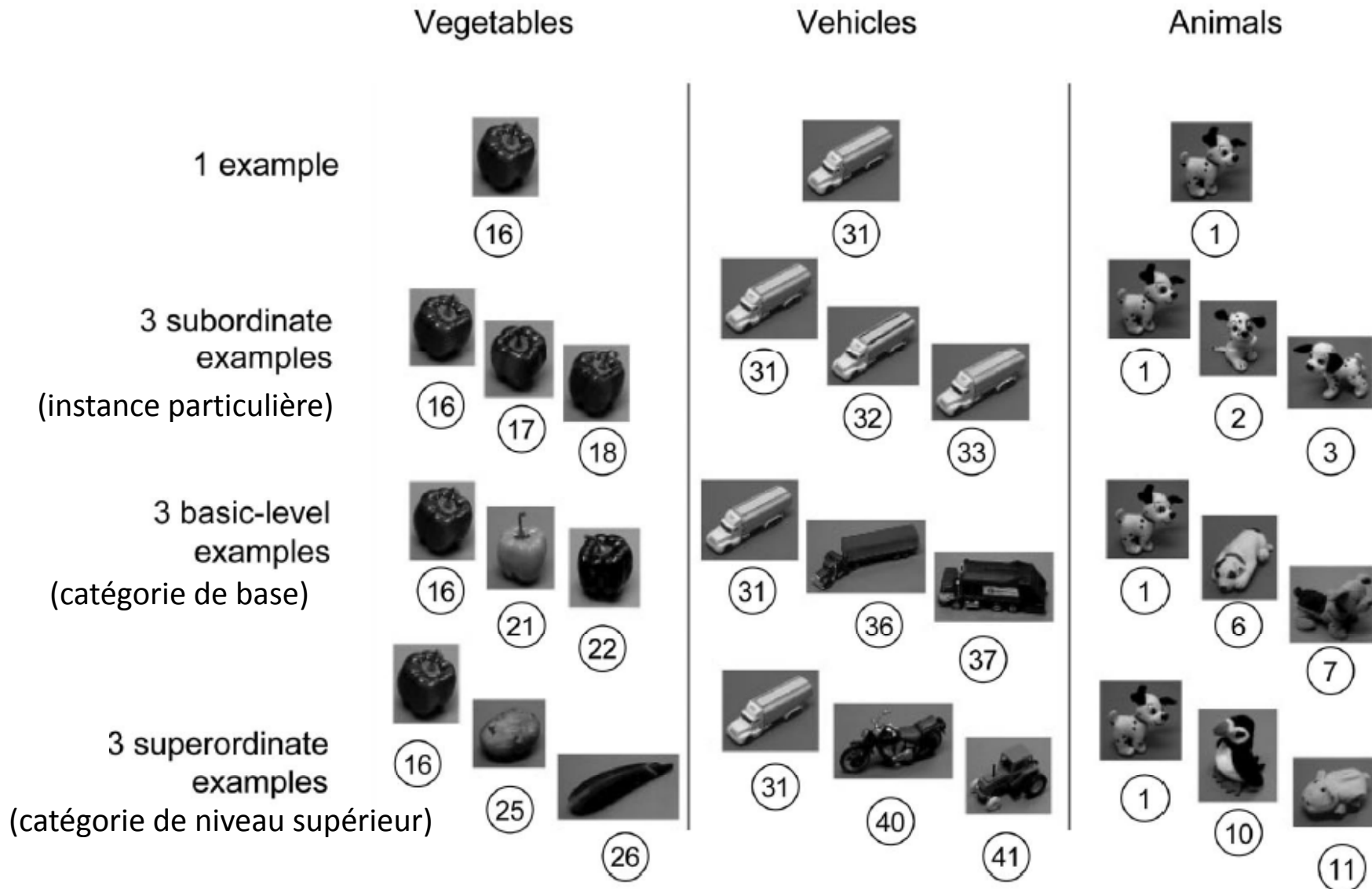
- Les hypothèses sont les branches d'un arbre des catégories de sens envisageables
- $P(X_i | \mathbf{H}) = 1/8$ $P(X_i | \mathbf{H}) = 1/14$ $P(X_i | \mathbf{H}) = 1/7$ mais $P(X_i | \mathbf{H}) = 0$ pour un X_i

Le mécanisme attribue une vraisemblance faible aux catégories les plus grandes. Il rend compte de l'idée que ce serait une coïncidence curieuse de ne voir que des dalmatiens quand on entend le mot « chien ».






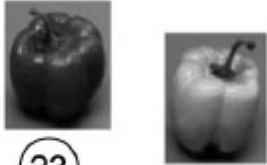
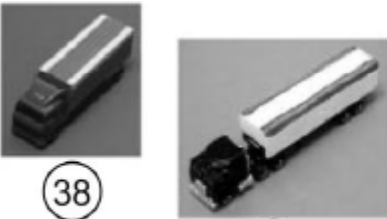





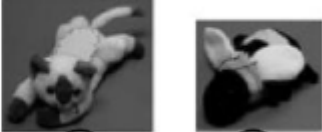

L'expérience de Xu et Tenenbaum (2007)

Les sujets sont exposés à divers objets, étiquetés par le même mot: « This is a fep ». La diversité des exemplaires varie selon les conditions:



L'expérience de Xu et Tenenbaum (2007)

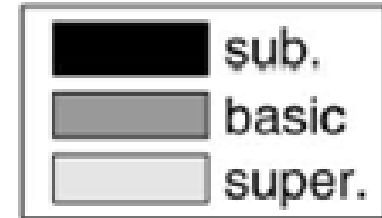
Puis on demande à la personne de généraliser: « Now tell me which of these items is a fep »

	Vegetables	Vehicles	Animals
Subordinate matches	 (19) (20)	 (34) (35)	 (4) (5)
Basic-level matches	 (23) (24)	 (38) (39)	 (8) (9)
Superordinate matches	 (27) (28)  (29) (30)	 (42) (43)  (44) (45)	 (12) (13)  (14) (15)

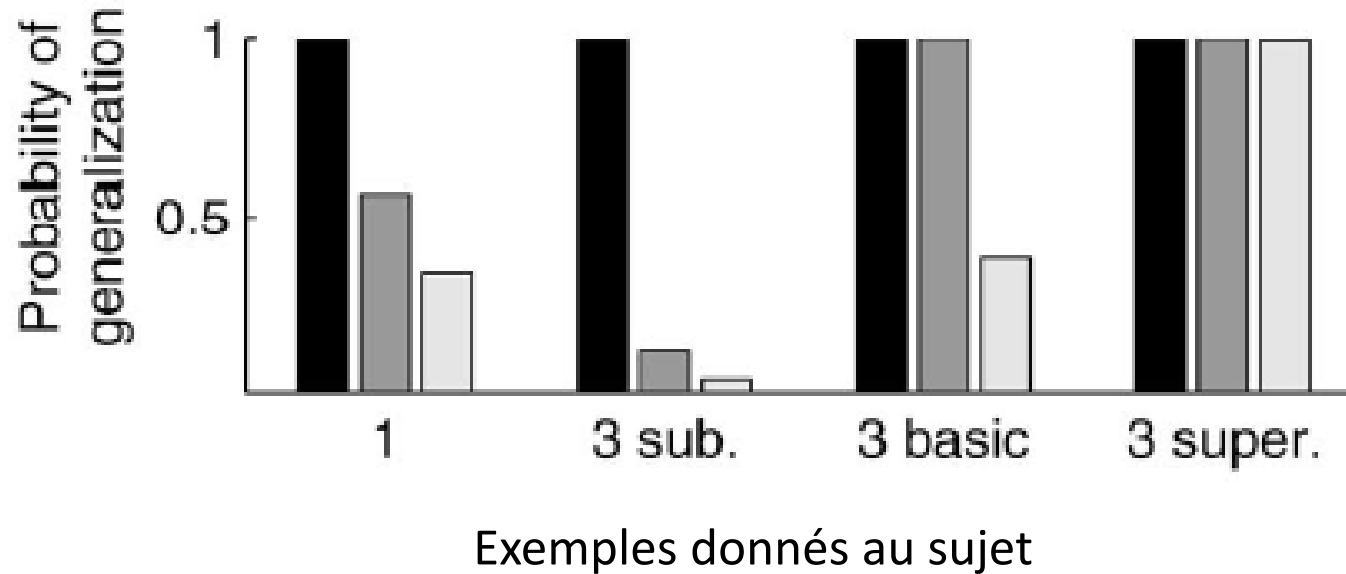
L'expérience de Xu et Tenenbaum (2007)

Prédictions du modèle bayésien:
(l'arbre des hypothèses est fondé sur une analyse en clusters des jugements de similarité entre les items).

Items de test

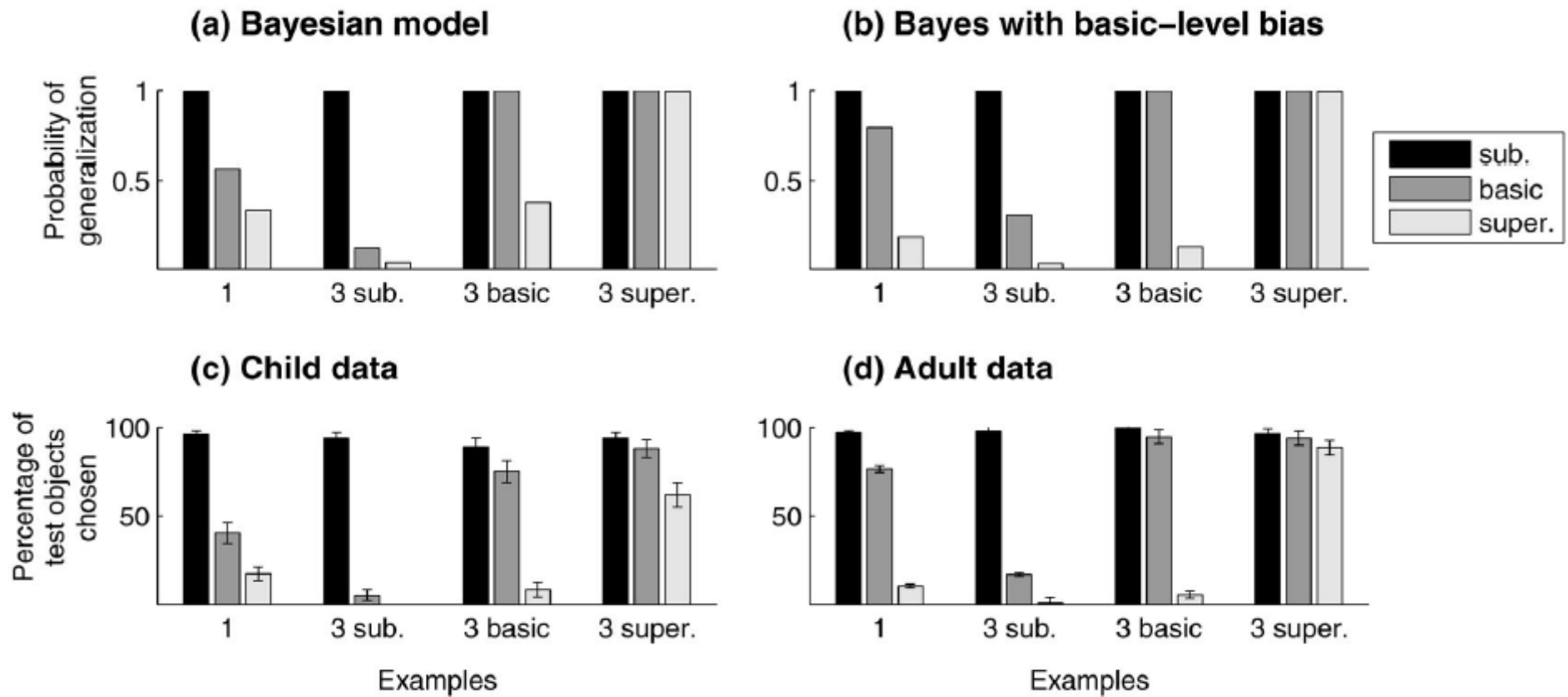


(a) Bayesian model



L'expérience de Xu et Tenenbaum (2007)

Comparaison avec les résultats empiriques chez l'enfant de 3-4 ans et chez l'adulte.



Un modèle bayésien de l'apprentissage du lexique

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychol Rev*, 114(2), 245-272.

Le modèle rend bien compte de la plupart des observations initiales:

1. Toutes les probabilités à posteriori $p(H|X)$ changent rapidement après chaque observation X_i .
2. Il n'est pas nécessaire d'entendre des contre-exemples pour que certaines hypothèses soient éliminées.
3. Les enfants, les adultes, et le modèle sont capables d'apprendre les noms qui étiquettent un espace complexe de catégories enchâssées.
4. Le modèle rend bien compte du caractère graduel des réponses des sujets.
 - Xu & Tenenbaum montrent que d'autres modèles discrets (ne gardant que la ou les règles compatibles avec les observations) s'ajustent moins bien aux données. De même, des modèles connexionnistes fondés sur la simple similarité des observations ne conviennent pas.

Reste le point 5: la flexibilité de l'apprentissage face au contexte pragmatique et aux intentions du locuteur. Plusieurs possibilités:

- l'enfant est sensible au caractère aléatoire de l'échantillonnage. Il se rend compte lorsqu'un échantillon n'est pas représentatif de l'ensemble (Xu & Tenenbaum, 2007)
- l'enfant développe des modèles hiérarchiques de la production verbale.

Le principe d'exclusivité dans l'apprentissage du lexique

L'expérience princeps de Carey et Bartlett (1978): un adulte montre à un enfant de deux ans deux plateaux, l'un d'une couleur bleue prototypique, l'autre d'un vert olive inhabituel. Il dit: "ramène moi le plateau chromé, pas le bleu, le chromé". La plupart des enfants infèrent que "chromé" renvoie à la couleur inconnue, et s'en souviennent 5 semaines plus tard.

Les enfants semblent postuler que chaque entité ne possède qu'un seul nom.

Dès 16 mois, ils évitent d'attribuer à la même entité des noms différents.

- Lorsqu'ils entendent un nom nouveau, ils postulent que celui-ci renvoie à une chose dont ils ne connaissent pas déjà le nom (Markman & Wachtel, 1988)
- Si, lorsqu'un nom nouveau est énoncé, ils ne voient qu'un objet dont le nom est déjà connu, ils cherchent activement un autre référent pour le nom nouveau (Markman et al., 2003).

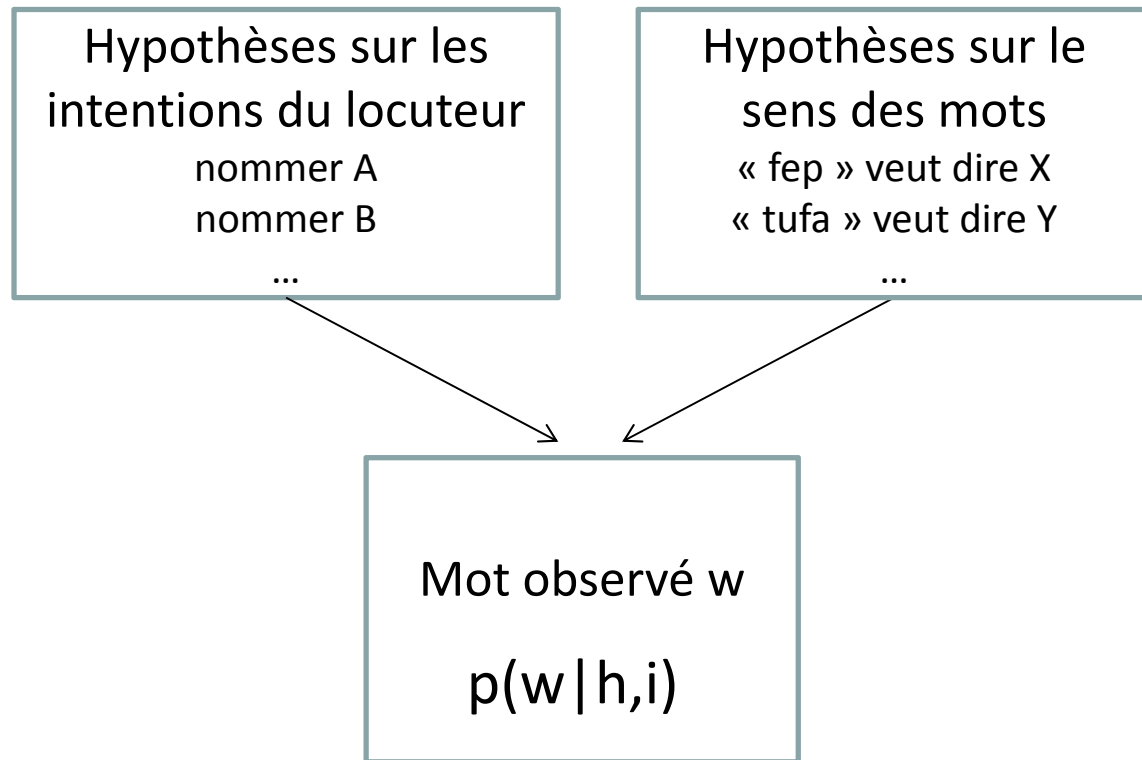
Ce comportement est élevé au rang de principe linguistique par Ellen Markman (1989) (« mutual exclusivity assumption », 1989) et Eve Clark (« principles of contrast and conventionality », 1988). Il pourrait faciliter l'apprentissage des noms de catégories, des noms de parties d'objets, des noms d'action, etc.

Mais... ne pourrait-il pas résulter d'un principe bayésien plus général?

Une origine possible du principe d'exclusivité

L'énoncé d'un mot ne dépend pas seulement de son sens, mais aussi de l'intention du locuteur (et de sa connaissance de ce mot).

Si l'enfant exploite ce modèle bayésien hiérarchique, il peut inférer simultanément l'intention du locuteur et le sens des mots – ou bien utiliser sa connaissance de l'un pour contraindre l'autre.

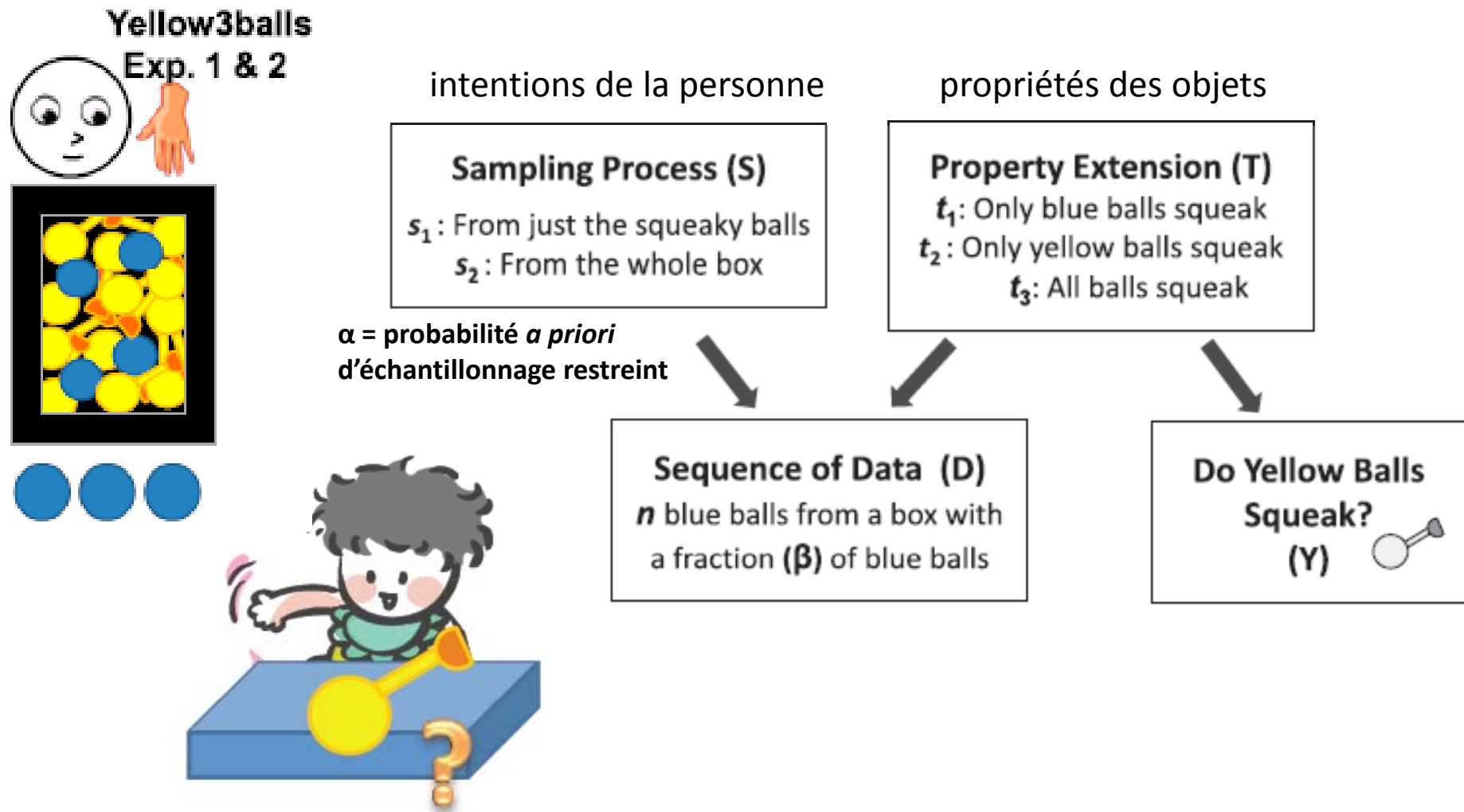


Exemples: Un adulte, en présence d'un chien, dit le mot « truffe ».

- l'enfant sait que, s'il avait voulu nommer le chien, il aurait probablement dit « chien »
- Il en déduit que la probabilité a posteriori qu'il ait voulu nommer le chien est faible
- il en déduit que la probabilité que « truffe » renvoie au chien est faible
- autres possibilités: Partie du chien? Catégorie de niveau supérieur? L'adulte joue? Il s'est trompé? etc...

Une inférence hiérarchique très similaire: les enfants de 13 à 18 mois sont sensibles au processus d'échantillonnage

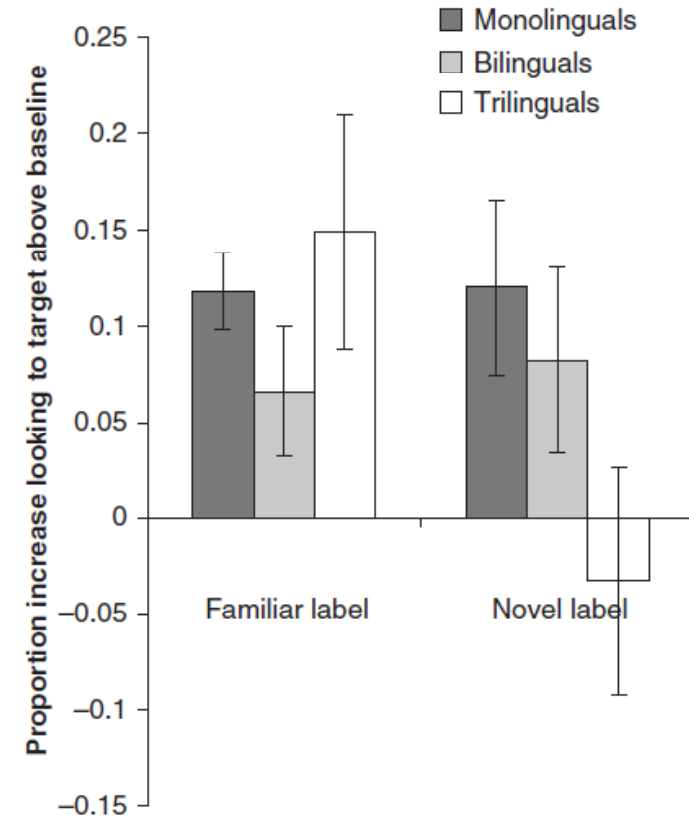
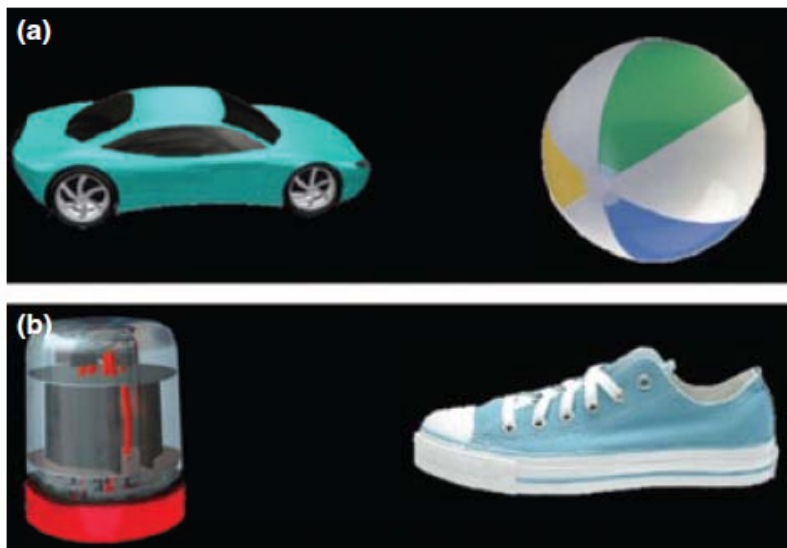
Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proc Natl Acad Sci U S A*, 107(20), 9066-9071.



L'usage du principe d'exclusivité varie avec le multilinguisme

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: infants' language experience influences the development of a word-learning heuristic. *Dev Sci*, 12(5), 815-823.

48 enfants de 17-18 mois, monolingues, bilingues ou trilingues, examinent des objets tandis qu'on leur dit « look at the car, ball, shoe, phototube »



Plus les enfants sont multilingues, moins ils utilisent le principe d'exclusivité.
Conclusion: le principe d'exclusivité n'est pas universel, il est modulé par des facteurs pragmatiques. L'enfant bilingue, qui présente d'ailleurs une meilleure « théorie de l'esprit » (Goetz, 2003; Kovcs, 2009), tient compte du fait que le même objet peut être appelé par des noms différents.

L'apprentissage rapide des mots pourrait faire appel à des mécanismes génériques d'apprentissage

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". *Science*, 304(5677), 1682-1683.

Un chien (Rico) a appris plus de 200 mots.

Lorsqu'on lui présente un mot nouveau, il infère correctement qu'il doit ramener l'objet qu'il ne connaît pas (apprentissage par exclusion).

Par la suite, plusieurs semaines plus tard, il semble se souvenir de la moitié des associations mot-objet ainsi apprises.

Il existe d'autres exemples d'usage de l'exclusion, notamment chez le chimpanzé (Beran, 2010) et le corbeau (Schloegl et al, 2009)



On demande d'abord à Rico deux objets connus: le "T-rex" (le dinosaure bleu), puis le "père Noël" (la petite poupée rouge). Puis un nom nouveau est employé ("sikirid"). Rico ramène le seul objet nouveau: le lapin blanc.

Un autre exemple d'apprentissage hiérarchique: Le biais pour la forme dans l'apprentissage des noms

Nous nommons la plupart des objets en fonction de leur forme (à un certain niveau d'abstraction), mais sans tenir compte de leur taille, de leur texture ou de leur couleur.

Les enfants connaissent-ils ce « principe »? Peuvent-ils l'apprendre?

marteau



chien



Un autre exemple d'apprentissage hiérarchique: Le biais pour la forme dans l'apprentissage des noms en anglais

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological Science (Wiley-Blackwell)*, 13(1), 13.

Etude longitudinale de l'apprentissage des mots chez l'enfant de 17 à 19 mois.

Apprentissage de mots associés à des formes, puis test de la généralisation (1) à des formes nouvelles; (2) à l'apprentissage de mots nouveaux.

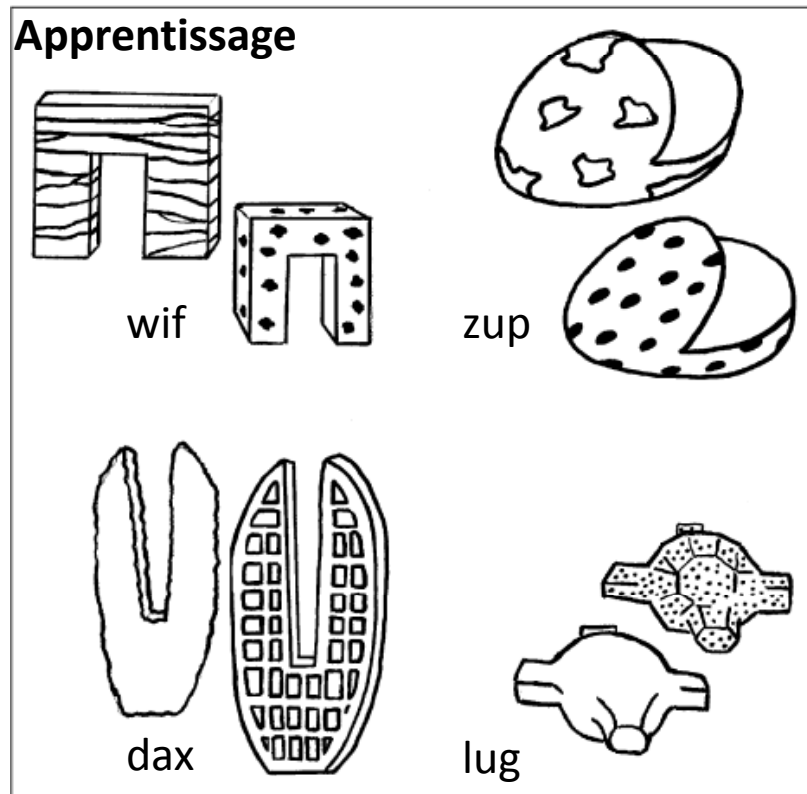


Fig. 2. Illustration of the training stimuli in Experiment 1. The stimuli included two exemplars for each of four novel object categories with novel names. Exemplars of the same category had the same shape, but differed in size, texture, and color.

Test 1. Mot ancien, objet nouveau

this is a lug  88 % de réussite



Where is the lug? Get the lug!

Test 2. Mot entièrement nouveau

this is a veet  70 % de réussite



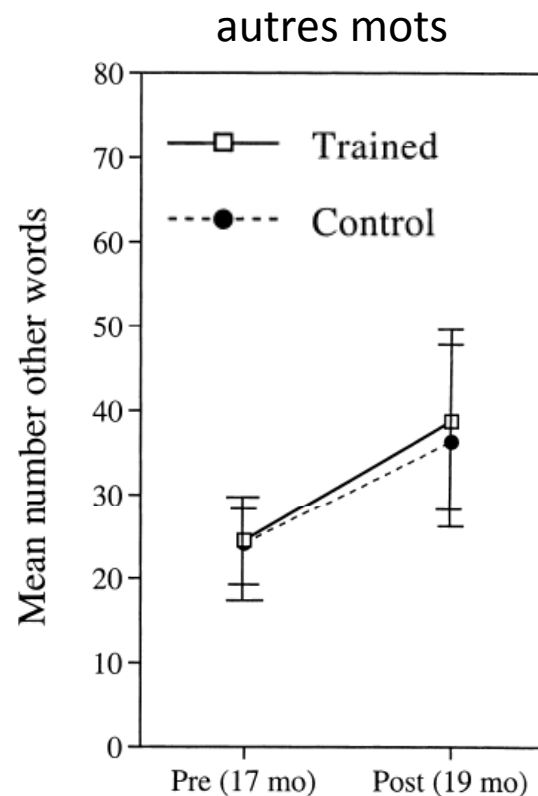
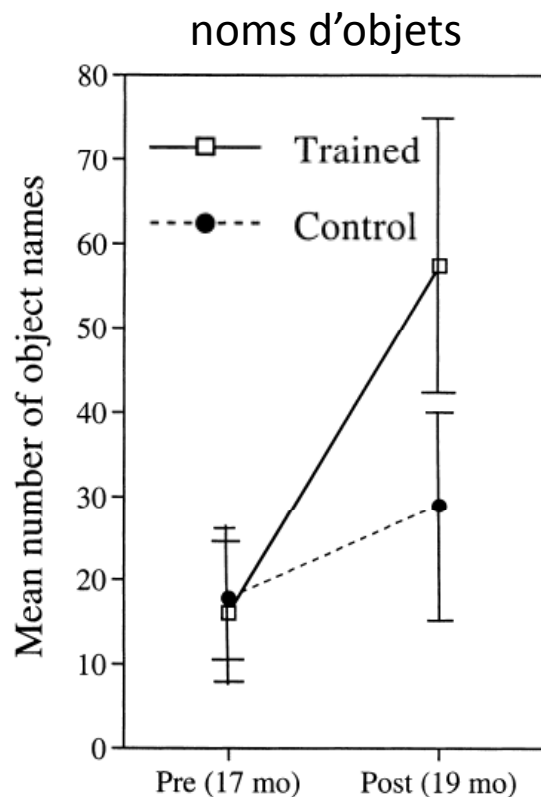
Where is the veet? Get the veet!

Un autre exemple d'apprentissage hiérarchique: Le biais pour la forme dans l'apprentissage des noms en anglais

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object Name Learning Provides On-the-Job Training for Attention. *Psychological Science (Wiley-Blackwell)*, 13(1), 13.

Après ce très bref entraînement à 17 mois, on étudie l'évolution du vocabulaire de l'enfant entre 17 et 19 mois.

Les enfants qui ont été "entraînés" apprennent plus de noms d'objets que les autres!



Conclusions:

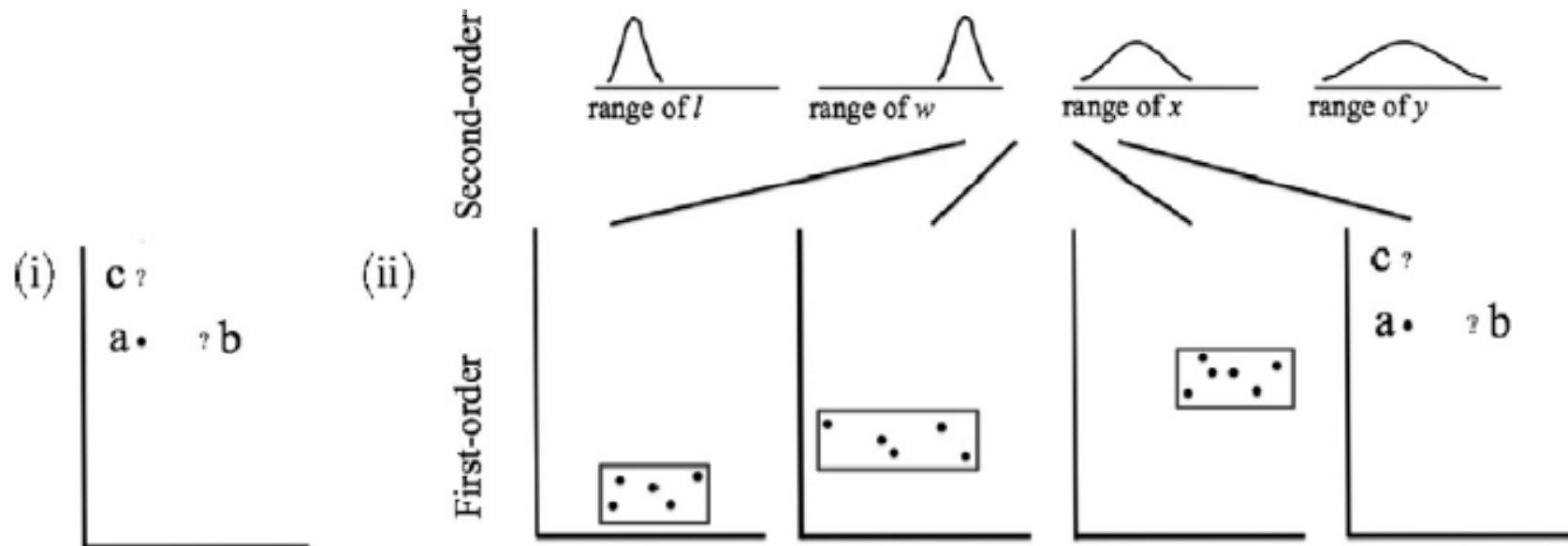
1. Lorsqu'ils apprennent un nom d'objet, les enfants restreignent l'espace des hypothèses à la forme des objets.
2. Ce type de principe peut lui-même faire l'objet d'un apprentissage – et, à son tour, faciliter l'apprentissage d'autres mots.

La modélisation bayésienne de l'apprentissage hiérarchique

L'espace des hypothèses sur le sens des mots peut être, à son tour, engendré par une fonction d'ordre supérieur, avec une probabilité « *a priori* » dont les **hyper-paramètres** sont susceptibles d'être appris.

Concept fondamental de **modèle bayésien hiérarchique**.

- étant donnée une observation **a**, peut-on généraliser aux points **b** et **c**?



- si d'autres catégories ont été apprises auparavant, cela permet de fixer certains hyper-paramètres sur la taille typique des régions de l'espace de premier ordre.
- l'apprentissage va alors généraliser très rapidement à de nouvelles instances: « Blessing of abstraction ».

Le lecteur bayésien

Hypothèse: Même à l'âge adulte, la reconnaissance des mots et l'accès au lexique continuent de faire appel à des inférences bayésiennes.

Une série d'articles fondamentaux par Dennis Norris et ses collaborateurs:

Norris, D., 2006. The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychol Rev* 113, 327-357.

Norris, D., 2009. Putting it all together: a unified account of word recognition and reaction-time distributions. *Psychol Rev* 116, 207-219.

Norris, D., Kinoshita, S., 2008. Perception as evidence accumulation and Bayesian inference: insights from masked priming. *J Exp Psychol Gen* 137, 434-455.

Norris, D., McQueen, J.M., 2008. Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev* 115, 357-395.

Kinoshita, S., Norris, D., 2010. Masked priming effect reflects evidence accumulated by the prime. *Q J Exp Psychol (Colchester)* 63, 194-204.

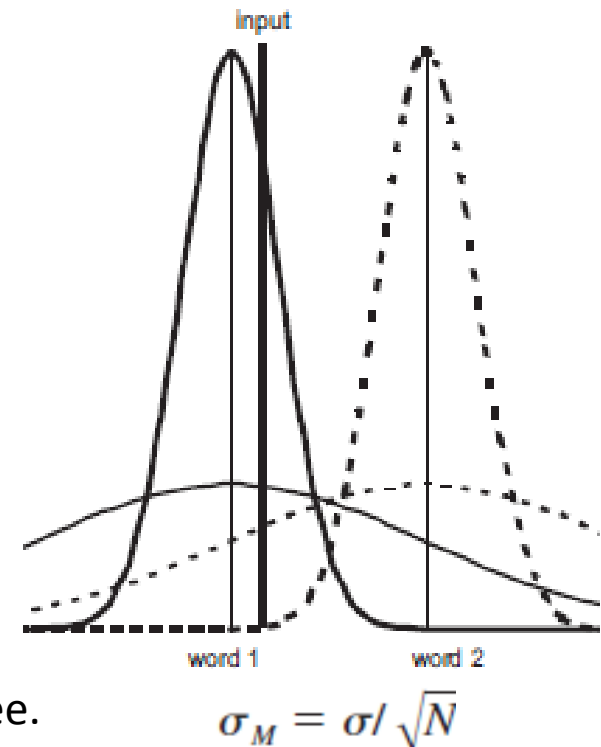
Une analyse rationnelle de la reconnaissance des mots

Le *Bayesian reader* s'inscrit dans le concept d' « analyse rationnelle du comportement » (Anderson, *The adaptive character of thought*, 1990).

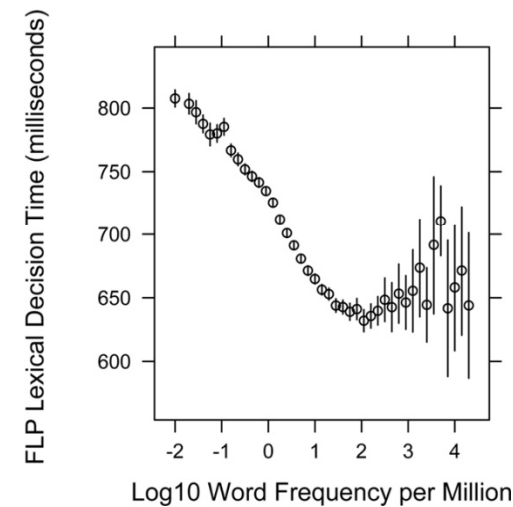
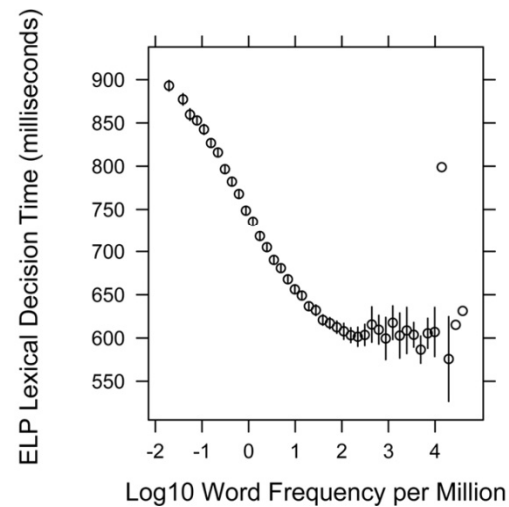
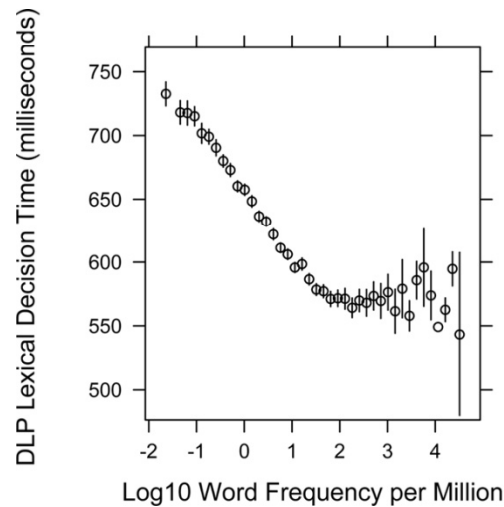
Il s'agit clairement d'une « analyse computationnelle » du problème au sens de la théorie des trois niveaux d'analyse proposés par David Marr.

Comment se pose le problème?

- Les entrées sensorielles sont nécessairement ambiguës.
- Dans ces conditions, la solution optimale consiste à adopter une perspective bayésienne.
- Calculer, pour chaque mot, la probabilité que ce mot ait été à l'origine des entrées sensorielles observées (écrites ou orales)
- Cela revient à calculer $p(\text{entrées} \mid \text{mot}) \times p(\text{mot})$
- Les mots sont représentés dans un espace multidimensionnel
- Le modèle accumule, au fil du temps, l'évidence issue d'échantillons centrés sur l'entrée réelle (par ex. le mot « chat »), et avec une certaine distribution de bruit.
- hypothèse simplificatrice : $p(\text{entrée} \mid \text{mot})$ peut être estimée empiriquement à partir de la distribution d'entrée.



L'effet du log de la fréquence dans l'accès au lexique



Keuleers,
Diependaele,
& Brysbaert,
(2010).
*Frontiers in
Psychology*

Dans de très nombreuses tâches (dénomination, décision lexicale...), le temps de réaction et le taux d'erreurs diminuent avec le logarithme de la fréquence.

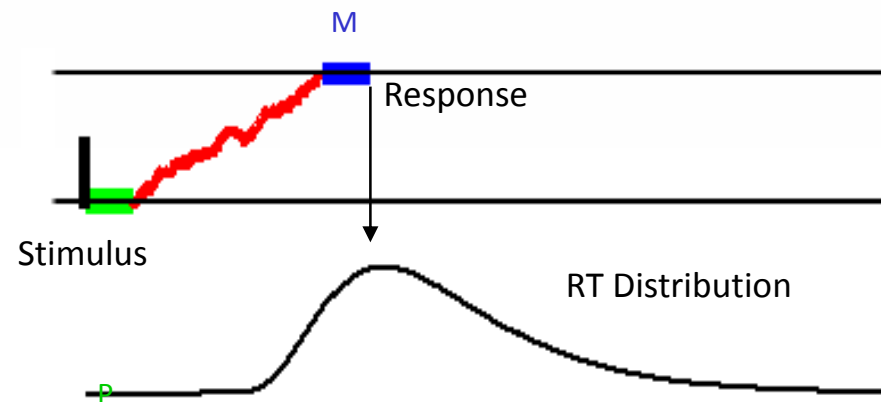
La plupart des modèles se contentent de le **stipuler** (par exemple, en supposant que l'activation de repos ou le seuil de déclencher des unités lexicales sont proportionnels à Log F)

D'autres (connexionnistes) stipulent qu'il résulte d'un **apprentissage** plus intensif des mots les plus fréquents. Mais...

- Pourquoi un logarithme?
- Pourquoi les mots les plus fréquents n'écrasent-ils pas les plus rares? (interférence catastrophique)
- Pourquoi, chez l'adulte, l'apprentissage ne converge-t-il toujours pas vers une asymptote indépendante de la fréquence de présentation des mots?

L'effet de fréquence découle immédiatement d'un modèle Bayésien

- Si l'on cherche la probabilité que l'entrée X corresponde à un mot particulier, il faut appliquer la règle de Bayes: $P(W|X)$ est proportionnelle à $P(X|W) P(W)$
 - On voit directement apparaître la probabilité *a priori* $P(W)$ qui, en l'absence de toute autre connaissance, est égale à la fréquence du mot.
- Si l'on dispose de plusieurs échantillons successifs et indépendants, $X_1, X_2, X_3 \dots$ alors le théorème de Bayes fournit une façon simple d'**accumuler ces données**:
 $P(W | X_1, X_2, X_3 \dots)$ est proportionnelle au produit $P(W) P(X_1 | W) P(X_2 | W) P(X_3 | W) \dots$
- En passant au logarithme, on obtient une sommation de l'évidence.
 $\text{Log}(P(W | X_1, X_2, X_3 \dots)) = a + \text{Log } P(W) + \text{Log } (P(X_1 | W)) + \text{Log } (P(X_2 | W)) + \text{Log } (P(X_3 | W)) + \dots$
- La principe de décision correspond à une **marche aléatoire interne**:
 - Chaque échantillon fait évoluer, en positif ou en négatif, l'évidence totale en faveur des différentes options de réponse.
 - Une décision optimale peut être prise lorsque l'évidence totale atteint un seuil
 - toutes choses égales par ailleurs, le seuil sera atteint d'autant plus vite que l'a priori est grand \rightarrow dépendance en $\text{Log}(P(W)) = \log$ de la fréquence du mot.



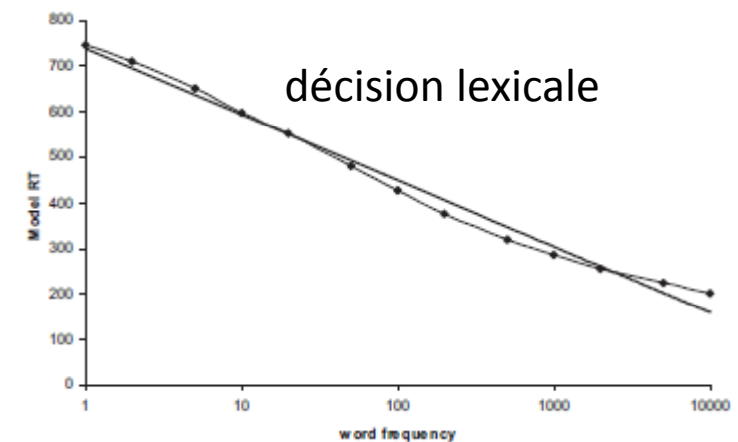
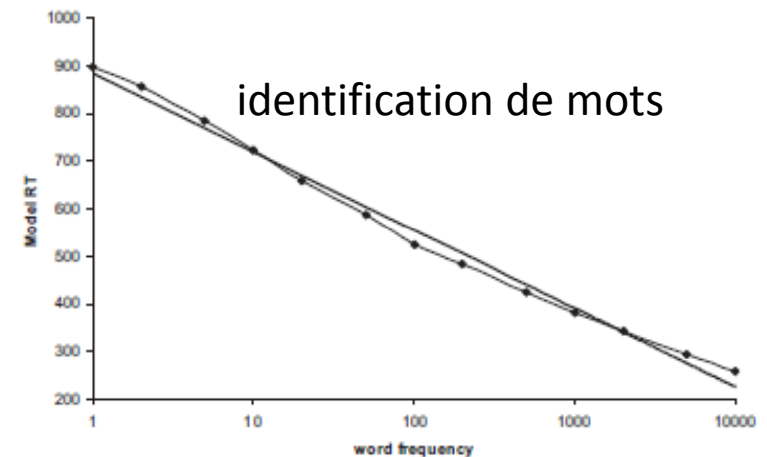
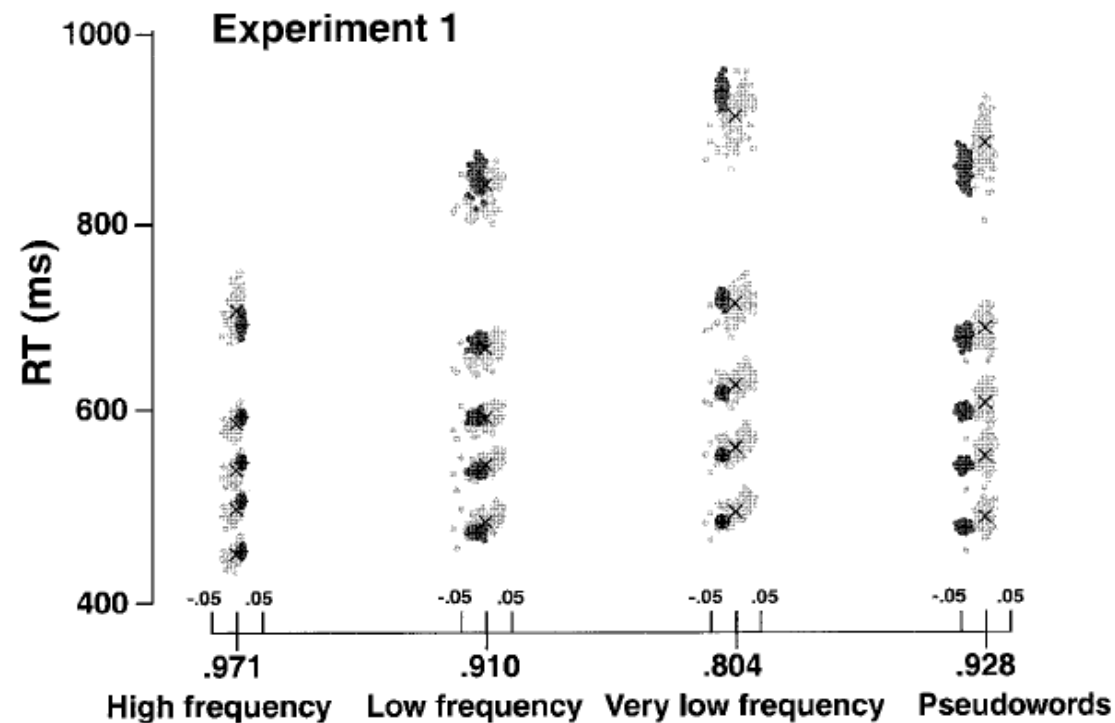
Simulations de l'effet de fréquence

Norris (2006) simule très facilement l'effet du log de la fréquence sur le temps de réaction.

Norris, D. (2006). The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychol Rev*, 113(2), 327-357.

Ratcliff et al. (2006) simulent un modèle phénoménologique de la marche aléatoire correspondant à la prise d'une décision lexicale.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychol Rev*, 111(1), 159-182.



Norris (2009) reprennent ces simulations, mais avec des paramètres non-arbitraires, issues du modèle bayésien (quasi-)optimal.

Norris, D. (2009). Putting it all together: a unified account of word recognition and reaction-time distributions. *Psychol Rev*, 116(1), 207-219.

Simulation de la tâche de décision lexicale

La tâche de décision lexicale exige de décider si une entrée est un mot ou pas.

Plusieurs points sont particuliers à cette tâche:

- Elle ne demande pas nécessairement d'identifier le mot [calculer $P(\text{mot}|\text{entrée})$], mais simplement de calculer $P(\text{réponse}|\text{entrée})$.
- Cela demande un modèle interne des pseudo-mots. Celui-ci peut varier selon le contexte de la tâche.
- Norris considère une simplification: seul compte le « pseudo-mot virtuel » le plus proche de l'entrée.
- Résultats:

Correlations of Model With Data From Balota et al. (1999)

Variable	1	2	3	4
1. Reaction time	—			
2. Log frequency	-.60	—		
3. Rank frequency	-.63	-.89	—	
4. N	-.11	.10	-.10	—
5. Model (.99)	.61	-.99	.91	-.14
6. Model (.95)	.56	-.89	.82	-.25

La modélisation des effets de voisinage: un succès du modèle bayésien

On appelle « voisins orthographiques » des mots qui ne diffèrent que d'une lettre

Par exemple: « chat » et « char ».

Le mot « froid » n'a pas de voisin (d'après www.lexique.org)

L'effet du nombre de voisins N varie selon la tâche:

- dans les tâches d'identification lexicale, la décision se ralentit lorsque N croît.
Pollatsek et al. (1999) retrouvent un tel effet « inhibiteur » en mesurant les mouvements oculaires dans une tâche « écologique » de lecture.
- dans la tâche de décision lexicale:
 - pour les mots, la décision s'accélère (typiquement) lorsque N croît
 - pour les non-mots, la décision se ralentit lorsque N croît.
 - avec des variantes: par ex, Johnson et Pugh (1994) trouvent, sur les mots, un effet facilitateur de N lorsque les non-mots sont non-prononçables, mais un effet inhibiteur de N lorsque les non-mots sont prononçables.

Cette inversion de l'effet selon la tâche pose de grands problèmes à la plupart des modèles de la lecture

– mais tombe tout naturellement dans l'escarcelle du modèle bayésien!

La modélisation des effets de voisinage: un succès du modèle bayésien

Pourquoi?

Les non-mots peuvent être vus comme des « trous » entre les mots dans l'espace de représentation lexicale.

Dans les zones à haute densité de mots, $P(\text{entrée} | \text{mot})$ est plus élevée que $P(\text{entrée} | \text{non-mot})$.

Bien entendu, cette conclusion dépend de la densité des non-mots et de leur proximité avec les mots.

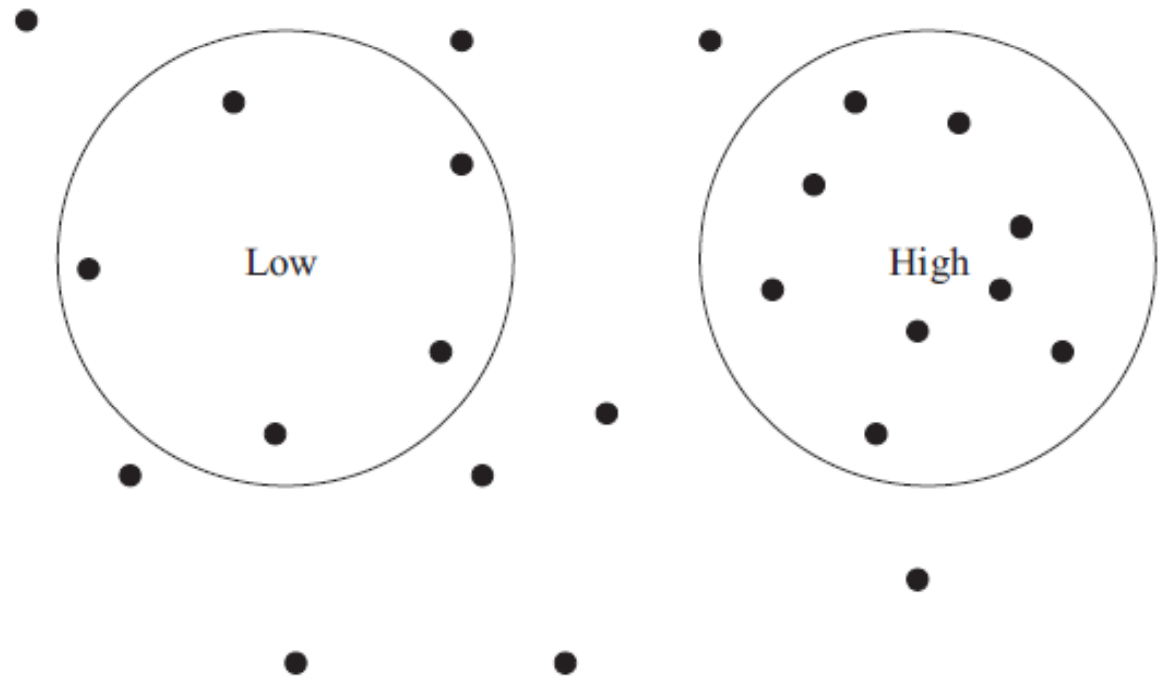


Figure 5. Illustration of areas of high and low neighborhood density in perceptual space. Each dot represents a word, and “High” and “Low” correspond to the centers of input distributions in high- and low-density areas of lexical space, respectively.

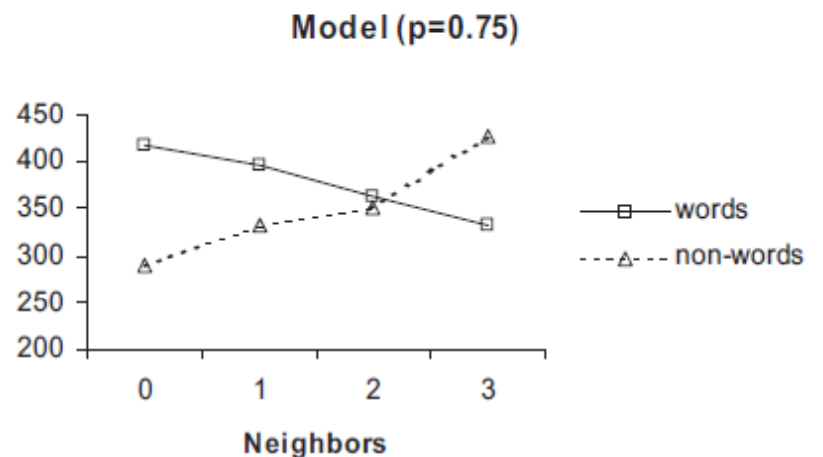
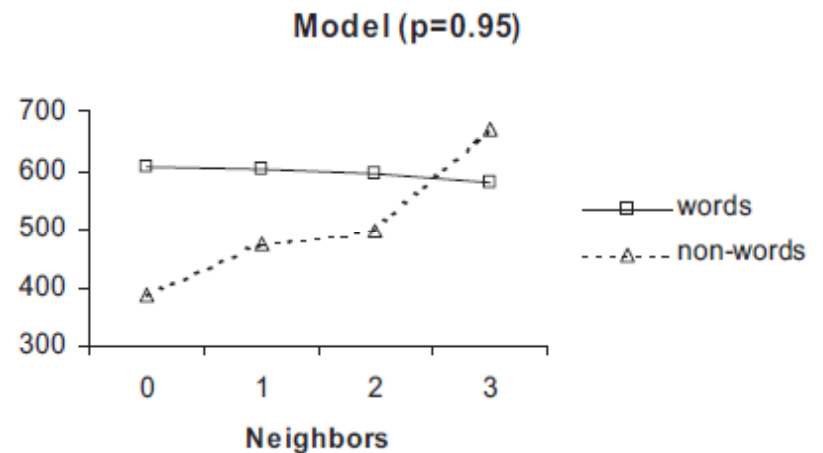
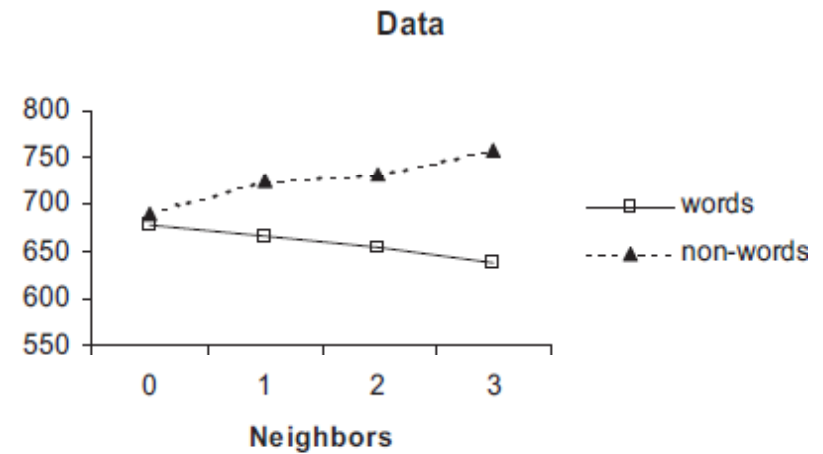
La modélisation des effets de voisinage: un succès du modèle bayésien

Simulation par Norris (1996) des données de Forster et Shen (1996).

(le paramètre p est le seuil de réponse)

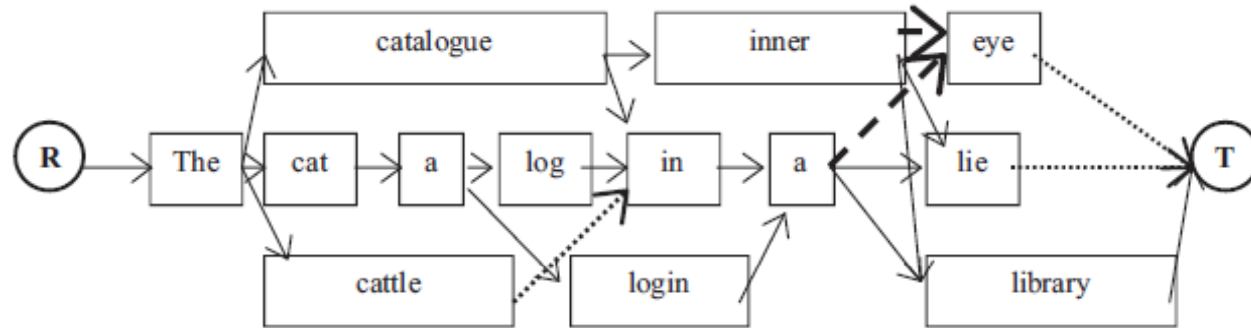
A noter que la simulation capture mal la vitesse relative des réponses « mot » et « non-mot ».

Mais cela résulte vraisemblablement des approximations faites au cours de la simulation.



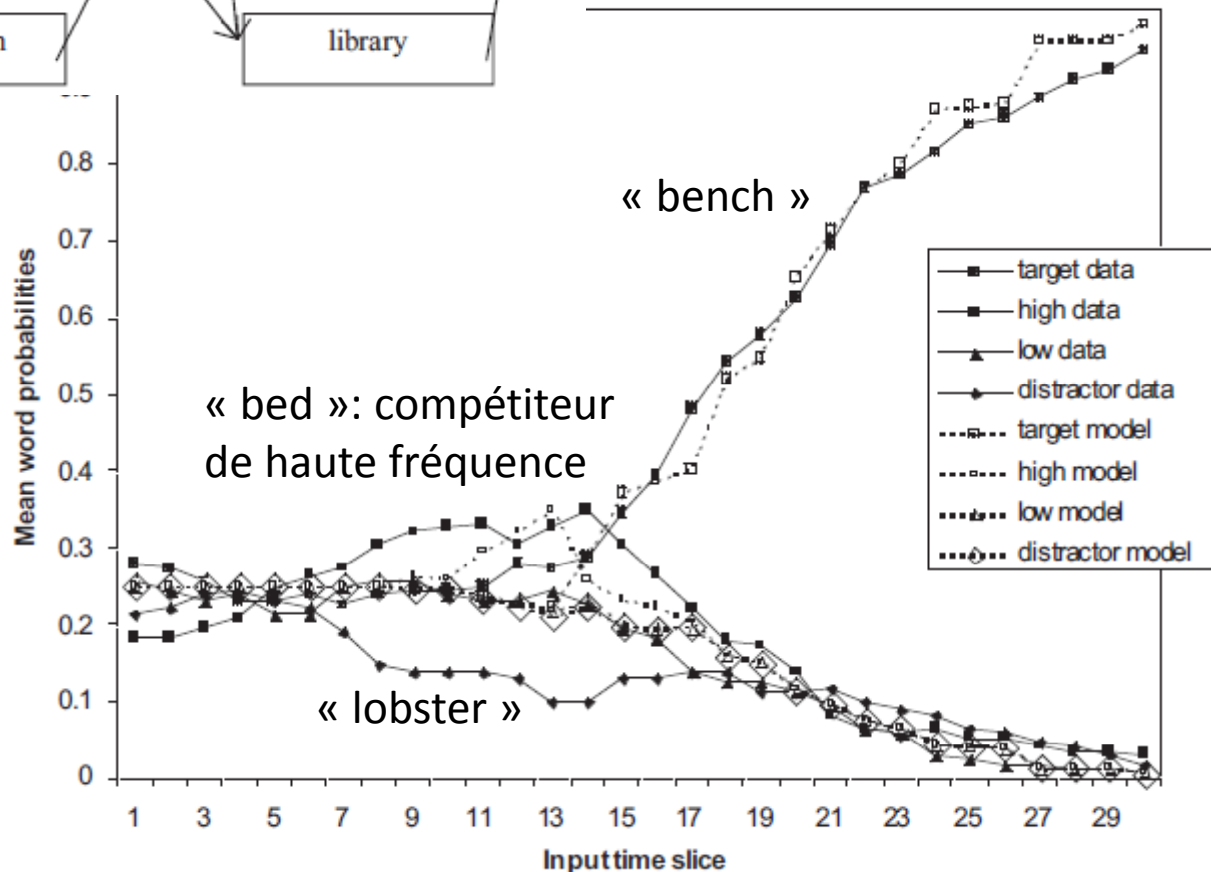
Shortlist B: généralisation à la reconnaissance des mots parlés

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev*, 115(2), 357-395.



L'entrée se déroule dans le temps, dans l'esprit des modèles Cohort (Marslen-Wilson) et Trace (McClelland & Elman)

Exemple: l'expérience de Dahan, Magnuson & Tanenhaus (*Cog Psych* 2001): On mesure l'orientation du regard alors qu'une personne entend « look at the bench ». Distracteurs = bed, bell, lobster. Le modèle rend bien compte des données: fixation d'abord vers le compétiteur de haute fréquence, puis vers le mot cible.



Conclusions: de nouveaux succès pour la théorie bayésienne

Au cours de la première année de vie, l'enfant compile des statistiques qui lui fournissent:

- une liste de candidats pour les mots de sa langue
- une hiérarchie d'hypothèses sur les objets et concepts auxquels ils réfèrent

Au cours de la seconde année, pendant la phase d'explosion lexicale ces deux ensembles vont être mis en relation.

La théorie bayésienne semble susceptible d'expliquer

- la **vitesse** remarquable avec laquelle cet apprentissage se produit
- les **contraintes** que l'enfant applique pour restreindre l'apprentissage
 - principe d'exclusivité
 - restrictions sur la nature des hypothèses lexicales (les noms réfèrent principalement à la forme des objets)
 - inférences sur l'attention, les intentions et les connaissances du locuteur

A l'âge adulte, la théorie bayésienne continue de rendre compte

- de l'apprentissage de mots nouveaux
- de la manière dont les mots sont reconnus en temps réel (effets de fréquence, de voisinage, de cohorte, d'amorçage...)