

Simulation de l'apprentissage des contextes nominaux/verbaux par n -grammes

Perrine Brusini¹ Pascal Amsili² Emmanuel Chemla³ Anne Christophe³

⁽¹⁾ Language, Cognition and Development Lab, Int. School for Advanced Studies (SISSA), Trieste

⁽²⁾ Laboratoire de Linguistique Formelle, CNRS & Université Paris Diderot

⁽³⁾ Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS & ENS, EHESS)

contact: pbrusini@gmail.com, amsili@linguist.univ-paris-diderot.fr

Résumé. On présente une étude d'apprentissage visant à montrer que les contextes locaux dans un corpus de parole adressée aux enfants peuvent être exploités, avec des méthodes statistiques simples, pour prédire la catégorie (nominale vs. verbale) d'un mot inconnu. Le modèle présenté, basé sur la mémorisation de n -grammes et sur une « graine sémantique » (un petit nombre de noms et verbes supposés connus et catégorisés) montre une excellente précision à toutes les tailles de graine sémantique, et un rappel plus faible, qui croît avec la taille de la graine sémantique. Les contextes les plus utilisés sont ceux qui contiennent des mots fonctionnels. Cette étude de faisabilité démontre que les très jeunes enfants pourraient exploiter les contextes de mots inconnus pour prédire leur catégorie syntaxique.

Abstract. A learning study is presented whose aim is to show that local contexts, in a child-directed speech corpus, can be exploited, with simple statistical methods, to predict the category (noun vs. verb) of unknown words. The model we present here is based on the memorisation of n -grams and on a “semantic seed” (a small number of nouns and verbs supposedly known and well categorised). It shows an excellent precision for every size of the semantic seed, and its recall grows along with the size of the semantic seed. The most useful contexts are the ones that include function words. This feasibility study shows that very young children could exploit the contexts of unknown words to predict their syntactic category.

Mots-clés : apprentissage, modélisation de l'acquisition du langage, n -grammes.

Keywords: learning, language acquisition modeling, n -gram.

1 Motivation

On sait que les enfants sont capables de distinguer dès 18 mois les contextes syntaxiques nominaux et verbaux : depuis Shipley *et al.* (1969), de nombreuses expériences comportementales montrent, par exemple, qu'un mot inconnu, associé à une situation où apparaissent à la fois une action nouvelle et un objet nouveau, est associé par les enfants à l'action s'il a été employé dans un contexte verbal (*Regarde, il dase !*) et à l'objet s'il a été employé dans un contexte nominal (*Regarde le dase !*) (Bernal, 2007, par exemple).

Un des aspects qui distinguent les contextes nominaux des contextes verbaux est la nature des mots fonctionnels présents dans ces contextes : le clitique *je* précède exclusivement un verbe (ou d'autres clitiques verbaux), alors qu'un déterminant comme *un* ne précède qu'un nom ou un adjectif. La situation est rendue compliquée, on le sait bien, par la grande quantité de mots fonctionnels homophones (*la* à la fois article (nominal) et clitique (verbal), *son* à la fois adjectif possessif (nominal) et nom plein, *etc.*).

Les mots fonctionnels se distinguent des mots de classe ouverte sur plusieurs plans : ils sont généralement courts et non accentués, ce qui a pu conduire à l'idée qu'ils étaient trop discrets pour être exploités par les enfants en cours d'acquisition (Pinker, 1984), mais ils sont également extrêmement fréquents et situés préférentiellement en bordure d'unité prosodique, ce qui les rend faciles à repérer (Shi *et al.*, 1998). De fait, de très nombreuses études démontrent que les très jeunes enfants reconnaissent les mots grammaticaux de leur langue avant un an (Shi, 2014, pour une revue) et les exploitent pour sélectionner des mots de la catégorie appropriée dès l'âge de 18 mois (Cauvet *et al.*, 2014; Zangl & Fernald, 2007).

L'objectif de l'étude présentée ici est de voir dans quelle mesure les propriétés statistiques de la parole adressée aux enfants peuvent être exploitées par les bébés en phase d'acquisition de leur langue. Il ne s'agit pas d'une hypothèse sur la façon dont le cerveau humain fonctionne, mais d'une investigation de faisabilité, avec quelques hypothèses volontaire-

ment simples et plausibles, correspondant à ce que l'on sait actuellement des capacités linguistiques des jeunes enfants. Il ne s'agit pas non plus de prétendre que d'autres indices, verbaux ou non, présents dans la situation de communication ne sont pas aussi exploités.

On suppose que les enfants, au moment où ils travaillent sur la catégorisation des noms et des verbes (potentiellement dans leur seconde année de vie), possèdent déjà un petit lexique qui peut les aider dans cette tâche — Bergelson & Swingley (2012, 2013) montrent que des enfants de 6 et 9 mois connaissent déjà un certain nombre de noms et verbes. On suppose de plus qu'ils sont capables de regrouper ces mots selon leur catégorie sémantique : par exemple, ils pourraient grouper ensemble *doudou*, *jouet*, et *voiture* parce que ces mots réfèrent à des objets, et *boire*, et *manger* qui réfèrent à des actions¹. Les noms référant en général à des objets et les verbes à des actions, ces classes sémantiques pourraient servir de point de départ pour la construction des catégories grammaticales nom et verbe.

À partir de cette hypothèse de base, le modèle que nous présentons ici exploite les quelques mots déjà catégorisés pour apprendre les contextes d'occurrence des noms et des verbes, puis utilise cette connaissance pour catégoriser des mots dont la catégorie n'est pas encore connue, sur la seule base de leur contexte.

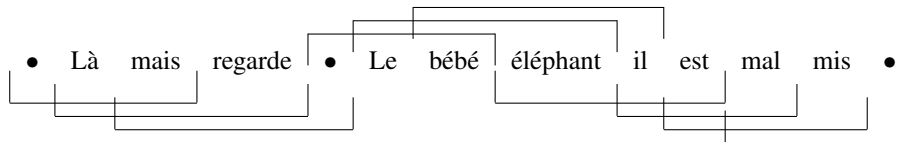
Les hypothèses que nous venons de résumer distinguent de façon assez radicale notre manipulation de la tâche classique de catégorisation morphosyntaxique telle qu'elle est pratiquée en TAL, où la forme de l'item à catégoriser est prise en considération dans le processus, et où l'inventaire des catégories est plus large et plus structuré.

D'autres tentatives de catégorisation non-supervisée sur la base d'indices distributionnels existent dans la littérature (Mintz, 2003; Redington & Finch, 1998). Ces modèles permettent une catégorisation bien meilleure que le hasard, mais souffrent de plusieurs problèmes : soit la catégorisation par le contexte ne fonctionne que sur les mots fréquents (Redington & Finch, 1998), soit elle fonctionne même sur les mots rares ou inconnus, mais crée un grand nombre de classes nominales et verbales distinctes, et surtout ne permet de catégoriser qu'un petit nombre de mots placés dans des contextes très spécifiques (Mintz, 2003). Dans la manipulation présentée ci-après comme dans les tentatives précédentes, le corpus utilisé est segmenté en mots et en tours de parole.

2 Manipulation

Corpus Les corpus sont issus de la base CHILDES 4 (MacWhinney, 2000). Ce sont des transcriptions écrites de paroles spontanées. Notre corpus (133 948 tokens) regroupe différentes transcriptions de deux couples mère-enfant. Nous avons sélectionné les prises de paroles des adultes. Le texte a été catégorisé avec l'étiqueteur de Cordial².

Apprentissage Le système construit son modèle en enregistrant les fréquences de tous les n -grammes de mots rencontrés dans le corpus d'apprentissage. Les ponctuations fortes (telles qu'elles sont transcrites dans le corpus) sont traitées comme des mots, mais les n -grammes comprenant une frontière ne sont comptés que si la frontière est le premier ou le dernier élément du n -gramme.



Prédiction Dans la phase de test, les fréquences apprises peuvent être exploitées pour faire une prédiction, à chaque position du corpus de test, sur la base du $(n - 1)$ -gramme précédent, qu'on appelle le *contexte*. Le mot prédit, dans le contexte (w_1, \dots, w_{n-1}) , est le mot w tel que le n -gramme formé avec le contexte est le plus fréquent, soit $w = \operatorname{argmax}_w \operatorname{freq}(w_1, \dots, w_{n-1}, w)$. Dans ce cas, le contexte est le contexte immédiatement à gauche de la position-cible. On envisage aussi des prédictions dans des configurations contexte/cible différentes : on parlera de *prédiction droite* pour le cas où le contexte est immédiatement à droite du mot cible, et de *contexte imbriqué* pour le cas où les mots du contexte sont répartis également à droite et à gauche du mot cible (n impair et ≥ 3).

Nous avons décidé de ne pas interpréter les fréquences comme des probabilités, et par conséquent il n'est pas nécessaire de recourir à un lissage au sens strict (de type Good-Turing par exemple), mais nous avons implémenté un repli minimal, pour traiter les cas où le contexte n'a jamais été rencontré dans le corpus d'apprentissage : s'il n'existe pas de mot w tel que le

1. Cette idée est compatible avec des résultats expérimentaux qui établissent que les enfants ont une représentation différente pour les agents et les artefacts d'un côté et les actions causales de l'autre (Carey, 2009, pour une revue).

2. Étant donné la nature du corpus, le taux de réussite de Cordial n'était pas excellent, même sur les noms et les verbes, où il restait environ 10% d'erreurs de catégorisation (d'après une évaluation faite à la main sur 500 phrases), mais nous avons pu réduire ce taux par un post-traitement semi-automatique.

modèle a rencontré le n -gramme, une prédiction est tentée en prenant comme contexte le $(n-2)$ -gramme (w_2, \dots, w_{n-1}) , et on recommence si nécessaire. Si cette procédure conduit à un contexte nul, le modèle ne fait aucune prédiction.

Projection L'objectif n'est pas à proprement parler de prédire des mots, mais des catégories (plus exactement l'une des deux catégories N ou V). C'est la raison pour laquelle on procède à ce que nous appelons la *projection* du vocabulaire : supposant que le système connaît déjà la nature d'un certain nombre de noms et de verbes, on remplace ces noms et verbes dans le flux d'apprentissage par leur catégorie. Ce faisant, on se distingue des approches classiques en étiquetage morphosyntaxique qui utilisent des chaînes de Markov cachées.

Nous avons sélectionné les noms et les verbes les plus fréquents du corpus comme mots supposés connus, afin de rendre ce processus le plus automatique possible. Comme point de départ, on considère la situation où le système connaît déjà 10% des occurrences de noms et de verbes dans le corpus ce qui correspond à 6 noms et à 2 verbes. Ensuite on a considéré 4 autres « états de vocabulaire » en doublant à chaque fois le nombre de noms et de verbes connus (types), et nous avons enfin considéré le cas où tous les noms et verbes sont connus. La figure 1 illustre l'évolution du corpus d'apprentissage selon le choix de vocabulaire connu.

FIGURE 1 – Illustration de la « projection » du vocabulaire (V_i : 6×2^i N et 2×2^i V connus)

			—	Là	mais	regarde	!	Le	bébé	éléphant	il	est	mal	mis	!
V_0	6 N	2 V	•	Là	mais	regarde	•	Le	N	éléphant	il	est	mal	mis	•
V_1	12 N	4 V	•	Là	mais	V	•	Le	N	éléphant	il	est	mal	mis	•
V_2	24 N	8 V	•	Là	mais	V	•	Le	N	éléphant	il	est	mal	mis	•
V_3	48 N	16 V	•	Là	mais	V	•	Le	N	N	il	est	mal	mis	•
V_4	96 N	32 V	•	Là	mais	V	•	Le	N	N	il	est	mal	V	•
V_m	1310 N	1253 V	•	Là	mais	V	•	Le	N	N	il	est	mal	V	•

Comme on fait l'hypothèse que les mots déjà catégorisés sont ceux dont l'enfant connaît le sens, on peut supposer qu'il identifie ces mots indépendamment de leur forme morphologique³. La table 1 donne une idée des mots supposés connus selon la taille du vocabulaire initial.

V_0	6N	doudou	bébé	livre	chose	micro	histoire
	2V	aller			faire		
V_1	V_0+6N	pied	poisson	peu ⁴	main	lait	nez
	V_0+2V	mettre			regarder		
V_2	V_1+12N	caméra	fleur	tête	eau	heure	côté
		oeil	bouche	biberon	assiette	éléphant	fois
	V_1+4V	voir			pouvoir		
		dire			falloir		

TABLE 1 – Table présentant quelques-uns des mots connus selon le vocabulaire initial

Test Le test est pratiqué sur un extrait nouveau du corpus, et consiste à comparer la prédiction faite par le modèle avec le mot/la catégorie effectivement présente, et ce pour des *positions cible* particulières, qui doivent vérifier les deux conditions suivantes : (1) le mot du contexte le plus proche de la cible doit avoir été rencontré dans le corpus d'apprentissage (ce qui revient à ne pas faire de repli jusqu'au contexte nul)⁵ ; (2) la cible à catégoriser doit être peu fréquente dans le corpus. Nous avons choisi le seuil de 0,05%⁶.

Mesures Le parti pris de faire l'apprentissage sur un flux unique dans lequel apparaissent aussi bien des mots que des catégories conduit à des réponses diverses de la part du système : il peut aussi bien prédire une catégorie N ou V qu'un mot quelconque, qui peut être un nom ou un verbe n'appartenant pas au vocabulaire connu.

Les réponses du système sont classées en trois catégories : N, V, ni N ni V ; et on définit une mesure de rappel et une mesure de précision pour les deux catégories N et V. On compte une *bonne réponse 'nom'* (BR_N) si le système a répondu N lorsqu'il y avait un nom dans le corpus de test. Ceci signifie que toutes les réponses où le modèle donne un nom spécifique (comme *girafe* ou *poupée*) sont comptées comme un *manqué* (MA_N), comme les cas où une autre réponse que

3. Pour les noms cette hypothèse est validée d'office puisque dans leur grande majorité ils ne distinguent pas à l'oral le singulier du pluriel ; et pour la plupart des verbes les formes fléchies varient peu dans notre corpus (les auxiliaires *être* et *avoir* ont été exclus).

4. *Peu* est un adverbe mais il a des emplois en fonction nominale.

5. Cette condition élimine environ 4% d'occurrences de noms et 9% d'occurrences de verbes.

6. Ce seuil garantit qu'on ne tente pas de prédiction sur les mots déjà connus ; il a de plus la propriété d'éliminer la quasi-totalité des mots fonctionnels, qui sont très fréquents, alors que la majorité des mots de classe ouverte sont moins fréquents que ce seuil (Brusini, 2012, p. 167).

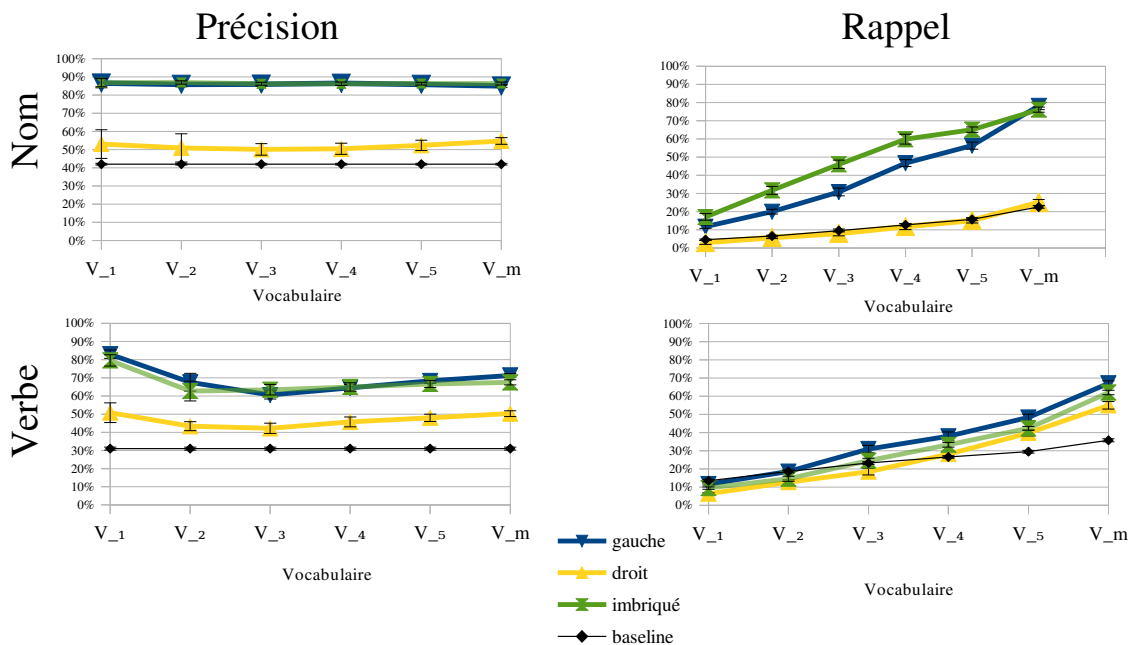
N est donnée. Si le système répond N alors qu'il n'y a pas de nom dans le corpus de test, on compte une *fausse alarme* 'nom' (FA_N). Les mêmes calculs sont effectués pour les réponses V. On peut alors définir de façon classique une précision et un rappel pour chaque catégorie X : $pre_X = \frac{BR_X}{BR_X + FA_X}$ $rap_X = \frac{BR_X}{BR_X + MA_X}$

Baseline Nous prenons comme baseline un modèle qui prédit N, V, ni N ni V, en fonction de la distribution de ces trois "catégories" dans le corpus d'apprentissage. Les distributions varient selon le vocabulaire connu.

Validation croisée Afin de déterminer l'influence de la taille du corpus d'apprentissage sur les performances du modèle et pour pouvoir tester la reproductibilité des résultats obtenus, nous avons procédé à une validation croisée : le corpus a été découpé en 10 sous-corpus de taille égale, et dans chaque cas 2/3 du corpus a servi pour l'entraînement, et 1/3 pour le test. Cette manipulation permet d'une part de calculer un écart-type des modèles (et donc d'évaluer la reproductibilité des résultats) et permet d'autre part d'évaluer l'impact de la taille du corpus d'apprentissage sur la performance. Les barres d'erreurs de la figure 2 correspondent à cet écart-type.

3 Résultats

FIGURE 2 – Performance des 3 modèles en précision et en rappel pour les catégories N et V



Les résultats des calculs pour $n = 3$, pour les contextes gauches, droits et imbriqués, en précision et rappel, comparés à la baseline⁷. On note tout d'abord que les 3 variantes qui exploitent le contexte ont de meilleures performances que la baseline⁷. On peut donc conclure qu'il y a de l'information dans les contextes d'occurrence de noms et de verbes qui permet de catégoriser des mots nouveaux. Le modèle utilisant le contexte droit (en jaune) est le modèle le moins efficace dans cette tâche de catégorisation, sa précision et son rappel sont plus bas que les modèles gauche (bleu) et imbriqué (vert) pour les noms comme pour les verbes⁸. Les deux autres modèles sont très efficaces tous les deux, avec un léger avantage pour le modèle utilisant les contextes imbriqués, notamment pour le rappel des noms où il est significativement meilleur que le modèle gauche⁹. De manière générale, les résultats sont meilleurs pour N que pour V.

Il est frappant de constater que la précision des différents modèles ne s'améliore pas lorsqu'on augmente le nombre de noms et de verbes connus lors de l'entraînement. Même à très petit vocabulaire, la précision des deux meilleurs modèles est excellente, autour de 90% pour les noms et 70% pour les verbes. Au contraire, le rappel augmente avec la taille de

7. Pour l'ensemble des comparaisons entre les 3 modèles et la baseline : Précision : $F(1, 9) > 40$; $p < 0.001$; Rappel : $F(1, 9) > 13$; $p < 0.01$.

8. Précision : $F(1, 9) > 180$; $p < 1.0 \times 10^{-6}$ pour l'ensemble des comparaisons entre les modèles gauche et imbriqué avec le modèle droit. Rappel : ensemble des comparaisons : $F(1, 9) > 50$; $p < 1.0 \times 10^{-4}$.

9. Comparaison des 2 modèles : $F(1, 9) = 10, 6$; $p = 0.01$ pour la précision et $F(1, 9) = 218, 6$; $p = 1.3 \times 10^{-7}$ pour le rappel.

vocabulaire¹⁰ : en partant d'une performance faible à petit vocabulaire (moins de 20% des noms et verbes identifiés), on arrive à une performance plus acceptable à vocabulaire moyen (avec 50% des verbes bien catégorisés lorsque 48 noms et 16 verbes sont connus (V_3), et 50% des noms à 24 noms et 8 verbes connus (V_2)).

On peut aussi constater que la variabilité est très faible, les barres d'erreurs étant à peine visibles parfois. Cela signifie que les performances de nos modèles sont très stables, alors qu'ils sont entraînés et testés sur des parties différentes du corpus. Le même test effectué sur le corpus entier (dix fois plus grand) fournit des résultats très similaires, ce qui indique que la taille du corpus d'apprentissage n'influe pas sur les performances du modèle.

4 Discussion

Le modèle présenté ici a pour particularité de présupposer des connaissances initiales plausibles — la graine sémantique — et de faire appel à des calculs très simples, puisqu'il se contente de comptabiliser la fréquence d'apparition des contextes et sélectionne l'élément le plus souvent rencontré dans un contexte donné. En dépit de cette simplicité, le modèle obtient une excellente performance en termes de précision, et ce même pour la plus petite taille de graine sémantique, seulement 6 noms et 2 verbes supposés connus — ce résultat est vrai pour les contextes gauche et imbriqué. Autrement dit, lorsque le modèle catégorise un mot, il le fait en général de manière correcte (80% pour les noms, 70% pour les verbes). Cette expérience montre que les contextes immédiats des noms et des verbes contiennent beaucoup d'information sur leur catégorie syntaxique.

Ce résultat est d'autant plus surprenant que le modèle opère sa catégorisation en regardant uniquement le contexte : le mot-cible lui-même n'est pas pris en compte lors de l'étape de catégorisation. Cet aspect du modèle est particulièrement désirable lors de l'acquisition d'un lexique, pour deux raisons principales. Premièrement, le modèle est tout à fait capable de catégoriser un mot-cible qu'il n'a encore jamais rencontré, pour peu qu'il soit entendu dans un contexte connu : en termes d'acquisition, un enfant qui rencontre un mot pour la première fois pourrait calculer sa catégorie syntaxique, nom ou verbe, et l'exploiter pour contraindre le sens de ce mot — si c'est un nom il réfère probablement à un objet, si c'est un verbe, à une action. Cette stratégie semble être utilisée par les enfants (Bernal, 2007) par contraste avec la stratégie de (Redington & Finch, 1998), où ce sont les mots connus qui sont catégorisés. Deuxièmement, le lexique contient de nombreux homophones qui appartiennent à des catégories syntaxiques différentes, comme *la fermel/je ferme*. Du fait que le modèle catégorise les mots grâce à leur seul contexte, il n'est pas gêné par de tels homophones, qui sont pour lui des mots comme les autres. Au contraire, un modèle (ou un enfant) qui entreprendrait de classer les mots en prenant en compte à la fois le contexte et la catégorie supposée du mot en cours de classification (telle que stockée dans le lexique), devrait éprouver des difficultés à apprendre le deuxième membre d'une paire d'homophones lorsque le premier membre est déjà connu.

L'excellente performance observée en précision, indépendamment de la taille de la graine sémantique, est en contraste frappant avec la performance en rappel : celui-ci est faible à la plus petite taille de graine sémantique, et augmente au fur et à mesure que la graine sémantique augmente. Intuitivement, lorsque le modèle connaît initialement peu de mots (petite graine sémantique), alors il apprend un petit nombre de contextes de noms et de verbes : lorsqu'il utilise ces contextes pour catégoriser un nom ou un verbe, sa performance est bonne (précision élevée), mais il « rate » beaucoup de noms et de verbes qui apparaissent dans des contextes non reconnus comme prédicteurs de noms ou de verbes. Cette analyse intuitive est confirmée par l'examen des manqués du modèle : pour les petites tailles de graine sémantique, les manqués correspondent le plus fréquemment à des cas où le modèle répond un nom spécifique au lieu de répondre avec la catégorie N (resp. V). En ce qui concerne l'acquisition, ce comportement ne pose pas de problème particulier : en effet, en l'absence d'information pertinente dans un contexte donné, l'enfant peut attendre de rencontrer le mot dans un contexte plus informatif. Au contraire, il serait désastreux de multiplier les fausses alarmes : si l'enfant pense qu'un mot donné est un verbe et réfère à une action, alors qu'en réalité c'est un nom qui réfère à un objet, il apprendra un mauvais sens. Autrement dit, il est préférable pour un apprenant d'avoir une bonne précision et un mauvais rappel, et c'est ce que fait le modèle aux petites tailles de graine sémantique.

L'examen des contextes utilisés le plus fréquemment pour la catégorisation montre qu'ils contiennent majoritairement des mots grammaticaux. Pour les contextes de nom, le mot-cible est le plus souvent immédiatement précédé d'un article ; pour les contextes de verbe, il s'agit de pronoms personnels ou relatifs, et d'auxiliaires. Ainsi, bien que le modèle ne contienne aucune hypothèse *a priori* sur l'utilité potentielle des mots grammaticaux, ceux-ci émergent spontanément dans les contextes utiles, du simple fait de leur fréquence et de leur distribution.

10. Effet de Vocabulaire pour chaque modèle : N : $F(6, 54) > 50$; $p < 1.0 \times 10^{-15}$; V : $F(6, 54) > 200$; $p < 1.0 \times 10^{-15}$.

La comparaison des modèles à contexte gauche, imbriqué et droit montre une performance bien pire sur le contexte droit (il faut noter que le modèle imbriqué se replie sur l'élément de gauche lorsque le contexte complet n'est pas connu, ce qui fait qu'il y a un recouvrement entre les modèles gauche et imbriqué). Ceci pourrait être une conséquence du fait que le français est récuratif à droite, avec en général les mots grammaticaux à gauche des têtes lexicales. Une autre possibilité serait que comme les mots de gauche sont entendus avant le mot-cible, et ceux de droite après, les langues s'organisent de manière à maximiser la quantité d'information disponible avant le mot lui-même. Pour trancher entre ces deux hypothèses, il faudrait effectuer ces mêmes calculs avec une langue récurative à gauche, comme le japonais.

Cette étude démontre la pertinence d'une approche de simulation dont les hypothèses sont contraintes par les résultats d'expérimentation psycholinguistique avec les très jeunes enfants. Par exemple, ce modèle démontre que l'utilisation des mots fonctionnels comme prédicteurs de catégorie pour les mots de contenu ne semble pas nécessiter de construire *a priori* des catégories de mots fonctionnels, telles que *déterminant* ou *clitique sujet* : la simple reconnaissance de l'item pourrait suffire. Ce résultat est très intéressant car l'homophonie entre mots fonctionnels rend leur catégorisation difficile. En retour, les résultats de modèles de simulation comme celui qui est présenté ici pourra permettre de faire des prédictions testables expérimentalement chez les très jeunes enfants, pour peut-être un jour parvenir à un modèle computationnel psychologiquement plausible de l'acquisition des catégories syntaxiques par les jeunes enfants.

Remerciements

Ce travail a bénéficié du soutien de l'Agence Nationale de la Recherche (grants n° ANR-2010-BLAN-1901, ANR-13-APPR-0012, ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC) et de la fondation de France. Nous remercions Benoît Crabbé pour son aide dans la réalisation de la première version du modèle.

Références

- BERGELSON E. & SWINGLEY D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(9), 3253–3258. doi:10.1073/pnas.1113380109.
- BERGELSON E. & SWINGLEY D. (2013). The acquisition of abstract words by young infants. *Cognition*, **127**(3), 391–397. doi:10.1016/j.cognition.2013.02.011.
- BERNAL S. (2007). *De l'arbre (syntaxique) au fruit (du sens) : Interactions des acquisitions lexicales et syntaxiques chez l'enfant de moins de 2 ans*. PhD thesis, Université Pierre et Marie Curie.
- BRUSINI P. (2012). *Découvrir les noms et les verbes : Quand les classes sémantiques initialisent les catégories syntaxiques*. PhD thesis, Université Pierre et Marie Curie.
- CAREY S. (2009). *The origin of concepts*. Oxford University Press.
- CAUVET E., LIMISSURI R., MILLOTTE S., SKORUPPA K., CABROL D. & CHRISTOPHE A. (2014). Syntactic context constrains lexical access in French 18-month-olds. *Language Learning and Development*, **10**(1), 1–18.
- MACWHINNEY B. (2000). *The CHILDES Project : Tools for analyzing talk*. Mahwah, NJ : Lawrence Erlbaum Associates. Third Edition.
- MINTZ T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, **90**(1), 91–117.
- PINKER S. (1984). *Language Learnability and Language Development*. Cambridge, MA : Harvard University Press.
- REDINGTON M. N. C. & FINCH S. (1998). Distributional information : A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**(425–469).
- SHI R. (2014). Functional morphemes and early language acquisition. *Child Development Perspectives*, **8**(1), 6–11.
- SHI R., MORGAN J. L. & ALLOPENNA P. (1998). Phonological and acoustic bases for earliest grammatical category assignment : a cross-linguistic perspective. *Journal of Child Language*, **25**, 169–201.
- SHIPLEY E. F., SMITH C. S. & GLEITMAN L. R. (1969). A study in the acquisition of language : Free responses to commands. *Language*, **45**(2), 322–342.
- ZANGL R. & FERNALD A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development*, **3**(3), 199–231.