



**Lidil**

Revue de linguistique et de didactique des langues

**31 | 2005**

**Corpus oraux et diversité des approches**

---

## Production naturelle précoce et acquisition du langage

L'exemple du développement des noms

**Dominique Bassano**



### Édition électronique

URL : <http://journals.openedition.org/lidil/136>

DOI : [10.4000/lidil.136](https://doi.org/10.4000/lidil.136)

ISSN : 1960-6052

### Éditeur

UGA Éditions/Université Grenoble Alpes

### Édition imprimée

Date de publication : 1 juin 2005

Pagination : 61-84

ISBN : 2-914176-12-0

ISSN : 1146-6480

### Référence électronique

Dominique Bassano, « Production naturelle précoce et acquisition du langage », *Lidil* [En ligne],

31 | 2005, mis en ligne le 03 octobre 2007, consulté le 01 mai 2019. URL : [http://](http://journals.openedition.org/lidil/136)

[journals.openedition.org/lidil/136](http://journals.openedition.org/lidil/136) ; DOI : [10.4000/lidil.136](https://doi.org/10.4000/lidil.136)

---

Ce document a été généré automatiquement le 1 mai 2019.

© Lidil

---

# Production naturelle précoce et acquisition du langage

L'exemple du développement des noms

Dominique Bassano

---

- 1 Le petit d'homme ne naît pas avec le langage. Pour communiquer avec son entourage, il doit apprendre la langue parlée autour de lui, qui est l'un des quelque 6000 systèmes symboliques conventionnels utilisés dans le monde. Et, compte tenu du haut niveau de sophistication des systèmes linguistiques, il y parvient en un laps de temps relativement court. L'enfant commence, en effet, à produire ses premiers mots conventionnels vers le début de sa seconde année, et, deux ou trois ans plus tard, il sera parvenu à mettre en place un système de communication efficace où il aura progressivement intégré la plupart des contraintes de base de la langue : il maîtrise alors les principaux aspects de la phonologie, il connaît à peu près le sens et les conditions d'emploi de plusieurs milliers de mots, et il sait utiliser correctement les principales règles morphologiques et syntaxiques de sa langue maternelle. Il y a donc tout lieu de se demander comment pareille performance est possible...
- 2 Le présent article apporte une contribution à cette interrogation. Nous espérons montrer que l'étude systématique des corpus de production naturelle des jeunes enfants est une voie d'accès féconde pour aider à comprendre les processus d'émergence et d'acquisition du langage, en particulier la mise en place des contraintes fondamentales de la langue. Nous illustrerons notre propos en présentant une étude où nous analysons, dans les productions langagières précoces des enfants français, le développement de la catégorie de mots appelés « noms », qui engage tout à la fois l'acquisition de contraintes lexico-sémantiques et grammaticales de base.  
Le renouvellement des études de production naturelle précoce
- 3 L'étude de la production spontanée des jeunes enfants, versant le plus visible de la capacité langagière, est le moyen le plus ancien auquel les chercheurs aient eu recours pour tenter de comprendre l'acquisition du langage à ses débuts. Les premières investigations – tels les travaux pionniers de W.F. Léopold (1939-1949) ou de A. Grégoire (1947) – reposaient sur les journaux que les auteurs tenaient en observant généralement

leurs propres enfants. L'entrée dans l'ère de la méthodologie moderne se produit vers les années soixante, avec l'innovation que représente la possibilité d'enregistrement des données sonores. Cette innovation va ouvrir la voie aux études qui, à l'instar du projet mené à l'Université d'Harvard par R. Brown et son équipe sur trois enfants devenus célèbres sous les noms d'Adam, Eve et Sarah (Brown, 1973), examinent le développement langagier à travers l'analyse intensive des productions réelles des jeunes enfants. Au cours des dernières décennies, les études de production se sont à nouveau profondément renouvelées, bénéficiant cette fois de l'essor des moyens audio-visuels et surtout informatiques. L'enregistrement vidéo modifie les conditions de recueil des données et permet l'analyse des indicateurs non linguistiques – situations, actions, gestes, regards – associés aux interactions verbales. L'informatisation des données permet des opérations de stockage, de traitement automatique et de transfert qui entraînent un véritable changement d'échelle dans la perspective de recherche.

- 4 Les études de production naturelle s'appuient ainsi maintenant sur des corpus diversifiés et systématiquement recueillis, constitués des productions verbales spontanées des jeunes enfants, enregistrées en audio ou préférentiellement en vidéo au cours de séances où l'enfant est en situation de vie quotidienne et en interaction avec son entourage. Les enregistrements sont ensuite exhaustivement transcrits et informatisés, les transcriptions comportant non seulement les productions de l'enfant et de ses interlocuteurs mais aussi des informations sur la situation et le contexte de ces productions. Les productions de l'enfant, voire de ses interlocuteurs, peuvent être ainsi diversement codées selon les dimensions souhaitées, et conformément aux objectifs des recherches. Comme on le voit, le recueil et l'analyse des données de production naturelle sont extrêmement coûteux en temps et en énergie, ce qui limite nécessairement la taille des échantillons. Mais en contre-partie, ces données sont d'une fiabilité élevée et d'une très grande richesse, donnant accès à l'analyse en contexte de tous les aspects du développement langagier, phonologique aussi bien que lexical, morphosyntaxique ou pragmatique.
- 5 L'une des manifestations les plus significatives de la révolution informatique dans l'étude de la production naturelle a été la création de réseaux et systèmes informatisés destinés à gérer la transcription, le codage et l'analyse des données. Le plus connu de ces réseaux est sans doute le *Child Language Data Exchange System* dit CHILDES, fondé en 1984 par B. MacWhinney et C. Snow grâce au financement de la Fondation MacArthur. Le CHILDES, dont la vocation est de favoriser les échanges entre chercheurs à travers la communauté scientifique internationale, fournit trois principaux instruments progressivement mis au point au cours des années et présentés en détail dans plusieurs ouvrages (MacWhinney, 2000) et mis à jour sur Internet. Le premier instrument, le *CHAT Manual*, est un guide pour l'informatisation des données de production naturelle et fournit un ensemble d'instructions ou normes destinées à assurer une standardisation des procédures de transcription et de codage. Les règles minimales concernent la présentation de la transcription, tandis que des conventions plus élaborées sont proposées pour le codage (phonologique, morphologique, codage des actes de parole par exemple). Le deuxième instrument est un ensemble de programmes informatiques permettant l'analyse automatique des données transcrites dans le format CHAT. Enfin le troisième instrument est la base de données elle-même, constituée d'un nombre croissant de corpus transcrits dans ce format. L'intérêt de principe d'une telle base de données est évident, puisqu'elle permet au chercheur qui veut examiner une question ou tester une hypothèse d'avoir

accès à un vaste ensemble de données d'origines variées et traitables automatiquement. Cette base de données s'étend dans diverses directions : au départ centrée sur les données d'enfants anglophones tout-venants, elle cherche à s'ouvrir aux données provenant d'autres langues, de langues secondes et d'enfants présentant des désordres du langage, et vise à associer de plus en plus aux transcriptions écrites les documents audio et vidéo digitalisés.

L'émergence du langage et les premières étapes de l'acquisition  
La constitution du système linguistique chez l'enfant français : une approche fonctionnaliste intégrative

- 6 Nous menons depuis plusieurs années un programme de recherche sur l'émergence du langage chez l'enfant et les premières étapes de l'acquisition, entre 1 et 3-4 ans environ, étude menée à travers l'analyse de la production naturelle. Ce programme, principalement centré sur l'acquisition du français comme langue maternelle (mais pouvant faire appel à des comparaisons inter-langues), a pour objectif général de dégager les processus de constitution du système linguistique chez l'enfant. Pareil objectif définit un domaine de recherche qui, s'il est bien représenté au plan international, est encore relativement peu exploré en France, où les travaux consacrés aux premières étapes du développement se sont surtout concentrés sur la communication prélinguistique et les interactions conversationnelles (Marcos, 1998 ; Veneziano, 2000) ou sur l'étude de la perception et de la production de la parole (Boysson-Bardies, 1996 ; Bertoncini & Boysson-Bardies, 2000). Par contraste, nos recherches sur la constitution du système linguistique ont pour but de déterminer les processus par lesquels l'enfant intègre les contraintes fondamentales de la langue, telles par exemple les contraintes lexicales et morphosyntaxiques, voire les contraintes pragmatiques. Nos travaux ont ainsi porté principalement sur la structuration du lexique (Bassano, 1998a ; 2000a ; Bassano, Maillochon & Eme, 1998 ; Bassano, Eme & Champaud, 2005), et sur certains processus de grammaticalisation, concernant notamment les noms et les verbes (Bassano, 1998b ; 1999 ; 2000b ; Bassano, Maillochon, Klampfer & Dressler, 2001a, b ; Bassano, Laaha, Maillochon & Dressler, 2004).
- 7 L'ensemble du programme développe une approche fonctionnaliste intégrative de l'acquisition du langage. Apparues en psycholinguistique autour des années quatre-vingt, les approches fonctionnalistes (Bates & MacWhinney, 1982 ; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996 ; MacWhinney, 1999 ; Tomasello, 2003) offrent une alternative aux modèles syntaxiques, modulaires et innéistes de l'acquisition du langage. Ces approches mettent l'accent sur les déterminants environnementaux et sociaux ou sur les déterminants cognitifs généraux, sans pour autant nier l'existence de prédispositions ni les particularités linguistiques de l'acquisition du langage. Elles conçoivent ainsi le développement langagier comme un phénomène global et interactif, auquel contribuent ensemble facteurs génétiques et facteurs environnementaux.
- 8 Éclairée par la théorie des systèmes dynamiques (Van Geert, 1991 ; 1994), notre approche se définit plus précisément par deux caractéristiques. La première est sa perspective intégrative. L'idée de base, au cœur des débats sur modularité et interaction, est que, dès le début, l'acquisition du langage est gouvernée, non seulement par le jeu des interactions entre facteurs génétiques, environnementaux et cognitifs, mais aussi par celui des interactions « internes » à l'œuvre entre les divers composants de la capacité langagière. L'une des questions centrales est, par exemple, celle des relations entre développement lexical et développement grammatical, que nous examinons de façon conjointe et dont nous recherchons les lieux d'interface et d'interdépendance. La deuxième caractéristique

est l'intérêt accordé à la variabilité : non seulement à la variation inter-linguistique, mais aussi à la variation inter- et intra-individuelle. L'analyse de ces différentes sources de variation permet de repenser la recherche des universaux, mais également, comme le montrent les modèles fondés sur la théorie des systèmes dynamiques, d'apporter un éclairage nouveau sur l'examen des changements, moments critiques et transitions dans les processus de développement (Bassano & Van Geert, 2005).

La base de données « Corpus français de productions langagières précoces »

- 9 Pour mener notre programme de recherche sur l'acquisition du système linguistique par les enfants francophones, nous avons entrepris de constituer une base de données sur le français, en relation avec le CHILDES mais de façon indépendante. Cette base est formée d'un ensemble de corpus de productions de jeunes enfants, enregistrés, transcrits et informatisés pour l'essentiel conformément aux standards du CHILDES (à titre d'exemple, un extrait de corpus transcrit est donné en annexe).
- 10 À ce jour, notre base de données *Corpus français de productions langagières précoces* comporte :
  - les corpus longitudinaux de Benjamin (de 21 à 48 mois) et de Pauline (de 14 à 39 mois)<sup>1</sup>. Dans les corpus longitudinaux, l'évolution du langage d'un même enfant est suivie au cours d'une longue période de temps -plusieurs mois ou plusieurs années- au moyen d'enregistrements régulièrement espacés et assez rapprochés (une ou deux sessions par mois, chacune d'une durée allant jusqu'à 2 heures).
  - les corpus transversaux ou mixtes de la base dite « TRL<sup>2</sup> » (transverso-longitudinale), constituée par les enregistrements de différents enfants étudiés à certains moments privilégiés du développement. Nos données sont recueillies à 20 mois, 30 mois, 39 mois, et 48 mois, chaque groupe d'âge comportant 20 enfants (les séances ont ici une durée de 30 à 45 minutes). Parmi ces enfants, dix d'entre eux se retrouvent d'un groupe à l'autre, formant la composante longitudinale du corpus. Ce corpus permet d'élargir la portée des résultats obtenus dans les corpus longitudinaux, et d'examiner la variabilité inter-individuelle au sein d'un groupe d'âge.
  - le corpus « Autisme<sup>3</sup> », constitué par les données langagières d'enfants atteints d'autisme. Ces données sont transversales (30 et 39 mois d'âge de développement verbal, 8 enfants par groupe d'âge), et longitudinales (4 enfants suivis depuis 30 mois d'ADV à 3 reprises, à intervalle de 9 mois).
- 11 Après le recueil des données, une étape essentielle dans l'étude de la production naturelle est celle de l'élaboration et de la mise en œuvre de systèmes de codage des productions des enfants. Comme nous l'avons indiqué, le CHILDES propose des systèmes de codage très complets. Nous avons cependant préféré élaborer nous-même des codages qui soient le plus possible appropriés aux objectifs et hypothèses de nos recherches. Pour chaque session transcrite, un fichier de codage comportant les productions de l'enfant est constitué ; en règle générale, il consiste en un échantillon formé d'un nombre constant d'énoncés extrait de la transcription, en l'occurrence 120 énoncés pour chaque enfant et chaque mois d'âge. Cet échantillon est choisi de manière la plus « écologique » possible, en respectant l'organisation des séquences discursives et en privilégiant les séquences les plus riches, tout en conservant l'idée que l'échantillon doit être représentatif du langage de l'enfant. Chaque énoncé est soumis à plusieurs séries de codages correspondant aux différents objectifs des recherches. Les codages sont réalisés en référence constante à l'ensemble de la transcription et aux informations fournies par les productions des interlocuteurs et les indications du transcripteur. Un programme informatique de

traitement des codages (PROCOD) <sup>4</sup> permet une première analyse automatique des données.

Le développement des noms en français Les noms dans la langue et dans l'acquisition

- 12 Nous nous proposons maintenant d'illustrer nos recherches sur l'étude de la production naturelle précoce et la constitution du système linguistique en présentant un ensemble d'analyses sur le développement des noms. On considère généralement que les noms constituent, avec les verbes, l'une des deux catégories linguistiques fondamentales qui structurent l'organisation d'un très grand nombre de langues, parmi lesquelles le français. Les noms servent à désigner les entités du monde, animées ou inanimées, tandis que les verbes permettent de faire référence aux événements, actions ou états attribués aux entités. Les noms, ancrés dans la dimension spatiale, remplissent le plus souvent des fonctions de sujet ou d'objet, tandis que les verbes, ancrés dans la dimension temporelle qu'ils expriment par des variations morphologiques, ont le plus souvent une fonction prédicative.
- 13 Figurant parmi les premiers mots conventionnels produits par les jeunes enfants, les noms ont un statut privilégié dans l'acquisition. En même temps que les enfants, au début de leur deuxième année, commencent à se servir de petits mots fonctionnels à valeur socio-pragmatiques tels qu'interjections, routines, particules d'interaction comme « oui » ou « non », ils se mettent aussi à utiliser des éléments qu'on désigne sous le terme de « noms » dans la langue adulte. Bien qu'elle ait été récemment questionnée par certaines études sur l'acquisition de langues comme le coréen ou le chinois (Choi & Gopnik, 1995 ; Tardif, 1996), l'idée que l'acquisition des noms est précoce et précède celle des verbes ne peut guère être mise en doute pour un bon nombre de langues qui partagent l'opposition verbo-nominale. L'existence du « noun bias », ou prédominance initiale des noms, a été tout particulièrement mise en évidence dans des études qui, en utilisant les questionnaires parentaux standardisés du *Communicative Development Inventories* (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994) ont examiné le développement lexical d'enfants acquérant comme langue maternelle l'anglais, l'italien, l'espagnol ou l'hébreu, par exemple (pour une revue, voir Bassano, 2000a et Bassano *et al.*, 2005). Diverses hypothèses ont été avancées pour expliquer la prédominance initiale des noms. La plus connue est l'hypothèse de la Partition Naturelle proposée par D. Gentner (Gentner, 1982 ; Gentner & Boroditsky, 2001) : les noms sont acquis précocement en raison de leur simplicité notionnelle et cognitive, parce que, contrairement aux verbes et aux autres termes relationnels, ils réfèrent généralement à des entités perceptives distinctes, d'une grande cohésion interne et hautement individualisées – des personnes, des objets concrets –, qui formeraient ainsi le marche-pied logique de l'acquisition lexicale. L'assignation du sens des noms est aussi beaucoup moins variable à travers les langues que celle du sens des verbes, moins dépendante des procédés de lexicalisation et des organisations spécifiques aux langues particulières, ce qui en faciliterait l'acquisition.
- 14 Nos études de production naturelle sur la structuration lexicale nous ont permis de mettre en évidence la place qu'occupent les noms dans le lexique des enfants français au cours des premières étapes du développement du langage, montrant en particulier leur précocité et leur haute fréquence initiale, ainsi que leur relatif déclin par la suite. Ainsi, l'analyse du corpus de Pauline (Bassano *et al.*, 1998) a fait apparaître que, parmi les quatre macro-catégories lexicales distinguées, deux prédominent largement en fréquence jusqu'à 18 mois au moins : les noms, qui forment 40 % des mots types, et les éléments que nous avons appelés « para-lexicaux » (interjections, particules oui/non, formules) qui en

forment plus de 30 %. En revanche, les prédicats (verbes et adjectifs) sont à cette étape nettement moins représentés (15 %), et les mots grammaticaux moins encore (10 %). Par la suite, au cours du développement, les proportions de noms et d'éléments para-lexicaux diminuent tandis que les proportions de prédicats et de mots grammaticaux augmentent, de sorte qu'autour de deux ans les quatre catégories lexicales occupent à peu près la même part du lexique, et qu'à 30 mois, ce sont les mots grammaticaux qui sont le plus représentés (40 % des mots-types), suivis par les prédicats et les noms (26 % et 23 %) et finalement les items para-lexicaux (11 %). Ces tendances développementales ont été confirmées par une étude transversale impliquant un grand nombre d'enfants (20 par groupe d'âge) observés à 20, 30 et 39 mois (Bassano *et al.*, 2005). La proportion moyenne de noms trouvée dans les productions des enfants diminue significativement entre 20 et 30 mois (de 31 % à 23 %), de même que la proportion d'items para-lexicaux (de 30 % à 17 %), tandis qu'augmentent significativement les proportions de prédicats (de 17 % à 23 %) et de mots grammaticaux (de 23 % à 38 %). Il est à noter que les changements développementaux observés entre 30 et 39 mois sont nettement moins drastiques qu'entre 20 et 30 mois : les changements entre 30 et 39 mois ne sont significatifs ni pour les noms (dont la proportion passe de 23 % à 26 % des mots-types), ni pour les prédicats (dont la proportion reste de 23 %), ni pour les mots grammaticaux (de 38 % à 41 %). Ces résultats indiquent que c'est au cours de la troisième année que se produisent les réorganisations les plus importantes dans la structuration du lexique, en particulier le déplacement d'accent de la production des noms vers celle des verbes, et l'explosion des mots grammaticaux.

- 15 Nous allons maintenant examiner plus précisément comment émerge et se développe la catégorie des noms, au plan lexico-sémantique d'abord, puis au plan grammatical, en nous appuyant principalement sur l'analyse des données issues du corpus de la petite Pauline.

#### Le développement lexico-sémantique des noms

- 16 L'évolution de la production des noms<sup>5</sup> chez Pauline entre 14 et 30 mois est présentée dans la figure 1 selon les deux indicateurs fournis par l'analyse de la production naturelle, le nombre de mots-types (les mots différents) et le nombre d'occurrences (tous les mots) relevés dans les échantillons mensuels. Cette analyse fait apparaître que la production des noms, en quantité relativement importante dès le début de la période analysée, reste, malgré des variations conjoncturelles, à peu près stable jusqu'à 25 mois, pour connaître à partir de 25 mois une assez nette augmentation, sensible au regard des deux indicateurs. Le décrochement qui se produit à cet âge dans la production des mots-types est particulièrement intéressant, parce qu'il dénote un phénomène d'accroissement non linéaire de la masse du lexique des noms.

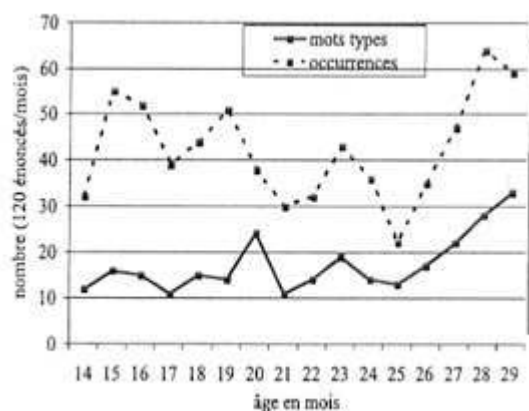


Figure 1 : Production lexicale des noms chez Pauline

- 17 Pour aller plus loin dans l'analyse du lexique des noms et de son développement, nous avons recherché quelles sous-classes sémantiques étaient produites par les jeunes enfants. Appliquant des distinctions classiques, nous avons considéré trois dimensions dans l'organisation sémantique de cette classe de mots : la dimension principale est la distinction « animé » vs « inanimé », qui se combine hiérarchiquement avec les dimensions « nom propre » vs « nom commun » et « concret » vs « abstrait ». Les six sous-classes ainsi obtenues par combinaison sont, pour les noms animés, des noms propres (Edouard, Babar), des noms communs concrets (papa, garçon, oiseau), et des noms communs plus abstraits, tels que, par exemple, ceux qui expriment une relation de parenté (père, sœur), une profession (policier), ou ont une valeur relationnelle (copain, mari) ou générique (foule, animal). Les sous-classes de noms inanimés sont des noms propres (Paris), des noms communs concrets, qui sont les classiques noms d'objet (pomme, chausson, pied), et des noms communs plus abstraits, tels que ceux qui codent des séquences d'action (dodo, histoire), des émotions et des sentiments (peur, faim) ou encore certains noms génériques ou collectifs (lumière, bruit).
- 18 Le développement des différentes sous-classes sémantiques de noms trouvées dans le corpus de Pauline est représenté dans la figure 2 selon une distribution en cinq périodes successives, la première constituée de 4 mois, et les suivantes formées chacune de 3 mois (nombre mensuel moyen de mots-types pour chacune des sous-classes et pour les échantillons de 120 énoncés).

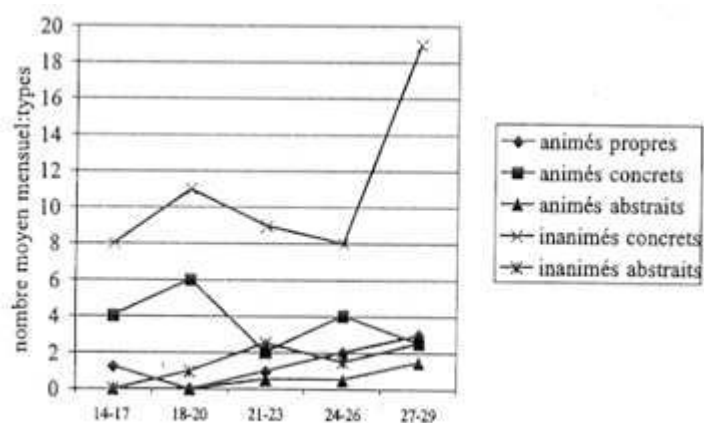


Figure 2 : Distribution des sous-classes sémantiques des noms dans le corpus de Pauline



- 19 Durant la première période, c'est-à-dire avant 18 mois, trois sortes de noms sont produits. On trouve en majorité des noms concrets inanimés, c'est-à-dire les fameux noms d'objet désignant des aliments, des vêtements, des parties du corps, des objets quotidiens. Mais on trouve aussi, en bonne fréquence, des noms communs animés concrets, qui désignent, soit des humains – et ce sont alors souvent des « quasi noms propres », comme « maman », « papa », « bébé » –, soit des animaux – et ce sont alors souvent des animaux sous forme symbolique représentés par des images ou des pièces de puzzle. On trouve aussi des noms propres animés, qui réfèrent de façon assez répétitive à l'entourage immédiat. Après 18-20 mois, les noms concrets inanimés restent les plus nombreux dans chacune des périodes, et ils augmentent en particulier dans la dernière période : ce sont eux qui sont responsables de « l'explosion lexicale » des noms observée à partir de 25 mois. En revanche, les noms communs animés tendent à diminuer au cours du développement. Les noms d'inanimés abstraits apparaissent vers 18-20 mois, référant essentiellement, au début, à des séquences d'action (dodo, histoire), et, ultérieurement, à des émotions et des sentiments, ce qui montre combien le début de l'abstraction est enraciné dans l'expérience. Les noms d'animés abstraits sont les plus tardifs et les plus rares. Et enfin, il n'y a pas de noms propres inanimés dans les échantillons de Pauline.
- 20 En conclusion, cette analyse accrédite en bonne partie l'idée classique que les noms d'objets concrets forment le substrat du lexique des noms chez l'enfant, dans la mesure où ce sont eux qui constituent la sous-classe majoritaire (au total 60 % des mots-types et 50 % des occurrences). Mais cette idée est seulement en partie vraie : nous avons souligné le rôle important joué dans les premiers mois, au moment de l'émergence du langage, par les noms d'animés, en particulier les noms propres et quasi noms propres, qui sont produits en bonne fréquence jusqu'à 20 mois ; ce n'est qu'après cet âge que la différence de fréquence entre animés et inanimés devient significative, à l'avantage des inanimés. Nous avançons l'idée que le développement de la fonction référentielle du langage a une double origine, puisée, d'une part, dans l'identification des individus personnalisés, et, d'autre part, dans l'identification des objets et classes d'objets. La fréquence initiale des noms d'animés, particulièrement des noms propres, s'accorde avec l'interprétation du « noun bias » comme une tendance naturelle de l'enfant à apprendre d'abord les termes qui désignent des entités fortement individualisées.

Le processus de grammaticalisation des noms

- 21 La question se pose maintenant de déterminer quand et comment les mots identifiés comme « noms » au vu de leur propriétés lexico-sémantiques acquièrent pour l'enfant les propriétés grammaticales de la classe des noms. Ce processus développemental, que nous appelons « grammaticalisation des noms », se manifesterait par l'intégration, dans les productions de l'enfant, de la contrainte d'emploi du déterminant. En effet, dans la langue française orale, la caractéristique grammaticale du nom est que celui-ci est, en règle générale, accompagné par un déterminant. Cette contrainte est double : le déterminant est antéposé (il précède le nom), et il porte les marques du genre et du nombre. Cependant, la contrainte n'est pas absolue, puisqu'il est assez fréquent que l'omission du déterminant soit acceptable et même obligatoire, par exemple devant les noms propres ou dans certaines expressions verbo-nominales.
- 22 Pour examiner cette question, nous avons défini un indice de grammaticalisation des noms, qui mesure la capacité de l'enfant à utiliser un déterminant prénominal dans les contextes où celui-ci est obligatoire. L'indice est donc le taux des noms employés avec déterminant par rapport à tous les noms pour lesquels le déterminant serait requis dans

la langue adulte. La figure 3 représente l'évolution de cet indice dans le corpus de Pauline entre 14 et 30 mois, calculé selon deux versions, l'une stricte et l'autre « accommodante ».

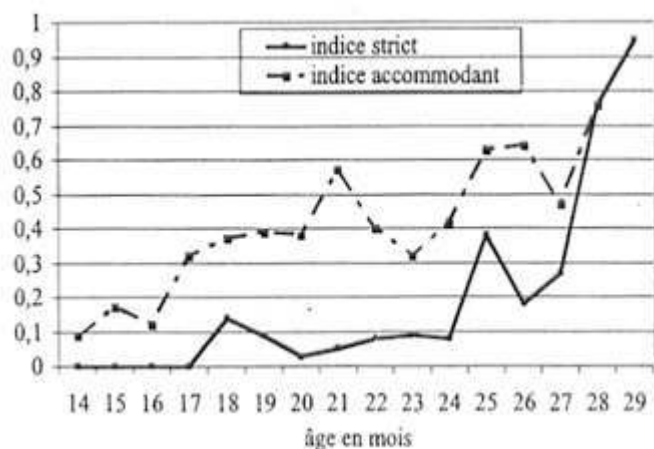


Figure 3 : Développement de l'indice de grammaticalisation des noms chez Pauline

- 23 La version stricte de l'indice mesure le taux d'utilisation de vrais déterminants, clairement spécifiés, comme dans « un chat », « le chien », « donne les chaussures », « c'est mon chausson ». Cette analyse indique que l'enfant ne commence pas à utiliser de déterminant en bonne et due forme avant 18 mois, et qu'entre 18 mois et deux ans, elle les utilise avec beaucoup de modération, puisque l'indice conserve une valeur moyenne d'environ 0,10. Ce n'est qu'après deux ans que sa valeur va augmenter très vite, de manière explosive, puisqu'il atteint la valeur de 0,75 à 28 mois et de 0,95 à 29 mois. Le taux de 95 % d'usage correct du déterminant permet de dire que le processus de grammaticalisation des noms est achevé à 2 ans et demi pour cette enfant-là. On remarquera que la version stricte de l'indice de grammaticalisation fait apparaître un typique phénomène d'explosion grammaticale, se manifestant par une accélération brusque du processus d'intégration de la contrainte, et localisé pour cette enfant entre 27 et 28 mois.
- 24 Mais nous avons établi aussi une autre version, appelée « accommodante », qui tient compte non seulement des vrais déterminants, mais aussi des *fillers*. Les *fillers* (ou « remplisseurs ») sont les éléments monosyllabiques qu'on observe le plus souvent devant les noms ou devant les verbes dans les productions des jeunes enfants, et dont on peut faire l'hypothèse qu'ils sont la préfiguration d'un morphème grammatical. Devant un nom – par exemple dans « e nez », « c'est a chat », « am lit », « mettre é chausson », « eum boîte » – le *filler* préfigure vraisemblablement un déterminant dont la place est bonne mais dont la nature et la forme ne sont pas clairement définies. Le calcul de la version accommodante de l'indice montre que le processus de grammaticalisation est plus précoce et plus progressif que ne le laissait penser la version stricte, et que l'explosion dans l'emploi du déterminant est en réalité préparée par des phénomènes précurseurs. Ce type d'analyse conduit ainsi à repenser les notions de continuité ou discontinuité de développement, en montrant que, selon la finesse de l'examen, un processus qui apparaît comme discontinu (explosif) à un certain niveau d'analyse peut se révéler en continuité à un autre niveau.

Interactions entre lexique et grammaire dans le développement des noms

- 25 L'examen d'une dernière question, celle des relations entre développement grammatical et développement lexical, apportera quelques arguments en faveur de la conception

intégrative de l'acquisition du langage que nous soutenons : nous souhaitons montrer que l'émergence de la grammaire, ou du moins de certains aspects de la grammaire, est liée à certains aspects du développement du lexique.

- 26 Une première série d'arguments concerne les aspects quantitatifs de la relation entre lexique et grammaire. Le point de départ de cet examen est la vérification de l'hypothèse avancée par Bates et ses collègues, l'hypothèse « de la masse lexicale critique », qui propose que la constitution d'un certain stock lexical est une condition nécessaire, voire suffisante, au déclenchement des acquisitions grammaticales (Marchman & Bates, 1994). Le développement grammatical serait de la sorte dans une relation de dépendance par rapport au développement lexical. Cette relation de dépendance est notamment montrée, dans des études fondées sur les données du CDI américain, par la mise en évidence de fortes corrélations non-linéaires entre l'accroissement du lexique et divers aspects de l'émergence de la grammaire, par exemple le début des combinaisons de mots ou de la complexité syntaxique (Fenson *et al.*, 1994 ; Bates & Goodman, 1999). Pour établir l'existence de telles relations de dépendance, plus spécifiques, dans nos données longitudinales sur les noms, il nous faut établir, d'une part, qu'il existe des moments « critiques » remarquables dans l'accroissement de la production lexicale et dans l'essor du processus de grammaticalisation, et, d'autre part, montrer que ces moments critiques sont en décalage temporel. Or nos analyses ont en effet montré que, dans le corpus de Pauline, la production lexicale des noms était marquée par un accroissement à partir de 25 mois, et qu'une explosion dans la grammaticalisation des noms se produisait entre 27 et 28 mois. Il existe donc entre les deux processus – expansion lexicale et grammaticalisation – un parallélisme d'évolution avec un léger décalage temporel, puisque l'explosion lexicale précède de deux mois l'explosion de la grammaticalisation. Cela permet de penser que, conformément à l'hypothèse de la « masse critique », l'accumulation d'un certain stock lexical est nécessaire pour que la grammaticalisation d'une classe de mots s'enclenche.
- 27 Nous voudrions maintenant avancer une hypothèse supplémentaire, qui concerne les aspects qualitatifs de la relation entre développement lexical et grammatical. Cette hypothèse est celle d'un « amorçage lexico-sémantique de la grammaticalisation » (Bassano, 2000b). C'est l'idée que la grammaticalisation pourrait être liée, non seulement à la taille du lexique, mais aussi à la structure et la nature de celui-ci : ainsi, certains types sémantiques de noms pourraient être plus propices que d'autres à former le substrat des processus de grammaticalisation. Cela revient à rechercher si certaines propriétés lexico-sémantiques sont susceptibles de favoriser la grammaticalisation des noms. Et en effet, l'analyse présentée dans la figure 4 montre que, dans le corpus de Pauline, c'est pour les noms animés que l'omission du déterminant est proportionnellement la plus forte (64 %), tandis que c'est pour les inanimés concrets que la proportion d'emploi d'un vrai déterminant est la plus forte (28 %) (les contrastes sont significatifs). Cela suggère que les noms d'objets concrets, s'ils ne sont pas les seuls à être produits précocement, pourraient constituer des candidats privilégiés pour entraîner la contrainte d'utilisation du déterminant et ainsi « amorcer » la grammaticalisation de la classe des noms.

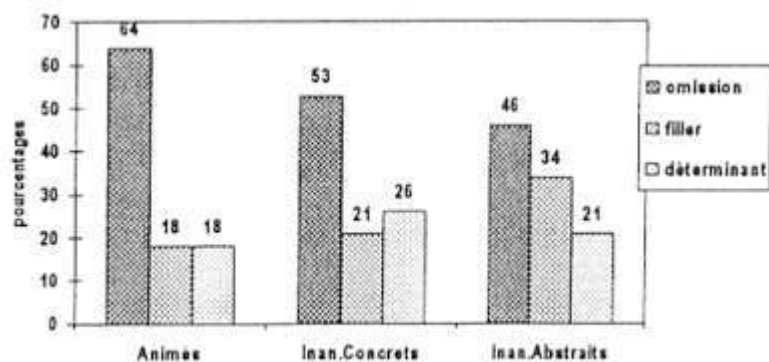


Figure 4 : Procédures d'emploi des noms dans les sous-classes sémantiques de noms (corpus de Pauline)

- 28 Nous espérons que les quelques analyses que nous venons de présenter sur le développement des noms auront illustré l'apport de l'étude de la production naturelle précoce à l'analyse de la constitution du système linguistique par le jeune enfant. Nous espérons aussi qu'elles illustrent notre approche fonctionnaliste intégrative de l'acquisition du langage. Nous avons, en effet, non seulement examiné les transformations qui s'opèrent, au cours des deuxième et troisième années, dans la production quantitative des noms et dans le développement de leurs propriétés lexico-sémantiques, mais aussi analysé le processus d'intégration de la contrainte grammaticale de base caractérisant cette catégorie linguistique, en l'occurrence l'emploi du déterminant antéposé. En analysant le corpus d'un enfant, nous avons apporté des arguments à l'appui de l'hypothèse de « la masse lexicale critique », d'une part, et de l'hypothèse de « l'amorçage lexico-sémantique de la grammaticalisation », d'autre part, et nous avons ainsi montré qu'il existait certaines relations de dépendance entre le développement grammatical et le développement lexical des noms, ce qui est directement au cœur de l'approche intégrative. Elargissant l'investigation à un nombre conséquent d'enfants, de nouvelles analyses, actuellement en cours, sur la production des noms dans la base de données transversale à 20, 30 et 39 mois, confirment l'incidence du facteur lexical sur la grammaticalisation et en précisent la portée au cours du développement. En outre, plusieurs recherches menées sur le français (Bassano, 1998b, 1999, 2000) et sur d'autres langues (Bassano et al., 2004 ; Bates & Goodman, 1999 ; Marchman & Bates, 1994) ont établi l'existence d'interactions similaires pour la production des verbes, ce qui suggère que la relation entre lexique et grammaire concerne des sous-domaines variés du développement langagier, et que les interdépendances entre ces deux dimensions constituent un phénomène robuste et bien étayé.

## BIBLIOGRAPHIE

BASSANO, D. (1998a) : L'élaboration du lexique précoce chez l'enfant français : structure et variabilité, *Enfance*, 4, 123-153.

- BASSANO, D. (1998b) : Sémantique et syntaxe dans l'acquisition des classes de mots : l'exemple des noms et des verbes en français, *Langue française*, 118, 26-48.
- BASSANO, D. (1999) : L'interaction lexicale/grammaire et l'acquisition des verbes, *Parole*, 1999, 9/10, 29-48.
- BASSANO, D. (2000a) : La constitution du lexique : le « développement lexical précoce », In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Vol. 1 : Le langage en émergence*, Paris, Presses universitaires de France, 137-168.
- BASSANO, D. (2000b) : Early development of nouns and verbs in French: Exploring the interface between the lexicon and grammar, *Journal of Child Language*, 27 (3), 521-559.
- BASSANO, D., MAILLOCHON, I. & EME, E. (1998) : Developmental changes and variability in early lexicon: A study of French children's naturalistic productions, *Journal of Child Language*, 25 (3), 493-531.
- BASSANO, D., EME, E. & CHAMPAUD, CC. (2005) : A naturalistic study on early lexical development: General processes and inter-individual variations in French children, *First Language* (sous presse).
- BASSANO, D., LAAHA, S., MAILLOCHON, I. & DRESSLER, W.U. (2004) : Early acquisition of verb grammar and lexical development: evidence from periphrastic constructions in French and Austrian German, *First Language*, 24 (1), 33-70.
- BASSANO, D., MAILLOCHON, I., KLAMPFER, S. & DRESSLER, W.U. (2001a et b) : L'acquisition de la morphologie verbale en français et en allemand autrichien (I) Les fondements théoriques, *Enfance*, 2001/1, 81-99, et (II) L'épreuve des faits, *Enfance*, 2001/2, 117-148.
- BASSANO, D. & VAN GEERT, P. (2005) : Modelling utterance length: Developmental changes, transitions and intra-individual variability, *Developmental Psychology* (en révision).
- BATES, E. & GOODMAN, J.-C. (1999) : On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language*, New Jersey, Lawrence Erlbaum Associates, 29-79.
- BATES, E. & MAC WHINNEY, B. (1982) : Functionalist approaches to grammar. In L. Gleitman & E. Wanner (Eds.), *Language acquisition: The state of the art*, Cambridge MA, Cambridge University Press, 173-218.
- BERTONCINI, J. & BOYSSON-BARDIES, B. (2000) : La perception et la production de la parole avant deux ans. In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Vol. 1 : Le langage en émergence*, Paris, Presses universitaires de France, 95-136.
- BOYSSON-BARDIES, B. (1996) : *Comment la parole vient aux enfants*, Paris, Éditions Odile Jacob.
- BROWN, R. (1973) : *A first language: the early stages*, Cambridge, MA, Harvard University Press.
- CHOI, S. & GOPNIK, A. (1995) : Early acquisition of verbs in Korean: A cross-linguistic study, *Journal of Child Language*, 22 (3), 497-529.
- ELMAN, J.L., BATES, E., JOHNSON, M.H., KARMILOFF-SMITH, A., PARISI D., & PLUNKETT, K. (1996) : *Rethinking innateness: A connectionist perspective on development*, Cambridge, MA, MIT Press.
- FENSON, L., DALE, P.S., REZNICK, J.S., BATES, E., THAL, D. & PETHICK, S.J. (1994) : Variability in early communicative development, *Monographs of the Society for Research in Child Development*, 59 (5, Serial N° 242, 189 p.).
- GENTNER, D. (1982) : Why nouns are learned before verbs: Linguistic relativity versus natural partitioning, In S. Kuczaj (Ed.), *Language development*, vol. 2, Hillsdale, NJ, Erlbaum, 301-334.

- GENTNER, D. & BORODITSKY, L. (2001) : Individuation, relativity and early word learning, In M . Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development*, Cambridge, UK, Cambridge University Press, 215-256.
- GRÉGOIRE, A. (1947) : *L'apprentissage du langage*. Faculté de philosophie et lettres de l'université de Liège.
- LEOPOLD, W. F. (1939-1949) : *speech development of a bilingual child: a linguist's record* (4 vols.), Evanston, II, Northwestern University Press.
- MAC WHINNEY, B. (2000) : *The CHILDES project: tools for analysing talk*, Hillsdale, NJ, Erlbaum.
- MAC WHINNEY, B. (1999) : *The emergence of language*, New Jersey, Lawrence Erlbaum Associates.
- MARCHMAN, V. & BATES, E. (1994) : Continuity in lexical and morphological development: a test of the critical mass hypothesis, *Journal of Child Language*, 21, 339-66.
- MARCOS, H. (1998) : *De la communication pré-linguistique à la communication linguistique : formes et fonctions*, Paris, L'Harmattan.
- TARDIF, T. (1996) : Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, 32 (3), 492-504.
- TOMASELLO, M. (2003) : *Constructing a language*, Cambridge, Massachusetts, Harvard University Press.
- VAN GEERT, P. (1991) : A dynamic systems model of cognitive and language growth, *Psychological Review*, 98, 3-53.
- VAN GEERT, P. (1994) : *Dynamic systems of development : Change between complexity and chaos*, New York, Londres, Harvester Wheatsheaf.
- VENEZIANO, E. (2000) : Interaction, conversation et acquisition du langage dans les trois premières années, In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Vol. 1 : Le langage en émergence*, Paris, Presses universitaires de France, 231-265.

## NOTES

1. Les enregistrements et transcriptions de Benjamin ont été effectués durant les années 1986-90 par D. Bassano, en collaboration avec J. Weissenborn et son équipe technique. Les enregistrements et transcriptions de Pauline ont été effectués durant les années 1991-94 par I. Maillolchon, sous la direction de D. Bassano.
2. Les enregistrements du corpus TRL ont été effectués pour leur plus grande part durant les années 1993-96 par I. Maillolchon, sous la direction de D. Bassano. Ont participé aux transcriptions : A.M. Blanco, N. Chénafi, R. Dahmen, S. Debarre, A. Fustec, C. Garnéro, I. Hévia, I. Maillolchon, K. Rougier, C. Ryckebusch.
3. Les enregistrements et transcriptions du corpus Autisme ont été effectués par M. Lavielle durant les années 2000-2003, sous la direction de D. Bassano.
4. Le programme Procod est réalisé par S. Mottet, informaticien au Laboratoire Cognition et Développement.
5. Nous considérons comme « noms » les mots considérés comme tels dans la production adulte.

---

AUTEUR

DOMINIQUE BASSANO

\* Laboratoire Cognition et Développement, CNRS-Université Paris 5.