

Développement et variabilité dans l'élaboration du lexique en français

DOMINIQUE BASSANO

ELSA EME

ISABELLE MAILLOCHON

L'élaboration du lexique, c'est-à-dire la constitution du stock des mots de la langue par l'enfant, est un processus plus complexe qu'il n'y paraît, impliquant non seulement l'apprentissage du sens des mots mais aussi l'apprentissage de leur catégorie grammaticale. C'est cette deuxième question qui est abordée dans la recherche présentée ici, qui examine quels types de mots – noms, verbes, mots grammaticaux, etc. – sont produits par les jeunes enfants français aux moments d'émergence du langage et comment la composition de leur lexique évolue au cours des premières étapes de l'acquisition.

Les études sur le développement lexical précoce, assez nombreuses mais pour l'essentiel encore concentrées sur l'acquisition de l'anglais, ont permis de mettre en évidence certaines tendances générales dans le processus de construction du vocabulaire. Ainsi, la récente étude transversale de E. Bates et de ses collègues (1994), qui utilise la méthode des rapports parentaux pour analyser le vocabulaire d'un très large échantillon d'enfants anglophones américains âgés de 8 à 30 mois, a montré, après d'autres, que les noms étaient initialement les éléments de loin les plus fréquents mais que leur part relative diminuait quand augmentait la taille du lexique. En revanche, la proportion de prédicats, initialement faible, s'accroissait régulièrement, et la proportion de mots de la classe fermée, presque inexistante tant que le vocabulaire n'avait pas atteint 400 mots, s'accroissait brusquement après ce seuil. Les auteurs en concluent qu'il existe trois vagues successives dans le processus de composition du vocabulaire entre 1 an et 2 ans et demi, l'accent se déplaçant de la référence vers la prédication, puis vers la grammaire.

Nous présentons ici une étude transversale où nous analysons la composition du lexique de deux groupes d'enfants français âgés de 20 mois et 30 mois respectivement, en nous appuyant sur un examen systématique de leurs productions naturelles. L'objectif de cette étude est double. Elle vise d'abord à dégager les changements les plus cruciaux s'opérant dans le développement du lexique entre ces deux âges clés. Nous espérons conforter ainsi les résultats obtenus dans une étude longitudinale menée sur les productions spontanées d'un enfant français entre 14 et 30 mois (Bassano, 1996; Bassano, Maillachon & Eme, à paraître; Bassano, sous presse). Cette étude longitudinale, en accord avec les principales conclusions de Bates *et al.*, montrait que la proportion de noms, très importante jusqu'à 20 mois au moins, diminuait progressivement avec l'âge au profit des prédicats et des mots grammaticaux. Elle montrait cependant aussi que, outre les noms, les éléments que nous avons appelés «para-lexicaux», comme les interjections et les particules élémentaires d'interaction, jouaient un rôle prépondérant dans l'émergence du langage avant de diminuer ensuite. Elle indiquait aussi que les mots grammaticaux présentaient un profil d'évolution beaucoup plus expansif chez notre enfant français que chez les enfants anglophones, allant jusqu'à former 40% des mots-types du lexique à 30 mois, contre environ 15% chez les anglophones.

Le deuxième objectif de l'étude transversale est d'évaluer l'ampleur de la variabilité inter-individuelle qui affecte la composition du lexique dans chacun des deux groupes d'âge. C'est en effet un des résultats anciens et récurrents de la littérature sur le lexique précoce que tous les enfants n'entrent pas dans le langage avec les mêmes mots. Cette variabilité dite «stylistique», établie par Nelson il y a plus de 20 ans (Nelson, 1973) et retrouvée par la suite (notamment par Bates *et al.*, 1994), qui oppose les enfants dits «référentiels» aux enfants dits «expressifs», est particulièrement manifeste aux alentours de l'âge de 20 mois et s'atténue ensuite au cours de la 3ème année. Pour être en conformité avec ces résultats classiques, nous devrions donc trouver une variabilité plus importante à 20 mois qu'à 30 mois.

MÉTHODE

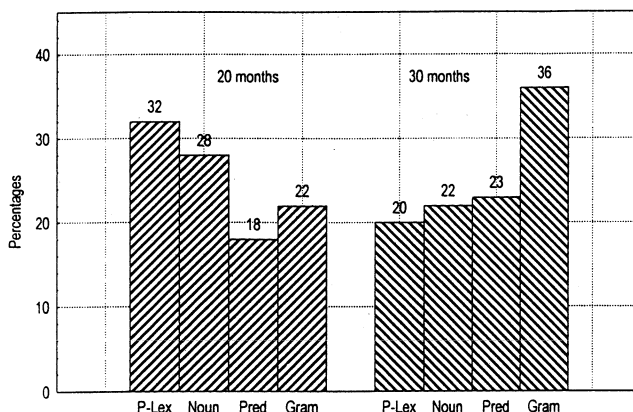
Notre étude transversale, comme l'étude longitudinale, repose sur l'analyse de données de corpus naturelles. Les productions spontanées de 12 enfants de 20 mois et 12 enfants de 30 mois (6 garçons et 6 filles dans chaque groupe) ont été recueillies au cours de séances d'enregistrement en vidéo. Chaque enfant a été enregistré individuellement une fois, dans une pièce aménagée à cet effet, au cours d'une séance semi-structurée de 30 à 45 minutes durant laquelle l'enfant est en interaction avec sa mère et deux animatrices. La session a été ensuite intégralement transcrite conformément aux standards du CHILDES (MacWhinney, 1995). Dans la session de chaque enfant, un échantillon de 120 énoncés a été extrait, sur lequel est menée l'analyse du lexique.

RÉSULTATS

1. Tendances moyennes dans la composition du lexique à 20 et 30 mois

Les différentes classes d'éléments lexicaux identifiés dans les productions des enfants ont été regroupées sous quatre catégories principales: les **noms**; les **prédicats**, qui incluent les verbes et les adjectifs; les **mots grammaticaux**, qui incluent les adverbes, les déterminants, les pro-

Figure 1. Mean proportions of word-types for each word-category in the 20 – and 30 – month – hold groups



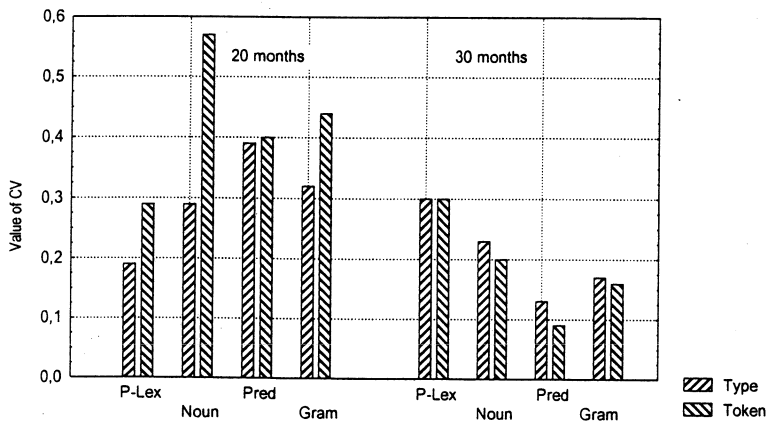
noms, les prépositions, les conjonctions, les auxiliaires; et enfin la catégorie additionnelle des éléments **para-lexicaux**, qui inclue interjections, *fillers*, formules, particules simples d'affirmation et de négation («oui» et «non»).

La figure 1 présente, pour les mots-types (*types*), la distribution en pourcentage des quatre catégories du lexique dans chacun des deux groupes d'âge. Cette analyse indique que le lexique des enfants de 20 mois est composé d'une majorité d'items para-lexicaux (32%) et de noms (28%), et d'une moindre proportion de mots-grammaticaux (22%) et de prédicats (18%). Chez les enfants de 30 mois, les items para-lexicaux ne représentent plus que 20% du lexique, les noms 22%, tandis que les prédicats forment maintenant 23% du lexique et les mots grammaticaux 36%. Il se dégage de cette analyse que, entre 20 et 30 mois, les proportions d'items para-lexicaux et de noms diminuent, tandis que, à l'inverse, les proportions de prédicats et surtout de mots grammaticaux s'accroissent. Notons que la même tendance développementale se retrouve, encore plus marquée, dans l'analyse des mots-occurrences (*tokens*). On trouve donc bien, dans ces changements, la confirmation des évolutions observées dans l'étude longitudinale.

2. VARIABILITÉ INTER-INDIVIDUELLE DANS LA COMPOSITION DU LEXIQUE À 20 ET 30 MOIS

Existe-t-il, entre enfants de même âge, des différences notables qui viendraient moduler les tendances moyennes mises en évidence précédemment? Notre analyse de la variabilité s'appuie sur l'examen des paramètres de dispersion qui caractérisent chaque distribution, et en particulier du **coefficient de variabilité**, calculé en divisant l'écart-type par la moyenne. Le coefficient de variabilité est donc un indice qui relativise l'écart-type et permet de comparer l'ampleur de la variation à travers les différentes distributions. Plus il est élevé, plus la variabilité est importante. C'est la valeur du coefficient de variabilité associé à chaque distribution qui est reporté dans le graphique de la figure 2. Cette représentation fait apparaître que la variabilité est, dans l'ensemble, passablement importante à 20 mois et bien moindre à 30 mois. En effet, le coefficient de variabilité est plus élevé à 20 mois qu'à 30 mois pour la proportion de

Figure 2. Coefficient of variability for distributions of each word – category in 20 – and 30 – month – old groups

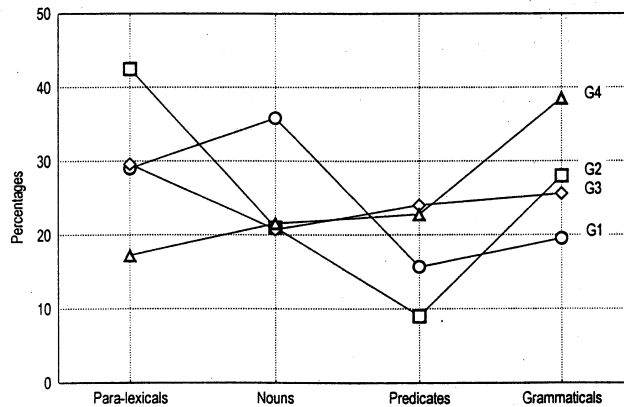


noms, de prédicats et de mots grammaticaux, et cela aussi bien pour les mots-types que pour les occurrences. La seule exception à cette évolution concerne la production des items para-lexicaux, pour lesquels la variabilité est plus importante à 30 mois qu'à 20 mois. Cette analyse confirme donc l'existence d'une assez grande variabilité dans la composition du lexique à 20 mois, variabilité qui se trouve très atténuée à 30 mois, sans cependant disparaître totalement.

3. RECHERCHE DES «PROFILS LEXICAUX» DES ENFANTS À 20 ET 30 MOIS

Une analyse en *clusters* (suivant la méthode de Ward) effectuée sur les quatre catégories de mots-types montre que les 24 sujets se répartissent en quatre groupes (74% de variance expliquée) correspondant à quatre différents profils d'organisation lexicale, représentés dans la figure 3. Le lexique du groupe G1 -qui comporte 6 sujets, dont 5 de 20 mois et un seul de 30 mois – se caractérise par une forte proportion de noms, qu'on ne retrouve dans aucun des autres groupes, contrebalancée par de faibles proportions de prédicats et de mots grammaticaux: aussi l'avons-nous appelé le «**profil référentiel**». Le groupe G2 – qui comporte seulement 2 sujets, tous deux âgés de 20 mois- compense sa faible proportion de noms par une très forte proportion d'items para-lexicaux, qui lui vaut l'appellation de «**profil para-lexical**», bien qu'il ait une proportion non négligeable de mots grammaticaux. Le groupe G3 – qui comporte

Figure 3. Frequencies of word – types categories for each cluster



7 sujets, dont 5 âgés de 20 mois et 2 de 30 mois – présente, quant à lui, une organisation lexicale assez diversifiée: il compense sa faible proportion de noms par des items para-lexicaux, des mots grammaticaux, mais aussi par une étonnante forte proportion de prédicats. Aussi est-il appelé le «**profil diversifié à tendance prédictive**». Enfin, le groupe G4 -constitué de 9 sujets, tous âgés de 30 mois- se caractérise par une organisation dans laquelle la proportion de mots grammaticaux domine largement, ce qui lui vaut l'appellation de «**profil grammatical**». Ainsi, les enfants de 20 mois se répartissent sous trois profils, alors que la grande majorité des enfants de 30 mois présente un même profil, typique de leur groupe d'âge.

En conclusion, nous espérons avoir établi par cette étude que des changements développementaux notables se produisent dans la composition du lexique des enfants français entre 20 et 30 mois: tandis qu'à 20 mois, ce sont encore les items para-lexicaux, tels qu'interjections ou

fillers, et les noms qui prédominent dans le lexique, à 30 mois ce sont les mots grammaticaux les plus fréquents, en mots-types et en occurrences. Ce remarquable essor des mots grammaticaux est apparemment une caractéristique de l'acquisition du français, puisqu'il ne semble pas se trouver dans celle de l'anglais. L'analyse de la variabilité intra-âge nous a permis de montrer qu'il existait de fortes différences individuelles dans la composition du lexique des enfants de 20 mois, qui présentent d'assez grandes variations dans les proportions de noms, de prédicats et de mots grammaticaux. Ces différences sont très amoindries chez les enfants de 30 mois, ce qui conforte l'idée répandue dans la littérature anglo-saxonne que, dans le domaine du lexique, la variabilité stylistique diminue dans le courant de la 3^{ème} année. L'analyse des profils lexicaux des enfants français confirme que la variabilité stylistique à 20 mois est un phénomène bien réel. Mais elle suggère en outre que les voies empruntées par les jeunes enfants français pour constituer leur lexique sont plus diverses ou complexes que ne le laisse penser l'opposition entre «style référentiel» et «style expressif», puisqu'il faut évoquer pas moins de trois profils lexicaux distincts pour rendre compte de la diversité d'organisation de leur lexique. A 30 mois, c'est un même profil «grammatical» qui s'impose chez les trois quarts des enfants, reflétant la forte grammaticalisation des productions déjà opérée à cet âge en français. Ceci invite à interpréter la diminution de la variabilité avec l'âge comme le reflet de la pression des contraintes linguistiques que le jeune enfant est conduit à intégrer pour acquérir le langage et le système de sa langue.

RÉFÉRENCES

- Bassano, D. 1996. Early lexical development in French: Formal and functional aspects of word classes acquisition. *Abstracts of the VIIth International Congress for the Study of Child Language*, 146-47
- Bassano, D. in press. Sémantique et syntaxe dans l'acquisition des classes de mots: l'exemple des noms et des verbes en français. *Langue Française*.
- Bassano, D.; Maillachon, I.; Eme, E. to appear. Developmental changes and variability in early lexicon: A study of French children's naturalistic productions. *Journal of Child Language*.
- Bates, E.; Marchman, V.; Thal, D.; Fenson, L.; Dale, P.; Reznick, J. S.; Reilly, J.; Hartung, J. 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, 21, 1, 85-124.
- MacWhinney, B. 1995. *The CHILDES project: tools for analysing talk*. Hillsdale, Erlbaum.
- Nelson, K. 1973. Structure and strategy in learning to talk. *Monographs of the society for Research in Child Development*, 38 (1-2), serial n° 149.