



HAL
open science

Emergence et développement du langage : enjeux et apports des nouvelles approches fonctionnalistes

Dominique Bassano

► **To cite this version:**

Dominique Bassano. Emergence et développement du langage : enjeux et apports des nouvelles approches fonctionnalistes. E. Demont & MN. Metz-Lutz. L'acquisition du langage et ses troubles, SOLAL Editeurs, pp.13-46, 2007, Collection Psychologie. hal-00601129

HAL Id: hal-00601129

<https://hal.science/hal-00601129>

Submitted on 30 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In E. Demont & M.N. Metz-Lutz (Eds.), *L'acquisition du langage et ses troubles*
(pp. 13-46), Solal, Marseille - 2007

**EMERGENCE ET DEVELOPPEMENT DU LANGAGE :
ENJEUX ET APPORTS DES NOUVELLES APPROCHES FONCTIONNALISTES**

Dominique Bassano

(Laboratoire *Cognition et Développement*, CNRS – Université Paris 5)

Laboratoire *Structures Formelles du Langage*, CNRS – Université Paris 8

Résumé

L'enfant acquiert normalement l'essentiel du système de sa langue maternelle entre un et quatre ans environ. S'appuyant sur les données fournies par la psychologie et la neuropsychologie développementales, la modélisation et la linguistique fonctionnelle, les approches fonctionnalistes développées en psycholinguistique au cours des deux dernières décennies contribuent à apporter des réponses nouvelles et documentées à la question de la constitution du système linguistique chez l'enfant. Dans ce chapitre, nous présentons d'abord les principaux fondements et enjeux de ces approches, illustrées notamment par les travaux de Bates et MacWhinney et, plus récemment, par les travaux menés dans le cadre des théories émergentistes et des théories « usage-based ». Cette présentation est ensuite illustrée avec l'examen, plus détaillé et empirique, de trois thèmes, explorés dans le cadre de notre programme sur l'émergence du système linguistique chez l'enfant français. Le premier thème concerne la structuration du lexique : nous examinons comment le lexique précoce se constitue autour des classes de mots, en prêtant une attention particulière à l'analyse de la variabilité inter-individuelle et de sa signification pour le développement. Le deuxième thème concerne les processus de grammaticalisation : en prenant l'exemple de la catégorie linguistique des noms, nous examinons comment est acquise la contrainte grammaticale qui caractérise cette classe en français et recherchons les lieux d'interaction entre développement grammatical et développement lexical. Pour finir, nous tentons une incursion dans l'examen de la modélisation de l'allongement des énoncés, analyse menée dans le cadre de la théorie des systèmes dynamiques et centrée sur la question du rôle de la variabilité intra-individuelle dans le développement.

Le langage est un système de communication complexe et élégant, spécifique à l'espèce humaine. Pour échanger avec son entourage, le petit d'homme doit apprendre la langue parlée autour de lui, et, compte tenu du haut niveau de sophistication des systèmes linguistiques, il y parvient en un laps de temps qu'on peut juger relativement court. L'enfant commence en effet, normalement, à produire ses premiers mots conventionnels vers le début de sa seconde année, et, trois ans plus tard environ, il aura mis en place un système de communication efficace intégrant la plupart des contraintes de base de la langue. Aux alentours de quatre ans, il maîtrise les principaux aspects de la phonologie, connaît à peu près le sens et les conditions d'emploi de plusieurs milliers de mots, et 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...

Au cours du dernier demi-siècle, diverses théories ont tenté de répondre à cette question, parfois appelée le « problème logique » de l'acquisition. L'un des argumentaires les plus connus est celui de l'approche générativiste initiée par Chomsky et développée par ses continuateurs (pour une revue des courants générativistes sur l'acquisition du langage, voir Clahsen, 1996). Les tenants de cette approche considèrent que, pour rendre compte de pareille performance, il faut faire l'hypothèse que l'être humain est doté d'une grammaire universelle innée, constituée de principes abstraits et de paramètres guidant le processus d'acquisition à la façon d'un bioprogramme (Pinker, 1994). Le principal argument évoqué à l'appui de cette conception est l'argument dit de « la pauvreté du stimulus » : l'input mis à la disposition de l'enfant serait trop inconsistant et incomplet pour fournir à celui-ci les éléments nécessaires à l'acquisition de la grammaire. L'insuffisance naturelle de l'input serait en outre renforcée par la rareté des corrections ou réactions négatives des parents (« *negative evidence* ») sur les erreurs de l'enfant, réactions auxquelles l'enfant semblerait d'ailleurs imperméable. Ces considérations conduisent à ce qu'on peut appeler un innéisme « représentationnel », postulant que des informations linguistiques abstraites – essentiellement grammaticales – sont codées au niveau des gènes et disponibles presque indépendamment de l'expérience.

Cependant, les progrès réalisés durant les deux dernières décennies dans les sciences cognitives – qu'il s'agisse de la psychologie et de la psycholinguistique développementales, des neurosciences, de la linguistique ou de la modélisation - peuvent largement inciter les chercheurs à ré-évaluer la situation et à chercher des solutions au problème de l'acquisition ailleurs que dans l'hypothèse de l'innéisme représentationnel. Telle est la perspective adoptée par les nouvelles approches fonctionnalistes de l'acquisition du langage, illustrées d'abord,

pour ne citer que les noms les plus médiatisés, par les travaux de E. Bates et B. MacWhinney avec leur Modèle de compétition (1979, 1982, 1987, 1989) ou par ceux de A. Karmiloff-Smith (1979, 1992), et, plus récemment, par les travaux menées dans le cadre des théories émergentistes (Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996 ; MacWhinney, 1999 ; 2000b ; 2004) et des théories dites « usage-based » (Tomasello, 1998, 2000, 2003 ; Tomasello & Brooks, 1999). Bien que présentant des points de divergence, toutes ces approches fonctionnalistes et constructivistes ont en commun que, comme le souligne N. Akhtar (2004), elles « croient » au développement et se proposent de l'expliquer. Sans pour autant nier l'existence des prédispositions ou contraintes biologiques ni les particularités linguistiques de l'acquisition du langage, elles soulignent le rôle central des déterminants environnementaux et sociaux et des déterminants cognitifs généraux dans le développement langagier.

1. Éléments d'une nouvelle donne

Les approches fonctionnalistes et émergentistes s'appuient sur les avancées de divers secteurs de recherche, allant de la psychologie et neuropsychologie développementales à la modélisation et à la linguistique théorique.

Développement du cerveau et plasticité. L'acquisition du langage suppose un équipement et un développement neuro-biologique approprié. Ainsi que le rappelle M. Kail (2004), des progrès considérables concernant le développement du cerveau ont été réalisés dans les dix dernières années grâce à l'essor des études d'imagerie et des études lésionnelles, bien que les connaissances dans ce domaine soient encore parcellaires. On trouvera une revue des travaux actuels dans la synthèse de Bates, Thal, Finley & Clancy (2003), en particulier ceux de Bourgeois (1997), et dans la synthèse de M. Kail (2004), dont nous ne rapportons ici que quelques éléments. Selon ces travaux, la séquence développementale de la synaptogenèse, qui semble être un événement crucial pour le développement précoce du langage, peut être décrite en cinq phases. Durant les deux premières phases, qui sont prénatales, les synapses sont présentes au niveau de la plaque neurale qui comprend les premiers neurones corticaux, puis générées au niveau de la plaque neurale elle-même. Dans la troisième phase, de la naissance à un an, la synaptogenèse est synchronisée par une explosion, tandis que la quatrième phase, jusqu'à douze ans, est une phase de stabilisation à haut niveau. Enfin, la cinquième phase, qui va de la puberté à la maturité, correspond à un déclin des synapses en

densité et en nombre absolu. Ainsi, « le développement du cerveau est marqué par une surproduction – un processus relativement soudain – suivi d’une élimination sélective – un processus plus lent et qui s’étend sur de plus longues périodes de temps » (Kail, 2004 : 54). Par ailleurs, les travaux récents de la neuropsychologie du développement indiquent que l’organisation cérébrale présente une plasticité qui ne se manifeste pas seulement dans les récupérations des aphasies mais aussi dans l’évolution des mécanismes neuronaux sous-tendant le développement normal. Selon cette conception, la plasticité constitue ainsi une propriété fondamentale des systèmes neuronaux et cognitifs.

Capacités communicatives et cognitives précoces. Il est aussi apparu, à la lumière des données fournies par la psychologie et la neuropsychologie développementales, que l’enfant avait à sa disposition, à la fin de sa première année de vie, un ensemble de mécanismes d’apprentissage beaucoup plus puissants qu’on ne le pensait à l’époque du behaviorisme et qui contribuaient à l’émergence du langage. Diverses capacités communicatives et cognitives se mettent en place de manière convergente pour former, vers 8-10 mois, un ensemble de conditions préalables et nécessaires à l’émergence du système linguistique proprement dit : par exemple la motivation à communiquer, les capacités de catégorisation, d’imitation, d’attention, de mémoire à court terme, et, plus particulièrement, les capacités de perception, de segmentation et de production des sons de parole. Ainsi, diverses recherches récentes ont montré que, dès l’âge de 8 mois, les jeunes enfants manifestent des aptitudes à détecter et extraire des régularités dans les stimuli auditifs – des séquences de sons de parole - qui leur sont présentés (Saffran, Aslin & Newport, 1996 ; Aslin, Saffran & Newport, 1999 ; Gomez & Gerken, 1999 ; voir aussi la synthèse de J. Bertoncini et B. de Boysson-Bardies, 2000, et celle de L. Naigles, 2002). Cette capacité perceptive et cognitive précoce à « l’apprentissage statistique », d’ailleurs observée aussi s’agissant de stimuli visuels, est certainement l’une des conditions nécessaires pour que le développement du langage, et en particulier le développement grammatical, puisse s’effectuer. Dans son livre « *Constructing a language* » (2003), M. Tomasello propose un regroupement des capacités précoces pré-requises en deux principaux ensembles, l’un recouvrant les capacités liées à la théorie de l’esprit, au sens large, (*intention reading*), et l’autre recouvrant les capacités liées à la catégorisation, conceptuelle ou perceptive (*pattern finding*). C’est le premier ensemble qui est à ses yeux le plus déterminant pour l’émergence du langage, dans la mesure où il inclut des aspects tels que l’attention conjointe, la compréhension des intentions de communication, l’imitation avec inversion des rôles, toutes capacités qui sont spécifiquement humaines, tandis

que certaines capacités de catégorisation perceptive semblent exister aussi chez des primates non humains tels que les singes tamarins (Ramus, Hauser, Miller, Morris & Mehler, 2000). M. Tomasello souligne ainsi combien le développement du langage se révèle intégré à celui des autres habiletés cognitives et socio-cognitives qui le rendent possible et l'accompagnent.

Modèles connexionnistes et systèmes dynamiques. Un autre secteur de recherches, en plein essor, qui aide le fonctionnalisme à penser les processus d'acquisition est celui de la modélisation. Deux modèles en particulier, le connexionnisme (Plunkett, 1995 ; Elman *et al.*, 1996) et la théorie des systèmes dynamiques (Thelen & Smith, 1994 ; 1998 ; Van Geert, 1991 ; 1994 ; 2003) mettent en œuvre des outils mathématiques et statistiques explicitement destinés à rendre compte du développement. Si les modèles connexionnistes ont été élaborés par des théoriciens travaillant sur les représentations mentales, les systèmes dynamiques avaient au départ comme terrain de prédilection le développement sensori-moteur, et c'est P. van Geert, notamment, qui a montré l'intérêt de cette approche s'agissant du développement cognitif et langagier. Comme l'indique l'article « autobiographique » co-écrit par E. Thelen et E. Bates (2003), ces deux types de modèles sont largement apparentés et complémentaires. Tous deux constructivistes et émergentistes, ils essaient d'expliquer pourquoi et comment des structures nouvelles émergent et changent, sous l'effet des interactions avec l'input et de l'auto-organisation des systèmes. Les modèles connexionnistes s'intéressent plutôt aux processus d'apprentissage, tandis que les systèmes dynamiques s'intéressent plus précisément à la trajectoire, la nature, la forme et les causes du changement.

Linguistiques fonctionnelles et cognitives. Dans les dernières décennies se sont développées des théories du langage souvent appelées « linguistiques cognitives – fonctionnelles », ou encore, « *usage based linguistics* », qui constituent des alternatives plausibles et rigoureuses au modèle générativiste formel (voir la présentation de ces courants linguistiques dans l'ouvrage édité par M. Tomasello, 1998, ainsi que dans celui édité par C. Fuchs, 2004). Ces théories, dont s'inspirent directement les approches fonctionnalistes en psycholinguistique, ont pour postulat central que la structure du langage est un produit de son usage. Elles proposent de caractériser la compétence linguistique adulte par un ensemble diversifié d'items, structures et représentations articulant syntaxe, sémantique et pragmatique, qui apparaît plus accessible à l'acquisition par l'enfant que ne l'est la grammaire formelle (voir par exemple, la « Grammaire de Construction » de A. E. Goldberg, 1999, ou la « Grammaire de Rôle et Référence » de R. D. Van Valin, 1998).

2. Enjeux des approches fonctionnalistes et émergentistes

L'hypothèse émergentiste : repenser l'innéisme et le développement. Comme le rappelle M. Kail (2000), l'étude du développement du langage est un lieu marqué par les débats sur des questions fondamentales telles que la nature des capacités initiales à la source du processus d'acquisition, la nature des mécanismes d'acquisition (modulaires ou interactifs, généraux ou spécifiques à chacun des domaines particuliers de la cognition), et la nature de l'input qui permet au système d'acquisition de progresser. Les approches fonctionnalistes, émergentistes en particulier, invitent à repenser ces questions en ouvrant une voie nouvelle qui remplace l'opposition traditionnelle entre innéisme et empirisme. Cet objectif a été précisé, d'abord dans l'ouvrage « *Rethinking innateness: A connectionist perspective on development* » écrit collectivement par J.L. Elman et ses collègues (1996), puis dans l'ouvrage édité par B. MacWhinney « *The emergence of language* » (1999). J.L. Elman et ses collègues (Elman *et al.*, 1996; Elman, 1999) proposent de distinguer, au niveau des réseaux de neurones cérébraux aussi bien que des réseaux de neurones artificiels des simulations connexionnistes, trois types de contraintes susceptibles de déterminer les comportements : les contraintes représentationnelles, architecturales et temporelles. En raison même de la plasticité du cortex cérébral, l'innéisme représentationnel, postulé par l'approche générativiste et affirmant l'existence de représentations spécifiques d'un domaine localisées quelque part dans le cerveau, est biologiquement peu plausible. L'innéité se situe plus vraisemblablement au niveau des contraintes architecturales et chronotopiques. Les premières concernent la structure du réseau neuronal, déterminant par exemple le type d'information qui peut être reçue et le type de représentation qui peut être stockée. Les secondes concernent la mise en œuvre temporelle de la maturation neuronale : elles semblent notamment jouer un rôle important dans la spécialisation de certaines régions corticales – selon qu'elles sont prêtes au bon moment ou pas - pour certaines fonctions cognitives.

B. MacWhinney (1999 ; 2000b) note que la nature abonde en exemples d'émergence : ainsi, la forme des rivages naît des interactions entre la géologie et les courants océaniques. La conception émergentiste du langage étend cette vision aux processus de formation, traitement et acquisition des structures linguistiques. Son enjeu est de rendre compte des mécanismes qui régissent les interactions entre processus biologiques et environnement dans les différents domaines du langage, tels que perception, phonologie, sémantique ou syntaxe. Dans sa version forte, l'hypothèse émergentiste propose que l'acquisition du langage n'est pas

le produit de règles innées spécifiques, mais résulte des interactions entre les capacités cognitives générales de l'enfant et l'environnement linguistique. Plus pratiquement, la tâche commune aux psycholinguistes qui adoptent cette approche est d'expliquer d'où provient un comportement langagier émergent donné, avec l'idée que, souvent, ce comportement émerge des capacités plus générales développées dans un autre domaine. Par exemple, Bates et Goodman (1999) proposent une analyse de l'émergence des premières capacités grammaticales de l'enfant à partir de ses apprentissages lexicaux.

Ainsi, dans la perspective émergentiste, la solution au problème de l'acquisition du langage est "multiple", selon le terme de B. MacWhinney (2004), qui recense un ensemble de mécanismes opérant de façon complémentaire ou conjointe - tels que conservatisme, apprentissage sur la base d'items (*item-based learning*), mécanismes de compétition, auto-correction, etc. – grâce auxquels l'enfant parvient à produire des énoncés corrects et à corriger ses erreurs. Plus généralement, il nous semble que les tenants des approches fonctionnalistes et émergentistes partagent, à des degrés divers, un certain nombre d'idées directrices que nous allons maintenant indiquer.

L'acquisition du langage vue comme un processus graduel et non linéaire. L'une des idées de base est que l'acquisition du langage est un processus graduel. Du point de vue le plus général, cette notion fait référence à l'idée que les catégories linguistiques ne sont pas données de manière innée à l'enfant, mais progressivement construites par celui-ci à partir de ses expériences langagières. Cette conception a cependant été développée plus spécifiquement par M. Tomasello, qui soutient, avec son « hypothèse des îles verbales » (*verb island hypothesis*) que la syntaxe du jeune enfant est d'abord organisée autour d'items lexicaux particuliers de façon assez canalisée. Ainsi, jusque vers trois ans, l'enfant emploie certains verbes dans certaines constructions (transitives par exemple), et d'autres verbes dans certaines autres constructions (intransitives), mais il transfère rarement les verbes d'une construction à l'autre. Ce n'est que très progressivement et à partir de la base concrète formée par les items particuliers que s'opèrent les processus de généralisation et d'abstraction marquant l'accès à un niveau élaboré de la grammaire (Tomasello, 1992, 2000 ; Akhtar & Tomasello, 1997 ; cf. Lieven, Pine & Baldwin, 1997). La notion d'acquisition graduelle ainsi conçue rejoint celle « d'apprentissage sur la base d'items » identifiée par MacWhinney (2004) comme l'un des processus centraux permettant la production correcte du langage. Cette vision de l'acquisition diffère de celle des recherches qui, en utilisant des paradigmes d'écoute préférentielle pour tester la compréhension grammaticale des très jeunes enfants, suggèrent que ceux-ci

manifestent une sensibilité précoce à certaines des contraintes de leur langue, telles que l'ordre des mots ou les dépendances morphosyntaxiques dans la phrase (Hirsh-Pasek & Golinkoff, 1996 ; Naigles, 2002). Les divergences dans l'analyse des performances grammaticales des enfants soulèvent diverses questions, relatives par exemple à la nature du décalage entre compréhension et production, ou à la nature des différences dans les processus d'abstraction et de généralisation impliqués par les structures, qui sont généralement des formes dénuées de contenu pour les très jeunes enfants et des formes associées à une signification pour les enfants un peu plus âgés (Naigles, 2002 ; Tomasello & Akhtar, 2003).

Une autre caractéristique du développement langagier est qu'il est souvent non linéaire, conformément aux attentes de la théorie des systèmes dynamiques. Sous l'effet des changements graduels, des transformations drastiques peuvent se produire sous certaines conditions et à certains moments critiques du développement. Ainsi, des explosions quantitatives, susceptibles de renvoyer à des transformations qualitatives sous-jacentes, ont été mises en évidence dans le domaine du développement lexical, telle la fameuse explosion de la dénomination se produisant vers le milieu de la seconde année, et plus encore dans le domaine du développement grammatical (cf. Bates & Goodman, 1999 ; Clark, 1993 ; cf. aussi Bassano, 2000a ; Bassano, Laaha, Maillochon & Dressler, 2004, ainsi que les analyses proposées ci-après). La question du statut de ces explosions et de leur rôle dans le processus de développement est devenue un thème central pour l'étude de l'acquisition du langage.

Le rôle de l'input et la variabilité inter-linguistique. Une autre idée directrice est que l'input, qu'il s'agisse du langage directement adressé à l'enfant ou du langage parlé autour de lui, joue un rôle déterminant dans le processus d'acquisition, et pas seulement un simple rôle de déclencheur. On sait depuis longtemps que le développement du langage est médiatisé par les pratiques de la communauté sociale et par les interactions de l'enfant avec son entourage immédiat, qui permettent par exemple la mise en œuvre de stratégies telles que les reprises imitatives (Veneziano, 2000). Les études sur la nature du langage adressé à l'enfant ont montré que les caractéristiques prosodiques et structurelles particulières de l'input maternel (élévation de la hauteur de la voix, intonation exagérée, vocabulaire simplifié) pouvaient faciliter l'apprentissage au tout début de l'acquisition du langage. Plus récemment, il a été souligné que la syntaxe de l'input reçu par l'enfant n'était pas si dégénérée qu'on le disait dans l'approche générativiste (MacWhinney, 2004).

Avec l'essor des études interlangues (Bates & MacWhinney, 1987 ; Slobin, 1985 – 1997 ; Kail & Bassano, 2000), la question s'est en outre posée de déterminer dans quelle

mesure les propriétés structurelles et statistiques de la langue en acquisition pouvaient jouer un rôle dans le processus de développement. A titre d'exemple, plusieurs études ont souligné que l'émergence des verbes et des formes verbales était influencée par certaines caractéristiques statistiques de l'input, telles que la fréquence des items, leur position plus ou moins saillante dans la phrase, ou leur plus ou moins grande transparence morphologique (Gentner, 1982 ; Gentner & Boroditsky, 2001 ; Choi & Gopnik, 1995 ; Tardif, Shatz & Naigles, 1997 ; Wijnen, Kempen & Gillis, 2001).

Le rôle de la variabilité inter- et intra-individuelle. Une troisième idée directrice des approches fonctionnalistes est que la variabilité inter-individuelle, voire intra-individuelle, est un phénomène signifiant et constitutif du développement. Tandis que la variabilité inter-linguistique reflète la diversité des langues naturelles, la variabilité inter-individuelle reflète la diversité des apprenants, et la variabilité intra-individuelle la diversité des stratégies utilisées par un même sujet au même moment du développement. La variabilité inter-individuelle, qui a été assez largement étudiée dans l'acquisition du langage, se manifeste par des différences dans les rythmes et les styles des enfants qui apprennent une même langue (Kail & Bassano, 2000). On parle de variabilité de rythme pour désigner le fait qu'une compétence langagière donnée n'émerge pas au même âge et ne se développe pas à la même vitesse chez tous les enfants dont on estime le développement normal. Ces variations ont été mises en évidence notamment par l'étude anglo-américaine de Fenson et collègues (1994), qui reste la référence dans ce domaine. En utilisant les informations obtenues grâce aux questionnaires du *MacArthur Communicative Development Inventories* (CDI), ces chercheurs ont étudié l'évolution des compétences communicatives et langagières de plus de 1800 enfants anglophones américains entre 8 et 30 mois et ont montré qu'il existait de substantielles différences individuelles dans l'apparition et le développement des capacités mesurant la compétence langagière des jeunes enfants : compréhension de mots, production de mots, combinaison de mots et expression de la complexité syntaxique. L'ampleur des variations inter-individuelles observées durant cette période défie ainsi l'idée d'un bio-programme universel qui réglerait le développement langagier précoce.

La variabilité dite « de style » fait référence aux différences plus qualitatives concernant le mode d'entrée des enfants dans le langage et la nature des séquences de développement. La variabilité stylistique est l'objet d'un intérêt assez ancien : c'est K. Nelson (1973, 1981) qui a attiré l'attention sur ce phénomène en formulant l'opposition entre « style référentiel » et « style expressif ». Depuis, l'analyse et la signification de la variabilité

stylistique ont souvent été repensées. Les travaux de E. Bates et de ses collègues, notamment, ont proposé la distinction plus générale entre « style analytique » et « style holistique », qui met l'accent sur les différences dans les stratégies et mécanismes de traitement du langage. Dans cette perspective, les variations individuelles de style – de même d'ailleurs que les variations intra-individuelles - sont considérées comme une fenêtre ouverte sur les corrélats, les facteurs endogènes et les lieux critiques des changements développementaux (Bates, Bretherton & Snyder, 1988 ; Bates, Dale & Thal, 1995 ; Bates, Marchman, Thal, Fenson, Dale, Reznick, Reilly & Hartung, 1995 ; voir aussi Bassano, 1998a ; Bassano, Maillochon & Eme, 1998 ; Bassano, Eme & Champaud, 2005, ainsi que les analyses proposées ci-après).

Le rôle des interactions internes. Enfin, une dernière idée centrale est que les interactions que nous appellerons « internes », c'est à dire les interactions entre les composants du langage (par exemple, entre lexique et grammaire), ont un rôle moteur dans le développement. Cette idée se situe dans le prolongement de la conception intégrative du traitement du langage avancée par le Modèle de compétition, selon laquelle le sujet a recours en même temps et de façon intégrée aux différents niveaux d'information, lexical, morphologique, syntaxique et prosodique (Bates et MacWhinney, 1987).

De manière similaire, des recherches récentes sur le développement précoce ont montré qu'il existait des relations d'interdépendance entre les différents composants du langage au cours de l'acquisition, notamment entre développement lexical et développement grammatical. Ainsi, l'asynchronie temporelle entre émergence du lexique et émergence de la grammaire, qui est l'un des phénomènes les mieux établis de la psycholinguistique du développement (Bates & Goodman, 1999), reflèterait en réalité, selon E. Bates et ses collègues, une puissante relation de cause à effet. Cette relation apparaît dans divers travaux qui ont mis en évidence de fortes corrélations non linéaires liant la taille du vocabulaire des enfants avec l'apparition des combinaisons de mots et de la complexité syntaxique. L'hypothèse proposée - confortée par les simulations connexionnistes et compatible avec une vision de la grammaire non dissociée du lexique - est que le développement grammatical précoce dépend du développement lexical, et, plus spécifiquement, dépend d'une certaine « masse lexicale critique », nécessaire pour que la grammaticalisation puisse s'effectuer (Bates *et al.* 1988 ; 1995 ; Bates & Goodman, 1999 ; Marchman & Bates, 1994 ; voir aussi Bassano, 1999 ; 2000b ; Bassano *et al.* 2004 ; cf. les analyses proposées ci-après).

3. Structuration du lexique : développement et variabilité inter-individuelle

Nous nous proposons maintenant d'illustrer plus directement notre propos avec quelques analyses provenant des recherches que nous menons au *Laboratoire Cognition et Développement* sur l'émergence et la constitution du système linguistique chez l'enfant français. La première illustration concerne le thème de la structuration du lexique (Bassano, 1998a, 2000a, 2005b ; Bassano, Maillochon & Eme, 1998 ; Bassano, Eme & Champaud, 2005), à propos duquel nous insisterons plus particulièrement sur le rôle de la variabilité inter-individuelle et sur l'articulation entre variabilité et tendances développementales.

Etude princeps sur le développement lexical. Le point de départ de nos travaux est l'étude princeps menée par E. Bates et ses collègues à partir des rapports parentaux du CDI (Bates *et al.*, 1994 ; 1995 ; Bates & Goodman, 1999). Cette étude montre notamment que le lexique des enfants anglophones américains est d'abord dominé par la production des noms, qui forment plus de 55% du vocabulaire quand ce vocabulaire comporte entre 100 et 200 mots (autour de 18-20 mois en moyenne dans cette étude), cette proportion diminuant ensuite. En revanche, la proportion de prédicats (verbes et adjectifs), initialement faible, s'accroît régulièrement. Et la proportion de mots de la classe fermée (ou mots grammaticaux), presque inexistante tant que le vocabulaire n'a pas atteint 400 mots, s'accroît rapidement après ce seuil. Trois étapes ont été ainsi identifiées dans l'évolution de la composition du vocabulaire entre 16 et 30 mois : l'accent se porterait d'abord sur les éléments à valeur référentielle –les noms-, puis sur les éléments à valeur prédicative –les verbes en particulier-, et enfin sur les éléments grammaticaux. Ces tendances de développement ont été retrouvées de façon largement similaire pour bon nombre d'autres langues examinées à partir d'adaptations du CDI, en particulier pour l'italien, pour lequel la comparaison avec l'anglais a été explicitement faite (Caselli, Casadio & Bates, 1999).

L'étude de E. Bates et collègues montre aussi qu'il existe, entre enfants disposant de la même quantité de mots, des différences dans la composition du vocabulaire initial. En particulier, il existe une forte variation concernant la proportion de noms au début de l'acquisition. Ainsi, chez les enfants qui ont un vocabulaire entre 10 et 50 mots, la proportion de noms varie de 20% à 60%. L'observation de cette variation confirme sur une grande échelle la découverte de Nelson (1973), qui avait examiné les 50 premiers mots des enfants qu'elle étudiait. Mais l'étude de Bates *et al.* fait apparaître en outre un résultat qui, nous semble-t-il, n'a pas été assez souligné ni expliqué par la suite : la variation inter-individuelle

dans la proportion de noms s'atténue progressivement au fur et à mesure que le vocabulaire s'accroît. La variabilité stylistique dans la composition du vocabulaire se manifesterait donc aux étapes précoces du développement pour diminuer par la suite.

Développement du lexique et production naturelle. Nos travaux sur le développement du lexique reposent sur l'étude de la production naturelle, méthode qui implique le recueil et l'analyse systématiques de corpus de productions spontanées de jeunes enfants enregistrés au cours de sessions (de 30 minutes au minimum) où ils sont en interaction avec leur entourage (voir Bassano, 2005a, pour des précisions sur cette méthode). La transcription informatisée de nos corpus se conforme pour l'essentiel aux standards du réseau CHILDES (*Child Language Data Exchange System*, MacWhinney, 2000), qui est le réseau international de gestion des données de production précoce le plus connu. La base de données que nous avons constituée sur le français associe différents types de corpus. Les plus classiques sont les corpus longitudinaux, où un même enfant est suivi au cours d'une longue période de temps, à raison d'une ou deux sessions par mois (corpus de Benjamin, enregistré et transcrit de 2 à 4 ans environ, corpus de Pauline, de 1 à 3 ans et demi environ). Un autre type de données est fourni par la base transverso-longitudinale dite « base TRL », 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. Parmi ces enfants, dix se retrouvent d'un groupe à l'autre, formant la composante longitudinale du corpus.¹

Principes de la structuration lexicale en français. L'une des principales questions examinée est la recherche des principes de structuration du lexique en termes de classes de mots. En analysant les productions spontanées des enfants français tout-venants, nous avons

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. Maillochon, sous la direction de D. Bassano. Les enregistrements du corpus TRL ont été effectués pour leur plus grande part durant les années 1993-96 par I. Maillochon, 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. Maillochon, K. Rougier, C. Ryckebusch.

identifié treize classes d'unités lexicales, ou mots, que nous avons regroupées en quatre macro-catégories. Les noms, qui servent à désigner des entités, comme « Popi », « bébé », « crayon », « histoire », sont des noms propres ou des noms communs, des noms animés ou inanimés, concrets ou abstraits. Les prédicats, qui servent à désigner des actions, des états ou des qualités attribués aux entités, sont essentiellement des verbes, comme « manger », « partir », « vouloir », mais aussi des adjectifs, comme « petit », « jaune ». Les mots grammaticaux, qui sont des éléments fonctionnels de liaison, regroupent les adverbes, les déterminants du noms, les pronoms, les prépositions, les conjonctions et les auxiliaires. Enfin, la dernière catégorie regroupe les éléments dits « para-lexicaux », qui correspondent en partie à ce que d'autres auteurs appellent « socio-pragmatiques », et qui sont des éléments langagiers rudimentaires tels qu'interjections, particules « oui » et « non », expressions syncrétiques comme « ça y est », et enfin *fillers*. Les *fillers* (remplisseurs) sont les éléments monosyllabiques que l'on trouve souvent devant les noms ou les verbes dans les productions des jeunes enfants, comme « /a/ chat » « /eu/ boire », et qui sont vraisemblablement des préfigurations de mots grammaticaux.

La figure 1, qui représente l'évolution de la proportion de chacune des quatre macro-catégories dans le corpus de Pauline, étudié ici entre 14 et 30 mois selon une segmentation temporelle en quatre périodes d'âge successives, permet d'illustrer les principales transformations se produisant dans la composition du lexique au cours de la deuxième année et de la première moitié de la troisième année (Bassano, 1998a ; Bassano *et al.*, 1998). L'évaluation est menée sur la base des mots-types, c'est-à-dire des mots différents produits par l'enfant, soit les différentes entrées lexicales. Cette représentation, qui montre l'évolution de la part relative de chacune des catégories dans la production de l'enfant, indique que deux catégories - les noms et les éléments para-lexicaux - prédominent largement en fréquence durant les premières étapes et tendent à diminuer avec l'âge. A l'inverse, les deux autres catégories - les prédicats et les mots grammaticaux - sont relativement modérées au départ (encore que non négligeables) et augmentent considérablement au cours du temps. De la sorte, aux alentours de deux ans, il s'opère une inversion des hiérarchies de fréquence : les prédicats d'abord, puis, et surtout, les mots grammaticaux deviennent les éléments dominants dans la production spontanée du français.

Insérer Figure 1

Nous avons analysé selon les mêmes principes la composition du lexique dans les productions des enfants de la base TRL à 20 mois, 30 mois et 39 mois (20 enfants par groupe d'âge, Figure 2). Les résultats obtenus à 20 mois et 30 mois confortent ceux de l'étude de

Pauline : entre 20 et 30 mois, les proportions de noms et d'items para-lexicaux diminuent significativement, tandis que les proportions de prédicats et de mots grammaticaux augmentent significativement. L'analyse à l'âge de 39 mois nous a permis de poursuivre l'examen au delà de 30 mois, qui est généralement l'âge butoir des études sur le développement lexical précoce. Nous avons trouvé que les changements observés dans la part relative de chacune des catégories de mots entre 30 et 39 mois sont de nettement moindre amplitude qu'entre 20 et 30 mois. En particulier, les proportions de noms et de prédicats restent à 39 mois à peu près identiques à celles qu'elles étaient à 30 mois. La proportion de mots grammaticaux augmente un peu entre 30 et 39 mois, mais pas de façon significative. La seule évolution significative concerne la proportion d'éléments para-lexicaux, qui continue à diminuer entre 30 et 39 mois.

Insérer Figure 2

Variabilité inter-individuelle dans la composition du lexique. Une autre question abordée dans ces études est celle de la variabilité inter-individuelle. Nous avons examiné dans quelle mesure et comment le processus de développement lexical était affecté par la variabilité inter-individuelle dans les trois groupes d'âge de la base TRL (pour plus de précisions, voir Bassano *et al.*, 2005). Nous avons traité cette question de plusieurs manières. Nous avons d'abord évalué l'ampleur de la variabilité en calculant les coefficients de variation associés aux proportions des quatre catégories de mots. Le coefficient de variation (CV) est un paramètre de dispersion calculé en divisant l'écart type par la moyenne : il renseigne donc sur la variabilité relative. L'analyse des coefficients de variation (figure 3) montre principalement que les coefficients de variation sont élevés à 20 mois – ce qui dénote une variabilité importante - et nettement moins élevés à 30 et 39 mois pour les trois principales catégories (noms, prédicats, mots grammaticaux), tandis que leur valeur ne change pas beaucoup pour la catégorie para-lexicale. On peut donc estimer que globalement, la variabilité relative dans la composition du lexique diminue fortement dans le courant de la troisième année, au moins pour les trois catégories de mots classiques.

Insérer Figure 3

Nous avons ensuite conduit une recherche plus qualitative de ce que nous avons appelé les « profils lexicaux » correspondant aux différents modes d'organisation lexicale apparaissant chez les enfants d'un même groupe d'âge (Bassano, 2005b ; Bassano *et al.*, 2005). Des analyses en clusters portant sur la répartition des quatre principales catégories de mots ont été effectuées pour chaque groupe d'âge. Ces analyses montrent que les enfants de

20 mois se répartissent en trois classes principales associées à trois profils d'organisation lexicale très différents. Le premier, correspondant à près de la moitié des enfants, comporte une forte dominante de noms (40%), qu'on n'observe dans aucun des deux autres profils : on peut donc considérer que c'est le « profil référentiel » de la littérature anglo-saxonne. Cependant, deux autres profils sont mis en évidence : un profil appelé « para-lexical » parce qu'il compense sa plus faible proportion de noms par une dominante d'éléments para-lexicaux ; et un profil appelé « diversifié » parce que la représentation des différentes catégories de mots y est équilibrée, et qui comporte une proportion de prédicats et de mots grammaticaux étonnamment forte. Chez les 20 enfants de 30 mois, quatre profils se dégagent, parmi lesquels l'un ne concerne qu'un seul enfant, en outre atypique, puisque c'est l'enfant qui a un langage peu élaboré par rapport à son groupe d'âge. Les trois profils restants sont très proches, et en particulier ils ont tous une forte dominante de mots grammaticaux (entre 32% et 42%). On peut donc penser que ce sont des variantes d'un profil général « grammatical », qui caractérise la production des enfants de 30 mois. A 39 mois enfin, on trouve trois profils, présentant tous une dominante de mots grammaticaux (entre 37% et 45%) et donc assimilables aussi à des variantes d'un profil fondamentalement « grammatical ».

En résumé, l'analyse des tendances du développement lexical montre que, pour les enfants français, c'est au cours de la troisième année que s'opère la structuration du lexique en termes de classes de mots. Ces transformations s'organisent selon deux phénomènes principaux : le déplacement de fréquence des noms vers les prédicats, en particulier vers les verbes, et le déplacement de fréquence des éléments para-lexicaux, interjections et *fillers* en particulier, vers les mots grammaticaux. Ces transformations vont globalement dans le même sens que celles établies - avec une méthode différente - par Bates et ses collègues pour l'anglais et l'italien (Bates *et al.*, 1994 ; Caselli *et al.*, 1999). Les régularités de développement se dégageant de ces recherches sur des langues européennes viennent à l'appui du modèle général « référence - prédication – grammaire », montrant en particulier une prédominance initiale des noms sur les verbes, et un délai dans le développement des mots grammaticaux. On peut cependant noter que, chez les enfants français, la proportion de verbes est légèrement plus élevée que chez les enfants anglophones américains au début du langage (une tendance mentionnée aussi par B. de Boysson-Bardies, 1996), et surtout que le développement des mots grammaticaux connaît un essor remarquable, sans commune mesure avec celui établi pour l'anglais. Ces différences conduisent à penser que les propriétés spécifiques de l'input linguistique peuvent avoir des incidences sur le développement du lexique, modulant ou infléchissant, d'une manière souvent subtile, les effets des contraintes

cognitives générales. L'analyse de la variabilité inter-individuelle au moyen des coefficients de variation montre que la variabilité dans la composition du lexique est importante à 20 mois et diminue au cours de la troisième année, ce qui rejoint les résultats de Bates et collègues sur l'anglais. En outre, l'identification des profils lexicaux permet de spécifier et d'interpréter le déroulement de la variabilité. Le resserrement qualitatif autour du profil de type « grammatical » observé dès 30 mois correspond au fait que tous les enfants ont alors bien développé leur stock de mots grammaticaux. Cela nous paraît refléter la nécessaire intégration des contraintes grammaticales qui s'opère déjà à cet âge en français, et qui est le tribut de l'acquisition de la langue. Et, en faisant un pas de plus, nous dirions que les différences entre enfants observées dans la composition du lexique au début du langage ne sont pas déterminantes pour le développement lexical ultérieur, puisque, finalement, la pression du développement et du système linguistique est plus forte que la variabilité.

4. Processus de grammaticalisation : l'exemple de la catégorie des noms

Le deuxième thème d'illustration concerne l'analyse des processus de grammaticalisation. En prenant l'exemple du développement de la catégorie des noms (Bassano, 1998b ; 2000b ; 2005a ; Bassano & Eme, 2001), nous espérons faire apparaître deux caractéristiques des processus de grammaticalisation : l'articulation entre explosion et progressivité, et la relation au développement lexical.

Explosion et progressivité dans la grammaticalisation des noms. Dans le prolongement direct de nos travaux sur le lexique, ces recherches visent à déterminer quand et comment les mots identifiés comme « noms » au vu de leurs propriétés lexico-sémantiques – parce qu'ils font référence à des entités animées ou inanimées, concrètes ou plus abstraites – acquièrent, dans la production de l'enfant, les propriétés grammaticales définissant la classe des noms. Ce processus, que nous appelons « grammaticalisation des noms », se manifeste, pour le français, par l'intégration de la contrainte d'emploi du déterminant. En effet, dans la langue orale adulte, la caractéristique grammaticale du nom est que celui-ci est, en règle générale, accompagné par un déterminant. La contrainte est double : le déterminant est antéposé (il précède le nom), et il porte les marques du genre et du nombre. Elle n'est cependant 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. Pour déterminer quand et comment cette contrainte est maîtrisée dans la production de l'enfant, nous avons défini un indice de grammaticalisation

des noms mesurant la capacité à utiliser un déterminant dans les contextes où celui-ci est obligatoire (Bassano, 2000b). Cet indice correspond au taux des noms employés avec déterminant par rapport à tous les noms pour lesquels le déterminant serait requis dans la langue adulte. Son évolution dans le corpus de Pauline entre 14 et 30 mois, calculée selon deux versions, l'une stricte et l'autre « accommodante », est présentée dans la figure 4.

Insérer Figure 4

La version stricte de l'indice mesure le taux d'utilisation de vrais déterminants, clairement spécifiés, comme dans « un chat », « le chien », « les chaussures », « c'est mon chausson ». Ce calcul 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. La version stricte fait donc apparaître un phénomène d'explosion grammaticale typique, se manifestant par une accélération brusque du processus d'intégration de la contrainte, localisé pour cette enfant entre 27 et 28 mois.

La version appelée « accommodante » tient compte, non seulement des vrais déterminants, mais aussi des *fillers*, c'est-à-dire des éléments monosyllabiques observés 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 et discontinuité de développement, en montrant que, selon la finesse de l'examen, un processus qui apparaît comme « discontinu » (explosif) à un certain niveau peut se révéler « en continuité » à un autre niveau.

Relations temporelles entre développement lexical et grammatical. L'examen des relations entre lexique et grammaire dans le développement des noms a pour objectif d'apporter des arguments en faveur de notre conception intégrative de l'acquisition du langage, en montrant que l'émergence de la grammaire - ou du moins de certains de ses

aspects - est liée à certains aspects du développement du lexique. Une première analyse examine les aspects quantitatifs et temporels de la relation entre lexique et grammaire dans les données de Pauline. Cette analyse vise à vérifier l'hypothèse « de la masse lexicale critique » (Bates & Goodman, 1999 ; Marchman & Bates, 1994), selon laquelle le déclenchement des acquisitions grammaticales est lié à la constitution d'un certain stock lexical, ce qui place le démarrage du développement grammatical en relation de dépendance par rapport au développement lexical. Pour établir l'existence de relations de dépendance dans nos données longitudinales sur les noms, il nous faut montrer, d'une part, qu'il existe des moments critiques remarquables dans l'accroissement de la production lexicale et dans l'essor du processus grammaticalisation, et montrer, d'autre part, que ces moments critiques sont en décalage temporel.

La Figure 5 met en regard, pour les données de Pauline, la courbe d'évolution de l'indice strict de grammaticalisation et les courbes d'évolution de la production lexicale des noms (nombres absolus de mots-types et d'occurrences dans les échantillons de 120 énoncés). Comme on peut le voir, la production lexicale des noms connaît, après un parcours assez erratique, un accroissement nettement marqué à partir de 25 mois, c'est-à-dire deux mois avant l'explosion dans la grammaticalisation, qui, comme nous l'avons déjà montré, se produit 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 qui place le moment d'explosion grammaticale en position différée par rapport au moment d'augmentation de la production lexicale, délai en faveur l'hypothèse de la « masse lexicale critique ».

Insérer Figure 5

Interactions entre propriétés lexico-sémantiques et grammaticalisation. 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 ; 2005a). Elle propose que la grammaticalisation pourrait être liée, non seulement à la taille du lexique, mais aussi à la structure et la nature de celui-ci. Selon cette hypothèse, certaines sous-classes sémantiques de noms pourraient être plus propices que d'autres à former le substrat des processus de grammaticalisation. Cela revient à rechercher si la grammaticalisation est influencée par des propriétés lexico-sémantiques des noms, telles que le caractère animé ou non animé. L'analyse de la répartition, toujours dans les données du corpus de Pauline, de la

procédure d'omission (absence incorrecte de déterminant) et de la procédure d'emploi d'un *filler* ou déterminant pour les noms communs animés, d'une part, et pour les noms communs inanimés, d'autre part, fait apparaître un effet significatif) du caractère animé / non animé. C'est pour les noms d'animés que l'omission du déterminant est proportionnellement la plus forte (pour 64% des animés et pour 52% des inanimés), tandis que c'est pour les noms d'inanimés que la proportion d'emploi d'un *filler* ou déterminant est la plus forte (pour 36% des animés et pour 48% des inanimés). L'analyse plus fine distinguant entre déterminant et *filler* (Figure 6) indique que les déterminants sont produits en proportion plus élevée avec les inanimés concrets, tandis que les *fillers* sont produits en proportion plus élevée avec les inanimés abstraits. Ces résultats suggèrent que les noms d'inanimés concrets, autrement dit les noms d'objets, pourraient constituer des candidats privilégiés pour entraîner la contrainte d'utilisation du déterminant et ainsi servir « d'amorceur sémantique » de la grammaticalisation de la classe des noms. En revanche les noms d'animés tendent à être plus souvent employés sans déterminant, sans doute parce que ce sont aussi eux qui, le plus souvent et comme les noms propres, ne nécessitent pas de déterminant.

Insérer Figure 6

En résumé, ces analyses sur la grammaticalisation des noms montrent que l'acquisition de la contrainte grammaticale caractéristique des noms - l'emploi du déterminant pré-nominal - est un processus tout à la fois, selon le niveau de l'analyse, « explosif » et « progressif », en d'autres termes un processus où coexistent discontinuité et continuité. Elles suggèrent en outre qu'il existe des relations de dépendance entre le développement grammatical des noms et leur développement lexical. Les arguments en faveur de cette conception proviennent, en premier lieu, de l'analyse des relations temporelles entre les deux processus, analyse qui étaye l'hypothèse de « la masse lexicale critique » (relation entre taille du lexique des noms et grammaticalisation). Ils proviennent en second lieu de l'analyse des interactions entre propriétés lexico-sémantiques des noms et emploi du déterminant, analyse qui nous conduit à avancer l'hypothèse d'un « amorçage lexico-sémantique » de l'emploi du déterminant, plus facilement associé aux noms inanimés, en particulier aux noms d'objets concrets (relation entre la nature du lexique des noms et la grammaticalisation).

5. Modélisation de la longueur des énoncés : moments critiques et variabilité intra-individuelle

Notre troisième et dernière illustration est issue d'un programme en cours, mené en collaboration avec P. van Geert, sur la modélisation du développement lexical et grammatical

précoce dans le cadre de la théorie des systèmes dynamiques et à partir de notre base de données de production naturelle. L'exemple présenté ici provient d'une recherche sur la modélisation de la longueur des énoncés (Bassano & Van Geert, 2005), qui nous permettra d'aborder la question du rôle de la variabilité intra-individuelle dans le développement.

Selon une conception classique en psycholinguistique, le processus d'allongement des énoncés chez le tout jeune enfant, caractérisé par le passage des énoncés de un mot aux énoncés de deux ou trois mots, puis aux énoncés plus longs, est considéré en large part comme un indice du développement grammatical. Dans une approche fonctionnaliste, l'acquisition de la grammaire est conçue comme une construction progressive où l'enfant passe peu à peu d'un mode de langage immature à un langage de type adulte selon un processus d'abstraction et de complexification croissantes. Notre but était donc de voir si, en analysant les données sérielles denses qu'offrent les corpus longitudinaux, il était possible de déterminer des moments critiques dans le processus de construction syntaxique.

Dans la théorie des systèmes dynamiques (van Geert, 1991, 1994, 2003), les moments critiques sont des périodes où se produisent dans le système des transformations qualitatives sous-jacentes aux changements quantitatifs. Ces périodes sont des transitions, caractérisées par un niveau élevé de fluctuations qui traduisent les hésitations de l'enfant entre différentes stratégies ou modes de production. Le niveau de fluctuation définit la variabilité intra-individuelle, c'est-à-dire la variabilité manifestée par un même sujet à un même moment du développement. Selon cette conception, la variabilité intra-individuelle n'est donc pas considérée comme du bruit, mais comme un phénomène signifiant : un pic de variabilité va être l'indicateur d'un moment critique et d'une transition dans le développement.

La figure 7, qui présente l'analyse simplifiée des données de Pauline entre 14 et 36 mois, montre l'évolution des énoncés de un mot, des énoncés de deux et trois mots, et des énoncés de quatre mots et plus (courbes lissées). Ces courbes ont des patrons de développement très différents les uns des autres. La production des énoncés de un mot a une fréquence maximale au début des observations, c'est-à-dire 14 mois, et suit une trajectoire déclinante jusqu'au 30^{ème} mois. La production des énoncés de deux et trois mots présente une courbe en U inversé avec un maximum de fréquence aux alentours du 26^{ème} mois, et ensuite un déclin qui se stabilise après le 32^{ème} mois. Enfin, la production des énoncés de quatre mots et plus débute autour du 23^{ème} mois seulement et augmente selon une courbe en S qui présente son maximum autour du 31^{ème} mois. Dans notre modèle, ces trois groupes d'énoncés constituent les attracteurs du système, qu'on voit donc passer de la dominance des énoncés de un mot, caractérisant le niveau globalement non combinatoire du langage, à la

dominance des énoncés de deux ou trois mots, caractérisant l'accès à un niveau combinatoire simple (où les catégories syntaxiques abstraites ne sont pas nécessairement requises, bien qu'on puisse trouver à cette période des précurseurs ou des mots grammaticaux simples), et finalement à la dominance des énoncés de quatre mots et plus, qui marque l'accès à un niveau grammatical plus élaboré où catégories et structures syntaxiques commencent à être présentes.

Insérer Figure 7

La figure 7 montre également la courbe correspondant au calcul de la variabilité intra-individuelle dans la production de ces trois groupes d'énoncés. On voit qu'elle présente une augmentation très nette - un pic, dont nous précisons qu'il est non aléatoire -, localisée au moment de plus forte croissance des énoncés de quatre mots et plus (aux alentours du 28^{ème} mois, juste au moment où ce type d'énoncés va dominer la production de l'enfant). Il se produit donc une augmentation de la variabilité au moment de la transition vers le niveau plus sophistiqué de la grammaire, augmentation qui désigne cette période comme un moment critique du développement grammatical. En analysant la structure des énoncés de quatre mots et plus produits par l'enfant aux alentours de cette période, on peut identifier trois phénomènes responsables de l'augmentation de la longueur des énoncés et de l'augmentation de la variabilité. Le premier est la systématisation des constructions transitives simples, qui deviennent à cette période les constructions simples les plus fréquentes, surpassant en nombre les constructions présentatives, impératives ou intransitives : les constructions transitives (comme « je veux une feuille », « tu me le donne ») sont celles qui, contrairement aux précédentes, requièrent la présence d'au moins deux arguments pour être correctes. Le deuxième phénomène est l'émergence des constructions complexes, représentées principalement par les propositions infinitives (« il faut ranger alors »), les propositions relatives (« c'est moi qui range mes cartes »), et les propositions circonstancielles (causales essentiellement, « oui, parce que t'as gagné ça »). Ces deux phénomènes sont bien représentatifs de l'accès à la complexité syntaxique. Le troisième est un peu différent : il s'agit de la production des dislocations, qui sont des constructions typiques du français oral avec redoublement du sujet ou de l'objet, se traduisant par une expansion du syntagme nominal (comme dans « ma viande, où elle est ma viande ? »).

En résumé, ces résultats montrent, d'une part, que la variabilité intra-individuelle est un phénomène signifiant pour le développement, un pic de variabilité non aléatoire étant annonciateur de réorganisations dans le système ; et d'autre part, que dans le développement du langage, un tel pic de variabilité se produit au moment de transition vers la combinatoire complexe et le niveau syntaxique du langage.

Conclusion

Nous avons présenté trois illustrations venant à l'appui d'une approche fonctionnaliste intégrative de l'acquisition du langage. L'analyse du développement lexical nous a permis de montrer comment la variabilité inter-individuelle dans la composition du lexique des enfants, importante au tout début du langage, s'atténue au cours de la troisième année, et comment ce décours s'intègre aux tendances générales du développement. L'analyse de la grammaticalisation des noms nous a servi d'exemple pour montrer que le développement de la grammaire chez le jeune enfant – s'agissant du moins des aspects étudiés ici, qui concernent principalement la morphologie - est un processus à la fois graduel et non linéaire (où continuité et discontinuité coexistent), et qu'il est en relation de dépendance avec le développement du lexique. Enfin, une étude exploratoire sur la modélisation de la longueur des énoncés, menée dans le cadre de la théorie des systèmes dynamiques, nous a permis de montrer que le passage du langage de l'enfant à un niveau de complexité syntaxique relativement élaboré est une période de transition, caractérisée par une forte augmentation de la variabilité intra-individuelle.

Références

- AKHTAR, N. (2004). Nativist versus constructivist goals in studying child language. *Journal of Child Language*, 31, 459-62.
- AKHTAR, N., & TOMASELLO, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, 33, 952-65.
- ASLIN, R. N., SAFFRAN, J. R., & NEWPORT, E. L. (1999). Statistical learning in linguistic and non-linguistic domains. In B. MacWhinney (Ed.), *The emergence of language*, New Jersey: Lawrence Erlbaum Associates, 359-380.
- 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 lexique/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. (2005a). Production naturelle précoce et acquisition du langage : l'exemple du développement des noms, *LIDIL*, 31 (sous presse).
- BASSANO, D. (2005b). Le développement lexical précoce : état des questions et recherches récentes sur le français. In F. Grossmann, M.A. Paveau & G. Petit (Eds), *Didactique du lexique : langue, cognition, discours*, (pp. 9-27). Grenoble : Ellug.
- BASSANO, D., EME, E. & CHAMPAUD, C. (2005). A naturalistic study on early lexical development : General processes and inter-individual variations in French children, *First Language*, 25 (1), 67-101.
- BASSANO, D. & EME, E. (2001). Development of noun determiner use in French children: lexical and grammatical bases. In M. Almgren, A. Barreña, M.J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney C., (Eds), *Research on Child Language Acquisition. Proceedings of the 8th Conference of the International Association for the Study of Child Language*, (pp 1207-1220). Cascadilla Press, Somerville, MA.
- 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. & 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. & VAN GEERT, P. (2005). Modeling utterance length: Evidence for changes, transitions and intra-individual variability in early grammatical development (soumis).
- BATES E., BRETHERTON I., & SNYDER L. (1988). *From first words to grammar: individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- BATES, E., DALE, P., & THAL, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (eds), *The Handbook of child language*, Cambridge MA: Blackwell.
- 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. (1979). A functionalist approach to the acquisition of grammar. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental pragmatics*, (pp. 167-211). New York: Academic Press.
- 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.
- BATES, E. & MAC WHINNEY, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, (pp. 157-193). Lawrence Erlbaum.
- BATES, E. & MAC WHINNEY, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds), *The crosslinguistic study of language processing* (pp. 3-73), Cambridge MA: Cambridge University Press.
- 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.
- BATES, E., THAL, D., FINLEY B., & CLANCY, B. (2003). Early language development and its neural correlates. In S. J. Segalowitz & I. Rapin (Eds.). *Handbook of neuropsychology* (vol. 8, Part II, 2nd Edition), Amsterdam, Elsevier.
- 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.
- BOURGEOIS, J.P. (1997). Synaptogenesis, heterochrony and epigenesis in the mammalian neocortex. *Acta Paediatrica Supplement*, 422, 27-33.
- BOYSSON-BARDIES, B. (1996). *Comment la parole vient aux enfants*, Paris : Editions Odile Jacob.
- CASELLI, M. C., CASADIO, P. & BATES, E. (1999). A comparison of the transition from first words to grammar in English and Italian, *Journal of Child Language*, 26(1), 69-111.
- CHOI, S. & GOPNIK, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study, *Journal of Child Language*, 22(3), 497-529.
- CLAHSEN, H. (1996) (Ed.). *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and cross-linguistic comparisons*. Amsterdam: Benjamins.
- CLARK, E. V. (1993). *The lexicon in acquisition*. Cambridge University Press.
- ELMAN, J. L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed), *The emergence of language*, New Jersey: Lawrence Erlbaum Associates.

- 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).
- FUCHS, C. (2004) (ed.). *La linguistique cognitive*. Ophrys / Maison des Sciences de l'Homme.
- 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.
- 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.
- GOMEZ, R. & GERKEN, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109-135.
- GOLDBERG, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. Mac Whinney (Ed), *The emergence of language*, New Jersey: Lawrence Erlbaum Associates.
- HIRSH-PASEK, K. & GOLINKOFF, R. (1996). *The origins of grammar*. Cambridge, MA: MIT Press.
- KAIL, M. (2000). Perspectives sur l'acquisition du langage. In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Vol. 1 : Le langage en émergence*, Paris: Presses Universitaires de France, 9-27.
- KAIL, M. (2004). Développement du langage et plasticité : l'innéisme en question. *Bull. Hist. Epistém. Sci. Vie*, 2004, 11, (2), 39-73.
- KAIL, M. & BASSANO, D. (2000). Méthodes d'investigation et perspectives heuristiques. In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Vol. 1 : Le langage en émergence*, Paris: Presses Universitaires de France, 29-59.
- KARMILOFF-SMITH, A (1979). *A functional approach to language acquisition*. Cambridge: Cambridge University Press.
- KARMILOFF-SMITH, A (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.
- LIEVEN, E. V., PINE, J. M., & BALDWIN, W. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 197-219.

- MAC WHINNEY, B. (1999) (ed.). *The emergence of language*, New Jersey: Lawrence Erlbaum Associates.
- MAC WHINNEY, B. (2000a). *The CHILDES project : tools for analysing talk*, Hillsdale, NJ: Erlbaum.
- MAC WHINNEY, B. (2000b). Emergence from what? *Journal of Child Language*, 27, 727-733.
- MAC WHINNEY, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 883-914.
- 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.
- NAIGLES, R. L. (2002). Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition*, 86, 157-199.
- 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.
- NELSON, K. (1981). Individual differences in language development: implications for development and language. *Developmental Psychology*, 17 (2), 170-187.
- PINKER, S. (1994). *The language instinct: How the mind creates language*. New York: Morrow.
- PLUNKETT, K. (1995). Connectionist approaches to language acquisition. In P. Fletcher & B. MacWhinney (eds), *The Handbook of child language*, Cambridge MA: Blackwell.
- RAMUS, F., HAUSER, M. D., MILLER, C., MORRIS, D. & MEHLER, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288, 349-351.
- SAFFRAN, J. R., ASLIN, R. N., & NEWPORT, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274 (5294), 1926-1928.
- SLOBIN, D. I. (1985-1997) (ed.). *The crosslinguistic study of language acquisition*. Vol. 1-5. Hillsdale: Erlbaum.
- TARDIF, T., SHATZ, M & NAIGLES, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian and Mandarin. *Journal of Child Language*, 24, 535-565.
- THELEN, B. & BATES, E. (2003). Connectionism and dynamic systems: are they really different? *Developmental Science*, 6:4, 378-391.
- THELEN, B. & SMITH, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

- THELEN, B. & SMITH, L. B. (1998). Dynamic Systems Theories. In W. Damon & R. Lerner (Eds.), *Handbook of Child Psychology* (pp. 563-634). New York: Wiley.
- TOMASELLO, M. (1992). *First verbs: A case study in early grammatical development*, Cambridge University Press.
- TOMASELLO, M. (1998) (ed.). *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, New Jersey & London: Lawrence Erlbaum.
- TOMASELLO, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209-253.
- TOMASELLO, M. (2003). *Constructing a language: A Usage-Based Theory of Language Acquisition*. Cambridge, Massachusetts: Harvard University Press.
- TOMASELLO, M. & AKHTAR, N. (2003). What paradox? A response to Naigles (2002). *Cognition*, 88, 317-23.
- TOMASELLO, M. & BROOKS, P. J. (1999). Early syntactic development: A Construction Grammar approach. In M. Barrett (Ed.), *The Development of Language* (pp. 161-190). Psychology 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.
- VAN GEERT, P. (2003). Dynamic systems approaches and modeling of developmental processes. In J. Valsiner & K. J. Conolly (Eds.), *Handbook of developmental psychology*. London : Sage.
- 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.
- VAN VALIN, R. D. (1998). The acquisition of WH-Questions and the mechanisms of language acquisition. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, New Jersey & London: Lawrence Erlbaum.
- WIJNEN, F., KEMPEN, M & GILLIS, S. (2001). Root infinitives in Dutch early child language: an effect of input? *Journal of Child Language*, 28, 629-660.

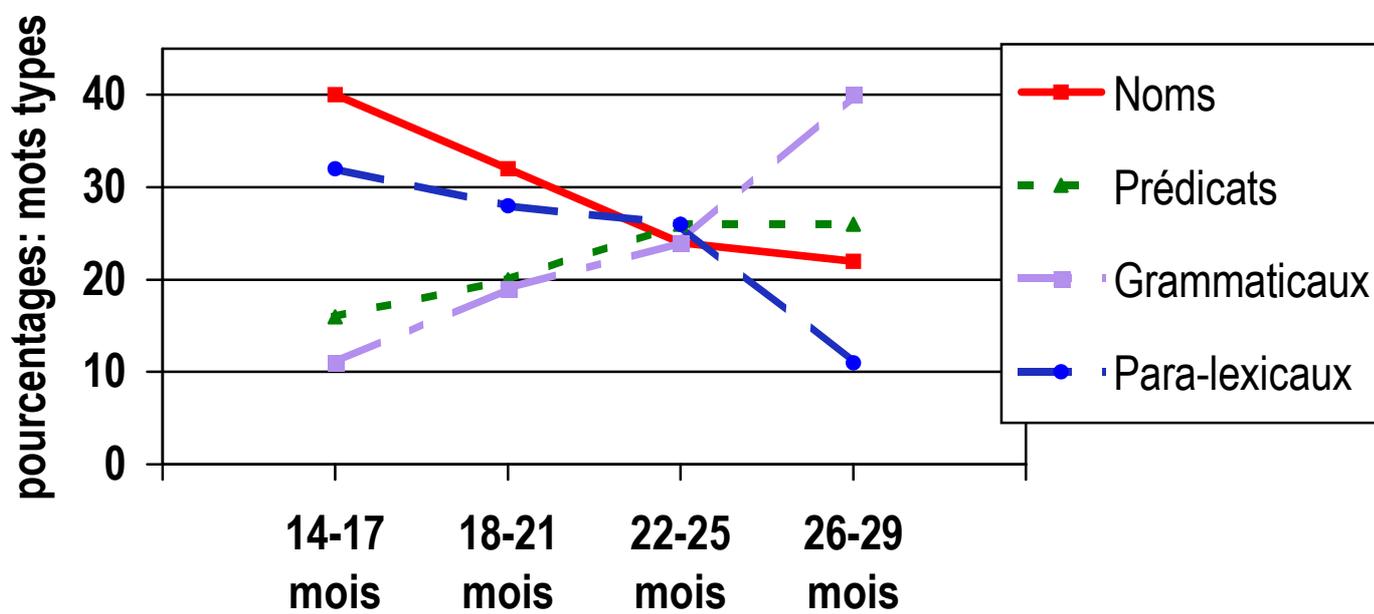


Figure 1.

Evolution de la composition du lexique dans le corpus de Pauline entre 14 et 30 mois :
proportion de mots-types pour chacune des 4 catégories

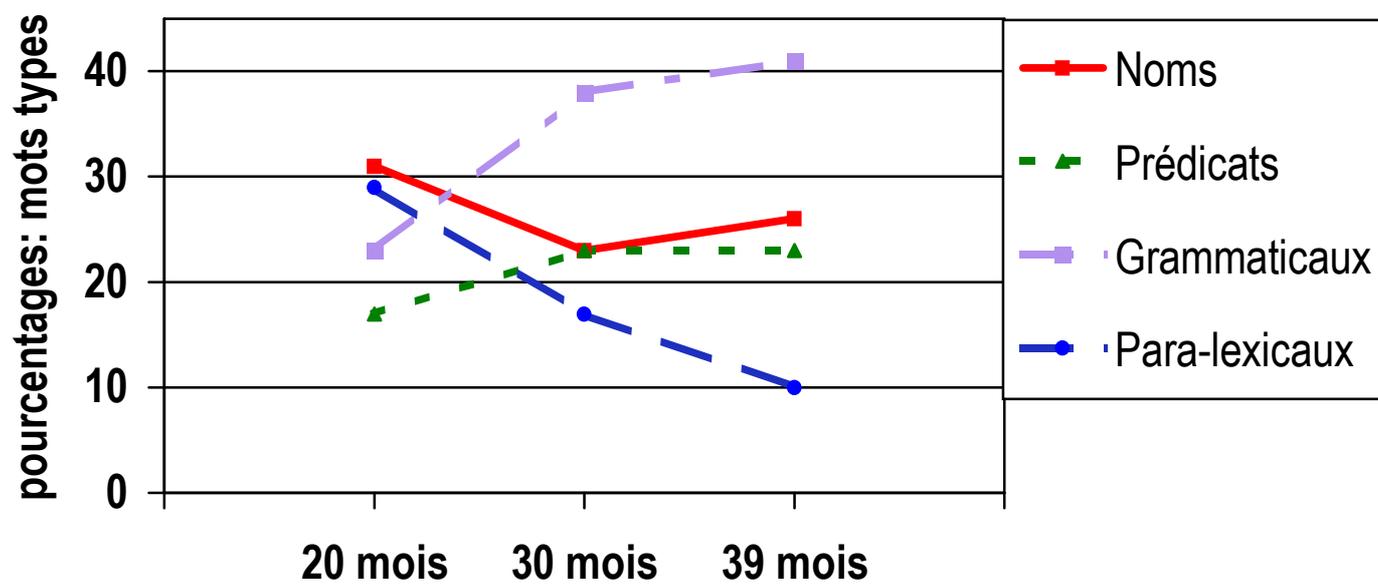


Figure 2.
Composition du lexique dans le corpus transversal à 20, 30 et 39 mois :
proportion de mots-types pour chacune des 4 catégories

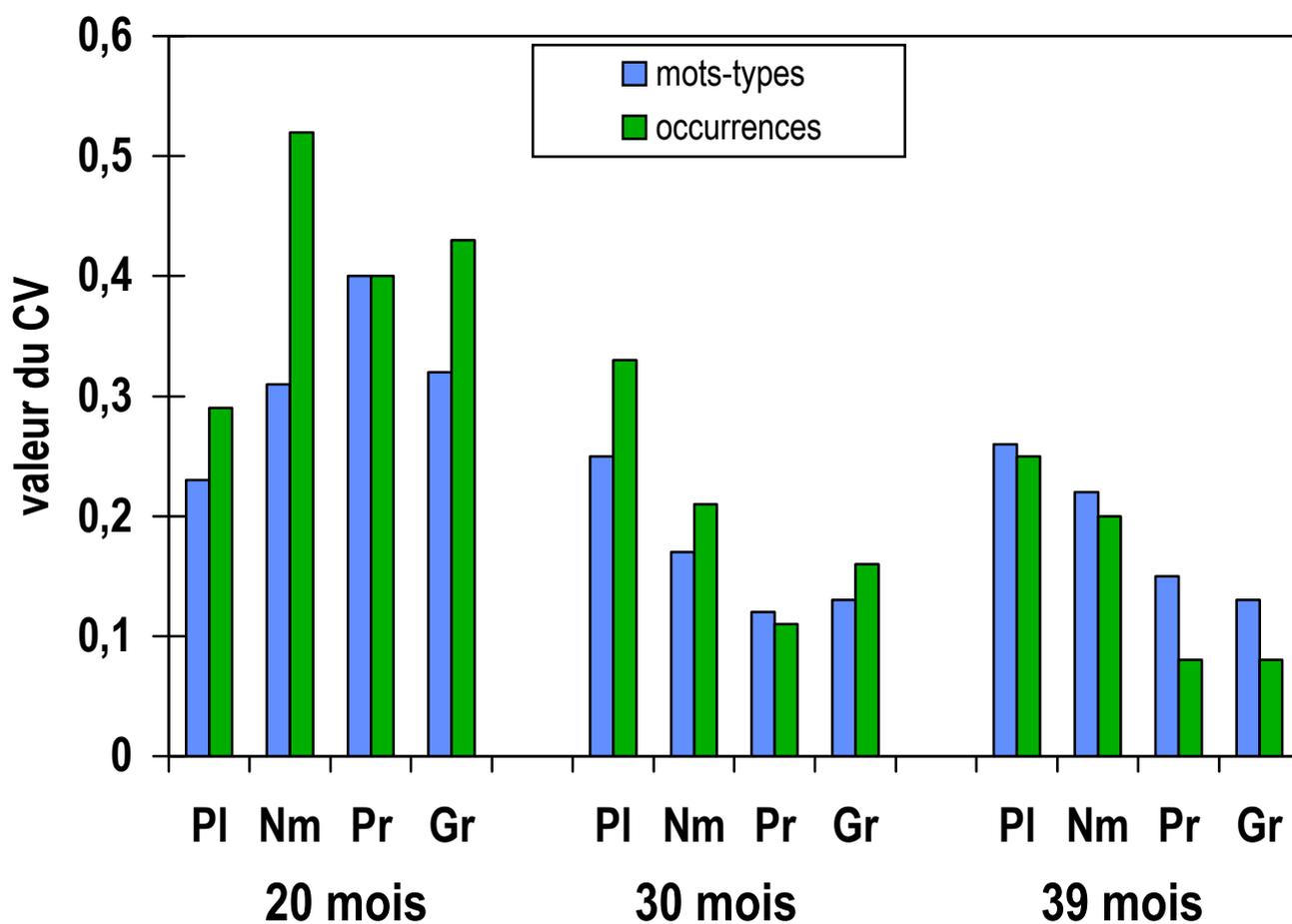


Figure 3.
 Variabilité inter-individuelle dans la composition du lexique :
 coefficients de variation pour chaque catégorie lexicale à 20 mois, 30 mois et 39 mois

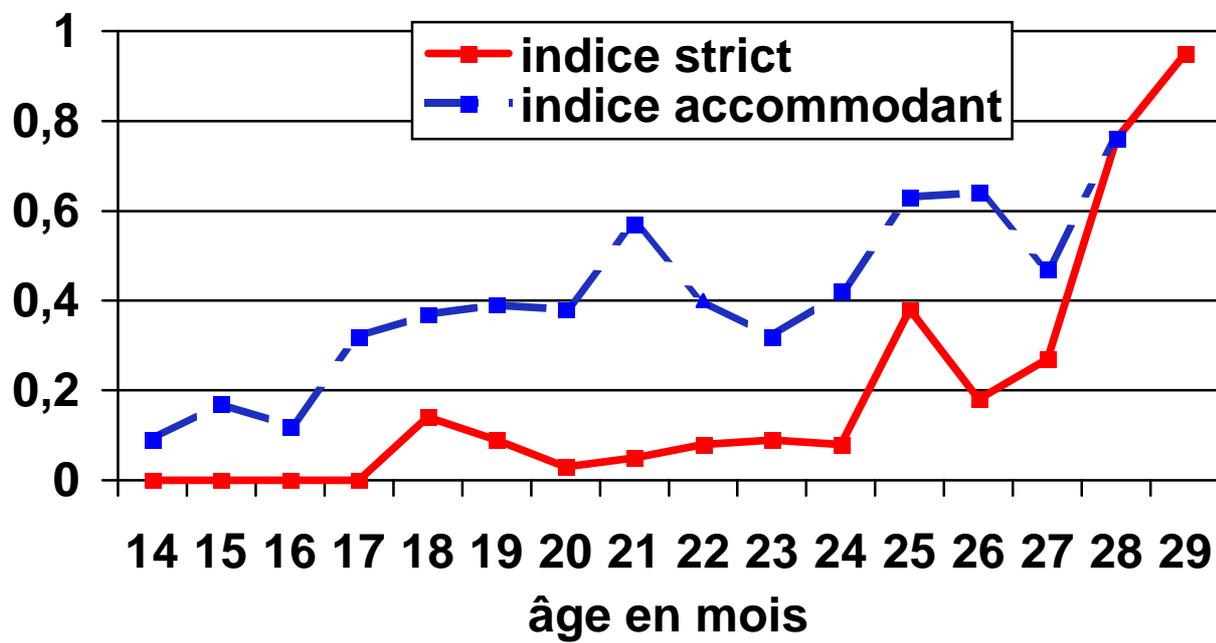


Figure 4.
Evolution de l'indice de grammaticalisation des noms chez Pauline entre 14 et 30 mois

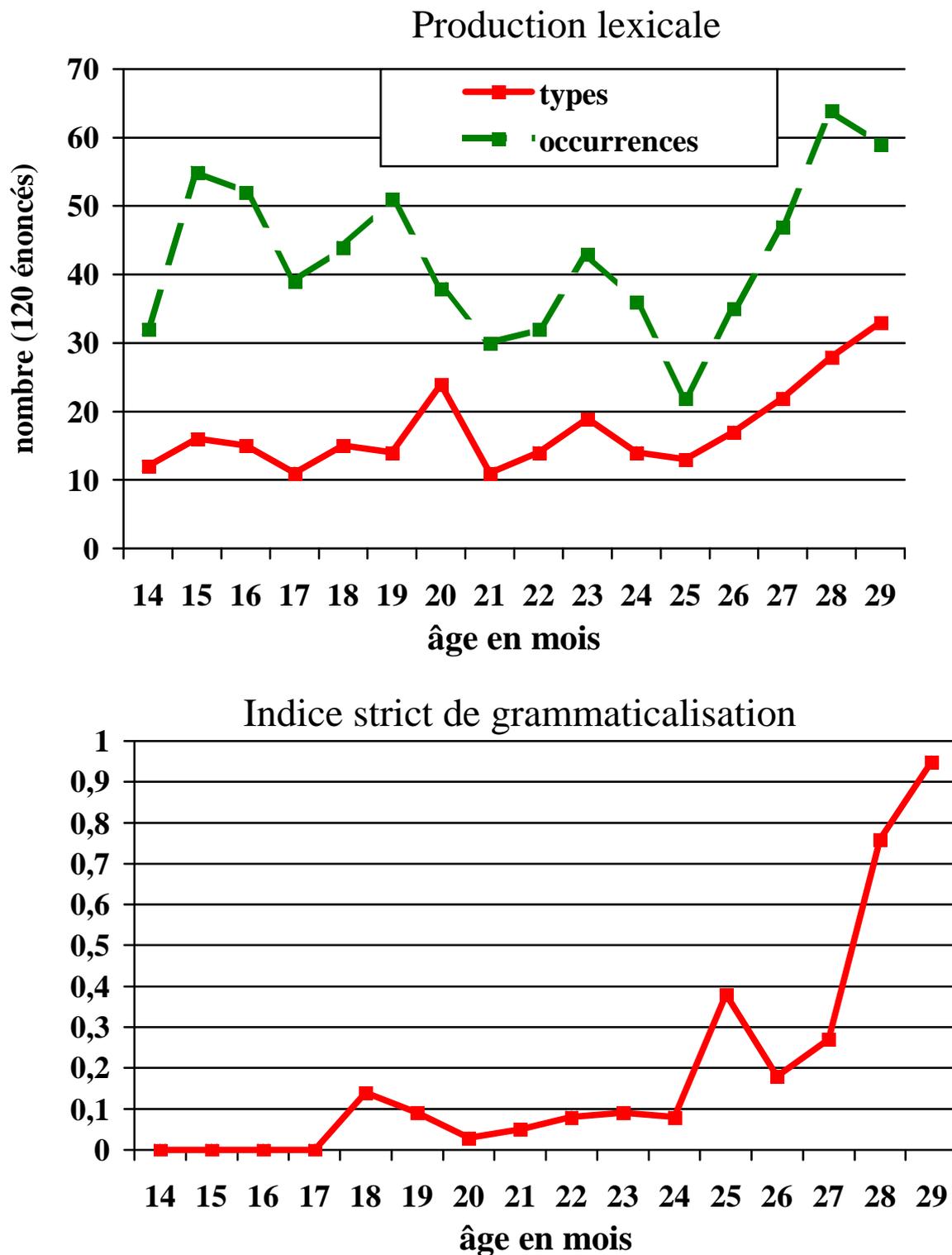


Figure 5.
Relations temporelles entre développement lexical et grammatical des noms (corpus de Pauline)

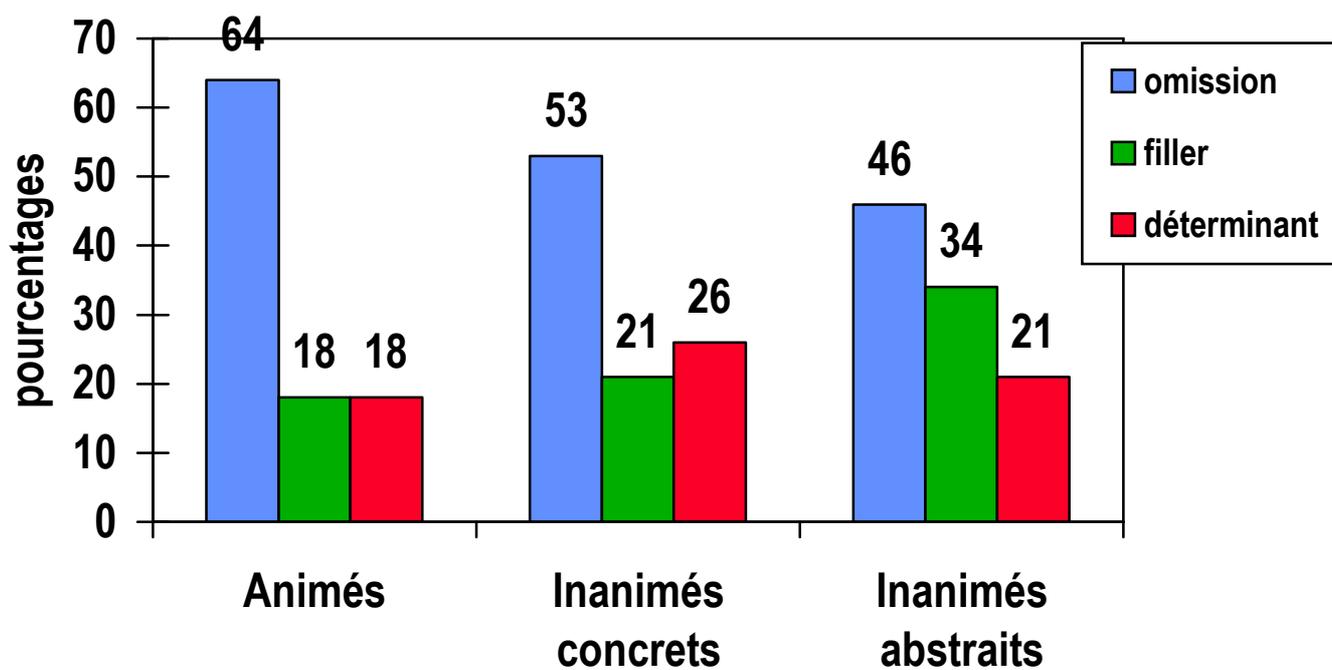


Figure 6.
Répartition des procédures d'emploi des noms dans les sous-classes sémantiques de noms (corpus de Pauline)

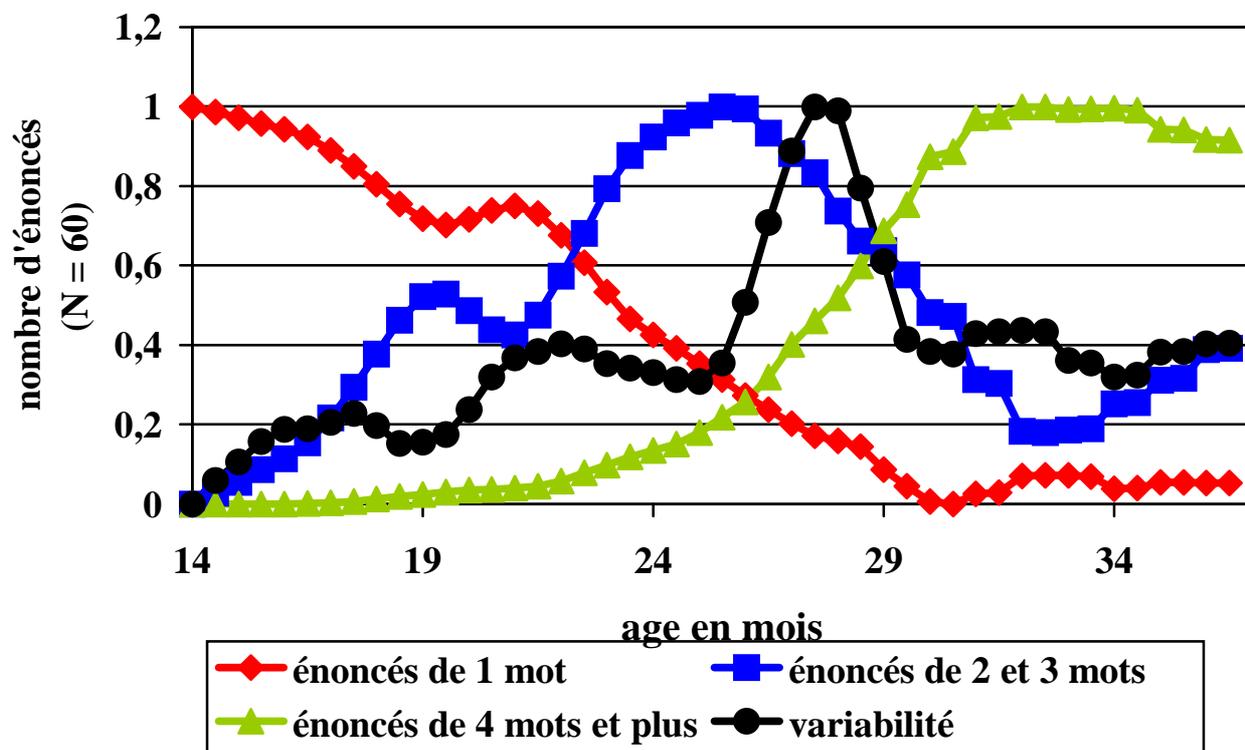


Figure 7.
 Courbes d'évolution pour la production des énoncés de 1 mot, 2 et 3 mots, 4 mots et plus et pour la variabilité (corpus de Pauline)