

Input et Acquisition du lexique en français : rôle de la fréquence et de la densité de voisinage

Sophie Kern, Christophe dos Santos

► **To cite this version:**

Sophie Kern, Christophe dos Santos. Input et Acquisition du lexique en français : rôle de la fréquence et de la densité de voisinage. Travaux de didactique du français langue étrangère, Université Paul Valéry, Montpellier III., 2011. halshs-02482633

HAL Id: halshs-02482633

<https://halshs.archives-ouvertes.fr/halshs-02482633>

Submitted on 18 Feb 2020

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.

Input et Acquisition du lexique en français : rôle de la fréquence et de la densité de voisinage

Sophie Kern, Laboratoire Dynamique du Langage, UMR5596 (CNRS-Lyon2)

sophie.kern@univ-lyon2.fr

Christophe dos Santos, Université François Rabelais de Tours; Unité

"Imagerie et Cerveau« Inserm UMR 930, CNRS ERL 3106; Tours

christophe.dossantos@univ-tours.fr

Introduction

Le vocabulaire a souvent été identifié comme une des composantes essentielles du langage (Hauser, Chomsky & Fitch, 2002; Pinker & Jackendoff, 2005) et de son acquisition : ce sont les mots qui font les langues, ou en d'autres termes : sans mot, pas de langue. Mieux comprendre le développement lexical est également important en raison du rôle que joue le lexique dans le développement linguistique ultérieur. D'une part, les développements grammatical et lexical précoce sont fortement corrélés : sans une masse critique de mots, les enfants ne sont pas capables de développer de bonnes compétences morphosyntaxiques (voir Bates & Goodman, 2001, pour une revue). D'autre part, les compétences lexicales sont corrélées à d'autres compétences linguistiques plus tardives, telles que l'écriture et la lecture. Par exemple, certaines études longitudinales ont montré que les performances syntaxiques et lexicales aux jardins d'enfants prédisent les performances en lecture (Catts, Fey, Zhang, & Tomblin, 1999; Muter, Hulme, Snowling, & Stevenson, 2004) et discriminent de façon fiable les bons lecteurs des mauvais lecteurs (Catts et al., 1999; Hagtvet, 2003; Share & Leikin, 2004).

Les adultes disposent d'environ 20000 à 50000 mots dans leur vocabulaire productif et encore bien davantage en compréhension. Le vocabulaire productif des jeunes enfants est quantitativement et qualitativement très différent de celui des adultes (Fenson, Dale, Reznick, Bates, Hartung, Pethick & Reilly, 1993). Entre 1 et 3 ans, le vocabulaire des enfants est limité

et les premiers mots peuvent être assez éloignés de la cible adulte ; éloigné phonologiquement, principalement par simplification (substitution ou élimination de sons ou de suites de sons, par exemple) et/ou éloigné sémantiquement, principalement par surgénéralisation (signification d'un mot généralisée à l'ensemble des membres de la même catégorie, par exemple). Avec l'avancée en âge, la taille du vocabulaire va augmenter, la production des mots et leur utilisation seront plus proche des standards adultes.

De nombreuses différences dans la trajectoire développementale du vocabulaire ont été notées, que ce soit en comparant des enfants monolingues du même âge apprenant des langues différentes ou des enfants monolingues du même âge apprenant la même langue. Néanmoins, les recherches s'accordent sur un certain nombre de similitudes concernant cette trajectoire développementale, notamment sur la production de mot d'une part, ainsi que sur la composition grammaticale et sémantique des premiers mots d'autre part. On observe en effet un développement non linéaire du vocabulaire productif. Les premiers mots sont produits aux alentours de 12 mois (Fenson et al., 1993 ; Kern & Gayraud, 2010). Une fois les premiers mots produits, on assiste à une augmentation très lente du vocabulaire avec un apprentissage de quelques mots par semaine. Dès que l'enfant dispose d'un vocabulaire composé d'environ 50 mots différents, il s'ensuit une période d'explosion lexicale consistant en une accélération du rythme d'acquisition. Même s'il existe un petit nombre de contre-exemples décrits dans la littérature, le vocabulaire de la plupart des enfants connaît une telle trajectoire développementale.

Pour ce qui est de la composition grammaticale, le vocabulaire passe par plusieurs étapes de restructuration entre 8 et 30 mois (Berglund & Eriksson, 1994 ; Caselli, Bates, Casadio, Fenson, Fenson, Sanderl, 1995 ; Jackson-Maldonado et al., 1993 ; Kern & Gayraud, 2010 ; Kauschke & Hofmeister, 2002 ; Maital, Dromi, Sagi, & Bornstein, 2000). Jusqu'à une taille de lexique d'environ 50 mots le vocabulaire des enfants est principalement composé de noms tandis que les prédicats (verbes et adjectifs) sont très minoritaires (5 % du vocabulaire). Quand le vocabulaire atteint 100 mots, les prédicats augmentent, signe de l'émergence de la capacité des enfants à encoder des significations relationnelles. Une autre catégorie très peu représentée dans le premier vocabulaire est la classe des mots grammaticaux qui ne représente pas plus de 5 % du vocabulaire total, et leur utilisation n'augmente que lorsque le vocabulaire productif de l'enfant atteint entre 300 et 500 mots. Longtemps proposées comme universelles, ces tendances sont aujourd'hui remises en question par l'étude de langues telles que le coréen par exemple. Il a été ainsi observée, pour ces quelques langues, la présence précoce (dès 14 mois) de verbes et de termes relationnels dans le répertoire du jeune enfant (Gopnik & Choi,

1995 ; Gopnik 1982 ; Gopnik & Melzhoff, 1985 ; Tomasello, 1992) ou une différence moins conséquente entre le nombre de noms et de verbes (Kern & Turkay, 2010) au début des acquisitions.

Il existe un nombre important de facteurs qui ont été envisagés comme facteurs explicatifs des caractéristiques quantitatives et qualitatives particulières des premiers vocabulaires des enfants (Bornstein, Cote, Maital, Painter, Park, Pascual, Pêcheux, Ruel, Venuti & Vyt, 2004). Parmi ces facteurs, les données langagières auxquelles sont exposés les enfants (appelés ci-après input) ont souvent été considérées.

L'objectif de ce travail est d'étudier le rôle de l'input et plus particulièrement de deux caractéristiques de cet input, à savoir la fréquence et la densité de voisinage sur le développement lexical chez l'enfant français au cours des 3 premières années de vie de l'enfant. Les questions auxquelles nous tentons de répondre sont : la fréquence des mots et leur densité de voisinage jouent-elles un rôle sur le développement lexical en français ? Ces facteurs jouent-ils le même rôle selon l'âge des enfants ou selon la taille du lexique des enfants ?

Arrière-Plan théorique

Fréquence, densité de voisinage et acquisition du vocabulaire

La plupart des études sur le développement lexical semble postuler un rôle important de la fréquence de l'input sur ce développement. Le plus souvent un mot est entendu dans une situation interactive, le plus tôt ce mot est acquis (Tomasello, 2003; Patterson, 2002; entre autres). On démontre par exemple une corrélation entre la taille du lexique des enfants et la quantité d'input entendu. Il a été établi que les enfants qui ont des parents qui leur parlent fréquemment ont une taille de lexique supérieure aux enfants qui ont des parents moins loquaces (Hart & Risley, 1995 ; Weizman & Snow, 2001). Par ailleurs, les mots nouveaux s'ils sont présentés fréquemment sont appris plus facilement, et ce, chez des enfants tout venants comme chez des enfants dysphasiques (Rice, Oetting, Marquis, Bode & Pae, 1994). Enfin, la fréquence d'exposition semble influencer la catégorie d'apprentissage. Les travaux de Choi & Gopnik (1995) et Gopnik & Choi (1990) ont montré que les enfants apprenaient une proportion plus importante de verbes que de noms lorsque le langage adressé à l'enfant comportait lui aussi une proportion plus grande de verbes que de noms.

Des études récentes ont remis en cause l'influence de la fréquence sur le développement en avançant l'hypothèse selon laquelle les arguments proposés pour étayer

l'effet de fréquence reposait principalement sur des preuves indirectes. Selon Goodman, Dale et Li (2008), par exemple, la plupart des études ont essayé d'évaluer l'effet de fréquence sur l'acquisition des catégories syntaxico-sémantiques, ces mêmes études sont donc dans l'incapacité d'expliquer pourquoi certains mots à l'intérieur d'une même catégorie étaient appris plus précocement que d'autres. Ainsi, la majorité des études, n'ont pas comparé la fréquence de mots individuels produit par la mère avec l'acquisition de ces mêmes mots produits pas l'enfant. Enfin, certaines rares études qui ont étudié la fréquence des mots individuels, ont comparé les productions des enfants avec des bases de données de fréquence adulte basées sur du discours écrit adressé à l'adulte (Kucera & Francis, 1967) plutôt qu'avec du langage oral adressé à l'enfant.

Goodman, Dale & Li (2008) ont réussi à montrer que la fréquence d'occurrences d'un mot dans l'input joue un rôle dans le développement lexical mais également qu'il était important de prendre en compte la nature grammaticale des mots étudiés ainsi que l'âge des enfants. Les auteurs ont mis en lien l'âge d'acquisition (compréhension et production) de 562 mots chez des enfants anglophones de 8 à 30 mois estimé à partir du MacArthur-Bates CDI (Fenson et al., 1993) et la fréquence de ces mots dans l'input, fréquence calculée à partir de 28 corpus de langage adressé à l'enfant disponible sur le site CHILDES (<http://childes.psy.cmu.edu/>). Les données révèlent plusieurs faits qui viennent contredire dans certains cas des conclusions avancées dans d'autres recherches ou qui viennent les compléter. Tout d'abord, les auteurs montrent que si l'on considère tous les mots ensemble, il existe une corrélation négative significative entre l'ordre d'acquisition des mots et leur fréquence : une haute fréquence est associée à une acquisition tardive. Ce résultat est mis en lien avec l'apparition tardive des mots fonctionnels qui sont peu diversifiés et très fréquemment utilisés. Par contre, à l'intérieur des catégories lexicales, une fréquence haute est liée à une acquisition précoce. Par ailleurs, la fréquence a un effet plus important sur le vocabulaire productif que sur le réceptif : pour la compréhension la fréquence parentale est corrélée avec l'âge d'acquisition seulement pour les noms communs. Enfin, les auteurs ont noté un effet de la fréquence différent selon la taille du vocabulaire en comparant le lien entre fréquence et les 100 premiers mots acquis et entre fréquence et les autres mots acquis plus tardivement. Les corrélations montrent des différences entre les deux groupes pour les noms et les items de classe fermée. Pour les noms, la corrélation est plus forte après 100 mots, alors que pour les mots grammaticaux, la tendance est inversée.

Un autre facteur à prendre en considération est la densité de voisinage, c'est-à-dire le nombre de voisins phonologiques que possède le mot à acquérir. Deux mots sont considérés comme des voisins phonologiques lorsqu'ils sont composés des mêmes phonèmes à l'exception d'un seul (Charles-Luce & Luce, 1990). En prenant comme exemple le mot « balle » /bal/, la différence peut être causée par 1. La substitution d'un phonème à la même position (substitution simple sans métathèse) /bal/ vs /baʁ/. 2. L'ajout d'un phonème /bal/ vs /balɛ/ ou 3. L'effacement d'un phonème /bal/ vs /ba/. Le mot « balle » fait partie d'un voisinage plus dense que le mot « baffle » puisque « balle » dispose de 42 voisins phonologique contre seulement 10 pour le mot « baffle ». Les mots d'une langue possèdent ainsi plus ou moins de voisins phonologiques. Les mots qui possèdent beaucoup de voisins phonologiques appartiennent à un voisinage dense alors que ceux qui en possèdent peu appartiennent à un voisinage épars.

Un certain nombre de travaux ont montré que la densité de voisinage influençait le développement lexical de l'enfant : les mots qui possèdent de nombreux voisins phonologiques sont appris plus précocement que les autres, l'enfant possédant alors un vocabulaire restreint. Storkel (2004 et 2009) par exemple a étudié la relation entre la densité de voisinage et l'âge d'acquisition des mots en utilisant la production d'enfants de 8 à 30 mois. Les données révèlent que le vocabulaire des enfants très jeunes est composé de plus de mots qui possèdent beaucoup de voisins phonologiques que de mots en possédant peu. De plus, la densité de voisinage jouerait un rôle moins important avec l'avancée en âge des enfants (Newman & German, 2005). Selon Storkel (2009), son rôle augmenterait entre 16 et 20 mois avant de diminuer à nouveau entre 20 et 30 mois.

Plus rares sont les travaux à avoir à la fois prise en compte les effets de fréquence et les effets de densité de voisinage sur l'acquisition du langage chez les mêmes enfants alors que la corrélation positive entre les deux facteurs est avérée (Landauer & Steeter, 1973). Storkel (2004) montre une interaction entre la fréquence des mots et la densité de voisinage dans le vocabulaire d'enfants anglophones. Plus précisément, un effet de la densité a été observé dans les mots de basse fréquence mais pas de haute fréquence, suggérant qu'une haute fréquence pouvait cacher l'effet de densité sur le développement du vocabulaire expressif. Storkel (2009) prolonge les analyses en avançant l'hypothèse que la fréquence des mots est une variable composite puisqu'elle concernerait trois différents domaines (domaines phonologiques, lexicales et sémantiques). Pour vérifier cela, et afin de prédire l'âge d'acquisition des mots, Storkel utilise, entre autres, deux prédicteurs phonologiques (pour chaque mot, moyenne des fréquences des segments en fonction de leur position dans le mot et

moyenne des fréquences des diphtongues en fonction de leur position dans le mot) et deux prédictors lexicaux (densité de voisinage et longueur de mots). Elle conclut à une influence des variables phonologiques sur l'ensemble de la période étudiée, à savoir de 16 à 30 mois et à une influence des variables lexicales uniquement durant la période d'acquisition des premiers mots (1;4 to 1;8).

Stokes en 2010 a également mesuré la part de chacun des facteurs dans l'acquisition des mots sur une population de 222 enfants britanniques de 27 mois (+/- 3 mois). Les données ont été collectées à partir de l'adaptation britannique du MacArthur CDI développée par Klee & Harrison (2001). Stokes montre que la densité de voisinage et la fréquence d'occurrence sont responsables de 61% de la variance présente dans la taille de vocabulaire avec respectivement 47% et 14%). Mais l'influence des deux facteurs est également différente en fonction de la taille du vocabulaire des enfants révélant par la même des stratégies d'acquisition différentes à travers les âges : le vocabulaire d'enfants avec une petite taille de vocabulaire est composé de mots de haute densité et de basse fréquence. Enfin, Stokes, Kern & dos Santos (sous presse) arrivent à des conclusions similaires sur une population française comparable de 208 enfants de 24 à 30 mois.

Méthodologie

Participants

Un groupe de 759 enfants francophones monolingues âgés de 8 à 30 mois ont participé à l'étude. Aucun de ces enfants ne présente de problèmes ORL, n'est prématuré ou issue de naissances multiples. Il s'agit d'un sous-échantillon de la population d'étalonnage des Inventaires Français du Développement Communicatif (Kern & Gayraud, 2010), adaptation française des MacArthur-Bates Communicative Development Inventories (Fenson et al., 1993).

L'étude porte donc sur deux groupes d'enfants : 551 enfants de 8 à 16 mois (Groupe 8-16) et 208 enfants de 24 à 30 mois (Groupe 24-30), répartis par tranches d'âge de 1 mois (Tableau 1 et Tableau 2).

Age	8	9	10	11	12	13	14	15	16	Total
Nbre	38	78	48	55	103	60	51	72	46	551

Tableau 1 : Groupe 8-16 : Age en mois et nombre

Age	24	25	26	27	28	29	30	Total
Nbre	26	38	29	23	32	33	27	208

Tableau 2: Groupe 24-30 : Age en mois et nombre

Collecte des données

Les compte-rendus parentaux IFDC « Mots et gestes » et « IFDC Mots et phrases » utilisés pour collecter les données se présentent sous la forme d'un questionnaire administré aux mères pour évaluer le développement linguistique de leur enfant. La version américaine sur laquelle nous nous sommes basées pour construire l'adaptation française a été étalonnée sur des enfants anglophones aux Etats-Unis dans les années 90 et a fait l'objet depuis lors d'une quarantaine d'adaptations de par le monde (voir le site <http://www.sci.sdsu.edu/cdi/> pour une liste exhaustive). La validité de l'instrument a été étudiée avec beaucoup d'attention et on lui attribue de multiples qualités métrologiques telles que la cohérence interne, la cohérence externe, la fidélité test-retest, la fidélité convergente, la fidélité concurrente et une certaine validité prédictive (Berglund & Eriksson, 1998 ; Camaioni et al., 1991 ; Fenson et al., 1993 ; Dale, 1991)

Les IFDC ont été distribués aux parents par le biais de pédiatres membres de l'Association Française de Pédiatrie Ambulatoire (AFPA). Les parents du Groupe 8-16 ont complété le questionnaire « mots & gestes » destiné aux enfants de 8 à 16 mois et les parents du Groupe 24-30 ont complété le questionnaire « mots & phrases » destiné aux enfants de 16 à 30 mois. L'IFDC « Mots & Gestes » vise l'évaluation de la production de gestes communicatifs (62 items) et de la production et de la compréhension de mots (414 items). Ces mots sont répartis en 18 catégories sémantiques et appartiennent à 4 grandes catégories grammaticales (Noms, Prédicats, Mots Fonctionnels et Autres (Onomatopées, jeux et routines)). L'IFDC « Mots & Phrases » se compose de deux parties principales. La première a pour but d'évaluer la taille du lexique productif de l'enfant. Elle se compose d'une liste de 690 mots répartis en 22 catégories sémantiques (telles que: cris d'animaux et sons, véhicules, objets d'extérieur ou encore mots d'action). La deuxième partie évalue le développement morphologique et syntaxique par le biais de questions à choix multiples. On y trouve 25 items répartis en 3 sections. La première section examine l'utilisation de morphèmes réguliers et irréguliers

(déterminants, pronoms, prépositions, génitif, structure interrogative, accord sujet/verbe, négation et utilisation d'auxiliaires). La deuxième section interroge l'utilisation des temps verbaux (infinitif, présent, passé composé, imparfait, futur et impératif). Enfin, la troisième section permet le calcul de la longueur maximale des énoncés à partir des 3 énoncés les plus longs produits par l'enfant et rapportés par les mères. Les deux questionnaires comportent également une série de questions sur l'enfant et son environnement (sexe, rang dans la fratrie, scolarisation, mode de garde, niveau d'éducation des parents).

Traitement des données et codage

Réduction des données

Suivant la méthodologie de Stokes (2010) et Stokes, dos Santos, Kern (sous presse), nous avons retenus pour le codage que les mots de contenus (noms, verbes, adjectifs et adverbes) monosyllabiques. Les mots retenus devaient comprendre au moins une voyelle et pouvaient comprendre également une ou plusieurs consonnes.

Ont également été considéré comme mot monosyllabique, les mots comportant deux voyelles dont un schwa en syllabe non accentuée qui n'est précédé que par une consonne. Ainsi, le mot « petit » [pəti] a été considéré comme un mot monosyllabique dont la prononciation est [pti]. Par contre, le mot « grenouille » [gʁənuij] a été considéré comme un mot dissyllabique car malgré le fait qu'il possède un schwa en position non accentuée, il est précédé par deux consonnes ([gʁ]). Enfin, pour les verbes, l'unique forme monosyllabique si elle existe a été pris en compte, si plusieurs formes monosyllabiques existent alors la plus fréquente a été pris en compte. Ainsi le verbe « marcher » ne présente qu'une forme phonologique monosyllabique : /maʁʃ/. Par contre le verbe « prendre » possède plusieurs formes phonologiques monosyllabiques. La forme /pʁɑ̃/ a été sélectionnée car c'est la plus fréquente.

Au final, une liste de 220 mots composés de 131 noms, 30 adjectifs, 3 adverbes et 56 verbes a été retenue.

Fréquence

La fréquence d'occurrence de chaque mot a été déterminée par le biais de la base de donnée Lexique3 (New, Brysbaert, Veronis, & Pallier, 2007). Cette base de données comprend plus de 50 millions de mots. La fréquence de chaque mot est donnée pour chaque catégorie à laquelle ce mot appartient. Par exemple, le mot « manger » appartient à plusieurs catégories grammaticales en fonction du contexte d'énonciation : il fait partie des verbes dans une phrase telle que « elle a fini par manger du chocolat » ou des noms dans « elle a apporté

son manger ». Dans la mesure où le mot « manger » fait partie des mots d'action dans l'IFDC, seule la fréquence du verbe « manger » a été considérée. Pour les noms, les fréquences des formes homophoniques du singulier et du pluriel sont ajoutées. Par exemple la fréquence de « lion » est ajoutée à celle de « lions » car les deux mots sont prononcés /ljɔ̃/. Chaque mot du vocabulaire productif des enfants a été codé selon sa densité de voisinage et sa fréquence d'occurrences.

Densité de voisinage

La densité de voisinage de chaque mot a été calculée à partir de la forme phonologique la plus fréquente. Par exemple, pour le verbe « chanter », la densité de voisinage a été calculée à partir de la forme phonologique /ʃɑ̃t/ seulement.

Résultats

La fréquence d'occurrence des mots et leur densité de voisinage ont été considérés comme prédicteurs possibles de la taille du vocabulaire dans les deux groupes d'enfants français : chez les 8-16 mois, les deux facteurs ont été considérés comme prédicteurs possibles du vocabulaire réceptif et productif, chez les 24-30 mois, comme prédicteurs du vocabulaire productif seulement du fait de l'absence dans l'IFDC 24-30 du volet vocabulaire réceptif.

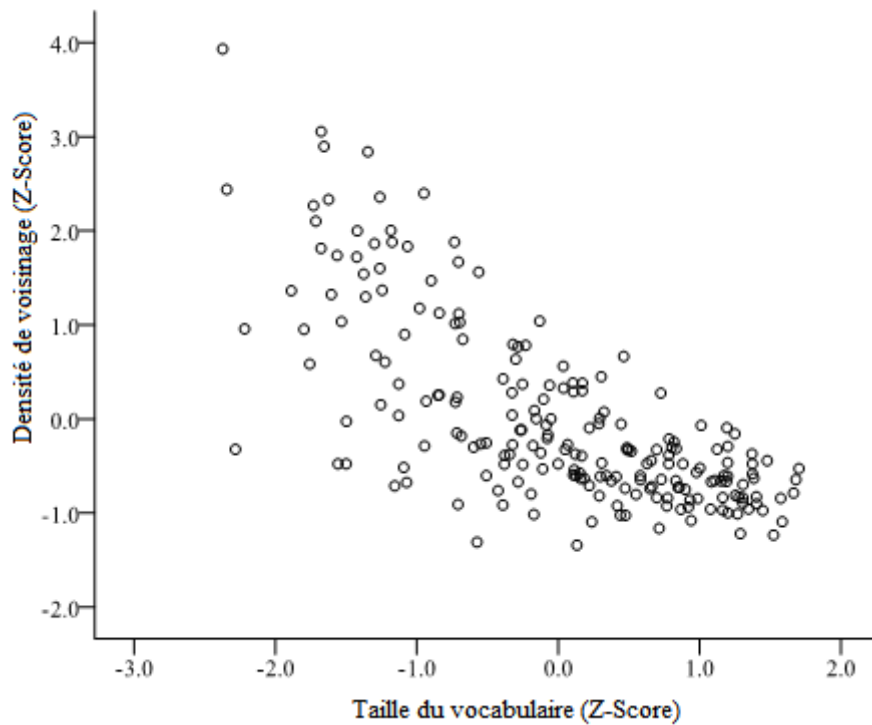
Les analyses de corrélations que nous avons menées ont révélées des corrélations positives entre l'âge et la taille du lexique. Ces corrélations nous ont conduits à transformer les scores en Z-scores. La normalisation des scores a été utilisée afin de centraliser l'asymétrie de la distribution présente dans les données et d'ajuster l'étendue relative entre variables afin de pouvoir comparer les variables et les enfants.

Après cette transformation, nous avons appliqué des régressions multiples à nos données, avec pour variable de résultat la taille du lexique réceptif (8-16 mois), la taille du lexique productif à 8-16 mois et à 16-30. Les résultats n'ont pas montré de lien entre la taille du lexique réceptif ou productif à 8-16 mois et les deux prédicteurs : fréquence et densité de voisinage.

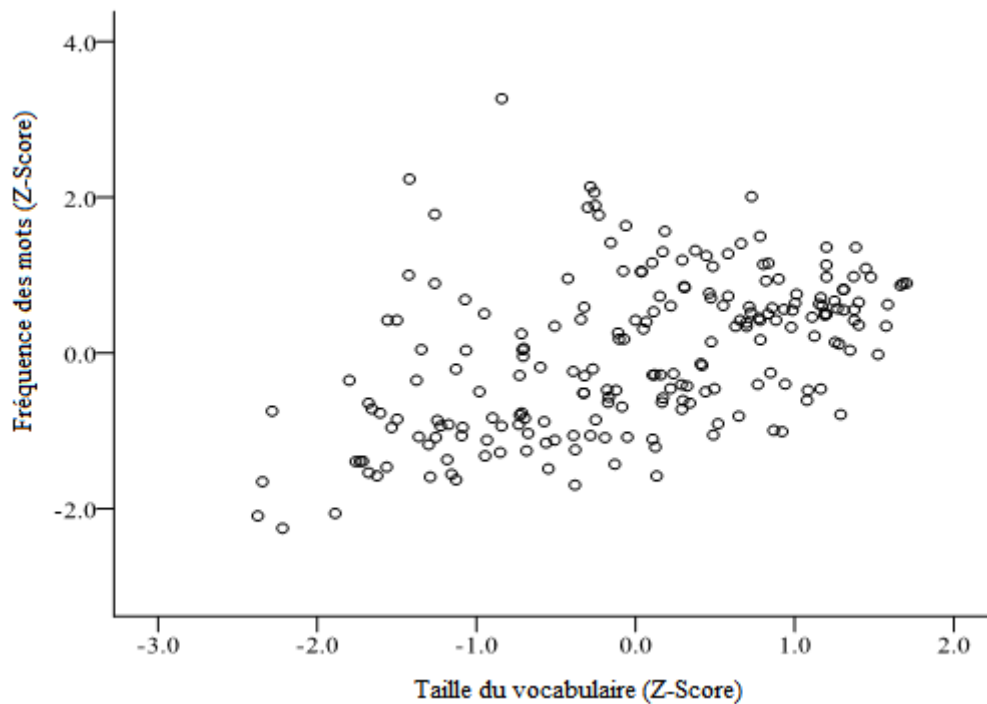
Par contre, la régression multiple avec taille du lexique productif de 24 à 30 mois comme variable de résultat et densité de voisinage et fréquence comme prédicteurs, présente un modèle linéaire significatif ($p < 0.001$) avec fréquence et densité et considérées ensemble comme responsables de 62% de la variance dans la taille du vocabulaire productif. Les

corrélations partielles font émerger une corrélation significative, haute et négative ($r=-.71$) pour la densité de voisinage (graphique 1) et une corrélation significative, moyenne et positive pour la fréquence ($r=.48$) (graphique 2).

Enfin nous avons réalisé une régression multiple hiérarchisée sur les données qui montrent que la densité de voisinage est responsable de 53% de variance unique dans la taille des vocabulaires ($p<0.001$) alors que la fréquence est responsable de 9% de variance additionnelle unique ($p<0.001$).

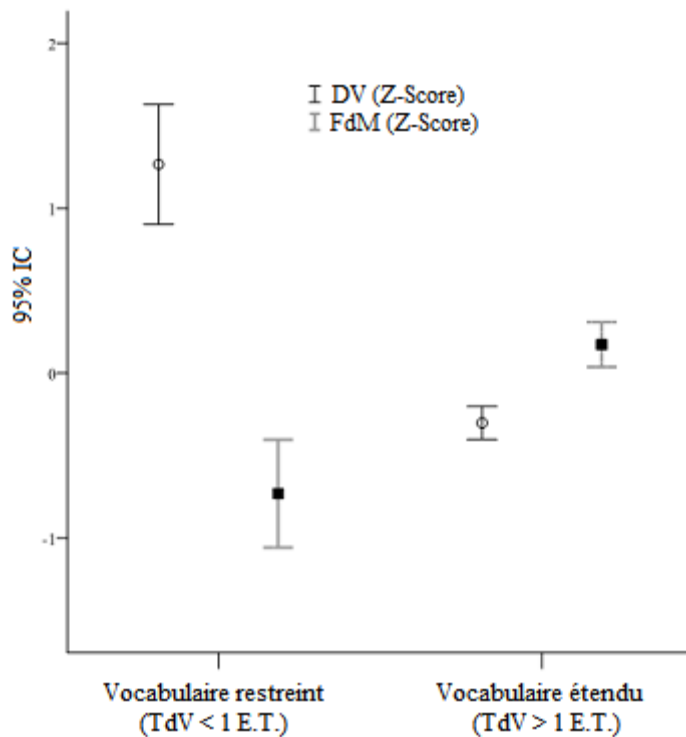


Graphique 1 : Corrélation entre densité de voisinage et taille du vocabulaire



Graphique 2 : Corrélation entre fréquence et taille du vocabulaire

Dans un second temps, nous avons comparé dans notre population de 24 à 30 mois la constitution des vocabulaires des enfants qui possèdent un vocabulaire réduit à celle des enfants qui en possèdent un large. Deux groupes d'enfants ont été constitués : un premier groupe composé de 168 enfants dont la taille du vocabulaire est au-dessus d'un écart type ($TdV > 1$ E.T.) et un second groupe de 40 enfants dont la taille du lexique est en-dessous d'un écart-type ($TdV < 1$ E.T.). L'utilisation du test non paramétrique de Mann-Whitney nous a permis de montrer que les enfants dont le vocabulaire est en dessous d'1 E.T. ont un score en densité de voisinage (DM) supérieur aux autres avec néanmoins des variabilités importantes au sein du groupe. Pour la fréquence des mots (FdM), une différence significative a également été montrée au sein des deux groupes : les enfants dont le vocabulaire est en dessous d'un E.T. ont un score en fréquence de mots inférieur aux autres (Graphique 3).



Graphique 3 : Densité de voisinage et fréquence en fonction de la taille du vocabulaire

Discussion, conclusion, perspectives

Nous ne trouvons aucune relation entre la taille du vocabulaire réceptif et productif, la densité de voisinage des mots et leur fréquence d'occurrences dans la base Lexique 3 pour le groupe des enfants les plus jeunes, à savoir entre 8 et 16 mois. Cette absence de relation peut être expliquée par le fait que les enfants à cet âge possèdent peu de mots, en particulier dans leur lexique productif. Dans le groupe des enfants de plus de 24 mois (24 à 30 mois), on observe une relation entre la taille du vocabulaire productif, la densité de voisinage et la fréquence d'occurrences : la densité de voisinage prédit fortement (53%) leur taille du lexique productif ; la fréquence quant à elle prédit également mais bien moins fortement (9%) la taille du lexique productif. Ces résultats vont dans le sens des travaux de Storkel (2009) et Stokes (2010) chez des enfants anglophones. Nos résultats soulignent également un effet plus important de la densité de voisinage que de la fréquence sur la taille du lexique. Si l'on compare les enfants anglophones aux enfants francophones, les premiers sont plus affectés par le critère de fréquence et moins par le critère de densité de voisinage que les seconds.

Nous ne sommes malheureusement pas en mesure de savoir si ces deux facteurs jouent également un rôle sur la taille du lexique réceptif dans la mesure où la taille de ce lexique n'est pas mesurée dans l'IFDC 'mots & phrases'.

Nos analyses ont également révélées que les vocabulaires les plus limitées sont davantage composés de mots avec beaucoup de voisins phonologiques et peu fréquent alors que pour les vocabulaires plus importants, la tendance est inversée. Là aussi nos résultats vont dans le sens des travaux ultérieurs, en particulier de Stokes (2010).

Plusieurs prolongements à ce travail sont prévus. Tout d'abord, nous allons réaliser le même type d'analyse sur l'ensemble des mots présents dans l'IFDC, et ce, quel que soit le nombre de syllabes qui les compose ou leur appartenance grammaticale. En effet, compte-tenu de l'importance des noms dissyllabiques dans la langue française (un tiers des mots présent dans la base de données Lexique3 sont des mots dissyllabiques, et 21% des mots que nous entendons le sont également d'après la même base), il est important de les intégrer dans les analyses afin de vérifier si les relations trouvées dans ce travail peuvent être généralisées à l'ensemble du lexique. Cette intégration est également importante compte-tenu de la corrélation négative qu'il existe entre la longueur des mots et la densité (voir par exemple Bard & Shillcock, 1993 ; Storkel 2004a) et la corrélation positive entre longueur et âge d'acquisition (Maekawa & Storkel, 2006). La même explication est valable pour l'intégration des mots grammaticaux dans l'analyse. L'intégration des mots grammaticaux est d'autant plus importante que les travaux de Goodman et collaborateurs ont montré un rôle différent de la fréquence sur l'acquisition du lexique en fonction de la nature grammaticale des mots appris.

Un troisième prolongement va consister à suivre la trajectoire développementale des enfants entre 8 et 30 mois. Pour le moment, seuls deux groupes entre 8 et 16 et 24 et 30 mois ont été étudiés. L'intégration d'enfants entre 16 et 24 mois va nous permettre de mieux situer à partir de quel moment la densité de voisinage, la fréquence et la taille du lexique sont liées.

À plus long terme, nous allons observer d'autres facteurs qui jouent un rôle sur le développement du lexique chez l'enfant. Stoel-Gammon soulignait récemment (2010) que les études sur l'acquisition du langage se focalisaient trop souvent sur un des sous-domaines (phonologie, morphologie ou syntaxe) sans prendre en compte les interactions entre domaines. On trouve un exemple clair de cette absence de considération des interactions entre domaine dans la relation entre développement phonologique et lexical. Même si le lien entre production et input sont intuitivement évident, le développement de la production et les effets de l'input ont pour le moment été majoritairement étudiés de manière disjointe dans l'acquisition de la langue maternelle. Un prolongement possible à ce travail sera de considérer ces deux facettes du développement de manière intégrative. Le développement lexical sera considéré comme émergent de l'interaction entre processus de perception et processus de

production, et comme le produit de l'interaction entre facteurs maturationnels (biologique) et facteurs d'apprentissage (environnement) (Hollich, Hirsh-Pasek, Golinkoff, 2000).

Références

- Bard, E. & R. Shillcock (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing* (pp. 235-275). Hove, U.K.: Erlbaum.
- Berglund, E. & M. Eriksson (1998). Communicative development in Swedish children 16-28 months old: The Swedish Early Communicative development Inventory--words and sentences. *Reports from the Department of Psychology*.
- Berglund E. & M. Eriksson (1994, 18 au 20 novembre). *Parental reports of early language skills. Presentation of inventories and a cross-linguistic comparison*. Communication présentée au Fente Nordiske symposium om barnesprak, Lysebu, Oslo.
- Bornstein, M., Cote, L., Maital, S., Painter, K., Park, Sung-Yun, Pascual, L., Pêcheux, MG., Ruel, J., Venuti, P. & A. Vyt. (2004). Cross-linguistic Analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75 (4), 1115-1139.
- Camaioni, L., Castelli, M., Longobardi, E. & V. Volterra (1991). A parent report instrument for early language assessment. *First language*. 11: 345-359.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159-199.
- Catts, H.W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, 3, 331-361.
- Charles-Luce, J. & P. Luce (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727-735.
- Gopnik A. & S. Choi (1995). Names, relational words and cognitive development in English and Korean Speakers: Nouns are not always learned before verbs. In M. Tomasello & W. Merriman (Eds.) *Beyond names for things: Young children's acquisition of verbs*. New Jersey: Erlbaum.
- Coady, J. A., & R.N. Aslin (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30, 441-469.

- Dale, P. (1991). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language* 16, 239-235.
- Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Hartung, J.P., Pethick, S., & Reilly, J.S. (1993). *MacArthur Communicative Development Inventories: user's guide and technical manual*. San Diego, CA: Singular Press.
- Goodman, J.C., Dale, P.S. and Li P. (2008) Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35 (3), 515–531.
- Gopnik, A. & S. Choi (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language*, 10, 199–215.
- Gopnik, A. & S. Choi (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (eds.), *Beyond Names for Things: Young Children's Acquisition of Verbs*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gopnik, A. & A.N. Meltzoff (1985). From people to plans to objects: Changes in the meaning of early words and their relation to cognitive development. *Journal of Pragmatics*, 9, 495-512.
- Gopnik, A. (1982). Words and plans: Early language and the development of intelligent action. *Journal of Child Language*, 9, 617-733.
- Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological skills. *Reading and Writing: An Interdisciplinary Journal*, 16, 505–539.
- Hart, B., & T. Risley (1995). Meaningful differences in the everyday experience of young American children. Baltimore: Brookes.
- Hauser M. D., Chomsky N., Fitch W. T (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An ECM for the origins of word learning. Society for Research in Child Development Monograph Series. Chicago: University of Chicago Press.
- Hollich, G., Jusczyk, P. W. & Luce, P. A. (2002). Lexical neighborhood effects in 17-monthold word learning. In B. Skarabela, S. Fish & A. H.-J. Do (eds), *Proceedings of the 26th annual Boston University Conference on Language Development*, 1, 314–23. Somerville, MA: Cascadilla.

- Jackson-Maldonado, D., Thal, D., Marchman, V. A., Bates, E. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, 20(3), 523-549
- Kauschke, C. & C. Hofmeister (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language*, 29, 735-757
- . & F. Gayraud (2010). « *L'IFDC* », Editions La Cigale, Grenoble.
- Kern, S. & F. Türkay (2010). Verb or noun dominance in caregiver speech ? A crosslinguistic and crosscontextual answer. 6th International Conference on Language Acquisition. 8-10 septembre, Barcelone, Espagne. (poster).
- Klee, T. & C. Harrison (2001). *CDI Words and Sentences validity and preliminary norms for British English*. Paper presented at Child Language Seminar, University of Hertfordshire, England.
- Kucera, H. & W.N. Francis (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Landauer T.K. & L.A. Streeter (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, 12:119–13.
- Maekawa, J. & H.L. Storkel, (2006). Dynamic assessment and word learning. *ACQuiring Knowledge in Speech, Language, and Hearing*, 8, 103-105.
- Maital, S. L., Dromi, E., Sagi, A., & Bornstein, M. (2000). The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, 27, 43–67.
- Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40, 665–681.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661-677
- Newman, R. S. & D.J. German (2005). Lifespan effects of lexical factors on oral naming. *Language and Speech*, 48 (2), 123-156.
- Patterson, J. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics*, 23, 493–508.
- Pinker S. & R. Jackendoff (2005). The faculty of language: what's special about it? *Cognition*, 95, 201–236.

- Rice ML, Oetting JB, Marquis J, Bode J, Pae S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, 37:106–122.
- Share, D. L. & M. Leikin (2004). Language learning impairment at school entry and later reading disability: Connections at lexical versus supralephical levels of reading. *Scientific Studies of Reading*, 8, 87–110
- Stoel-Gammon, C. (2010). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38, issue 1, 1-34.
- Stokes, S. (2010). Neighborhood Density and Word Frequency Predict Vocabulary Size in Toddlers. *Journal of Speech, Language, and Hearing Research* Vol.53 670-683 June 2010.
- Stokes, S., Kern, S. & dos Santos, C. (2012). Extended Statistical Learning as an Account for slow vocabulary growth. *Journal of Child Language*, 39, 105-129
- Storkel, H.L. (2004). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research*, 47, 1454-1468.
- Storkel, H.L. (2004). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research*, 47, 1194 -1212.
- Storkel, H.L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201-221
- Storkel, H.L. (2008). First utterances. In G. Rickheit and H. Strohner (eds), *Handbook of communication competence* (pp. 125-147). Berlin: Mouton de Gruyter.
- Storkel, H. L. (2001). Learning new words, phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44, 1321-1337.
- Storkel, H.L. (2009). Developmental differences in the effects of phonological, lexical, and semantic variables on word learning by infants. *Journal of Child Language*, 36, 291-321.
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Weizman, Z.O. & C.E. Snow (2001). Lexical output as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology*, 37, 265–79.