

Université Côte d'Azur
ED 86 - SHAL

Variabilité et acquisition du français hexagonal :
Comment la syntaxe vient aux enfants

Katerina Palasis
Maître de conférences en Sciences du langage

Habilitation à Diriger des Recherches – CNU 07
Volume 2 : CV et sélection de publications

2024

Table des matières

I. Curriculum Vitae *(format CNU 07)*

1. CV synthétique (3 pages) 5
2. Descriptif rédigé (maximum 4 pages) 8
3. Liste exhaustive des publications et travaux 12

II. Sélection de publications

1. Palasis, K. (2010). Introducing New French Child Data: Thoughts on their Gathering and Coding. *Corpus, 9, La syntaxe de corpus*, 33-51,
<https://journals.openedition.org/corpus/1801> 19
2. Palasis, K. (2011). Français spontané et français normé : bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In O. Bertrand & I. Schaffner (Eds.), *Variétés, variations & formes du français* (pp. 55-72). Palaiseau: Les Éditions de l'École Polytechnique,
<https://www.researchgate.net/profile/Katerina-Palasis/publications> 41
3. Palasis, K. (2012). Economy Principles within Child Speech: When the Nominative Clitic does not Surface. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 261-283). Newcastle: Cambridge Scholars Publishing,
<https://www.researchgate.net/profile/Katerina-Palasis/publications> 59
4. Palasis, K. (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies, 23*(1), Numéro spécial *L'hypothèse d'une diglossie en France*, Massot, Benjamin & Paul Rowlett (Eds.), 17-35,
<https://doi.org/10.1017/S0959269512000348> 83
5. Palasis, K. (2014). Étude morphosyntaxique des pronoms sujets à partir d'un corpus de maternelle. In R. Mir-Samii (Ed.), *Du sujet et de son absence dans les langues* (pp. 259-270). Le Mans: Université du Maine,
<http://doi.org/10.13140/2.1.1586.9440> 103
6. Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua, 161*, 125-143,
<http://dx.doi.org/10.1016/j.lingua.2014.11.012> 115

7. Palasis, K., Faure, R. & Lavigne, F. (2019). Explaining variation in wh-position in child French: A statistical analysis of new seminaturalistic data. *Language Acquisition*, 26(2), 210-234, <https://doi.org/10.1080/10489223.2018.1513004>. 135
8. Faure, R. & Palasis, K. (2021). Exclusivity! Wh-fronting is not optional wh-movement in Colloquial French. *Natural Language & Linguistic Theory*, 39(1), 57-95, <https://doi.org/10.1007/s11049-020-09476-w> 161
9. Palasis, K., Faure, R. & Meunier, F. (2023). Wh-in-situ in child French: Deictic triggers at the syntax-semantics interface. *Journal of French Language Studies*, 1-26, <https://doi.org/10.1017/S0959269523000030> 201

I. Curriculum Vitae

1. CV synthétique (3 pages)

Diplômes

- 2015 et 2010 : **Qualification** aux fonctions de Maître de Conférences pour la **Section 07** (N° 10- et 15-207207898)
- 2009 : **Doctorat de Sciences du langage** (mention très honorable avec les félicitations du jury à l'unanimité ; thèse publiée en 2010)
- 2005 : Master de Sciences de l'homme et de la société, à finalité Recherche, Mention Sciences du langage (mention très bien)
- 2004 : Maîtrise de Sciences du langage (mention très bien)
- 2003 : Licence de Sciences du langage (mention bien)
Licence de Sciences du langage, Mention Français Langue Étrangère (mention bien)
- 2002 : DEUG Lettres et langues, Mention Sciences du langage (mention très bien)
- 1987 : BTS Secrétariat trilingue (français, anglais, allemand)
- 1985 : Baccalauréat A2 Philosophie, lettres et langues étrangères (allemand, espagnol, russe)

Parcours professionnel

- 01/09/2015- : **Maître de conférences en Sciences du langage** à Université Côte d'Azur, rattachée au Laboratoire BCL, UMR 7320
- 02/2015-06/2015 : **Chargée de cours** à l'Université Nice Sophia Antipolis
- 02/2012-01/2015 : **Chercheure contractuelle** à temps plein dans le cadre du projet ANR-DFG, ANR 11 FRAL 007 01, « Données et analyses dialectologiques, acquisitionnelles et diachroniques des pronoms sujets en Gallo-Roman » (DADDIPRO), sous la responsabilité scientifique de Michèle Oliviéri (Université Nice Sophia Antipolis, CNRS, BCL UMR 7320) et Georg Kaiser (Université de Konstanz)
Chargée de cours à l'Université Nice Sophia Antipolis
- 10/2011-12/2011 : **Chargée de cours** à l'Université de Konstanz (Allemagne)
- 08/2011-10/2011 : **Coordinatrice technique** pour la rédaction du Projet LABEX 2011 CoSySoLife « Systèmes Cognitifs et Formes de Vie Sociale » sous la responsabilité scientifique de Tobias Scheer, BCL, UMR 7320 (ex-6039)
- 09/2010-07/2011 : **Chargée de cours** à l'Université de Konstanz (Allemagne, 28 heures)
Chargée de cours à l'Université Nice Sophia Antipolis (18 heures)
Post-doctorante interne à BCL, UMR 6039, Université Nice Sophia Antipolis
Membre du partenariat de recherche PROCOPE (2010-2012, N° 22016ZL) entre la France (BCL, UMR 6039) et l'Allemagne (Université de Konstanz) sur la micro-variation syntaxique à travers l'étude des pronoms clitiques dans différents systèmes linguistiques romans (notamment le français oral, l'ancien français, les dialectes occitans, l'espagnol dominicain, le portugais brésilien et le galicien)

- 09/2008-08/2010 : **Attachée Temporaire d'Enseignement et de Recherche** à temps plein à la Faculté des Lettres de l'Université Nice Sophia Antipolis
- 10/2006-06/2008 : **Tutrice** pour le département des Sciences du langage de l'Université Nice Sophia Antipolis : encadrement pédagogique en L1 et soutien pour toutes les années
- 2006-2011 : **Vacataire au CNRS** : relectures et traductions en anglais de communications et d'articles pour le Laboratoire BCL, UMR 6039
- 1995-2006 : **Soutien scolaire** en français, français langue étrangère et anglais pour enfants du CP à la Terminale et cours pour adultes
- 03/1995-04/1995 : **Rédactrice bilingue** (français, anglais) pour Fraser Yachts Monaco (vente, location, construction et gestion de grands yachts) (CDD)
- 10/1987-01/1995 : **Secrétaire et rédactrice bilingue** pour Associated Yacht Brokers Monaco (vente, location, construction et gestion de grands yachts) (CDI, licenciement économique)

Administration et responsabilités collectives

- 01/2023- : **Responsable de l'Équipe de Recherche 'Acquisition, apprentissage, adaptation'** (AAA) du Laboratoire Bases, Corpus, Langage (BCL, UMR 7320 CNRS), <https://bcl.cnrs.fr/rubrique461?lang=fr>
- 10/2017- : **Responsable pédagogique de la Mention/Licence 3 Sciences du langage**, Université Côte d'Azur, <https://univ-cotedazur.fr/formation/offre-de-formation/licence-sciences-du-langage>
- 10/2017-09/2022 : **Responsable pédagogique de la Licence 1 et Licence 2 Sciences du langage**, Université Côte d'Azur
- 10/2017-11/2020 : **Directrice du Département de Sciences du langage**, Université Côte d'Azur
- 10/2017-11/2020 : **Co-responsable du Master Linguistique**, traitements informatiques du texte et processus cognitifs, Université Côte d'Azur

Publications et travaux

Articles		21
Revue internationale à comité de lecture	8	
Communications publiées dans des actes de colloques internationaux	7	
Communications publiées dans des actes de colloques nationaux	1	
Chapitres d'ouvrages	4	
Autres supports	1	
Communications sans actes		39
Colloques internationaux à comité de sélection	13	
Colloques nationaux à comité de sélection	2	
Conférences sur invitation	8	
Conférences dans des séminaires	12	
Conférences de vulgarisation scientifique	4	
Autres		2
Total		62

Évaluatrice pour revues et maisons d'édition

Canadian Journal of Linguistics/Revue canadienne de linguistique ; Corpus ; Isogloss. Open Journal of Romance Linguistics (et membre du Comité éditorial) ; Language Acquisition: A Journal of Developmental Linguistics ; Presses Universitaires de Rennes ; Revue de linguistique romane

Encadrement d'étudiants

Conseillère externe de thèse (2023-)	KU Leuven, Directrices : Karen Lahousse et Cécile De Cat, <i>The emergence of syntactic complexity</i>	1
Jury de thèse (2020)	Laboratoire des Sciences Cognitives et Psycholinguistique, ENS Paris	1
Comité de suivi de thèse	Laboratoire BCL	1
Directions ou codirections de Mémoires de Master 2 (2015-)	Masters Nice et Lyon : Linguistique théorique, Orthophonie, Psychologie du développement	20
Codirections de Mémoires de Master 1 (2020-)	Master Nice : Psychologie du développement	7
Rapporteuse, consultante, personne ressource (2012-)	Masters Nice et Lyon : Orthophonie	4

Enseignements universitaires

Enseignements (HeqTD)			MCF								CC					ATER		TOT	
	Intitulés	Départements	2023-24	2022-23	2021-22	2020-21	2019-20	2018-19	2017-18	2016-17	2015-16	2014-15	2013-14	2012-13	2011-12	2010-11	2009-10	2008-09	
L1	Découverte des SDL	Portail LLAC, Psychologie	3	3	3	3	6	3						3					24
L1	Introduction aux SDL	Orthophonie	52.5	52.5	76.5	76.5	76.5	76.5	76.5	76.5							24	24	688.5
L1	Grammaire française	SDL															30	30	60
L1	Introduction à la syntaxe	SDL															30	30	60
L2	Syntaxe française	SDL															30	30	60
L2	Développement du langage et	Orthophonie	25.5	25.5	25.5	25.5	22.5	22.5	22.5	22.5									214.5
L2	Connaissances SDL appliquées à	Orthophonie					8												8
L3	Acquisition du langage	SDL, Psychologie, Lettres, Ethnologie	42	42	42	42	44.5	44.5	44.5	44.5	42	30	30		56	28			532
L3	Problèmes d'acquisition	SDL, Psychologie, Lettres	42	42	42	42	30	30	30	30	30	30	30						378
L3	Syntaxe générative	SDL		15	15	15	18.8		11.3		15					18	78	78	264.1
M1	Syntaxe générative	SDL											15						15
M1	Acquisition du langage et apprentissage des	SDL, FLE	30	30	30	30	30	30	30	30	30	30							300
M1	Enonciation développementale	SDL				7.5													7.5
M1	Variation et changement linguistique	SDL											10	7.5					17.5
M2	Rédaction du mémoire	SDL										15							15
TOTALUX			195	210	234	242	236	207	215	204	216	105	85	10.5	56	46	192	192	2644.1

2. Descriptif rédigé (maximum 4 pages)

Titulaire d'un Doctorat de Sciences du langage (obtenu en 2009 avec la mention très honorable avec les félicitations du jury à l'unanimité ; thèse publiée en 2010) et après deux années d'ATER, des vacances à l'étranger et trois années en tant que chercheure contractuelle dans un projet ANR-DFG, j'ai été recrutée le 01/09/2015 en tant **maître de conférences au Département de Sciences du langage à Université Côte d'Azur et affiliée au Laboratoire BCL, UMR 7320.**

Recrutée entre autres afin de **créer des liens entre les Départements de Sciences du langage, d'Orthophonie et de Psychologie** à Université Côte d'Azur, l'ensemble de mes activités a pour objectifs de faire connaître à un plus large public les sciences du langage et de promouvoir l'interdisciplinarité entre sciences du langage, psychologie et orthophonie, tant au niveau de l'enseignement que de la recherche. La mutualisation de cours de sciences du langage avec les départements de Psychologie, Lettres, Anthropologie, ainsi que des publications dans des revues internationales avec des collègues psycholinguiste et neurolinguiste représentent des aboutissements concrets de ces objectifs (Arslan, Palasis & Meunier, 2020 ; Palasis, Faure & Meunier, 2023).

Mes **activités pédagogiques** se répartissent principalement sur trois départements (Sciences du langage, Orthophonie, Psychologie), sur trois campus et sur les niveaux de la Licence 1 au Master 1. La **thématique principale est l'acquisition du langage**, qui se décline autour des axes suivants :

- **Développement typique du langage chez l'enfant entre 0 et 6 ans** : description des étapes du développement linguistique et découverte de protocoles expérimentaux fondateurs en psycholinguistique.
- **Problèmes d'acquisition du langage** : description de facteurs qui peuvent impacter le développement linguistique chez l'enfant (déficiences auditives, aphasies, dysphasies, etc.) et analyse de données enfantines atypiques.
- **Approches expérimentale et clinique** : mise en regard du développement typique avec des trajectoires atypiques prises en charge en orthophonie (dysphasies notamment) et présentation de protocoles expérimentaux et de tests qui permettent de recueillir et évaluer les données atypiques.
- **Comparaison d'approches théoriques qui rendent compte du processus acquisitionnel** : formation et sensibilisation des étudiants à l'existence de cadres théoriques différents (grammaire générative vs théories basées sur l'usage) afin qu'ils aient les connaissances pour choisir, décrire et utiliser un cadre théorique s'ils décident de faire leur mémoire de recherche M2 dans le domaine de l'acquisition du langage.

Depuis mon recrutement en 2015, j'ai d'ailleurs **(co-)encadré 20 Mémoires de Master 2** soutenus en Linguistique théorique, Orthophonie et Psychologie développementale et **7 Mémoires de Master 1** en Psychologie développementale.

Très rapidement après mon recrutement, j'ai également été très impliquée au niveau administratif, notamment par le biais des fonctions cumulées de **directrice du département de**

Sciences du langage pendant trois années (d'octobre 2017 à octobre 2020) et de celles de **responsable pédagogique L1/L2 pendant cinq ans** (d'octobre 2017 à septembre 2022) et **responsable de la Mention/L3 Sciences du langage** (depuis octobre 2017).

Plus récemment, j'ai pris des **fonctions axées sur la recherche au sein du Laboratoire BCL UMR 7320**, en assurant notamment la **responsabilité d'une équipe de recherche** créée au 1^{er} janvier 2023 : « **Acquisition, Apprentissage, Adaptation** » (<https://bcl.cnrs.fr/rubrique461?lang=fr>). Cette équipe réunit des enseignants-chercheurs et chercheurs en linguistique, psycholinguistique, neurolinguistique, didactique des langues, psychologie expérimentale et neurosciences. Grâce à cette richesse pluridisciplinaire, nous étudions le développement des langues et le traitement des différents types d'informations linguistiques (phonologiques, morphologiques, syntaxiques, lexicales et pragmatiques), en nous intéressant aux processus dynamiques qui caractérisent les usages linguistiques. Plusieurs publications pluridisciplinaires illustrent d'ores et déjà ce travail d'équipe (Arslan, Palasis & Meunier, 2020 ; Edmonds, Gerbier, Palasis & Whyte, 2021 ; Whyte, Edmonds, Palasis & Gerbier, 2022 ; Palasis, Faure & Meunier, 2023).

Mes principaux thèmes de recherche depuis ma thèse en 2009 ont été la **morphosyntaxe du français oral, sa variabilité et son acquisition**, grâce à une approche intégrative de l'acquisition du langage qui combine certains aspects théoriques de la Grammaire générative et un socle empirique important via la base de données enfantines de maternelle créée à l'occasion de ma thèse et partiellement disponible sur la plateforme internationale CHILDES (<http://childes.talkbank.org/access/French/Palasis.html>). Mes thèmes de recherche se sont articulés autour des trois axes suivants :

- **Formes du sujet** : à partir de la base de données de maternelle, **toutes les formes que peut prendre le sujet dans les énoncés enfantins** sont décrites, quantifiées et discutées dans le cadre de la Grammaire générative. Ainsi, le phénomène très documenté du « **sujet nul enfantin** » est revisité et décrit comme une absence du clitique nominatif et une vue d'ensemble de toutes les autres formes du sujet est proposée. Ces recherches permettent alors de dresser un état des lieux des productions enfantines pendant leurs trois années de maternelle et d'établir un comparatif avec les structures du français adulte oral et du français standard. Les propositions principales sont que **le clitique nominatif est un marqueur d'accord préverbal** et que **le français oral est une langue à sujet nul** (comme certains dialectes nord-italiens).
- **Structures interrogatives *wh-*** : une étude descriptive et statistique des différents éléments interrogatifs *wh-* (*qui, quoi*, etc.), de leurs positions (in-situ : *tu vois qui ?* vs ex-situ : *qui tu vois ?*) et des verbes avec lesquels ces *wh-* apparaissent (présentatif *c'est* vs autres formes du verbe *être* vs verbes lexicaux) pendant les trois années de maternelle a mis en évidence une **corrélation entre la position in-situ des items *wh-* et le présentatif *c'est*** (*c'est quoi ça ?*), les autres formes verbales étant davantage produites avec des *wh-ex-situ* (*combien ça coûte ?*). Il semblerait

alors que le déplacement ex-situ soit déclenché par la **sémantique du verbe**. Des facteurs développementaux, tels que l'augmentation de l'utilisation de la forme ex-situ *qu'est-ce que* au détriment du pronom in-situ *quoi* et la diversification du système verbal enfantin, sont compatibles avec les taux de questions *wh*-ex-situ plus élevés chez les adultes que chez les enfants. En français oral adulte, par le biais d'une approche plus formelle de la variabilité du placement des items *wh*-, Richard Faure et moi-même avons proposé que le **français oral est une langue à *wh*-in-situ** et que **le mouvement ex-situ n'est pas déclenché par le statut interrogatif de la phrase mais par un trait sémantique** (le trait *Exclusion*, forme de contraste non-exhaustif). Les deux hypothèses, sur les questions enfantines et adultes, se rejoignent donc quant à leur **interface entre syntaxe et sémantique**. Enfin, 2024 a vu le démarrage de **recherches expérimentales** chez des adultes en collaboration avec Seçkin Arslan et Lena Baunaz.

- **Formalisation de la variation en français** : les descriptions du français enfantin produites à partir de l'analyse de productions de seize enfants pendant leurs trois années de maternelle ont permis d'établir que **le français enfantin présente des caractéristiques proches du langage adulte adressé aux enfants (LAE)**, quant au doublement très fréquent des syntagmes nominaux (*le chien il aboie*), l'élision du clitique *il* devant consonne (*i dort*) et l'absence de la particule négative préverbale *ne* (*le chien i dort pas*). Les données LAE et enfantines ont également montré qu'aucun de ces phénomènes n'est néanmoins présent dans 100 % des occurrences. Il existe donc une **variation** pour chacun de ces phénomènes, avec de rares occurrences de syntagmes nominaux non-doublés (*le chien aboie*), de clitiques non-éolidés devant consonne (*il dort*) et de négations discontinues (*le chien ne dort pas*).

Par ailleurs, les données enfantines ont permis d'établir que **l'émergence d'une variante par rapport à l'autre ne semble pas aléatoire** : lorsqu'un enfant produit un énoncé, toutes les caractéristiques de cet énoncé renvoient soit au français non-standard (*le chien i dort pas*), soit au français standard (*le chien ne dort pas*). Le **contexte de production** a été relevé comme facteur déclencheur de l'une ou l'autre variante, avec un **effet de l'activité** (lecture partagée, comptines, chansons) et un **possible effet d'amorçage entre enfants**. Sur la base de ces éléments non-aléatoires, la variation est décrite comme relevant de **deux grammaires séparées** (bien qu'apparentées), nommées « **français spontané** » (Grammaire 1) et « **français normé** » (Grammaire 2) dans le cadre de **l'hypothèse de la diglossie**. Le français spontané est acquis par tous les enfants qui bénéficient d'un environnement francophone quotidien et précoce. Le français normé est acquis ou appris, en fonction de l'environnement linguistique de l'enfant. Un adulte francophone n'est donc pas forcément un locuteur diglosse.

Ces dernières années, mes recherches initiales sur l'acquisition du français oral enfantin étudié à partir de données de productions spontanées se sont élargies à plusieurs autres thématiques et types de données, dans le cadre de **collaborations avec des collègues psycholinguistes, neurolinguistes, psychologues et didacticiens**. Les recherches suivantes ont ainsi été initiées :

- **Effet de l'âge sur le traitement des pronoms anaphoriques et cataphoriques** : le traitement des pronoms anaphoriques (*Jules_i a persuadé des amis de parler parce qu'il_i était aphone*) et cataphoriques (*Parce qu'il_i était aphone, Jules_i,...*) en français a été étudié chez des adultes de 19-35 ans et 57-88 ans grâce à une tâche de jugement d'acceptabilité de phrases et de mesures EEG. Les résultats suggèrent que le traitement des pronoms est similaire dans les deux groupes d'âges, mais que davantage de régions cérébrales sont impliquées dans le traitement chez les sujets plus âgés, notamment pour les anaphores (Arslan, Palasis & Meunier, 2020).
- **Effet de l'espacement des séances d'apprentissage sur la rétention des informations lexicales et grammaticales en L2** : publication d'une revue de littérature sur l'effet d'espacement sur l'apprentissage du lexique en L2, montrant une meilleure rétention des informations lexicales lorsque les séances d'apprentissage sont temporellement plus espacées (« pratique distribuée » ; Edmonds, Gerbier, Palasis & Whyte, 2021), puis mise en place d'une plateforme pour l'apprentissage de l'anglais L2 en ligne, avec un enseignement explicite d'items lexicaux et une exposition implicite à une structure grammaticale, et avec un espacement des séances d'apprentissage massé ou distribué (Whyte, Edmonds, Palasis & Gerbier, 2022 ; plateforme <https://lexlearn.org/>). Les données pré- et posttests collectées en 2023 auprès de 145 **étudiants** sont à présent analysées et il est prévu de soumettre un article à la revue *TESOL Quarterly* en 2024. Le développement le plus récent de ce projet a étendu les investigations à des participants plus jeunes dans le cadre d'une **collaboration Éducation-Recherche** initiée avec un collègue à Nice. La plateforme a ainsi été adaptée à un public de **collégiens** et un premier jeu de données a été recueilli auprès de deux classes en 2024. Il est prévu de prolonger la collaboration et la collecte de données en 2024-2025 afin d'élargir notre échantillon et de publier l'ensemble des résultats en 2025.

Cette dernière thématique a bénéficié de **financements obtenus localement en 2020, 2023 et 2024** (Projet ESPACE L2 : *ESpacement dans l'Apprentissage de la Complexité En L2*, 5 000€ ; Projet ETALE : *Espacement Temporel dans l'Apprentissage des Langues Étrangères*, 7 000€ ; Projet AVEC : *Amélioration de l'apprentissage du Vocabulaire Enseigné au Collège*, 5 000€). En 2024, les membres de ces projets (Amanda Edmonds, Shona Whyte, Émilie Gerbier, et moi-même) ont déposé une **demande de financement ANR-FNS** (projet franco-suisse) sur l'effet d'espacement dans l'apprentissage de l'anglais L2 chez des personnes plus âgées : Projet SAGE (*Spacing and AGing in L2 English*) ; Simone Pfenninger (PI Suisse) et Amanda Edmonds (PI France).

3. Liste exhaustive des publications et travaux

3.1 Articles		21
3.1.1 Revues internationales à comité de lecture	8	
3.1.2 Communications publiées dans des actes de colloques internationaux	7	
3.1.3 Communications publiées dans des actes de colloques nationaux	1	
3.1.4 Chapitres d'ouvrages	4	
3.1.5 Autres supports	1	
3.2 Communications sans actes		39
3.2.1 Colloques internationaux à comité de sélection	13	
3.2.2 Colloques nationaux à comité de sélection	2	
3.2.3 Conférences sur invitation	8	
3.2.4 Conférences dans des séminaires	12	
3.2.5 Conférences de vulgarisation scientifique	4	
3.3 Autres		2
Total		62

3.1 Articles : 21

3.1.1 Articles dans des revues internationales à comité de lecture

- 2023, Palasis, K., Faure, R. & F. Meunier, "Wh-in-situ in child French: Deictic triggers at the syntax-semantics interface", *Journal of French Language Studies*, 33(3), 273-298. <https://doi.org/10.1017/S0959269523000030>
- 2021, Edmonds, A., Gerbier, E., Palasis, K., & S. Whyte, "Understanding the distributed practice effect and its relevance for the teaching and learning of L2 vocabulary", *Lexis*, 18, *Lexical learning and teaching*. <https://doi.org/10.4000/lexis.5652>
- 2021, Faure, R. & K. Palasis, "Exclusivity! Wh-fronting is not optional wh-movement in Colloquial French", *Natural Language & Linguistic Theory*, 39, 57-95. <https://doi.org/10.1007/s11049-020-09476-w>
- 2020, Arslan, S., Palasis, K., & F. Meunier, "Electrophysiological differences in older and younger adults' anaphoric but not cataphoric pronoun processing in the absence of age-related behavioural slowdown", *Scientific Reports*, 10(19234). <https://doi.org/10.1038/s41598-020-75550-3>
- 2019, Palasis, K., Faure, R., & F. Lavigne, "Explaining variation in wh-position in child French: A statistical analysis of new seminaturalistic data", *Language Acquisition*, 26(2), 210-234. <https://doi.org/10.1080/10489223.2018.1513004>
- 2015, Palasis, K., "Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony", *Lingua*, Grohmann, Kleanthes K. & T. Neokleous (éds.), Vol. 161, p. 125-143. <http://dx.doi.org/10.1016/j.lingua.2014.11.012>
- 2013, Palasis, K., "The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French", *Journal of French Language Studies*, N° 23/1, Numéro thématique « L'hypothèse d'une diglossie en France », Massot, B. & P. Rowlett (éds.), p. 17-35. <https://doi.org/10.1017/S0959269512000348>
- 2010, Palasis, K., "Introducing New French Child Data: Thoughts on their Gathering and Coding", *Corpus*, N°9 « La syntaxe de corpus », Oliviéri, M. (éd.), p. 33-51. <https://journals.openedition.org/corpus/1801>

3.1.2 Communications publiées dans des actes de colloques internationaux

- 2022, Whyte, S., Edmonds, A., Palasis, K., & E. Gerbier, "ESPACE L2: Exploring spacing effects in explicit and implicit online learning of L2 English", In B. Arnbjörnsdóttir, B. Bédi, L. Bradley, K. Friðriksdóttir, H. Garðarsdóttir, S. Thouësny, & J. Whelpton (Eds.), *Intelligent CALL, granular systems, and learner data: Short papers from EUROCALL 2022* (pp. 386-391): Research-publishing.net, <https://doi.org/10.14705/rpnet.2022.61.1489>
- 2020, Olivieri, M., Kaiser, G. A., Palasis, K., Zimmermann, M., & R. Faure, « Quand la dialectologie, la diachronie et l'acquisition se parlent : Étude comparative des pronoms sujets en occitan et en français », In J. Sibille (Ed.), *La microvariation syntaxique dans les langues romanes de France. Actes du colloque Symila, Toulouse, 11 et 12 juin 2015* (pp. 81-98). Limoges: Lambert-Lucas, <https://www.researchgate.net/profile/Katerina-Palasis/publications>.
- 2020, Broc, L., Palasis, K., & T. Olive, « Écrire des narrations permet aux élèves dysphasiques de mobiliser leurs connaissances syntaxiques », In F. Neveu, B. Harmegnies, L. Hriba, S. Prévost, & A. Steuckardt (Eds.), *SHS Web of Conferences: 7^e Congrès Mondial de Linguistique Française* (Vol. 78, pp. 10003). Montpellier: EDP Sciences, <https://hal.archives-ouvertes.fr/hal-02929793/file/CMLF-Montpellier-avec-noms.pdf>
- 2012, Palasis, K., Economy Principles within Child Speech: When the Nominative Clitic does not Surface. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 261-284). Newcastle: Cambridge Scholars Publishing, <https://www.researchgate.net/profile/Katerina-Palasis/publications>
- 2009, Palasis, K., On the Variable Morpho-Syntactic Status of the French Subject Clitics. In G. A. Kaiser & E.-M. Remberger (Eds.), *Null-subjects, Expletives and Locatives in Romance* (Vol. Arbeitspapier 123, pp. 47-62). Konstanz: Konstanz Working Papers of the Department of Linguistics. <http://ling.uni-konstanz.de/pages/publ/PDF/ap123.pdf>
- 2007, Palasis, K. & M. Olivieri, Beyond Deficiency. Reconsidering the Null Subject Phenomenon. In S. Baauw, J. Van Kampen & M. Pinto (Eds.), *The Acquisition of Romance Languages. Selected Papers from The Romance Turn II 2006* (Vol. LOT Occasional Series 8, pp. 127-143). Utrecht: LOT Publications. <https://dspace.library.uu.nl/handle/1874/296778>
- 2006, Palasis, K. & M. Olivieri, On the Null-Subject Stage in Non-Null Subject L1 Acquisition. In A. Belletti, E. Bennati, C. Chesi, E. Di Domenico, & I. Ferrari (Eds.), *Language Acquisition and Development. Proceedings of GALA 2005* (pp. 393-398). Newcastle: Cambridge Scholars Publishing.

3.1.3 Communications publiées dans des actes de colloques nationaux

- 2014, Palasis, K., Étude morphosyntaxique des pronoms sujets à partir d'un corpus de maternelle. In R. Mir-Samii (Ed.), *Du sujet et de son absence dans les langues* (pp. 259-270). Le Mans: Université du Maine, <http://doi.org/10.13140/2.1.1586.9440>

3.1.4 Chapitres d'ouvrages

- 2017, Heap, D., Olivieri, M. & K. Palasis, Clitic Pronouns. In A. Duffer & E. Stark (Eds.), *Manual of Romance Morphosyntax and Syntax* (Vol. 17, pp. 183-229). Berlin: De Gruyter. ISBN 978-3-11-037693-7
- 2013, Kaiser, G. A., Olivieri, M. & K. Palasis, Impersonal Constructions in Northern Occitan. In E. Carrilho, C. Magro & X. Álvarez (Eds.), *Current Approaches to Limits and Areas in Dialectology* (pp. 345-366). Newcastle upon Tyne: Cambridge Scholars Publishing. ISBN 978-1-4438-4757-5

- 2011, Palasis, K., Français spontané et français normé : bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In O. Bertrand & I. Schaffner (Eds.), *Variétés, Variations & Formes du français* (pp. 55-72). Palaiseau: Les Éditions de l'École Polytechnique. ISBN 978-2-7302-1586-2.
<https://www.researchgate.net/profile/Katerina-Palasis/publications>
- 2008, Palasis, K., Syntaxe, acquisition et dialectologie : la pluridisciplinarité appliquée à l'étude du sujet nul enfantin. In Lidilem (Ed.), *Autour des langues et du langage. Perspective pluridisciplinaire* (pp. 241-248). Grenoble: Presses Universitaires de Grenoble. ISBN 978-2-7061-1427-4

3.1.5 Autres supports

- 2020, Palasis, K., compte rendu de l'ouvrage *L'interrogative en français* de Béguelin, M.-J., Coveney, A., & Guryev, A. (Eds.), 2018, Berne : Peter Lang, dans *Revue de Linguistique Romane (LLiR)*, 84(335-336), p. 584-589.

3.2 Communications sans actes : 39

3.2.1 Communications sans actes présentées dans des colloques internationaux à comité de sélection

- 2024, Palasis, K., "Acquiring evidentiality in French: A syntactic hypothesis", *(Un)Reality and Evidentiality in Language, Cognition, Society*, Nicosie, 24-26 octobre 2024.
- 2024, Palasis, K., Angelopoulos, N., Baunaz, L. & S. Arslan, "Information-source judgement experiments in French syntactic evidentiality", *(Un)Reality and Evidentiality in Language, Cognition, Society*, Nicosie, 24-26 octobre 2024.
- 2024, Wensink, T., Lahousse, K., de Cat, C., Palasis, K., & B. Busson, "The Development of the Syntactic Tree: Evidence from Cleft Sentences in L1 French", *The 16th Generative Approaches to Language Acquisition Conference (GALA 2024)*, Lisbonne, 12-14 septembre 2024.
- 2024, Wensink, T., Lahousse, K., de Cat, C., Palasis, K., & B. Busson, "The Syntactic Development of Cleft Sentences in L1 French", *The 57th Annual Meeting of the Societas Linguistica Europaea (SLE 2024)*, 21-24 août 2024, Helsinki.
- 2024, Edmonds, A., Whyte, S., Palasis, K. & E. Gerbier, "The distributed practice effect in L2 vocabulary learning: A close replication to look at long(er)-term retention", *The 33rd Conference of the European Second Language Association (EuroSLA33)*, 3-6 juillet 2024, Montpellier.
- 2024, Cordero-Rull, M.-D.-M., Badier, J.-M., Kanzari, K., Palasis, K., Alario, F.-X., Meunier, F., & S. Arslan, "Tracing Pronouns in the Brain: a MEG study in French", *The Annual Meeting of the NeuroMod Institute*, 3-4 juillet 2024, Antibes.
- 2023, Lahousse, K., Jourdain, M., Thonissen, K. & K. Palasis, "The acquisition of the syntax-information structure interface: On the emergence of c'est 'it is' in child French L1", *56th Annual meeting of the Societas Linguistica Europaea*, 29 août-1 septembre 2023, Athènes.
- 2023, Edmonds, A., Whyte, S., Palasis, K. & Gerbier, E., "The distributed practice effect with an incidental grammatical target", *The 32nd Conference of the European Second Language Association (EuroSLA 32)*, Birmingham, UK, 31 août 2023.
- 2023, Whyte, S., Edmonds, A., Palasis, K. & E. Gerbier, "The distributed practice effect in L2 vocabulary learning: an ecologically-designed online experiment", *The EUROCALL 2023 Conference*, Reykjavik, 15-18 août 2023.

- 2017, Palasis, K. & F. Meunier, « Coût cognitif des morphèmes silencieux », *58ème Congrès annuel de la Société Française de Psychologie, Symposium 9 : Processus cognitifs et structure morphologique des mots*, 1 septembre 2017, Nice.
- 2014, Faure, R. & K. Palasis, “French *wh*-questions: Crossing diglossia and topicality”, *The 44th Linguistic Symposium on Romance Languages (LSRL)*, 2-4 mai 2014, London (Ontario).
- 2014, Palasis, K. & R. Faure, “Grammatical and interpretive constraints on the development of *wh*-questions in L1 French”, *The Romance Turn VI, International Conference on the Acquisition of Romance Languages*, 18-20 septembre 2014, Palma de Mallorca, <https://doi.org/10.13140/2.1.4634.3683>
- 2009, Palasis, K., “From Morphological Affixes to Syntactic Arguments: What Forces Children to Reinterpret Nominative French Clitics”, Colloque international *Going Romance XXIII*, 3-5 décembre 2009, Nice.

3.2.2 Communications sans actes présentées dans des colloques nationaux à comité de sélection

- 2011, Palasis, K., « Pour une meilleure représentativité des données orales en grammaire générative : problèmes et solutions à partir d’un nouveau corpus enfantin », Colloque Jeunes Chercheurs *Praxiling Corpus, données, modèles : approches qualitatives et quantitatives*, 9-10 juin 2011, Montpellier.
- 2006, Palasis, K., « Une perspective transversale entre acquisition de la syntaxe et dialectologie », 3^{ème} Colloque Jeunes Chercheurs en Sciences du Langage *Intra-disciplinarité et extra-disciplinarité en sciences du langage*, 14-15 juin 2006, Université Paris X Nanterre.

3.2.3 Conférences sur invitation

- 2024, *The acquisition path for clefted wh-questions in preschool French: Outline of the project and first results*, Conférence hybride du Réseau Listen (Karen Lahousse & Jorina Brysbaert), KU Leuven, 17 juin 2024, <https://sites.google.com/view/listen-is/talks-events>
- 2023, *Syntactically encoded evidentiality in French? How do preschool children talk about non-witnessed events?*, Rencontres “A Grammar of Lies: Evidentiality and Deception”, Istanbul, 2 novembre 2023.
- 2023, *Wh-in-situ ou wh-ex-situ ? Étude d’un corpus de maternelle entre 3 et 6 ans*, Atelier « L’interrogation en français et son acquisition », Universität Konstanz, 3 février 2023.
- 2023, avec Richard Faure, *Wh-fronting in adult Colloquial and child French : Triggers at the syntax-semantics interface*, Département de Linguistique, Universität Konstanz, 2 février 2023.
- 2022, *La compétence socio-pragmatique en français L2 entre 3 et 6 ans : Étude de la variabilité morphosyntaxique chez deux enfants non-natifs dans leurs interactions avec des enfants natifs à la maternelle*, Journées d’Études « L’acquisition de la compétence pragmatique en L2 », 20-21 octobre 2022, Université Côte d’Azur, <https://doi.org/10.13140/RG.2.2.30223.05281>
- 2022, avec Richard Faure et Fanny Meunier, *Wh-in-situ in child French: Presentational triggers at the syntax-semantics interface*, 13 juin 2022, Journée d’Études « Les questions *wh*-in-situ en français », <https://univ-poitiers.webex.com/univ-poitiers/ldr.php?RCID=e54059638d8c19b2e8f63055aa7fbd28> (password: bJWiJpq5)
- 2022, *Wh-questions in Colloquial French: Empirical and theoretical perspectives on child and adult data*, University College London (UCL), Division of Psychology and Language

Sciences, Linguistic Seminar Series, 25 Mai 2022,
[https://ucl.zoom.us/rec/share/9GKAma0h20ANOR-BBcZoy6M7mhoC9eXPJMxLlul1D5nV-RxuCFeqBGFCK9BjQV0.1hY - rdtu4SG AdS?startTime=1653491094000](https://ucl.zoom.us/rec/share/9GKAma0h20ANOR-BBcZoy6M7mhoC9eXPJMxLlul1D5nV-RxuCFeqBGFCK9BjQV0.1hY-rdtu4SGAdS?startTime=1653491094000) (password: rJ.WQ+2r)

2010, *Syntax around 3: Evidence from 20 French children and theoretical hypotheses within generative grammar*, conférence donnée au Département de Linguistique de l'Université de Konstanz (Allemagne), 9 décembre 2010.

3.2.4 Conférences dans des séminaires

2021, « Spécificités des informations linguistiques ? Aperçu de réponses en acquisition typique et atypique », présentation faite à l'équipe « Langage et Cognition » de BCL, 20 mai 2021.

2013, « Des enregistrements en maternelle aux hypothèses en acquisition du langage », conférence donnée au Département de Linguistique de l'Université de Konstanz (Allemagne), 13 juin 2013.

2012, avec Charlotte Luson, « Comment combiner syntaxe formelle et approche expérimentale. Etude d'un article », présentation faite à l'équipe « Langage et Cognition » de BCL, 20 décembre 2012.

2012, « L'hypothèse d'une diglossie en France défendue du point de vue acquisitionnel », présentation faite à l'équipe « Dialectologie, Diachronie, Phonologie » de BCL, 26 janvier 2012.

2010, « Français L1 : formaliser son acquisition pour réfléchir différemment à son enseignement », présentation faite à l'équipe « Didactique et Linguistique » de BCL, 18 novembre 2010.

2010, « Syntaxe générative et acquisition : le sujet dans le développement du système linguistique du jeune enfant », présentation faite aux étudiants de Master 2 de l'Université Nice Sophia Antipolis, séminaire de syntaxe, 15 octobre 2010.

2008, « La notion de gabarit en syntaxe », présentation faite à l'équipe « Dialectologie, Diachronie, Phonologie » de BCL, 9 octobre 2008.

2008, « La syntaxe de *Qu'est-ce que* et l'hypothèse du Remnant IP Movement », présentation faite en séminaire de syntaxe à BCL, 26 février 2008.

2007, « Quel statut pour les pronoms clitiques sujets en français ? », présentation faite en séminaire de syntaxe à BCL, 6 février 2007.

2006, « La grammaire générative et les pronoms : aperçu de traitement », présentation faite en séminaire de syntaxe à BCL, 2 mai 2006.

2006, "Linguistic Duality and the Null Subject Stage", conférence donnée au Département de Linguistique de l'Université de Konstanz (Allemagne), 7 mars 2006.

2005, « Les corpus destinés à l'étude du langage enfantin : méthodologie et utilisation », présentation faite en séminaire sur les bases de données dialectales à BCL, novembre 2005.

3.2.5 Conférences de vulgarisation scientifique

2024, *Bilan 2023-2024 de la Collaboration Éducation-Recherche avec le Collège Alphonse Daudet*, Présentation à l'ensemble du personnel à l'Assemblée générale du Collège, Nice, 5 juillet 2024.

2024, avec Philippe Del Giudice et Shona Whyte, *Que disent de nous les accents et les intonations du langage ?* Conférence grand public sur les accents régionaux et étrangers, Festival Trajectoires, Carros, 2 février 2024.

2023, *Quiz et discussion autour du langage*, Animation de stand à la Fête de la Science, Journée Grand public, Nice, 15 octobre 2023.

2023, *Quiz et discussion autour du développement du langage chez l'être humain*, Animation d'ateliers à la Fête de la Science, Journée des scolaires (collégiens et lycéens), Nice, 12 octobre 2023.

3.3 Autres : 2

2012, Palasis, K., *Romance French Child Corpus 1 & 2 (2;5-4;11)*. Base de données enfantines de maternelle, partiellement disponible sur la plateforme CHILDES : <http://childes.talkbank.org/access/French/Palasis.html>

2010, Palasis, K., *Syntaxe générative et acquisition : le sujet dans le développement du système linguistique du jeune enfant*. Villeneuve d'Ascq: ANRT Diffusion, ISBN 978-2-7295-7923-4, <https://doi.org/10.13140/2.1.2722.0806>

II. Sélection de publications

1. Palasis, K. (2010). Introducing New French Child Data: Thoughts on their Gathering and Coding. *Corpus, 9, La syntaxe de corpus*, 33-51, <https://journals.openedition.org/corpus/1801> 19
2. Palasis, K. (2011). Français spontané et français normé : bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In O. Bertrand & I. Schaffner (Eds.), *Variétés, variations & formes du français* (pp. 55-72). Palaiseau: Les Éditions de l'École Polytechnique, <https://www.researchgate.net/profile/Katerina-Palasis/publications> 41
3. Palasis, K. (2012). Economy Principles within Child Speech: When the Nominative Clitic does not Surface. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 261-283). Newcastle: Cambridge Scholars Publishing, <https://www.researchgate.net/profile/Katerina-Palasis/publications> 59
4. Palasis, K. (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies*, 23(1), Numéro spécial *L'hypothèse d'une diglossie en France*, Massot, Benjamin & Paul Rowlett (Eds.), 17-35, <https://doi.org/10.1017/S0959269512000348> 83
5. Palasis, K. (2014). Étude morphosyntaxique des pronoms sujets à partir d'un corpus de maternelle. In R. Mir-Samii (Ed.), *Du sujet et de son absence dans les langues* (pp. 259-270). Le Mans: Université du Maine, <http://doi.org/10.13140/2.1.1586.9440> 103
6. Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, 161, 125-143, <http://dx.doi.org/10.1016/j.lingua.2014.11.012> 115
7. Palasis, K., Faure, R. & Lavigne, F. (2019). Explaining variation in wh-position in child French: A statistical analysis of new seminaturalistic data. *Language Acquisition*, 26(2), 210-234, <https://doi.org/10.1080/10489223.2018.1513004>. 135
8. Faure, R. & Palasis, K. (2021). Exclusivity! Wh-fronting is not optional wh-movement in Colloquial French. *Natural Language & Linguistic Theory*, 39(1), 57-95, <https://doi.org/10.1007/s11049-020-09476-w> 161
9. Palasis, K., Faure, R. & Meunier, F. (2023). Wh-in-situ in child French: Deictic triggers at the syntax-semantics interface. *Journal of French Language Studies*, 33(3), 273-298, <https://doi.org/10.1017/S0959269523000030> 201

1. Palasis, K. (2010). Introducing New French Child Data: Thoughts on their Gathering and Coding. *Corpus, 9, La syntaxe de corpus*, 33-51, <http://corpus.revues.org/index1801.html>

Introducing New French Child Data: Thoughts on their Gathering and Coding

Katerina PALASIS

Laboratoire BCL, UNSA, CNRS, MSH de Nice¹

Abstract: This article browses through a number of significant steps which face an investigator while compiling a corpus for acquisitional research. It ranges from thoughts when gathering the information (the choice of the informants, their number, representativity, etc.) to its final coding in order to eventually share an efficiently constructed database with the scientific community.

Keywords: acquisition, generative syntax, spontaneous data, data representativity, linguistic and methodological variables, coding, French, subject, null subject.

1. Introduction

It is well-known that collecting, transcribing and labelling child data are time-consuming and demanding tasks. However, it is also acknowledged that a reliable and appropriately transcribed and coded corpus turns out to be worthwhile and enlightening. For instance, thanks to the software provided by the CHILDES database (Child Language Data Exchange System)², a total of 30,542 spontaneous French utterances from three adults and twenty-two children between 2;5 and 4;0 have now been analysed with regard to the acquisition of the syntactic subject, with interesting outcomes (Palasis (2009b)).

However, the way from child spontaneous utterances through to acquisitional theory is treacherous and many intermediary steps must be achieved before one or several

1. 98 bd E. Herriot, 06200 Nice, France.

2. Cf. MacWhinney (2000a, 2000b) and the website at <http://childes.psy.cmu.edu/>.

Katerina PALASIS

hypotheses can emerge from child recordings. Hence, the aim of this article is to account for my own experience while compiling two corpora (Corpus N°1 as analysed in Palasis (2005) and Corpus N°2 as analysed in Palasis (2009b)) by sharing thoughts on what are perceived as cornerstones in this domain. However, it is quite obvious that all the matters can not be addressed within these pages. Consequently, two particular topics have been chosen, i.e. data gathering and data coding, and since my experience stems from work with children above 2;5, I refer the reader to Morgenstern & Parisse (2007) in order to complement the picture insofar as these scholars mainly address complementary issues such as data interpretation and transcript with children between 1;0 and 3;0.

The first section of this article goes through the different issues dealt with before and while gathering the two above-mentioned corpora. Questions such as data reliability and representativity are hence addressed. The aim of these thoughts is to identify the different variables linguistic data can display in order to eliminate the undesirable ones and master the others. This methodology, in turn, provides the investigator with an empirical background that will allow him/her to forward reliable theoretical generalisations. The second step which is addressed is data coding and the two coding schemes put together for Corpus N°2 in order to obtain a fine-grained description and analysis of the syntactic notion of “subject” are detailed. These two tiers, i.e. %mor (morphosyntactic) and %err (non-target), are part of the many suggested by CHILDES.

2. What are reliable data?

“L’analyse ne vaut que ce que vaut le corpus”.³ Indeed, there is no doubt that, when a sample is of poor quality, whatever the field of research, the subsequent analyses and generalisations are also likely to be flawed. Since linguists work from samples, i.e. corpora, this qualitative issue is also central in this particular domain. Consequently, the first question that needs to be addressed is: what does “good quality” mean with regard to

3. “The analysis is only worth what the corpus is worth” (Dalbera (2002:94)).

Introducing New French Child Data

linguistic data? I pin down the answer to “representativity”. However, it is stating the obvious to say that working from a representative corpus gives more breadth to the theory which then stems from it. So the crucial concern lies elsewhere. Indeed, the problem with this statement is its circularity: how can we get to know that a corpus is representative of a linguistic system as, if we knew this language so well, we would not be studying it? Such a corpus can hence not be defined empirically. Rather, its criteria must pertain to more general principles which I claim are linked to the notion of variation.

2.1 Individual variation

Each individual has his/her own characteristics, each of us having his/her own personality which in turn is rooted in a particular era, place, culture, education, etc. Quite obviously, all these differences surface in language, whether we consider adults or children. Consequently, as far as acquisitional research is concerned, I argue that working from data uttered by a single child can present pitfalls if the scientific aim is to forward generalisations with regard to language acquisition. Indeed, the individuality of child speech is a well-established fact: De Boysson-Bardies (1996) for instance insists on this type of variation and on its importance within syntactic research. Hence, in order to compensate this individual variation, it then sounds safe to assume that working from data which come from different children reduces this risk. Cohen (1924) for instance already favours such an approach:

Un fait observé chez l'enfant n'est bien utilisable
[...] que si la part originale de l'individu peut y
être délimitée, ce qui ne se réalise bien que par
des comparaisons nombreuses.⁴

The study of the verbal system of seventeen children between 2;3 and 3;1 (Palasis (2005)) confirms the impact of this individual variation on data analysis. Indeed, Table 1 hereunder

4. Cohen (1924:34): “A fact observed with a child can be correctly interpreted [...] only if the original part of the individual can be delimited, which can only be correctly realised with numerous comparisons”. More recently, the same kind of approach is also encouraged in Demuth (1996) for instance.

Katerina PALASIS

displays the occurrences for the four most uttered verbs in a particular type of sentence within this corpus, i.e. *être* ‘be’, *adorer* ‘love’, *avoir* ‘have’, and *faire* ‘do/make’. The penultimate column illustrates that Malcolm is the only child out of the eleven present in this extract who utters the verb *adorer*. If the analysis had relied on that one child or if the statistics had been exploited globally (as shown in the second column), this particular verb would have been ranked second within these standings. However, the other columns hereunder illustrate that Malcolm’s utterances are not representative of the whole group of children at least at two levels. First, Malcolm is the only child who utters the verb *adorer*. Secondly and more generally, the whole corpus points out that children within this age group only utter very few first-group verbs (5.1% of all their verbs).

Verbs	Tot	Mat	Jul	Te	Th	Ma	Ra	No	Al	Ali	Mal	Tho
<i>être</i>	81	1	2	1	11	4	46	1	3	10		2
<i>adorer</i>	14										14	
<i>avoir</i>	14				1	2	7	1	1	1	1	
<i>faire</i>	13		1	1	4		6			1		

Table 1: Verbal occurrences (extract, Corpus N°1)

Consequently, I argue that collecting data from different children is an important methodological step towards representativity. On the other hand, the above data also illustrate that this method is insufficient and that a *a posteriori* filtering out step must be added in order to master and eliminate possible individual characteristics. As far as the above-mentioned corpus is concerned, Malcolm’s utterances of *adorer* were hence eliminated of the final and general statistics on the verbal system.

Moreover, further reasons imply that quantity *per se* does not guarantee representativity. Indeed, gathering numerous data in a random fashion can be counterproductive too as the collected samples can also include a number of other variables which can render an analysis partial or wrong.

2.2 Collective variation

Besides the above-mentioned individual variation, four types of collective variation are traditionally referred to in order to account for the differences which exist between the various linguistic systems. Indeed, variation can be diachronic (as far as acquisition is concerned, intergenerational evolutions can be considered), geographic (e.g. French in France vs. French in Quebec), diastratic (e.g. the influence of the different parental socioprofessional categories), and diaphasic (e.g. when a child speaks with a non-familiar adult vs. when he/she speaks with a family member or with another child). If a set of data used as one whole corpus cumulates two or more different sources of variation, it then makes it particularly difficult to identify which characteristics of the data are to be ascribed to which type of variation. This situation can then lead to a dead end in terms of analysis. Let's illustrate this claim with a tentative comparison of two child corpora.

For her study on the acquisition of wh-questions in French, Crisma (1992) relies on the data of one child, namely Philippe.⁵ Some of his questions are repeated in (1) hereunder.

- (1) Some of Philippe's questions:
- (a) *Où il est le fil ?* "where is the thread"
 - (b) *Où elle est la petite aiguille ?* "where is the little needle"
 - (c) *Où elle est maman ?* "where is mummy"
 - (d) *Où elle est la voiture grosse voiture ?* "where is the big car"

With these French data, Crisma (1992) aims at testing the "clausal truncation" hypothesis as forwarded by Rizzi (1992) for English according to which a child who omits a subject in an utterance (i.e. the [Spec, IP] position) also omits all the elements which could appear above IP in the same clause (e.g. the [Spec, CP] position which hosts the moved wh-elements). In other words, if IP is not projected, nothing is projected above IP. As far as French is concerned, Crisma (1992:117)

5. Data from Suppes, Smith & Léveillé (1973) available on the CHILDES database.

concludes: “As the data clearly show, the co-occurrence of a wh-element and a null subject is impossible”.

Within my own first study of the so-called “null-subject” phenomenon in French (Palasis (2005)), I attempted to compare Philippe’s questions with questions from Corpus N°1 in order to test Rizzi’s hypothesis against a different set of French data. Consequently, I used the 30 wh-questions uttered by one of the children of this corpus, namely Raphaël. Some of his questions are repeated in (2) hereunder.

- (2) Some of Raphaël’s questions:
- (a) Il est *où* ? “he is where”
 - (b) *Où* il est la toute ? “where it is all of it”
 - (c) Elle est *où* ? “she is where”
 - (d) Je le mets *où* celui-là ? “I put it where that one”

Wh-elements can either remain in their base position (as in (2a) above) or move to [Spec, CP] (as in (2b)). Philippe and Raphaël are about the same age, nevertheless their utterances illustrate the two possibilities with regard to wh-placement, i.e. systematic movement for Philippe vs. the *in situ* strategy for Raphaël (only 7 questions out of his 30 imply movement, 6 of which display *pourquoi* ‘why’; *où* ‘where’, *quoi* ‘what’, and *comment* ‘how’ overwhelmingly remain *in situ*). Consequently, trying to falsify the hypothesis according to which wh-movements and null subjects are incompatible is an impossible task when working from Raphaël’s data. Indeed, the purported incompatibility between wh-elements in [Spec, CP] and the so-called child “null subject” phenomenon can not be studied with Raphaël since his three null-subject questions do not display a wh-movement to [Spec, CP], as illustrated in (3) hereunder.⁶

- (3) Raphaël’s “null subject” wh-questions:
- (a) Est *où* cachée ? “is where hidden”
 - (b) Est *pourquoi* ? (x2) “is why”

His utterances hence neither infirm nor confirm the above-mentioned hypothesis. So, on the one hand, we have Philippe’s

6. Note however that (3a) seems to display some kind of wh-movement since the wh-element does not surface in its base position after the participle.

Introducing New French Child Data

data – with wh-movement – that confirm Rizzi’s hypothesis and, on the other hand, we have Raphaël’s data – which display the opposite, *in situ* strategy – whose study is pointless with regard to this particular issue. Nevertheless, the discrepancy between these two children arouses curiosity.

Indeed, it then becomes interesting to identify the reasons of such a difference with regard to wh-placement between Philippe and Raphaël. However, comparing these two sets of data turns out to be vain as several explanations are possible and no particular conclusion can be reached. Firstly, since these two corpora were collected in 1971 and 2003 respectively, their difference with regard to the wh-elements could be ascribed to diachronic (intergenerational) variation since the *in situ* strategy is often associated with current oral French. Secondly, the difference between Philippe and Raphaël could also be ascribed to diastratic or diaphasic variation. The accumulation of several variables can hence be considered as parasitic insofar as these different variables cloud the issue. Moreover, comparing only two children brings us back to the first matter addressed in this section with regard to speech individuality. Consequently, I claim that diachronic variation must be avoided within a corpus and that diastratic and diaphasic variation must be counterbalanced by a large number of informants in order to forward as representative as possible a survey.

As far as geographic variation is concerned, it is established that French in France substantially differs from Belgian, Swiss, or Canadian French. Indeed, this geographic variation is illustrated at all levels since phonology, morphology, syntax, and lexicology are concerned. With regard to syntax, Auger (1994) for instance describes the interrogative marker *-tu* as specifically Canadian and De Cat (2005) illustrates the different interrogative strategies in Belgium, Canada, and France with regard to subject-verb inversion and insertion of *est-ce que*. This is why I do not work from data which come from different French-speaking countries and that it is claimed that geographic variation also represents a parasitic factor within a corpus.

2.3 Methodological variation

On top of these linguistic sources of variation, methodological diversity can also arise between corpora. Indeed, many factors can interfere in the way data are collected, transcribed, coded, and analysed: the era, the country, the theoretical background, the matter of the study, the type of investigation, the equipment, the persons in contact with the children, the protocols of research, the transcription conventions, etc. However, all these differences seem quite easy to identify and hence to cancel. It then sounds quite straightforward to postulate that data collected by one person or one team will not present these kinds of differences.

2.4 Outcome

To summarise so far, it has been argued that some sources of variation must absolutely be prohibited within a corpus if the aim is to provide broad generalisations on language acquisition. These sources of variation have been identified as being diachronic, geographic, and methodological. On the other hand, other types of variation are necessary within a set of data in order to guarantee its representativity. Indeed, it has been mentioned that individual variation (which in turn can include diastatic and diaphasic variation) can also represent a pitfall. Consequently, I favour an approach which takes a broad number of informants into account, since, as illustrated with Malcolm in Section 2.1, this methodology enables to identify peculiarities and filter them out.

Nevertheless, these conclusions are insufficient since many questions still need to be addressed, e.g. how many children to interview, how to select them in order to be representative, etc.? At this stage, the issue could then be summarised as follows: what kind of child group could be linguistically representative of a defined geographic area at a precise moment in time? And quite a satisfactory answer to this issue was eventually found. Indeed, it was considered to work from an already-constituted group, homogeneous in age, time, and space, and often regarded as generally representative of the adult population, i.e. an entire kindergarten class. Consequently, Corpus N°2, as analysed in Palasis (2009b), stems from

Introducing New French Child Data

interviews with a total of twenty-two children from a first-year class in a French *maternelle*. These children together with their two caretakers and myself were audio-recorded and video-taped in school over a period of seven months, as detailed in Table 2 hereunder.⁷

Language	L1 French (spontaneous)
Place	South of France
Dates	From Nov. 2006 to June 2007
Longitudinal study	7 months (to be continued)
Sessions	13
Intervals between sessions	Min.: 10 days; Max.: 40 days
Children	22
Ages first session	Between 2;5.5 and 3;4.24
Ages last session	Between 3;0.13 and 4;0.1
Adults	3
Recordings	Audio and video
Transcript tier	Orthographic
Coding tiers	Morphosyntactic (%mor) Non-target (%err)
Utterances (total)	30542
Utterances (children/adults)	17500/13042

Table 2: Database specifications (Palasis (2009b))

Additionally, all these recordings were transcribed and coded by one unique person along one single transcript and coding protocol, i.e. CHAT (Codes for the Human Analysis of Transcripts) as provided by the CHILDES database. Hence, these facts guarantee further homogeneity at the methodological level.

2.5 Towards an alternative to the competence vs. performance dichotomy

It is well-known that generative grammar does not favour such an approach based on language observation, i.e. “performance”.

7. These children were recorded and filmed over a total period of three years however Corpus N°2 in Palasis (2009b) only displays the data corresponding to the first year.

Katerina PALASIS

A statement such as the one found in Chomsky (1965:4) illustrates this rationale:

Observed use of language may provide evidence but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline.

On the other hand, the relative absence of data within generative research can be considered as one of the drawbacks to this theory:

[...] generative grammar has produced many explanatory hypotheses of considerable depth, but is increasingly failing because its hypotheses are disconnected from verifiable linguistic data. Issues of frequency of usage are by design made external to matters of syntax, and as a result categorical judgments are overused where not appropriate, while a lack of concern for observational adequacy has meant that successive versions have tended to treat a shrinking subset of data increasingly removed from real usage.⁸

After having read the above extract, one could imagine that Manning (2003) rather favours corpus linguistics such as the approach chosen by Tomasello (2003) with regard to acquisition. However, this is not the case either:

On the other side, corpus linguistics [...] or “usage-based models of grammar” [...] has all the right rhetoric about being an objective, falsifiable, empirical science interested in the totality of language use, but is failing by largely restricting itself to surface facts of language, rather than utilizing sophisticated formal models of grammar, which make extensive use of *hidden structure* (things like phrase structure trees and other abstract representational levels).⁹

As far as Tomasello (2006:5-6) is concerned, he claims that generative grammar is “more adult-centered” than usage-based

8. Manning (2003:296).

9. Manning (2003:296).

Introducing New French Child Data

theories of acquisition, which he contrastively defines as “more child-centered” approaches. These different statements illustrate that there is a deep methodological dichotomy between these two approaches, which in turn seems to hinge upon one main contrast, i.e. formal, symbolic-based theories such as generative grammar do not stem from the observation of many empirical facts whereas less symbolic theories, e.g. connectionism, emerge from broad data observation. Hence, as far as scientific conclusions are concerned, abstraction seems to manage without reality and *vice versa*.¹⁰ Moreover, this mutual exclusion is not just a current trend in linguistics; rather it is deeply rooted in theory since it brings us back to the seminal Chomskyan “competence” vs. “performance” and Saussurian “langue” vs. “parole” dichotomies.

Although theoretically essential and indispensable, it is argued that such dichotomies could be slightly qualified in order to avoid the mutual exclusion they entail. Indeed, I claim that both could be associated within a unique scientific reasoning thus taking up a third, intermediary stance which combines the formal, generative model of grammar with the observation of many empirical facts. More precisely, it is argued that abstract, theoretical hypotheses can arise from language observation.¹¹ The close scrutiny of Corpus N°2 for instance leads to split the Pro-drop Parameter into two morphosyntactic parameters and to argue that the so-called “null-subject” phenomenon is generated by principles of economy applied by the child to his/her linguistic system.¹² However, in the first part of this article, it was also argued that not all kinds of data can lead to such conclusions and that a corpus must meet certain criteria which were subsumed with one term, i.e. representativity. Therefore, such data can not be strictly related to “performance” anymore since, as mentioned above, these linguistic facts are controlled and filtered. One

10. Also see Scheer (to appear:chapter 13) for a historical overview on this opposition.

11. Also see Oliviéri (2009, 2010) on this matter with regard to syntax and dialectology.

12. See Palasis (submitted) and Palasis (2010) respectively for further details.

Katerina PALASIS

then obtains an intermediary level of access which includes a part of abstraction. Such an intermediary tier reminds us of the intermediary representational level within the Government & Binding framework, i.e. S(hallow)-Structure, insofar as this level was not the deepest in the representation but it was not the surface level either since it displayed abstract items such as traces.

3. Coding child data

The second step addressed in this article is the importance of data coding. Indeed, one obvious peculiarity of a large corpus is the high number of its utterances which makes it impossible to handle manually or by memory. Consequently, in order not to lose the benefit of the quantity, the information must then be processed in an efficient manner so that any aspect of the data can be extracted and analysed. Within the CHAT files, once the investigator has transcribed his/her data on a main tier, as illustrated in (4a) hereafter, he/she can then choose from a broad variety of dependent tiers which represent different coding domains, e.g. phonology, morphosyntax, syntax, intonation, etc. It is hence expected that each investigator will select one or two tiers from this list according to his/her domain(s) of research. The study of the acquisition of the syntactic subject has led to favour two particular tiers, i.e. %mor and %err, which are illustrated in (4b) and (4c) hereunder and which are then described in the following two sections.

(4) The transcript (Alan, 2;8.18):

(a) Main tier: et moi je suis [*] fait un escargot .

(b) First coding: %mor: conj|et,
pro:ton:dg:nom|moi&1S,
pro:cli:d:nom|je&1S,
v:aux|être&PRES&1SV [*],
v:mdllex|faire&PP&_MASC&_SING,
det|un&MASC&SING,
n|escargot&_MASC .

(c) Second coding: %err: suis = ai \$LEX \$\$SUB \$AUX

3.1 The morphosyntactic coding tier (%mor)

The first coding tier is morphosyntactic (%mor): it displays information on the nature of the items (nouns, pronouns, verbs, etc.) as well as information on various features such as person, gender, or case for instance.

More specifically, particular attention was devoted to the coding of the subject clitic and non-clitic items in this database, as illustrated in Table 3 hereafter for the first person forms *je* and *moi*. Indeed, these different codes show that these items were coded differently depending on the following conditions: (i) they are either clitic or strong pronouns (lines a and b vs. c, d, e, and f), (ii) they appear on their own (lines a and c) vs. along with a coindexed item (line b: ‘d’ in the code stands for ‘doubling’), (iii) the coindexed element is at the left of the clitic (line d: ‘g’ stands for *gauche*) vs. at its right (line e: ‘d’ for *droite*), or both (line f).

Forms	Codes & Examples
<i>j(e)</i> 'I'	a pro:cli:nom je&1S <i>j'</i> ai perdu l(e) chien . “I have lost the dog”
	b pro:cli:d:nom je&1S attends <i>je l'</i> ai <i>moi</i> . “wait I have it <i>I_{strong}</i> ”
<i>moi</i> 'I _{strong} '	c pro:ton:nom moi&1S <i>moi</i> 0 [*] veux ça les cartes . “ <i>I_{strong}</i> want that the cards”
	d pro:ton:dg:nom moi&1S <i>moi j(e)</i> suis jaune ! “ <i>I_{strong}</i> I am yellow!”
	e pro:ton:dd:nom moi&1S <i>j'</i> ai perdu <i>moi</i> . “I have lost <i>I_{strong}</i> ”
	f pro:ton:dg:nom moi&1S and pro:ton:dd:nom moi&1S <i>moi j'</i> ai pas encore fini <i>moi</i> . “ <i>I_{strong}</i> I have not yet finished <i>I_{strong}</i> ”

Table 3: The %mor coding of *je* and *moi*

Katerina PALASIS

Similar codes were also applied to the five other persons, hence accounting for a total of thirteen different clitics, i.e. *je, tu, il, expletive il, elle, on, ce, ça, nous, vous, polite vous, ils, and elles* whether these elements appear on their own (a total of fourteen different codes since expletive *il* appears preverbally as well as postverbally) or are coindexed with another item (a total of thirteen codes).¹³

3.2 The error coding tier (%err)

As illustrated with line c in the above table, young children's utterances include a number of non-target structures and elements. The database consequently also displays an "error" line (%err) which labels all the non-target utterances symbolised with [*] on the main tier and for which an accurate and levelled coding system was also devised. In order to be of interest to acquisitionists whatever their field of research, all the different types of non-target utterances were tagged on the main tier and further described on %err. Hence, this database provides the investigator with a total of 130 different codes which pertain to syntax (\$SYN) and morphology (\$MOR) as well as phonology (\$PHO) and lexicology (\$LEX). Each of these main codes is then complemented with one or more indications which describe the phenomena as precisely as possible, as illustrated in (5) hereafter.¹⁴

(5) Some of the 130 code combinations available on %err:

(a) Determiner omission:

NOE: oh la [: y+a] 0 [*] oiseau dedans !

"oh there's 0 bird inside!"

%err: 0 = un \$SYN \$LOS \$DET \$INDEF

(b) Gender substitution on a subject clitic:

ZOE: xxx moi il_i [*] a pris ça à moi Noémie_i.

"me he_i [*] has taken that from me Noemie_i."

%err: il = elle \$MOR \$SUB \$AGA \$PRO

13. See Palasis (2009b:228 et sqq) for the detail.

14. The whole list together with illustrations are available in Palasis (2009b:372-377).

Introducing New French Child Data

- (c) Unexpected liaison:
 NOE: les mamans zarrivent [*] .
 %err: zarrivent = arrivent \$ALL \$ADD
- (d) Addition of a consonant applying perseveration:
 ZOE: i(l) va arriver le Papa_Nonël [*] [: Noël] .
 “he will arrive Father Christmas”
 %err: Nonël = Noël \$PHO \$ADD \$PER \$CON
- (e) Substitution of an auxiliary:
 ALA: et moi je suis [*] fait un escargot.
 %err: suis = ai \$LEX \$SUB \$AUX

As far as the nominative clitic omission is concerned, Table 4 hereafter displays the total range of codes available. The examples in (6) further illustrate each possibility.

%err levelled codes	Persons	Missing elements
\$SYN \$LOS \$PRO \$SUBJ	unidentified	?
\$SYN \$LOS \$PRO \$SUBJ \$1S	1	je
\$SYN \$LOS \$PRO \$SUBJ \$2S	2	tu
\$SYN \$LOS \$PRO \$SUBJ \$3S	3 ref.	il, elle
\$SYN \$LOS \$PRO \$SUBJ \$IMPRS	3 expl.	il
\$SYN \$LOS \$PRO \$SUBJ \$DEM	3 dem.	ce
\$SYN \$LOS \$PRO \$SUBJ \$3P	6	ils, elles

Table 4: Coding of the subject clitic omissions

- (6) Examples of %err coding: the nominative clitic omissions:
- (a) TOM: 0 [*] veux [=? faut] remett(r)e ça .
 “(?) want to put that back”
 %err: 0 = ? \$SYN \$LOS \$PRO \$SUBJ
- (b) MAX: euh 0 [*] sais pas .
 “(I) don’t know”
 %err: 0 = je \$SYN \$LOS \$PRO \$SUBJ \$1S
- (c) ALA: 0 [*] as vu j' ai rangé .
 “(you) have seen I have cleaned up”
 %err: 0 = tu \$SYN \$LOS \$PRO \$SUBJ \$2S
- (d) TOM: 0 [*] veut pas manger .
 “(he) doesn’t want to eat”
 %err: 0 = il \$SYN \$LOS \$PRO \$SUBJ \$3S

Katerina PALASIS

- (e) CEL: 0 [*] faut pas la casser .
“(one) mustn’t break it”
%err: 0 = il \$SYN \$LOS \$PRO \$SUBJ \$IMPRS
- (f) EMA: 0 [*] est des feuilles .
“(these) are leaves”
%err: 0 = c' \$SYN \$LOS \$PRO \$SUBJ \$DEM
- (g) EKT: 0 [*] [?] mangent [?] des herbes !
“(they) eat grass”
%err : 0 = ils \$SYN \$LOS \$PRO \$SUBJ \$3P

3.3 Computerized Language Analysis (CLAN)

The two coding tiers described above then enable the investigator to submit very accurate requests to the CLAN programs (Computerized Language Analysis) in order to extract data. Moreover, the combination of both these tiers furnishes the researcher with efficient tools that allow him/her to access a broad range of phenomena directly in two concomitant ways. First of all, a host of characteristics in the child linguistic system can be grasped by analysing the frequency or co-occurrence of any of the items coded on the %mor tier (types of nouns, verbs, adjectives; which grammatical persons are uttered; which gender, number, or tenses are favoured; mastery of negation, etc.). Secondly, the %err coding then provides direct information on all the non-target utterances which bear on the selected matter, whether the utterances result from addition, substitution, or loss of an item and whether the mechanism resorts to phonology, morphology, syntax or the lexicon, as illustrated in (5) above.

4. Conclusions

By and large, as far as language acquisition is concerned, two different types of approaches are usually contrasted. Indeed, formal linguistics, e.g. generative grammar, are usually opposed to usage-based theories, e.g. connectionism, following two main criteria: (i) theoretical representations (abstract vs. actual) and (ii) observed use of data (scarce vs. broad). This article has argued in favour of a third, intermediary approach which associates formal and corpus linguistics insofar as I aim at formalising abstract, theoretical hypotheses within the latest

Introducing New French Child Data

generative framework using what can be considered as “broad” corpora from a generativist point of view. However, going through the different variables linguistic data can display (individual, collective, and methodological), it has also been claimed that quantity *per se* does not guarantee quality which is pinned down to representativity (Section 2). Moreover, quantity then requires efficient data processing. Section 3 hence details the morphosyntactic (%mor) and non-target (%err) coding grids applied to Corpus N°2 (Palasis 2009b) in order to furnish the investigator with as many holds as possible on the data, whether coarse or fine-grained.

The amount of data (30,542 utterances) together with the conditions under which this information was collected, transcribed, and coded provide the researcher with what is assumed to be a sound foundation stone for theoretical investigation. I hence feel that crossing these two coding tiers on information gathered from twenty-two different children has already allowed to shed new light on various long-standing matters such as the child null-subject phenomenon (Palasis (2010)) and the status of the nominative clitics in oral French (Palasis (2009a)). However, many other issues remain to be studied in order to forward an in-depth study and formalisation of the French child grammar which I name “spontaneous” French (Palasis (submitted)).

References

- Auger J. (1994). *Pronominal Clitics in Québec Colloquial French: A Morphological Analysis*. PhD Dissertation, University of Pennsylvania.
- Chomsky N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Cohen M. (1924). "Sur les langages successifs de l'enfant", *L'acquisition du langage oral et écrit* 31 (1993), 33-43.
- Crisma P. (1992). "On the Acquisition of Wh-Questions in French", *Geneva Generative Papers* 0, 1-2: 115-122.
- Dalbera J.-P. (2002). "Le corpus entre données, analyse et théorie", *Corpus* 1, 89-104.

Katerina PALASIS

- De Boysson-Bardies B. (1996). *Comment la parole vient aux enfants. De la naissance jusqu'à deux ans*. Paris: Editions Odile Jacob.
- De Cat C. (2005). "French Subject Clitics are not Agreement Markers", *Lingua* 115, 9: 1195-1219.
- Demuth K. (1996). "Collecting Spontaneous Production Data", in McDaniel D., McKee C. & Cairns H.S. *Methods for Assessing Children's Syntax*, Cambridge, MA: The MIT Press, 3-22.
- MacWhinney B. (2000a). *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume I: Transcription Format and Programs*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney B. (2000b). *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume II: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Manning C.D. (2003). "Probabilistic Syntax", in Bod R., Hay J. & Jannedy S. *Probabilistic Linguistics*, Cambridge, MA: The MIT Press, 289-341.
- Morgenstern A. & Parisse C. (2007). "Codage et interprétation du langage spontané d'enfants de 1 à 3 ans", *Corpus* 6 "Interprétation, contextes, codage", 55-78.
- Olivieri M. (2009). "Syntactic Parameters and Reconstruction", in Kaiser G.A. & Remberger E.-M. *Null-subjects, Expletives and Locatives in Romance*, Konstanz: Konstanz Working Papers of the Department of Linguistics, 27-46.
- Olivieri M. (2010). "From Dialectology to Diachrony", in Upton C. & Heselwood B. *Proceedings of Methods XIII: Papers from the 13th International Conference on Methods in Dialectology*, Frankfurt: Peter Lang, 42-52.
- Palasis K. (2005). *Problèmes d'acquisition et le Paramètre du Sujet Nul*. Mémoire de Master, Université de Nice-Sophia Antipolis.
- Palasis K. (2009a). "On the Variable Morpho-Syntactic Status of the French Subject Clitics", in Kaiser G.A. & Remberger E.-M. *Null-subjects, Expletives and Locatives*

Introducing New French Child Data

- in Romance*, Konstanz: Konstanz Working Papers of the Department of Linguistics, 47-62, <http://ling.uni-konstanz.de/pages/publ/arbeitspapiere.html>.
- Palasis K. (2009b). *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*. Thesis, Université de Nice-Sophia Antipolis.
- Palasis K. (2010). "Principles of Economy within the Child Speech: when the Nominative Clitic does not Surface", poster presented at *The Romance Turn IV. Workshop on the Acquisition of Romance Languages*, Tours, 25-27 August 2010.
- Palasis K. (submitted). "From Spontaneous to Normed French: *il* and *ne* as Conclusive Indicators of two Differently Parameterized Grammars", in *Selected Proceedings of Going Romance 23*, Amsterdam: John Benjamins.
- Rizzi L. (1992). "Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives", *Geneva Generative Papers* 0, 2: 102-114.
- Scheer T. (to appear). *A Lateral Theory of Phonology. Volume 2: How Morpho-syntax Talks to Phonology. A Survey of Extra-phonological Information in Phonology since Trubertzkoy's Grenzsignale*. Berlin: Mouton de Gruyter.
- Suppes P., Smith R. & Léveillé M. (1973). "The French Syntax of a Child's Noun Phrases", *Archives de Psychologie* 42, 207-269.
- Tomasello M. (2003). *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Cambridge, MA & London: Harvard University Press.
- Tomasello M. (2006). "Acquiring Linguistic Constructions", in Kuhn D. & Siegler R. *Handbook of Child Psychology: Cognition, Perception and Language, Volume I*, Hoboken, NJ: John Wiley & Sons, 255-298.

Corpus n°9 “Corpus Syntax” (2010), 33-51

2. Palasis, K. (2011). Français spontané et français normé : bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In O. Bertrand & I. Schaffner (Eds.), *Variétés, Variations & Formes du français* (pp. 55-72). Palaiseau: Les Éditions de l'École Polytechnique.

Français spontané et français normé : bien plus que deux variétés d'un même français.

Implications acquisitionnelles et didactiques

Katérina Palasis

Université de Nice-Sophia Antipolis, BCL UMR 6039 CNRS

Introduction

La présente étude s'intéresse à l'acquisition de la syntaxe chez les jeunes enfants francophones natifs; le cadre théorique est celui de la grammaire générative. Cet article présente quelques-uns des résultats de ces recherches dans la mesure où il est proposé que ces analyses éclairent sous un angle différent certaines des problématiques qui sont traditionnellement liées à l'étude et à l'enseignement de la variation en français.

1. La problématique

Nous savons tous que « le français » se réalise en fait sous des formes extrêmement diverses dans les productions des locuteurs francophones (qu'il s'agisse d'enfants, d'adolescents ou d'adultes) et que, par conséquent, la notion de « variation » occupe une place centrale dans l'étude, l'enseignement et la formalisation du processus acquisitionnel de cette langue. Ainsi, quel que soit le cadre théorique et la perspective adoptés, les descriptions de cette variation abondent. Il est néanmoins intéressant de noter que, malgré la diversité des points de vue, ces descriptions exploitent majoritairement une seule et même approche: la définition contrastive. Les « dimensions de la variation » proposées par Zribi-Hertz (2006), et partiellement reprises ci-dessous, illustrent notamment ce propos.

Tableau 1 : quelques dimensions de la variation (Zribi-Hertz (2006:3))

Dimensions	Dichotomies	
Norme	Standard	Non Standard
Medium	Écrit	Parlé
Sociologie	?Bourgeois	Populaire
Pragmatique	Formel	Informel

Dans le cadre de cet article, nous nous intéresserons à une dimension en particulier de cette variation : il s'agit de celle qui est basée sur la notion de norme et qui oppose les variétés dites « standard » et « non standard ». Cette opposition, qui rejoint parfois celle qui contraste l'« écrit » et le « parlé », a fait l'objet de nombreuses formalisations, tant

dans le cadre des grammaires traditionnelles (Wagner & Pinchon (1962) ou Grevisse & Goosse (2008), entre autres) que dans celui des études linguistiques (Lambrecht (1981) ou Blanche-Benveniste (2000) par exemple). Néanmoins, malgré ces perspectives différentes, toutes ces formalisations possèdent de nouveau un socle commun : celui de toujours décrire la variété « non standard » du français en prenant comme point de référence la variété « standard ». Ainsi, systématiquement, une implication est posée au niveau de la capacité du locuteur francophone à manier cette variation. En effet, si celui-ci utilise le français « standard », il peut alors également produire ce français « non standard ». L'existence de cette implication est illustrée ci-après par deux extraits de travaux en linguistique. Prenons tout d'abord la description de Lambrecht (1981). Cet auteur décrit le français « non standard » sur la base d'un certain nombre de caractéristiques qui définissent en premier lieu le français « standard » comme illustré en (1) ci-dessous. La terminologie (« élimination », « érosion », etc.) ainsi que les symboles (flèches unidirectionnelles) marquent clairement la relation d'implication relevée ci-dessus.

(1) Quelques traits du français « non standard » d'après Lambrecht (1981:6-7) :

- a. Élimination du clitique négatif *ne*:
il ne vient pas → *il vient pas*
- b. Érosion phonétique des pronoms clitiques :
il me voit → *i-m-voit*
- c. Remplacement de *nous* par *on*.
- d. Baisse de l'aspect « marqué » des dislocations.
- e. Élimination des déplacements dans les interrogations :
où vas-tu ? → *où tu vas ?* → *tu vas où ?*

Cette implication est également perceptible dans de nombreuses autres descriptions des français non-standard et parlé. Ainsi, la définition de Jakubowicz & Rigaut (1997:58) postule que la seconde variété est utilisée par des locuteurs qui maîtrisent la première :

Le premier [le français standard] correspond, *grosso modo*, au français écrit productif, le deuxième [le français parlé] est la langue utilisée familièrement, dans les échanges verbaux ordinaires, par les locuteurs qui maîtrisent par ailleurs le français standard.

2. Une discussion en trois points

Sur la base de l'étude de données spontanées enfantines (cf. Section suivante pour le détail), une discussion en trois points est alors proposée. Tout d'abord, la perspective acquisitionnelle envisagée ici oblige à considérer les descriptions traditionnelles de la variation et la relation posée entre les différentes variétés sous un angle radicalement différent. En effet, l'étude des données enfantines met en évidence la relation exactement inverse à celle évoquée ci-dessus puisque, chronologiquement, dans sa vie de locuteur, l'enfant produit d'abord des énoncés qui correspondent à la variété « non standard » du français et, seulement dans un deuxième temps, à des moments extrêmement divers, l'enfant produit des énoncés qui correspondent à la variété « standard » du français. Ce propos est illustré dans la Section II puis, dans la Section III, ces observations empiriques sont transposées vers le cadre théorique de la grammaire générative à des fins de formalisation. Je rejoins alors des préoccupations telles que celles décrites

par Zribi-Hertz (2006) sur la difficulté à formaliser la ou les « grammaire(s) interne(s) » d'un locuteur face à une telle hétérogénéité des réalisations. Dans ce cadre, je défendrai alors l'hypothèse que la variation provient du fait que le locuteur francophone active non pas une grammaire unique au sein de laquelle il y a de la variation mais deux grammaires distinctes, chacune avec ses paramètres propres. Cette hypothèse, basée sur des données acquisitionnelles, rejoint alors celles formulées par ailleurs par Zribi-Hertz (2009) et Massot (2010) sur l'existence d'une « diglossie francophone ». Ces réflexions récentes sur la formalisation de la variation en français ont alors des répercussions extrêmement intéressantes dans les domaines de la didactique et de l'enseignement du français, tant au niveau de la transmission des connaissances grammaticales en français à des jeunes natifs que dans le cadre d'un enseignement du français langue seconde. Une extension aux discussions proposées notamment dans Bertrand & Schaffner (2010) sur la norme est alors proposée en clôture de cet article.

3. Le corpus de référence

Les réflexions proposées ici sont issues de l'analyse de données enfantines recueillies auprès de vingt enfants francophones sur une période de sept mois (Palasis (2010a, 2010b)). Les caractéristiques principales de ce corpus sont reprises dans le Tableau 2 ci-dessous puis l'exemple (2) illustre la façon dont ces données ont été transcrites et codées selon les normes et grâce aux logiciels proposés par CHILDES (*Child Language Data Exchange System*)¹.

Tableau 2: le corpus de référence

Langue	« Français » (productions spontanées)
Suivi longitudinal	7 mois
Nombre de séances	13
Nombre d'enfants retenus	20
Âges à la première séance ²	De 2;5.5 à 3;4.24
Âges à la dernière séance	De 3;0.13 à 4;0.1
Nombre d'adultes	3
Enregistrements	Audio et vidéo
Transcription	Orthographique
Codage	Morphosyntaxique et agrammaticalités
Nombre total d'énoncés	30542
Nombre d'énoncés verbaux retenus	19568
Répartition énoncés enfantins/adultes	9748/9820

2

1. « Système d'échange de données linguistiques enfantines ». Cf. MacWhinney (2000a, 2000b) et le site internet à l'adresse <http://chilides.psy.cmu.edu/>.

2. Les âges sont données en années;mois.jours.

(2) Transcription et codage des données :

TOM: 0 [*] veut pas manger. (I.2, 135; 2;11.9)³

%mor: v:mdl|vouloir&PRES&3SV [*] adv:neg|pas v:inf|manger.

%err: 0 = il \$SYN \$LOS \$PRO \$SUBJ \$3S⁴

Le premier objectif consiste à présent à confronter les descriptions traditionnellement proposées de la variation en français avec la réalité de ces données enfantines.

I. Prendre en compte le processus acquisitionnel

Pour des raisons de place, uniquement deux des traits décrits par Lambrecht (1981) seront confrontés au corpus enfantin dans le cadre de cet article. Il s'agit des caractéristiques (a) « Élimination du clitique négatif *ne* » et (b) « Érosion phonétique des pronoms clitiques ». L'étude des autres caractéristiques a néanmoins également été effectuée et celle-ci permet d'arriver aux mêmes conclusions⁵.

1. « Élimination du clitique négatif *ne* »

Commençons par considérer la négation chez les jeunes enfants francophones. Le Tableau 3 ci-après indique que le corpus comporte 1311 énoncés verbaux négatifs; ce total représente 13,4 % de l'ensemble des énoncés verbaux du corpus. Trois marqueurs différents sont utilisés par les enfants mais le marqueur *pas* est de loin le plus fréquent des trois. En ce qui concerne le clitique négatif *ne*, les enfants à ce stade ne l'associent que sporadiquement avec *pas* et *plus*. En effet, *ne* n'apparaît que seize fois dans l'ensemble du corpus, ce qui représente un taux extrêmement bas de 1,2 % de l'ensemble des énoncés verbaux négatifs. Ces données enfantines permettent donc de constater empiriquement la quasi-absence du morphème discontinu *ne... X* chez les enfants francophones natifs de moins de 4 ans.

Tableau 3 : la négation dans le corpus de référence

Marqueurs	<i>pas</i>	<i>plus</i>	<i>jamais</i>	Totaux	%
X	1164	126	5	1295	98,8
ne X	7	9	0	16	1,2
Totaux	1171	135	5	1311	100,0

Intéressons-nous à présent de plus près aux détails des seize occurrences qui comportent le clitique négatif *ne*. Les énoncés correspondants sont repris en (3) ci-dessous.

3. Les indications données entre parenthèses à la suite d'une occurrence indiquent les numéros de séance, de fichier, de ligne puis l'âge de l'enfant.

4. \$SYN = syntaxe, \$LOS = omission, \$PRO = pronom, \$SUBJ = sujet, \$3S = troisième personne du singulier. Un total de 130 combinaisons différentes est proposé afin de rendre compte de toutes les occurrences non-cibles dans ce corpus (Palasis (2010a:372-377)).

5. Le détail de l'analyse se trouve dans Palasis (2010a).

- (3) Les seize occurrences de *ne* dans le corpus de référence :
- a. EMA: 10 occurrences du type :
 - EMA: il *ne* chante plus. (XI.41d, 1404; 3;2.14)
 - EMA: il *ne* saute plus. (XI.41d, 1411; 3;2.14)
 - EMA: il *ne* danse plus. (XI.41d, 1418; 3;2.14)
 - b. EKT: 1 occurrence :
 - EKT: le facteur *n'est pas* passé. (IX.36, 2126; 3;7.27)
 - c. LUS: 3 occurrences :
 - LUS: ta photo *n'est pas* là? (II.5, 145; 2;10.5)
 - LUS: non non ce *n'est pas* 0 [*] moi. (IV.19f, 1290; 2;11.30)
 - LUS: maman elle *n'est pas* là. (VIII.31a, 724; 3;2.7)
 - d. EVA: 2 occurrences :
 - EVA: lui i veut prendre la poule i mange il *n'a pas* gagné pour [/] pour [/] pour les oeufs. (XIII.46a, 1097; 3;10.13)
 - EVA: *ne* te cache pas avec la photo ! (XIII.46e, 1086; 3;10.13)

L'analyse de ces données entraîne la série de remarques suivante. Tout d'abord, il faut souligner le fait que, des vingt enfants enregistrés, seuls quatre en tout produisent ce clitique négatif ; il s'agit d'Emma, Hector, Lucie et Eva. Ainsi, entre trois et quatre ans, seul un enfant sur cinq a produit au moins une fois cette particule au cours des enregistrements. La deuxième remarque concerne le type d'énoncé dans lequel *ne* émerge. Il est en effet important de mentionner que, par rapport à l'ensemble des données de ce corpus (données qui relèvent dans leur très grande majorité d'énoncés produits spontanément par les enfants dans un contexte informel de jeu), onze des seize occurrences en (3) ci-dessus ne **représentent pas du véritable discours spontané**. En effet, les dix occurrences d'Emma ont été produites par la petite fille en chantonnant et l'unique occurrence d'Hector relève d'une comptine enfantine bien connue. Leur mode de production est donc particulier. Par ailleurs, ces occurrences possèdent une seconde spécificité par rapport à l'ensemble des données. Dans deux énoncés, l'émergence du clitique négatif *ne* est concomitante avec un autre phénomène très rare dans le corpus : l'absence d'un clitique nominatif. Il s'agit de l'énoncé d'Hector et du premier énoncé de Lucie. En effet, comme l'illustre le Tableau 4 ci-après, la très grande majorité des énoncés enfantins (les deux premières lignes du tableau, soit 80,9 % des énoncés) comporte un clitique nominatif. Le taux de 0,4 % sur la dernière ligne du tableau confirme que les sujets sans clitiques sont donc extrêmement rares à ce stade de l'acquisition. Or, il se trouve que cette structure particulière est concomitante, par deux fois, avec la production de *ne*. Bien que statistiquement inexploitable (car quantitativement bien trop limitée), cette corrélation de deux phénomènes rares représente néanmoins une particularité dans le langage enfantin qu'il convient de relever.

Tableau 4: les types de sujets potentiels dans le corpus

Types	N	%	Exemples
+ Clitique	5831	54,6	ah là j'ai fait un rêve
+ Clitique + Tonique ou Nom	2804	26,3	moi je prends le rouge

Types	N	%	Exemples
Impératif	1275	11,9	<i>regardez ma p(e)tite [/] p(e)tite cabane</i>
- Clitique - Tonique ou Nom	348	3,3	<i>euh 0 [*] sais pas</i>
+ qu rel	278	2,6	<i>c' est moi qui pioche !</i>
+ qu int	99	0,9	<i>c'est qui qu'a l(e) chat ?</i>
+ Nom	39	0,4	<i>alors Max prend la poule en chocolat</i>
Totaux	10674	100,0	

Si nous complétons à présent ces remarques basées sur l'étude transversale des données par une observation longitudinale du corpus (sept mois se sont écoulés entre la première et la dernière date), une troisième série de remarques vient alors compléter l'étude du clitique négatif *ne* chez ces jeunes enfants. Les données ont été enregistrées au cours de treize séances différentes. Le décompte par séance indique alors que **douze** des seize *ne* sont produits **au cours des trois dernières séances**. Ce fait confirme donc qu'il s'agit bien d'un phénomène naissant à ce stade du développement infantin⁶. De plus, si l'on regarde les données des quatre enfants qui ont produit ce *ne*, on constate qu'il y a un décalage systématique entre la production du marqueur négatif *X*, qui survient toujours en premier lieu, et celle de *ne... X*. Le Tableau 5 ci-après illustre ce propos en présentant les âges et les numéros des séances correspondants. Il convient néanmoins de préciser que ces données ne permettent pas de calculer le délai qui existe entre la production des deux items dans la mesure où les âges suivis du symbole (I) indiquent simplement que les enfants produisaient déjà *X* au cours de la première séance. Il n'y a donc aucun moyen de savoir depuis combien de temps, avant les enregistrements, ils produisaient *X*. Les éléments du Tableau 5 permettent néanmoins de circonscrire le processus acquisitionnel : (i) il existe une chronologie des événements entre une production initiale (*X*) et une production subséquente (*ne... X*) ; (ii) le décalage entre l'émergence de *X* et celle de *ne... X* est d'au moins un mois chez Lucie, quatre mois chez Hector, six mois chez Emma et sept mois chez Eva.

Tableau 5 : émergence de *ne* par rapport à *ne... X*

Marqueurs	Emma	Hector	Lucie	Eva
<i>X</i>	2;8.0 (I)	3;3.14 (II)	2;9.22 (I)	3;3.5 (I)
<i>ne... X</i>	3;2.14 (XI)	3;7.27 (IX)	2;10.5 (II)	3;10.13 (XIII)

Plaçons à présent les résultats de cette étude acquisitionnelle du français L1 dans la perspective descriptive de Lambrecht (1981). D'après les données étudiées, il semble

6. Ces mêmes enfants ont ensuite été enregistrés pendant deux années supplémentaires (jusqu'à leur entrée à l'école primaire). L'analyse de ces données subséquentes sera donc très intéressante et permettra de proposer une description complète non seulement de l'émergence du phénomène mais également de son développement avant l'entrée à l'école primaire.

donc que les jeunes enfants francophones produisent très majoritairement la variété « non standard » du français, la variété « standard » apparaissant uniquement dans seize énoncés chez quatre enfants sur vingt. Néanmoins, si l'on revient sur la terminologie utilisée par Lambrecht (1981) pour décrire ces variétés, nous remarquons alors que la formulation « Élimination du clitique négatif *ne* » ne peut pas être utilisée pour décrire le type de français produit spontanément par les enfants. En effet, nous avons vu que la terminologie ne correspond pas à la réalité du processus acquisitionnel dans la mesure où les enfants produisent d'abord *pas* et seulement ensuite *ne... pas*; il n'y a donc aucun processus cognitif enfantin qui « élimine » *ne* lorsqu'un enfant dit *pas*. Au contraire, le développement linguistique montre le processus exactement contraire : celui « d'ajouter » *ne* à partir d'un certain moment du développement et dans certaines conditions que nous avons commencé à cerner avec la concomitance du sujet sans clitique.

Par ailleurs, la prise en compte de la chronologie des acquisitions linguistiques oblige également à revenir sur le lien qui est traditionnellement posé entre ces deux variétés du français. Nous venons de constater que le jeune enfant semble produire la variété « non standard » du français. Par contre, le processus acquisitionnel vient totalement falsifier l'implication supposée entre les deux variétés dans la mesure où les jeunes enfants ne sont pas des locuteurs « qui maîtrisent par ailleurs le français standard ». L'étude de données acquisitionnelles oblige ainsi à inverser l'implication traditionnellement posée et il semble alors plus adéquat de postuler que le français « standard » est produit par des locuteurs francophones qui maîtrisent par ailleurs le français « non standard ». Se posent alors des problèmes de définition et d'appellation des variétés car on ne peut définir un système sur la base de caractéristiques qui, chronologiquement dans la vie du locuteur, ne lui sont pas encore connues (cf. Section III sur ces problèmes). Voyons à présent ce qu'il en est de la seconde caractéristique proposée par Lambrecht (1981).

2. « Érosion phonétique des pronoms clitiques »

L'« érosion phonétique » des clitiques nominatifs est une autre des particularités souvent mentionnée dans la littérature pour caractériser la variété « non standard » du français. Ainsi, l'une de ses illustrations les plus symboliques est *Personne i(l) m'aime* présentée avec des parenthèses autour du *l* indiquant l'aspect facultatif de la consonne. Puisque nous avons établi que le français produit par les jeunes enfants semble relever de la variété « non standard », les transcriptions du corpus enfantin devraient alors contenir de tels éléments entre parenthèses. Les 1688 occurrences du clitique nominatif *il* dans le corpus ont donc été étudiées, que le clitique soit pleinement réalisé [il] ou qu'il présente sa forme élidée [i].

Les données montrent alors un modèle d'ajustement phonétique de ce clitique nominatif extrêmement contraint chez les enfants dans la mesure où ceux-ci utilisent systématiquement **la forme pleine *il* devant voyelle et la forme élidée *i* devant consonne**. Les données reprises en (4) ci-dessous illustrent cette alternance. De plus, les données en (b) montrent que le phénomène n'est pas lié à la stricte adjacence entre le clitique nominatif et le verbe fléchi puisque l'ajustement est le même quel que soit le nombre d'items entre les deux.

(4) L'ajustement phonétique du clitique *il* (1688 occurrences) :

a) *il* + voyelle :

ANT: *il* est à moi. (I.3b, 68; 2;9.13)

b) *i* + consonne :

LUS: *i* veut manger. (I.3d, 1077; 2;9.22)

ROM: *i* nage. (VI.26d, 432; 3;0.7)

TOM: *i* m'a griffé! (II.9a, 1520; 2;11.23)

EMA: *i* t'en manque un. (III.10b, 964; 2;9.5)

EKT: *i* se met les chaussons. (IV.19c, 99; 3;5.8)

ALA: après *i* m le redonne. (X.40a, 1131; 3;2.18)

c) Les deux :

NOE: *i* pleure parc + que *il* est très malheureux. (X.40b, 194; 3;3.15)

Les données montrent ainsi que les formes *il* et *i* apparaissent concomitamment chez les jeunes enfants et que ce qui déclenche l'élision de la consonne n'est pas lié à la notion de variation mais au contexte phonique de cette consonne. Ainsi et de nouveau, la terminologie utilisée traditionnellement pour décrire la variété « non standard » du français n'est pas applicable à la description du français infantin. En effet, les enfants ne produisent pas initialement *il* pour ensuite élider la consonne; il n'y a donc aucun processus cognitif « d'érosion phonétique ».

Dans le corpus, il y a néanmoins onze exceptions au mécanisme d'ajustement qui vient d'être décrit, c'est-à-dire que l'on a onze cas où le clitique *il* est produit dans sa forme pleine devant consonne. Ces onze formes sont reprises en (5) ci-dessous.

(5) Les onze formes non élidées de *il* :

a) EMA: *il* ne chante plus. (XI.41d, 1404-1472; 3;2.14; 9 occurrences du même type)

b) EVA: lui *i* veut prendre la poule *i* mange *il* n'a pas gagné pour [l] pour [l] pour les œufs. (XIII.46a, 1097; 3;10.13)

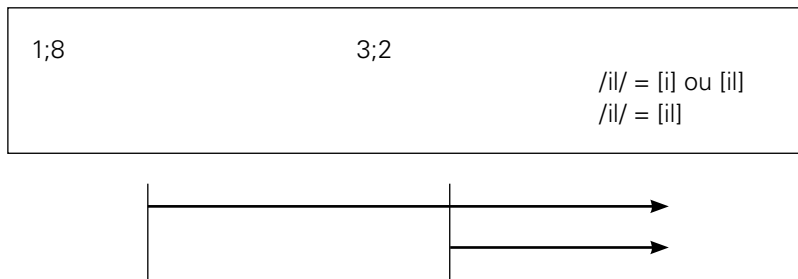
c) LUS: et après oh *il* dit <qui est là> [«] ! (VIII.31a, 947; 3;2.7)

Dix fois sur onze, *il* n'est pas élidé devant consonne lorsque cette consonne appartient à la particule négative *ne*. À la corrélation déjà dégagée entre l'émergence de *ne* et celle de sujets sans clitiques nominatifs, on peut alors ajouter une seconde particularité qui semble également présenter un lien avec la particule négative : la non-élision du clitique *il* devant consonne (cette non-élision n'est pas en rapport avec la consonne/n/comme le montre le second exemple en (4b) ci-dessus).

On a donc: (i) dans la très grande majorité des énoncés, de la première à la dernière séance, un clitique *il* qui présente un mécanisme d'ajustement phonétique extrêmement contraint et (ii) onze exceptions, mais à partir de la huitième séance seulement : dix devant la particule négative *ne* et une devant un verbe fléchi. L'hypothèse qui se dégage de cette partition est que l'on n'a pas affaire à des exceptions en (ii) mais à un deuxième constituant *il*, distinct du premier. Le premier constituant présente un ajustement phonétique contraint et est présent dès les premiers énoncés infantins (dès les premières productions de verbes fléchis, vers 1;8, comme illustré dans la Figure 1 ci-dessous). Le

second constituant, distinct du premier, ne présente pas de mécanisme d’ajustement et émerge plus tard dans le développement de l’enfant : dans ce corpus, l’âge le plus précoce est celui de Lucie (3;2.7), immédiatement suivie par Emma (3;2.14). Notons que Lucie et Emma font également partie des enfants qui commencent à produire la particule négative *ne* avant l’âge de quatre ans.

Figure 1 : deux éléments /il/ distincts



La dichotomie proposée ci-dessus rejoint alors un débat ancien en linguistique : celui du statut morphosyntaxique des clitiques nominatifs en français (cf. par exemple Heap & Roberge (2001)). En effet, deux hypothèses concurrentes coexistent : soit les clitiques sont des constituants syntaxiques, c’est-à-dire qu’ils sont considérés comme de véritables arguments du verbe au même titre que les syntagmes nominaux (dans ce cas, le clitique est alors un véritable pronom au sens étymologique du terme) ; soit les clitiques sont morphologiques, c’est-à-dire qu’ils sont considérés comme des préfixes verbaux et ils ne sont donc pas sujets syntaxiques mais marqueurs d’accord. Mon hypothèse, basée sur les données que l’on vient de voir, explore alors une troisième piste : celle de l’existence concomitante des deux ty pes de clitiques. Le clitique morphologique est le *il* qui émerge en premier dans le développement enfantin ; c’est celui qui présente le mécanisme d’ajustement phonétique contraint. Le clitique syntaxique est le *il* qui émerge en second et qui ne présente pas d’ajustement particulier.

Des mécanismes acquisitionnels se dégagent ainsi progressivement de l’observation de ce corpus. Formalisons à présent toutes ces remarques afin d’en tirer profit dans le cadre de la didactique.

II. Formaliser le processus acquisitionnel du français

1. Deux grammaires distinctes : français spontané et français normé

Un certain nombre de caractéristiques du français produit par les jeunes enfants francophones a ainsi été mis en avant, notamment en ce qui concerne la négation et les clitiques nominatifs. Ces deux traits sont repris sur les deux premières lignes du Tableau 6 ci-après. D’autres traits ont également été étudiés et leurs développements suivent le même type de chronologie que précédemment : le statut morphologique du clitique est ainsi corrélé avec une présence obligatoire de cet élément en position préverbale, ce qui

exclut la possibilité « d'inverser » le verbe et le clitique dans une interrogation.

Tableau 6 : deux faisceaux distincts de traits corrélés

Caractéristiques	Données initiales	Données plus tardives
Négation	X	ne... X
Clitiques	morphologiques	syntactiques
Présence obligatoire d'un clitique	oui	non
Inversion des clitiques	non	oui
Syntaxe	parataxe (à étudier)	hypotaxe
Quantification	rare (à étudier)	oui
Français	« spontané »	« normé »

Bien sûr, d'autres traits demandent encore à être étudiés afin d'étoffer cette description. Néanmoins, dès ce stade des investigations, il est notable qu'il existe des corrélations entre ces différentes caractéristiques. Ce fait est alors extrêmement important en vue d'une formalisation du processus cognitif, tant dans une perspective purement acquisitionnelle (enfants natifs avant l'entrée à l'école) que dans une perspective didactique (explicitation de la grammaire à des enfants natifs ; enseignement L2 à des enfants et adultes non-natifs). Cet aspect fondamental de toute grammaire est d'ailleurs mis en avant par Zribi-Hertz (2006:2) :

Les corrélations entre propriétés sont solidaires de la cohérence interne qui préside nécessairement à toute grammaire interne. Le principe de cohérence est la clef de l' « apprenabilité » des grammaires.

Les deux faisceaux de traits dans le Tableau 6 indiquent donc que la « variation » présente dans le discours des enfants francophones dès leur quatrième année peut en fait être formalisée sous la forme de deux grammaires distinctes que j'appelle français « spontané » (produit spontanément sans enseignement explicite) et français « normé » (issu d'une pression normative). Les appellations et les définitions de ces deux grammaires sont alors chacune positive et la grammaire initiale n'est plus définie sur la base de caractéristiques de la seconde. Je rejoins ainsi l'hypothèse défendue par Zribi-Hertz (2006, 2009) et Massot (2008, 2010) sur l'existence non pas d'un seul système linguistique qui présenterait des variations profondes mais plutôt de deux systèmes linguistiques à part entière, ce que ces chercheurs appellent la « diglossie francophone ».

2. Deux systèmes dont les concepts ne se superposent pas

Le Tableau 7 ci-après illustre alors le fait qu'il s'agit bien de deux véritables systèmes linguistiques, de richesses égales. En effet, toutes choses égales par ailleurs (comme l'illustrent les paires minimales x-x' dans le tableau), le nombre de combinaisons est le même pour les deux grammaires. Néanmoins, il est visible que les concepts ne se

superposent pas. En ce qui concerne la notion de sujet syntaxique, le français spontané a en effet recours à deux procédés qui ne s'appliquent pas en français normé : la préfixation verbale et les sujets silencieux (Palasis (2010a) pour le détail et les motivations sous-jacentes à ces mécanismes).

Tableau 7 : les sujets en français spontané et français normé

		Dislocation	Sujet	Verbe fléchi
Français spontané	a			idort
	b		lui	idort
	c		l'escargot	idort
	d		un escargot	idort
	e	l'escargot		idort
	f	*un escargot		idort
	g	l'escargot	lui	idort
Français normé	a'		il	dort
	b'		lui	dort
	c'		l'escargot	dort
	d'		un escargot	dort
	e'	l'escargot	il	dort
	f'	*un escargot	il	dort
	g'	l'escargot	lui	dort

Le double statut du clitique nominatif explique alors notamment pourquoi les jeunes enfants produisent des associations de syntagmes nominaux indéfinis et de clitiques (cf. (6b) ci-après), combinaisons qui sont considérées comme agrammaticales lorsque le clitique relève de la syntaxe. La production induite ci-dessous est alors très intéressante à cet égard. En effet, l'enseignante raconte et mime une courte comptine (6a) puis elle demande à un enfant de reproduire ce qu'il vient d'entendre et de voir (6b). En termes de sujet syntaxique, le français en (a) présente toutes les caractéristiques du français normé : syntagmes nominaux sans clitiques concomitants, clitiques syntaxiques et ellipses du pronom en coordination. La production d'Alan est alors révélatrice de la grammaire que le jeune enfant active lors de l'exercice puisque les sujets dans sa production relèvent tous du français spontané. À ce stade de son acquisition, Alan interprète donc correctement les clitiques syntaxiques qu'il entend mais ne produit que des clitiques morphologiques.

(6) Production induite d'une comptine mimée (I.1b) :

a) Modèle donné par l'enseignante :

SOF: *l'escargot* dort parc + qu'il fait beau dehors.

SOF: *un nuage* arrive et fait tomber la pluie !

SOF: *l'escargot* sort une corne deux cornes et s'en va se promener.

b) Production subséquente d'Alan (2;8.18) :

ALA: *l'escargot i* dort parc + qu'*i* fait beau dehors.

ALA: un *nuage il* arrive et xx la pluie !

ALA: deux cornes.

ALA: *i* s'en va se promener.

3. Paramétrage en grammaire générative

Une formalisation en grammaire générative trouve souvent son aboutissement dans le cadre de la théorie des Principes et des Paramètres (depuis Chomsky (1981)). Les Principes sont les caractéristiques communes à toutes les langues du monde (le Liage des pronoms par exemple). Les Paramètres sont les traits qui divergent d'une langue à une autre ; ils sont formulés sous forme binaire. Les différences entre les deux systèmes ainsi que la cohérence au sein de chaque système sont ainsi formalisées dans le cadre de cette théorie. Par contre, il est par ailleurs évident que ces deux français présentent également de nombreux points communs ; ceci est exemplifié par un autre paramètre, le premier du tableau, qui indique qu'aucun de ces deux systèmes ne possède une suffixation verbale distinctive (à la différence du grec par exemple).

Tableau 8: formaliser les différences en termes de Paramètres

Paramètres	Français spontané	Français normé
Suffixation verbale distinctive	–	–
Préfixation verbale	+	–
Tête syntaxique <i>ne</i>	–	+

Dans ce cadre, il est donc supposé qu'un locuteur francophone va activer l'une ou l'autre grammaire lors d'une production. La cohérence interne évoquée ci-dessus implique alors que les combinaisons d'éléments qui relèvent de deux grammaires différentes n'apparaîtront pas au sein d'une même production⁷. Deux prédictions sont alors proposées pour les énoncés négatifs. Ces prédictions, illustrées en 0 ci-dessous, partent donc du principe qu'un locuteur ne va pas utiliser des valeurs qui appartiennent à des grammaires différentes et, en effet, le corpus ne comporte aucun énoncé du type astérisqué ci-dessous.

7. La question de la définition de « production » est laissée en suspens ici mais cf. la discussion proposée dans Palasis (2010a:194-200) sur la difficulté à segmenter le continuum phonique de l'oral et cf. (5b) ci-dessus et l'hypothèse de Massot (2010) de considérer la proposition comme l'unité à l'intérieur de laquelle cette cohérence grammaticale s'applique.

(7) Prédications pour les énoncés négatifs :

a) Le clitique morphologique et la tête syntaxique *ne* ne peuvent pas apparaître ensemble :

* *ne i-chante pas*

✓ *i-chante pas* (Français spontané)

b) Le clitique syntaxique et l'absence de la tête syntaxique *ne* ne peuvent pas apparaître ensemble :

* *il chante pas*

✓ *il ne chante pas* (Français normé)

4. Récapitulons

Sur la base de l'observation de données enfantines, je postule que : (i) la variation, largement constatée par ailleurs en français, provient de l'existence de deux grammaires distinctes : celle du français spontané et celle du français normé ; (ii) la chronologie du processus acquisitionnel chez les enfants francophones indique que le premier français produit est le français spontané ; (iii) le français normé émerge dans un second temps mais à des moments très différents d'un enfant à un autre. La Section suivante vise à expliquer pourquoi.

III. La transmission de ces deux français

En termes de transmission, plusieurs cas de figure doivent être distingués. Commençons par nous pencher sur le français initial, le français spontané.

1. Le français spontané : « acquisition »

Le français spontané est donc la première grammaire à être acquise par tous les locuteurs francophones natifs dès le plus jeune âge. Les locuteurs francophones, ayant tous été des enfants, ont donc tous acquis les valeurs des paramètres de cette grammaire. **L'enfant grandit alors dans un premier temps en manipulant un premier système linguistique que je qualifie de « spontané » dans la mesure où les paramètres de cette grammaire sont fixés de façon inductive par l'enfant, grâce à son environnement linguistique mais sans que celui-ci ne reçoive jamais la moindre instruction explicite (ce qui est donc qualifié d'« acquisition » en grammaire générative et qui est opposé à « apprentissage »).** Les propos de Pinker (1994), repris ci-dessous, illustrent parfaitement le point de vue génératif sur cette question :

La connaissance tacite de la grammaire chez le petit enfant avant l'entrée à l'école est plus élaborée que le plus gros manuel de style ou que le langage informatique le plus avancé. (Pinker (1994:17))

Débarrassons-nous du folklore selon lequel ce sont les parents qui apprennent à parler à leurs enfants. [...] C'est aux enfants eux-mêmes qu'il faut imputer la plus grande partie du langage qu'ils acquièrent. En fait, nous pouvons démontrer qu'ils savent des choses qu'on ne peut pas leur avoir enseignées. (Pinker (1994:38-39))

En effet, nous avons vu que tous les enfants dans le corpus de référence produisent la négation avec un morphème unique (*pas*, *plus* ou *jamais*) et le clitique morphologique (*i/ll*). À ces deux caractéristiques, on peut également ajouter les faits suivants : les enfants ne produisent jamais le clitique *nous* mais utilisent *on* et ils formulent les interrogatives avec les éléments placés *in situ* (cf. le trait (1°) de Lambrecht (1981)).

Comment l'enfant acquiert-il cette langue ? Par immersion dans un bain linguistique particulier. Il est évident qu'un enfant qui grandit dans un milieu francophone ne va pas s'exprimer spontanément en chinois par exemple. Ainsi, tout est mis en œuvre, plus ou moins consciemment par l'environnement de l'enfant, pour que celui-ci « décode » les stimuli qu'il reçoit. Des études montrent en effet que les parents parlent d'une façon spécifique à leur nouveau-né en insistant notamment sur la mélodie des phrases (cf. par exemple De Boysson-Bardies (1996)). Ce type de langage fait partie de ce qui est appelé le L.A.E. (Langage Adressé aux Enfants). Veneziano (2000:246) le définit comme suit :

[...] le LAE diffère par plusieurs des caractéristiques du langage utilisé entre locuteurs déjà compétents. [...] les mères s'adressent à leurs enfants dans un langage clairement articulé, produit avec un ton de la voix plus élevé et une intonation exagérée. Le LAE utilise un vocabulaire plus limité et est en général syntaxiquement plus simple : les énoncés de la mère contiennent moins de pronoms de troisième personne, sont plus courts, contiennent moins de formes verbales composées et moins de subordonnées. Du point de vue de son contenu, il porte le plus souvent sur le « ici et maintenant » [...].

2. Français normé : « acquisition » ou « apprentissage » ?

Mais qu'en est-il pour le français normé ? Nous avons vu, qu'entre trois et quatre ans, il y a des différences importantes entre les enfants puisque, dans le corpus de référence, seuls quatre enfants sur les vingt produisent le morphème discontinu de la négation et que ces productions restent très parcimonieuses à ce stade. Nous avons également évoqué l'aspect très marqué dans le discours enfantin de la présence d'un syntagme nominal sans clitique en position sujet et, effectivement, seul un enfant sur les vingt utilise cette structure de façon relativement productive (il s'agit de Lucie qui, à elle-seule, produit dix-huit des trente-neuf énoncés). Une proportion très basse de l'ensemble des enfants active donc cette seconde grammaire à ce stade de leur développement. La question qui s'impose alors est bien évidemment : pourquoi une telle disparité ? Considérons les deux cas de figure séparément.

2.1. 1^{er} cas de figure : l'activation précoce des deux grammaires

À cet âge, vu que l'enfant n'a pas encore reçu d'enseignement explicite que ce soit pour la grammaire du français spontané ou celle du français normé, on peut supposer que cette deuxième grammaire est acquise de la même façon que la première. Le processus serait alors le suivant : l'enfant détecte qu'il existe une seconde grammaire et en fixe les paramètres grâce à son environnement linguistique. La condition *sine qua non* à l'acquisition du français normé est donc que l'enfant soit en contact avec cette langue, comme il l'est avec le français spontané. Néanmoins, la disparité entre les enfants du corpus montre qu'ils ne le sont pas tous. L'acquisition du français normé se ferait alors notam-

ment lorsque les parents de l'enfant (ou plus largement les personnes de son environnement quotidien) ont eux-mêmes une connaissance du français normé et lorsque l'enfant est exposé à un certain type d'émissions de télévision ou de radio et à des séances de lecture. En effet, je considère que la lecture d'histoires a un impact fondamental à ce niveau dans la mesure où le français que l'enfant entend au cours de ces séances par un adulte est clairement du français normé. De plus, lorsque l'on lit une histoire à un enfant, le rythme est généralement beaucoup plus lent que celui utilisé en français spontané et le lecteur a souvent tendance à exagérer l'intonation, l'emphase, etc. On retrouve alors dans ce type d'activité les mêmes moyens que ceux qui sont mis en œuvre dans le L.A.E. décrit ci-dessus pour la transmission du français spontané. Le français normé possède donc également un mode de transmission privilégié qui facilite la détection de ses caractéristiques. Les enfants qui ont accès à ce type d'informations peuvent alors paramétrer cette seconde grammaire de la même façon qu'ils paramètrent la première. Nous sommes alors dans une situation de « diglossie » ou de « bilinguisme »⁸.

2^e cas de figure: l'acquisition plus tardive du français normé

A contrario, un enfant qui n'aura pas été exposé à ce type d'environnement linguistique « diglosse » ou « bilingue » au cours de ses toute premières années arrivera à l'école maternelle avec une production du français normé moindre voire inexistante. Les données du corpus de référence le montrent. Ce type d'enfants se trouve donc dans une situation différente puisque le français normé fera effectivement partie de leur environnement linguistique de façon régulière et soutenue mais à partir d'un moment plus tardif de leur développement et limité au cadre bien précis de l'école. L'étude des enregistrements effectués pendant les trois années de maternelle permettra d'étudier avec précision le développement langagier de chacun. **Ainsi, si tous les enfants démarrent leur acquisition avec un système spontané, le moment auquel ils paramètrent les valeurs du système normé, la proportion et l'aptitude avec lesquelles ils manipulent l'un et l'autre sont extrêmement variables d'un enfant à un autre.**

En termes d'acquisition, sommes-nous encore dans une situation de « diglossie »/« bilinguisme » ou basculons-nous alors dans une situation d'« apprentissage d'une langue seconde » ? À ce stade, vu le jeune âge des enfants et vu le mode oral et encore très implicite de transmission, on peut postuler qu'il s'agit encore d'une forme d'acquisition. Néanmoins, on ne peut s'empêcher de supposer que les enfants de cette catégorie ne vont peut-être pas tous discriminer les caractéristiques du français normé avec la même acuité.

IV. Des enjeux didactiques majeurs

Des enjeux didactiques émergent alors. Serait-il possible, sur la base des productions d'enfants de quatre ou cinq ans, de repérer dès ce stade les enfants qui pourraient déjà éprouver des difficultés face au français normé ? Les enfants ne manipulant pas le fran-

8. Étymologiquement, ces deux termes sont similaires mais, pratiquement, le premier est plus restrictif dans la mesure où il implique notamment que les deux langues en question sont apparentées et qu'il y a une répartition sociolinguistique de l'utilisation des deux systèmes (cf. Ferguson (1959)).

çais normé à la maternelle seraient-ils plus enclins à éprouver des difficultés face aux enseignements à l'école primaire ?

Ces recherches visent à montrer que les réalités linguistiques du français spontané, langue manipulée par tous les locuteurs francophones dès leur plus jeune âge, ne se superposent pas avec les structures du français normé. Français spontané et français normé présentent donc des divergences profondes du point de vue cognitif à propos de notions qui *a priori* peuvent sembler très anodines telles que celles de sujet syntaxique, de pronom ou de négation. À mon sens, la simple reconnaissance du fait qu'il existe bien deux systèmes distincts serait un premier pas qui pourrait aider à rendre compte de la difficulté de certains enfants face à une langue qui est supposée être leur langue « première ». En effet, pourquoi certains enfants sont-ils plus à l'aise que d'autres avec les concepts grammaticaux qui sont explicités à l'école primaire ? La réponse de cette étude est qu'il ne s'agit pas de la langue « première » de tous les enfants francophones, malgré les apparences. Il peut donc y avoir un décalage profond entre ce que l'on postule comme étant « acquis » par tous les enfants qui entrent à l'école primaire et ce qui l'est réellement.

En effet, l'enseignement tel qu'il est proposé actuellement dispense des formalisations et une terminologie qui correspondent à un français bien particulier, le français normé. C'est ainsi que les enseignants de français langue première sont confrontés de façon récurrente à un paradoxe majeur. En effet, il est largement constaté que les enfants, les adolescents puis les jeunes adultes, de l'école à l'université, éprouvent des difficultés parfois non négligeables lorsqu'il s'agit d'intégrer et de manipuler des notions qui relèvent de l'analyse grammaticale du français qui, pourtant, est *a priori* leur langue « première ». Ces deux dernières années, ma propre expérience d'enseignante d'une unité de « Grammaire française » destinée aux étudiants de première année à l'université m'a confirmé que des étudiants pourtant francophones, pourtant en Faculté de Lettres, peuvent se retrouver à ce stade de leurs études avec une compréhension pour le moins « nébuleuse » de la grammaire du français, compréhension qui va souvent de pair avec un désintérêt parfois relativement prononcé pour ce type de disciplines et, au final, des résultats pour le moins médiocres et démotivants. Pour l'enseignant, cette situation peut alors déboucher sur un sentiment profond d'échec et d'impuissance qui peut trivialement être subsumé sous une formule du type : « c'est comme si je leur parlais chinois ». Les démonstrations effectuées en linguistique théorique sur les réalités d'une « diglossie francophone » tendent alors à montrer que cette formule n'est peut-être pas si triviale.

Conclusions

La « variation » en français représente depuis de longues années une problématique centrale tant dans la description grammaticale de cette langue, que dans son étude linguistique et son enseignement en tant que L1 ou L2. À travers l'observation minutieuse d'un corpus de français enfantin (vingt enfants entre 2;5 et 4;0), ces travaux visent à montrer que la linguistique (en l'occurrence, la syntaxe générative et l'acquisition du

langage) peut contribuer à la réflexion qui est d'ores et déjà menée par la didactique autour des thèmes liés de la variation et de la norme (cf. Bertrand & Schaffner (2010)).

Ces recherches ambitionnent alors de formaliser cette variation et elles permettent d'accorder un statut linguistique égal aux productions généralement définies négativement comme étant « non standard ». Dans ce cadre, il est alors proposé que la variation en français provient de l'existence de deux grammaires distinctes : le français spontané et le français normé. Les problématiques liées à la cohérence interne de chacune de ces grammaires, à leur acquisition/apprentissage et à leur mode de transmission sont également traitées. Ces recherches mettent alors au jour des pistes de réflexions aux enjeux didactiques importants notamment pour l'enseignement du français L1 à l'école primaire. Ainsi, en matière d'enseignement, la reconnaissance du français spontané en tant que système linguistique à part entière représente peut-être l'un des maillons manquants dans la chaîne de la transmission du savoir grammatical aux élèves en difficulté. Il est également espéré que la présente formalisation suscitera l'intérêt des enseignants et des didacticiens du français L2.

Références bibliographiques

- Bertrand Olivier, Schaffner Isabelle, (2010), *Quel français enseigner? La question de la norme dans l'enseignement/apprentissage*, Palaiseau, Les Éditions de l'École Polytechnique.
- Blanche-Benveniste Claire, (2000), *Approches de la langue parlée en français*, Paris, Éditions Ophrys.
- Chomsky Noam, (1981), *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht, Foris.
- De Boysson-Bardies Bénédicte, (1996), *Comment la parole vient aux enfants. De la naissance jusqu'à deux ans*, Paris, Éditions Odile Jacob.
- Ferguson Charles A., (1959), « Diglossia », *Word* 15, p. 325-340.
- Grevisse Maurice, André Goosse, (2008), *Le bon usage*, Bruxelles, De Boeck & Duculot (14^e édition).
- Heap David, Yves Roberge, (2001), « Cliticisation et théorie syntaxique, 1971-2001 », *Revue québécoise de linguistique* 30 (1), p. 63-90.
- Jakubowicz Célia, Catherine Rigaut, (1997), « L'acquisition des clitiques nominatifs en français », *Les pronoms: morphologie, syntaxe et typologie*, Zribi-Hertz, Anne (ed.), Saint-Denis, Presses Universitaires de Vincennes, p. 57-99.

- Lambrecht Knud**, (1981), *Topic, Antitopic and Verb Agreement in Non-standard French*, Amsterdam, John Benjamins.
- MacWhinney Brian**, (2000a), *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume I: Transcription Format and Programs*, Mahwah, NJ, Lawrence Erlbaum Associates.
- MacWhinney Brian**, (2000b), *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume II: The Database*, Mahwah, NJ, Lawrence Erlbaum Associates.
- Massot Benjamin**, (2008), *Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie: arguments morphosyntaxiques*, Thèse, Saint-Denis, Université Paris 8.
- Massot Benjamin**, (2010), « Le patron diglossique de variation grammaticale en français », *Langue française* 168 (4), p. 87-106.
- Palasis Katérina**, (2010a), *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*, Villeneuve d'Ascq, ANRT Diffusion.
- Palasis Katérina**, (2010b), « Introducing New French Child Data: Thoughts on their Gathering and Coding », *Corpus 9, La syntaxe de corpus*, p. 33-51.
- Pinker Steven**, (1994), *L'Instinct du langage*, Paris, Éditions Odile Jacob, 1999.
- Veneziano Edy**, (2000), « Interaction, conversation et acquisition du langage dans les trois premières années », *L'acquisition du langage. Le langage en émergence. De la naissance à 3 ans*, Kail, Michèle & Michel Fayol (éd.), Paris, Presses Universitaires de France, p. 231-265.
- Wagner Robert, Jacqueline Pinchon**, (1962), *Grammaire du français*, Paris, Hachette.
- Zribi-Hertz Anne**, (2006), *Français standard et français dialectal: pour une reconnaissance générale du phénomène diglossique pour la description et l'enseignement du français*, Manuscrit.
- Zribi-Hertz Anne**, (2009), « Pour un modèle diglossique de description du français: quelques implications théoriques, didactiques et méthodologiques », *Journal of French Language Studies* 2011

3. Palasis, K. (2012). Economy Principles within Child Speech: When the Nominative Clitic does not Surface. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 261-284). Newcastle: Cambridge Scholars Publishing, <https://www.researchgate.net/profile/Katerina-Palasis/publications>

ECONOMY PRINCIPLES WITHIN CHILD SPEECH: WHEN THE NOMINATIVE CLITIC DOES NOT SURFACE

KATÉRINA PALASIS

Throughout the years, one central objective of Generative Grammar has been to improve the descriptive and explanatory forces of its successive frameworks (e.g. Chomsky, 1993 for Minimalism). While doing so, economy has proven particularly pervasive, impacting on the evolution of representational, derivational, and computational devices (Chomsky, 1998). As far as acquisition is concerned, it is stated that “Economy Principles guide the language learner in constructing his or her internal grammar” (Gelderen, 2009:99). Relying on recently collected data from French children, this contribution presents an illustration of this process. Indeed, it is argued that the null subject phenomenon found in L1 typical acquisition of French can be accounted for in terms of computational economy. In turn, this hypothesis supports the morphological analysis of the nominative clitics in Colloquial French (e.g. Culbertson, 2010) and proposes to recast the Pro-drop Parameter (Chomsky, 1981).

The outline is as follows. Section 1 lays the theoretical and empirical background to this study. Section 2 examines the new data, focussing on the children’s verbal paradigms and their null subject utterances. Section 3 then details the analysis accounting for the null subject phenomenon in terms of quantitative and qualitative economy principles applied by the children in correlation with the morphological richness of certain verbal paradigms. Finally, Section 4 discusses these results in connection with the morpho-syntactic status of the nominative clitics and the Pro-drop Parameter.

1. Background

The child null subject phenomenon can be described as follows: children acquiring a non-null subject language such as French, English or German nevertheless go through a period until the approximate age of 3;0 when

they utter subjectless sentences, such as *veux pas ça* ‘(I) don’t want that’ (Zoé, 2;5).¹ As far as French is concerned, the omission rate amounts to circa one quarter of the children’s utterances (e.g. 26.2% in Rasetti, 1996 and 22.4% in Palasis, 2005).

1.1 Theoretical assumptions

Since the 1980s, a lot of seminal work has been carried out on this phenomenon (e.g. Hyams, 1986 and Rizzi, 1992 among many other references) within the very inspiring Principles & Parameters theory (Chomsky, 1981). Indeed, interlinguistic variation was then formalised in terms of Parameters and a partition was established with regard to subjects in finite clauses between the null subject languages, such as Italian, that license and identify a phonologically unrealized pronoun named *pro* and the non-null subject languages, such as French, that do not (e.g. Rizzi, 1997). Within this framework, French children thus fix the Null Subject/Pro-drop Parameter negatively.² However, the morpho-syntactic status of the nominative clitics represents a debated issue in Colloquial French, and hence in child French. Indeed, these elements are analysed either as proper arguments (Rizzi, 1986b and De Cat, 2005, among others) or as agreement markers (Roberge, 1990; Pierce, 1992; Auger, 1994; Culbertson, 2010, among others). In turn, the latter analysis questions the status of Colloquial French with regard to the Pro-drop Parameter since it classifies this type of French as a null subject language, on a par with Northern Italian dialects, such as Trentino, that also display preverbal clitics (Brandi & Cordin, 1989). Various aspects of the actual acquisitional process have also been debated, e.g. the initial value of the parameters and the moment they are set. This work follows Wexler (1998)’s hypothesis on early and correct setting. Alongside these points come the possible but debated correlation between the absence vs. the presence of a syntactic subject in a language and the richness vs. the poorness of its verbal morphology (Taraldsen, 1978) and general economy principles such as the “Avoid Pronoun” principle (Chomsky, 1981).

Minimalism has then particularly focused on economy and, as stated in Chomsky (1998:120), “excess baggage is shed”, e.g. parameters “are restricted to formal features of functional categories” (Chomsky, 1995:6) and parametric variation is limited “to a narrow category of morphological properties, primarily inflectional” (Chomsky, 1999:2); a crucial distinction is devised between interpretable and uninterpretable features (Chomsky, 1995); and the status of *pro* is questioned (e.g. Roberts & Holmberg,

2010). In a nutshell, it hence seems theoretically relevant to reconsider the child null subject phenomenon in the light of these assumptions.

1.2 Data

The second motivation to this investigation is empirical. Indeed, whenever new child data are available for analysis, it then becomes somewhat pressing to collate them with previous hypotheses. This research hence relies on two recent sets of spontaneous data collected with two different groups of French-speaking children from the South of France. Corpus #1 stems from five sessions of audio recordings with seventeen children between 2;3 and 3;1 and represents a total of 1,072 verbal utterances (Palasis, 2005). Corpus #2 comes from audio and video recordings with twenty other children between 2;5 and 4;0; it displays 9,748 verbal utterances and was gathered over a period of seven months during thirteen different sessions (Palasis, 2010a). Corpus #2 was further transcribed and coded along the CHILDES formats.³ It displays a main orthographical transcript tier associated with a morpho-syntactic coding tier (%mor) as well as an “error” tier (%err) accounting for all the children’s non-target utterances (Palasis, 2010b). Example (1) provides an illustration of these tiers together with an instance of the transcript and coding of the child null subject phenomenon in this French corpus.

- (1) The transcript and coding tiers in Corpus #2:
 MAX: euh 0 [*] sais pas .
 ‘er (I) don’t know’
 %mor: co|euh [*] v:mdllex|savoir&PRES&1SV adv:neg|pas .
 %err: 0 = je \$LOS \$PRO \$SUBJ \$1S

The present study analyses Corpus #2 in order to test and, if accurate, refine the initial conclusions forwarded thanks to Corpus #1 (Palasis, 2005; Palasis & Oliiviéri, 2007). The following sub-section hence briefly recalls what these conclusions were.

1.3 Previous outcome (Corpus #1)

Quite a few assumptions are relatively uncontroversial with regard to the acquisition of French syntax. For instance, it has been widely acknowledged that the first combinations of words emerge around age 18 to 20 months, that these sentences display finite verbs starting around age 20 months, that nominative clitics represent the majority of the subjects in

the early stages, that these clitics are mainly third person singular forms, and that they only surface together with finite verbs (e.g. Prévost, 2009 for details and the corresponding references).

Corpus #1 conforms to these characteristics, thus showing that this set of data is indeed representative of child French. Nevertheless these data also point to a characteristic that has not been particularly emphasized elsewhere, i.e. the peculiar architecture of the verbal system at this stage of acquisition compared to the French adult system. Indeed, despite the fact that the main verbs already correspond to those found in adult speech (e.g. Muller, 1979), Table 1 also shows that the children's full paradigms, being reduced ones in terms of different persons, display an overwhelming rate of morphologically distinctive verbal forms (72.9%), on a par with pro-drop languages such as Italian and hence contrary to what is generally assumed for non-pro-drop systems such as French. It is also highlighted that children actually utter very few Group 1 (regular) verbs at this stage, although approximately 90% of French verbs belong to this group.⁴ The subject omission rate corresponding to these paradigms is 22.4%.

Table 1: The child verbal system in Corpus #1 (IPA)⁵

		<i>être</i> 'be'	<i>vouloir</i> 'want'	<i>faire</i> 'do'	<i>avoir</i> 'have'	<i>savoir</i> 'know'	<i>voir</i> 'see'	Group 1 Verbs
Person	%	48.7	17.2	16.2	7.0	3.6	2.2	5.1
1sg	35.0	su i	v ø	fe	e	se	vwa	mãʒ
3sg	60.9	e	-	fe	a	se	vwa	mãʒ

In Palasis (2005) and Palasis & Oliiviéri (2007), the conclusions stemming from these data were: (i) the Pro-drop Parameter is set early to the correct value for French, i.e. <non-pro-drop language> (this accounts for the 77.6% sentences displaying a subject); (ii) however since the majority of the verbs in this reduced system display morphologically distinctive forms with regard to [Person], it is morphologically and computationally uneconomical to add systematic preverbal clitics to these verbal forms.

As far as the disappearance of the phenomenon is concerned, it was related to the fact that children gradually grow out of this reduced system, verb-wise and person-wise. Indeed, more and more different verbs are acquired, including group 1 verbs, which do not display distinctive 1sg and 3sg forms, and it is acknowledged that 2sg and 3pl forms then also gradually emerge. Children hence leave the null subject period when their

linguistic system gets closer to the adult's system. Corpus #2 will now give us indications as to whether these hypotheses are on the right track since the children are slightly older than in Corpus #1 (up to age 4). It is therefore expected that this second set of data will display the following characteristics: more different verbs, more different persons in the paradigms, and fewer null subjects. Additionally, the %mor and %err tiers should give us fine-grained indications as to the characteristics of the null subject utterances in this second corpus.

2. Further data (Corpus #2)

2.1 The verbal system

As far as verbs are concerned, Table 2 illustrates several discrepancies with Corpus #1. Indeed, whereas almost all the verbs present in Table 1 are still at the top of the list in Table 2, it can be noticed that their rates are different. This can be interpreted as the sign of a broader lexical diversification for this group of children, i.e. *être*, initially uttered as a superordinate verb, appears less often because it is replaced with an array of more specific verbs. Indeed, the table also shows that these children utter more different verbs than in Corpus #1, e.g. *aller*, *pouvoir*, *falloir*, and *mettre* enter the list. As far as persons are concerned, the facts are also different since 2sg forms appear in these data (6.3%) whereas Corpus #1 displays a low rate of 2.9% for 2sg. On the other hand, both sets of data are very alike as far as 3sg person is concerned: it remains the most frequently produced form in both corpora. Additionally, plural is equally scarce for both groups of children (Corpus #1: 1.2%; Corpus #2: 1.4%).

Table 2: The child core verbal system in Corpus #2 (IPA)

		<i>être</i> 'be'	<i>avoir</i> 'have'	<i>vou-loir</i> 'want'	<i>aller</i> 'go'	<i>faire</i> 'do'	<i>pou-voi</i> 'can'	<i>falloir</i> 'have to'	<i>voir</i> 'see'	<i>mettre</i> 'put'
Per	%	36.7	21.6	8.9	3.8	3.2	2.0	1.8	1.8	1.7
1	26.4	su i	e	vø	ve	fe	pø	-	vwa	me
2	6.3	e	a	vø	va	fe	pø	-	vwa	me
3	65.9	e	a	vø	va	fe	pø	fo	vwa	me

If we now contrast these verbal paradigms with those in Table 1, it can be noticed that the presence of 2sg forms makes a significant difference with regard to the richness of the verbal morphology within each

paradigm. Indeed, only three 1sg forms remain distinctive, i.e. *suis*, *ai*, and *vais*, together with the 3sg form of the defective verb *falloir*. The rate of distinctive forms hence drops to 14.01% in this system. In short, it can be concluded that the overall linguistic system of the children in Corpus #2 is broader than the system in Corpus #1, verb-wise and person-wise. Following the hypotheses mentioned in Section 1.3 above, it is then expected that the null subject phenomenon will have partly faded away in Corpus #2.

2.2 The null subject phenomenon

2.2.1 Extracting the relevant data

Table 3 shows the different types of constructions that can display a nominative clitic in Corpus #2 (subject relatives, subject interrogatives, and imperatives are hence left aside).⁶ These data are consistent with other child French corpora since nominative clitics represent the dominant category for subjects at this stage. Indeed, they are present in 95.7% of the utterances and break down into two subtypes: they can be uttered either alone (+Clitic: 64.6%) or with another element (+Strong Pronoun or +Noun: 31.1%). Within this system, the average rate corresponding to null subject sentences amounts to 3.9%. This figure is low, as expected from the above morphological analysis of the verbal system.

Table 3: The different types of subjects

Types	N	%	Examples	Glosses
+Clitic	5,831	64.6	ah là <i>j'</i> ai fait un rêve	oh there I had a dream
+Clitic +Strong Pronoun or Noun	2,804	31.1	<i>moi je</i> prends le rouge <i>c'</i> est une chenille <i>ça</i> et <i>le chat i</i> joue <i>i</i> veut venir <i>Téo</i>	me I take the red it is a caterpillar that and the cat it plays he wants to come Teo
–Clitic –Pron / –Noun	348	3.9	euh <i>0</i> [<i>*</i>] sais pas	uh don't know
+DP – Clitic	39	0.4	alors <i>Max</i> prend la poule en chocolat	so Max takes the chocolate hen
Total	9,022	100		

Table 4 provides the average rates of null subject sentences per session, showing a clear decrease of the average rate between the first and the last recordings. This is also an expected outcome.

Table 4: The decrease in null subjects, from first to last session (+7 months)

Session	1	2	3	4	5	6	7	8	9	10	11	12	13
Age range	2;5 3;4	2;5 3;5	2;6 3;5	2;7 3;7	2;7 3;7	2;8 3;8	2;9 3;8	2;9 3;9	2;10 3;9	2;11 3;10	2;11 3;11	2;11 3;11	3;0 4;0
Finite verbs	438	704	874	785	838	807	825	870	1022	1084	993	504	1037
Null clitics	23	25	54	35	34	26	31	29	24	22	29	5	11
%	5.25	3.55	6.18	4.46	4.06	3.22	3.76	3.33	2.35	2.03	2.92	0.99	1.06

While coding the data, particular attention was paid to the notion of subject. Consequently, fine-grained codes were devised for this particular corpus in order to be able to distinguish the following characteristics, as shown in (2) hereunder: (a) nominative clitic alone, (b) strong pronoun alone, (c) doubled clitic, (d) doubling strong pronoun to the left of the clitic, and (e) doubling strong pronoun to the right of the clitic. At this stage, no particular analysis is given for these elements with regard to the

morpho-syntactic status of the clitic and the notion of dislocation, “doubled/doubling” only meaning “not alone”.

- (2) The %mor coding scheme for subject related pronouns (examples for 1sg items only):
- pro:cli:nomlje&1S: *j' ai perdu l(e) chien* . ‘I have lost the dog’
 - pro:ton:nomlmoi&1S: *moi 0 [*] veux ça les cartes*. ‘Me want the cards’
 - pro:cli:d:nomlje&1S: attends *je l' ai moi* . ‘Wait I have it me’
 - pro:ton:dg:nomlmoi&1S: *moi j(e) suis jaune !* ‘Me I am yellow!’
 - pro:ton:dd:nomlmoi&1S: *j' ai perdu moi* . ‘I have lost me’

These %mor codes can then be crossed with the %err tier. This handling allows us to now scrutinize the phenomenon.

2.2.2 Identifying the contexts

First of all, the different types of sentences that are non-target as far as the subject is concerned can be described. Five contexts thus stem from these extractions, as shown in Table 5, and two main contexts emerge: the total absence of any kind of constituent (71.6%) and the ungrammatical presence of a preverbal strong pronoun (26.1%). The latter structure mainly features the 1sg form of the verb *vouloir* “want”, as exemplified in the table.

Table 5: Non-target subjects

Descriptions	<i>N</i>	%	Example	Glosses
No constituent	249	71.6	euh sais pas	hm (I) don't know
Preverbal strong pronoun	91	26.1	<i>moi</i> veux partir	me (I) want to leave
Postverbal strong pronoun	5	1.4	veux <i>moi</i>	(I) want me
Postverbal noun	2	0.6	oh est pas là <i>la grenouille</i>	oh (it) is not there the frog
Pre- and postverbal strong pronouns	1	0.3	<i>moi</i> veux regarder ça aussi <i>moi</i>	me (I) want look at that too me
Total	348	100.0		

In order to provide an appropriate description of the ungrammaticality in the above utterances, it is important to compare these sentences with their adult equivalents. Since children this age produce *spoken* language – as opposed to written language (Miller & Weinert, 1998; Palasis, to appear)– these utterances have to be compared with *oral* French only (e.g. Blanche-Benveniste, 1994). Their target equivalents hence stand as illustrated in (3).

- (3) Target equivalents to the child non-target subjects:
- a. euh *je* sais pas
 - b. moi *je* veux partir
 - c. *je* veux (ça) moi
 - d. oh *elle* est pas là la grenouille
 - e. moi *je* veux regarder ça aussi

In all cases, the phenomenon under scrutiny can hence be described as the absence of a nominative clitic. This is consistent with the oral adult system and the rest of the child system, as illustrated in Table 3. The traditional “null subject” appellation of the phenomenon then calls for two remarks. Firstly, it describes the absence of a syntactic function, hence not mentioning anything about the type of constituent involved. Secondly, it implies a specific theoretical analysis with regard to the status of the clitic (syntactic) and the position of the strong pronoun/noun phrase in the structure (dislocated). At this stage of this investigation, the appellation is deliberately kept neutral with regard to these long-standing debates in French linguistics (ever since Kayne, 1975) and the term “nominative null clitic” is preferred due to its descriptive adequacy whatever theoretical stance is taken on these matters.⁷

2.2.3 Identifying the persons

Pursuing the investigation thanks to the double-levelled coding, a second characteristic can be established with regard to the kind of person involved in this null clitic phenomenon. Indeed, the seven different %err codes provide the breakdown reported in Table 6. It is hence established that the omissions mainly imply 1sg and 3sg clitics and that they only affect singular forms (with only one very uncertain exception for 3pl).

Consequently, as far as the [Plural] feature is concerned, a correlation could be envisaged between the emergence of the plural clitics and the disappearance of the omissions. However, after having tested this correlation within the data, it was abandoned since it appears that children

can utter plural clitics and still omit singular ones. The fact that clitic omission is limited to singular forms could then simply be due to the fact that the whole system is mainly singular. Other leads were also followed and abandoned in Palasis (2010a), such as the correlation between clitic omission and child egocentrism (e.g. Piaget, 1923) since 62.93% of these omissions are 1sg. However, this hypothesis does not explain the remaining 37.07%.

Table 6: Breakdown of the omissions per person

Persons	Missing clitics	<i>N</i>	%	<i>N</i>	%
1sg	je	219	62.93	219	62.93
2sg	tu	11	3.16	11	3.16
3sg ref	il, elle	35	10.06	115	33.05
3sg expl	il	71	20.40		
3sg dem	ce	9	2.59		
3 pl	ils, elles	1?	0.29	1?	0.29
Indeterminate	?	2	0.57	2	0.57
Total		348	100.00	348	100.00

The initial hypothesis linking verbal morphology and nominative clitics hence remains plausible. We turn now to the verbal forms actually affected by clitic omission in Corpus #2.

2.2.4 Identifying the verbs

Table 7 presents the list and figures relating to the verbal forms that are actually involved in null clitic sentences. These figures are detailed per form and per child. It is interesting to note that *aller* and *faire*, which are ranked before *pouvoir* and *falloir* in **Erreur! Source du renvoi introuvable.** 2, are not affected by the phenomenon (this matter is addressed in Section 3.4 hereafter).

Table 7: Breakdown of the omissions per child and per verbal form

Verb	Children																	
	Q	L	LE	E	EK	T	I	TO	J	C	LU	N	A	R	AL	EM	M	Z
<i>veux</i>	9			1	1	17	1	42	3	1	3	2	2	23	20	6	10	2
<i>faut</i>		2	1	1	1		1	7	1	2	8	2	10		5	15		1
<i>peux</i>					1	1		12			1	2			3	1	5	
<i>est</i>						7	1	1			2			2	1	5	1	1
<i>a</i>					1	2		1				2	1		6	1		2
<i>sais</i>											4		1				2	1
Total	9	2	1	2	4	27	3	63	4	3	18	8	14	25	35	28	18	7

These figures are summarized in Table 8, which gives the proportions for the different verbal forms affected by clitic omission.

Table 8: Proportions per verbal form involved

Verbs	<i>veux</i> 1sg	<i>faut</i>	<i>peux</i> 1sg	<i>est</i>	<i>a</i>	<i>sais</i> 1sg
%	52.8	21.0	9.6	7.7	5.9	3.0

3. Outcome

3.1 The verbal paradigms: Overall analysis

Table 9 hereunder presents two types of information. First, it displays the total occurrences per verbal form (whether a clitic is associated with the form or not) for the verbs involved in clitic omission (as presented in Table 8). Secondly, the bold font in the table highlights the forms actually involved in clitic omission within each paradigm. For instance, within the 590 occurrences of *veux*, some of them appear with subject clitic omission.

Table 9: Total occurrences per verbal form (with and without a clitic)

Person	<i>vouloir</i> 'want'	<i>falloir</i> 'have to'	<i>pouvoir</i> 'can'	<i>être</i> 'be'	<i>avoir</i> 'have'	<i>savoir</i> 'know'
1sg	590	-	72	79	951	76
2sg	67	-	57	33	151	59
3sg	123	163	48	3110	803	8

Table 9 brings up two important observations: (i) clitic omission only affects one form per paradigm, (ii) the affected form corresponds to the most frequently produced form within each paradigm, e.g. for *vouloir* the most frequently produced form within the paradigm is 1sg *veux* and it is precisely this form, and only this one, that is affected by clitic omission. There is only one exception to this first generalisation, i.e. the *avoir* paradigm, since the most frequently produced form is the 1sg one, whereas the null clitic form is the 3sg one.⁸

On the whole, the clitic omissions correspond to the most frequently produced verbal forms within each paradigm. This first generalisation can be interpreted as a quantitative economy principle applied by the children to their system. This is consistent with the initial hypothesis as far as economy principles are concerned. We turn now to the morphological aspect of our hypothesis.

3.2 The verbal paradigms: Morphological analysis

As far as verbal morphology is concerned, three types of verbal paradigms are affected by clitic omission: one defective “paradigm” (*falloir*), two paradigms presenting two homophonous singular forms (*être* and *avoir*), and three paradigms displaying three homophonous singular forms (*vouloir*, *pouvoir*, and *savoir*).

3.2.1 *faut*

Being defective, verbal “paradigm” for the verb *falloir* ‘have to’ provides one perfectly distinctive and interpretable 3sg form [fo]. A straightforward economy principle can hence apply here, i.e. “avoid unnecessary material”, formulated as the “Avoid Pronoun” principle in Chomsky (1981). The clitic can easily be dropped and indeed Table 7 shows that this expletive is omitted by nearly all the children and that, for the older children (in the left-hand part of the table, apart for Q: Quentin), it is the

last clitic to remain omitted. This is not surprising owing to adult speech which also displays such an economy strategy, e.g.:

- (4) KAT: *0 faut fermer les yeux pour pas tricher* . (adult, Corpus #2)
 ‘(One) must close his eyes in order not to cheat’

Nevertheless, it has to be noted that not all the expletives are omitted in this system. Indeed, children only drop the proper “expletive” clitics (in the sense of Rizzi, 1986a), not the “quasi-argumental” ones found with the meteorological verbs hence illustrating the fact that children correctly analyse their language as a non-pro-drop system (according to the assumptions given in Section 1.1 and discussed in Section 4).⁹

3.2.2 *est* and *a*

The singular paradigms for *être* ‘be’ and *avoir* ‘have’ both display two homophonous forms, i.e. *est* (3sg) / *es* (2sg) [e] and *a* (3sg) / *as* (2sg) [a]. Table 10 shows that the children all omit the same clitic within each paradigm and that these omissions only occur with the most frequently produced of the two verbal forms.

Table 10: Homophonous verbal forms with corresponding null clitic rates

<i>être</i> ‘be’	Total <i>N</i>	Null Clitics %	<i>avoir</i> ‘have’	Total <i>N</i>	Null Clitics %
<i>t(u) es</i>	33	0.00	<i>t(u) as</i>	151	0.00
<i>(il) est</i>	3110	0.68	<i>(il) a</i>	803	1.99

Within each paradigm, the two homophonous forms are hence distinguished thanks to an economy strategy, i.e. producing a nominative clitic with only one of the two forms. This can be characterized as a qualitative economy principle on top of the quantitative one since it results into two distinctive utterances as far as [Person] is concerned: [te] ‘be 2sg’ ~ [e] ‘be 3sg’ and [ta] ‘have 2sg’ ~ [a] ‘have 3sg’. Indeed, this is sufficient and more economical than adding clitics to both forms (e.g. Oliiviéri, 2010 for similar conclusions in Occitan dialects).

3.2.3 *veux, peux* and *sais*

The singular paradigms for *vouloir* ‘want’, *pouvoir* ‘can’, and *savoir* ‘know’ all display three homophonous forms, i.e. *veux/veux/veut* [vø], *peux/peux/peut* [pø] and *sais/sais/sait* [se]. Table 11 shows that the behaviour of *savoir* matches that of *avoir* and *être*, i.e. only one clitic omission is found within this paradigm and it corresponds to the most frequently produced verbal form (1sg). In the output, the distinctiveness of the three forms is hence also preserved as within the smaller paradigms of *être* and *avoir*.

Table 11: Homophonous verbal forms with null clitic rates

<i>vouloir</i> ‘want’	Total <i>N</i>	Null Clitic %	<i>pouvoir</i> ‘can’	Total <i>N</i>	Null Clitics %	<i>savoir</i> ‘know’	Total <i>N</i>	Null Clitic %
<i>(je) veux</i>	590	24.24	<i>(je) peux</i>	72	36.11	<i>(je) sais</i>	76	10.53
<i>tu veux</i>	67	0.00	<i>(tu) peux</i>	57	10.53	<i>tu sais</i>	59	0.00
<i>(i) veut</i>	123	6.50	<i>(i) peut</i>	48	8.33	<i>i sait</i>	8	0.00

On the other hand, the data are less clear-cut for *vouloir* since two different clitics are omitted within the same paradigm (1sg and 3sg). As far as *pouvoir* is concerned, neither the quantitative nor the qualitative hypothesis seems tenable against the data insofar as all three clitics present omission rates within this paradigm. These data hence seem to falsify the initial hypothesis. Nevertheless, pursuing with a child-by-child survey of the data for *pouvoir* sheds different and interesting light on the matter and eventually enables us to disentangle the apparent falsification.¹⁰

3.3. The verbal paradigms: Child-by-child analysis

Looking into the detail of the *pouvoir* ‘can’ forms, it can be noted that 1sg null clitics are produced by eight different children (but essentially one, i.e. Tom) whereas the 2sg and 3sg omissions are exclusively due to one child, i.e. Maxime and Lucie, respectively (hence their absence in the preceding tables). Table 12 details the *pouvoir* paradigms for these three children.

Table 12: Three individual paradigms for *pouvoir*

Person	Tom		Maxime		Lucie	
	Total <i>N</i>	Null clitics <i>N</i>	Total <i>N</i>	Null clitics <i>N</i>	Total <i>N</i>	Null clitics <i>N</i>
1sg	20	12	8	5	5	1
2sg	3	0	9	6	7	0
3sg	5	0	2	0	11	4
Total	28	12	19	11	23	5

Tom's data exactly match the overall economy strategy proposed in the previous sections. Indeed, within his data, only one clitic is omitted (1sg) and this clitic corresponds to the most frequently produced verbal form within his paradigm (1sg). These findings are hence consistent with the previous overall analysis. Data also become clear-cut for Lucie in Table 12 since, within the *pouvoir* paradigm, she produced a majority of 3sg forms and it is shown that this is the most frequently produced form that lacks the clitic (with one exception).

On the other hand, Maxime's data still look rather confusing, since he omits 2sg clitics, which again correspond to the most frequently produced verbal form, but he also omits 1sg clitics nearly as often. Thus, the distinctiveness of each form does not seem preserved in Maxime's paradigm since [pø] can mean 'can 1sg' as well as 'can 2sg'. His data were then double-checked and an explanation for this duality was found in the corresponding %err codes. Indeed, person reversing is also coded within this database, as illustrated in (5) hereunder. It can be hypothesized that Maxime produces the verb inverting the theta-roles since the context clearly indicates that he is asking the addressee to give him the box, whereas he eventually utters a *je* 'I' and a null dative clitic. Consequently, what is initially coded as a 2sg clitic omission turns out to be a 1sg clitic omission.

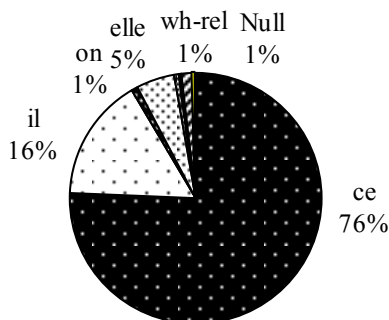
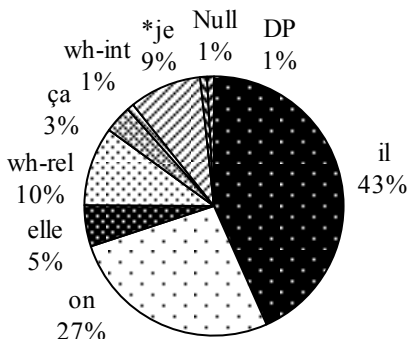
- (5) MAX: 0 [*] peux [/] *je* [*] peux 0 [*] donner la boîte ?
 '(You) can [/] I can give the box ?'
 %err: 0 = tu; je = tu; 0 = me

To sum up so far, quantitative and qualitative economy principles account for the null clitic phenomenon in the child system. This strategy is quantitatively economical as it is what would be the most frequently produced clitic that is dropped. This strategy is also qualitatively economical since it enables each child to provide only the strictly

necessary information, and no more, in order to preserve the distinctiveness of each verbal form within a paradigm. Apparent exceptions to these economy principles were found and accounted for above. In order to complete the analysis, a step is still missing though. Indeed, comparing Table 2 and Table 8 points to the fact that not all the verbs of the children's system are affected by the null clitic phenomenon.

3.4 The non-affected verbs

A comprehensive hypothesis has to account for the application of a mechanism as well as its non-application, if need be. Table 8 shows that *aller* 'go' and *faire* 'do' are not involved in the null clitic phenomenon. In order to account for this absence, two verbs and their nominative clitics are contrasted. The verbs are *être* 'be' and 'go' *aller* since, according to Table 2, they correspond to the most frequently produced affected and non-affected verbs in the child system. Table 8 further establishes that *être* displays clitic omission with its 3sg form. Clitics accompanying a 3sg verbal form are manifold, contrary to other persons. Indeed, [Person 3] and [Singular] features can surface as *il* 'he', expletive *il* 'it', *elle* 'she', *on* 'one', *ce* 'that', and *ça* 'that'.¹¹ Additional non-clitic pronouns can also surface with 3sg verbal forms, such as the wh- interrogative and the wh-relative form *qui*. Finally, DPs start to emerge at this stage (Table 3). Chart 1 hereunder provides the exact breakdown for these different elements with the 3rd person (Pe3) verbal form *est* 'is'. It can be seen that this particular verbal form is overwhelmingly associated with one clitic, i.e. *c(e)* (76%). Referential *il*, *elle*, *on*, and the relative *qui* come well behind in terms of frequency. It can then be stated that *c(e)* in this particular context is not very informative and that, whenever this clitic is dropped, it is quite easy to retrieve it owing to the rest of the system.

Chart 1: Pe3 elements with *est*Chart 2: Pe3 elements with *va*

This first picture is in stark contrast to the one displayed in Chart 2 for 3rd person (Pe3) *va* ‘goes’. Indeed, the latter presents the following discrepancies with the former: (i) the breakdown is more even (the maximum rate is 43% for *il* vs. 76% for *ce*), (ii) there are more different elements associated with this verbal form, (iii) children frequently associate *va* and 1sg **je*, and (iv) DPs emerge. Under these conditions, if one of these items were dropped, the loss of information would be much higher than with *est* ‘is’. Consequently, it is hypothesized that verbs like *aller* ‘go’ and *faire* ‘do’ are not affected by the null clitic phenomenon because they appear in contexts displaying a broader array of clitics (and hence of different features) as well as DPs. Such verbal forms, without any preverbal information, would hence not be fully “interpretable” and would cause the derivation to crash.

4. Discussion

The architecture of the child system around 3;0 is quite different from the adult one since children still produce only a few different persons and verbs, and few Group 1 verbs at this stage. A crucial effect of this architecture is the morphological richness of the verbal forms within the child system (Tables 1 and 2), which contrasts with that in adult French. This characteristic is given particular attention here since it is argued that the traditionally labelled “null subject” phenomenon is triggered by this morphological richness. Furthermore, these omissions are tagged “null nominative clitic” omissions upon comparison with their equivalents in the *spoken* adult language. They are accounted for by positing that quantitative and qualitative economy principles are in effect in the child

system, i.e. when a nominative clitic conveys morphological information already interpretable on the verb, the clitic can be omitted since it carries redundant information.

It is hence expected that these economy principles evolve with the architecture of the overall child system and, indeed, several stages can be identified: (i) Stage 1 (Corpus #1, up to circa 3;0): the verbal paradigms often reduce to two singular forms, i.e. 1sg vs. 3sg.. They are hence often fully distinctive with regard to [Person], the economy principles can apply quite often, and many clitics are dropped (average omission rate: 22.4%); (ii) Stage 2 (Corpus #2, between circa 3;0 and 4;0): the verbal paradigms broaden to include 2sg forms and additional verbs. The verbal paradigms are not as distinctive as at Stage 1. The economy principles can hence not apply as often since many paradigms now display two or three homophonous verbal forms. However, these principles remain active and some clitics are still dropped (average omission rate: 3.9%). Indeed, only one clitic is necessary to distinguish two homophonous verbal forms so one clitic out of the two is dropped and it is quantitatively more economical to drop the clitic corresponding to the most frequently produced verbal form. Applying the same principles, two clitics are necessary to distinguish three homophonous verbal forms so one clitic out of the three is dropped, as summarised in Table 13 hereunder.

Table 13: The actual paradigms (clitics and verbs, IPA)

<i>falloir</i>	<i>être</i>	<i>avoir</i>	<i>vouloir</i>	<i>pouvoir</i>	<i>savoir</i>
-	ʒəsqi	ʒe	ʒəvø	ʒəpø	ʒəse
-	tye	tya	tyvø	typø	tyse
ifo	ile	ila	ivø	ipø	ise

Furthermore, it is interesting to note that these economy principles are also active whenever a new paradigm enters a child system. For instance, when Antoine starts to use the three homophonous forms for *vouloir* ‘want’ in the past tense (at 2;10), he only omits one clitic out of the three and this clitic corresponds to the most frequently produced verbal form (1sg), as illustrated in Table 14 hereunder.

Table 14: Antoine and the verb *vouloir* ‘want’ in past tense

Utterances	Total	Null Clitics
<i>(je) voulais</i> ‘(I) wanted’	30	11
<i>tu voulais</i> ‘you wanted’	1	0
<i>i voulait</i> ‘he wanted’	3	0
Total	34	11

Finally, (iii) Stage 3 (prediction¹²): the null clitics disappear almost totally when the children’s verbal system broadens further to include a regular use of [Plural] forms as well as additional first-group verbs, hence resembling the adult system (with possible exceptions, as in (4)).

This investigation hence confirms the tight link observed in Corpus #1 between the morphological information regarding [Person] conveyed by the verbal form in a paradigm and the preverbal clitic. When the latter is redundant, it can be omitted. In Minimalist terms, a form such as *suis* ‘am’ for instance is thus perfectly “interpretable” with regard to [Person] as soon as it is merged into the derivation. Such a configuration thus presents two advantages, i.e. economy –less material and computation than with an additional subject– and optimality –fewer uninterpretable features. The latter consequence pertains to the discussion on the level of perfection of the system with the existence and diversity of these uninterpretable elements (e.g. Martin, 1999) and, more generally, of the Language Faculty (e.g. Chomsky, 2007). It therefore sounds sensible to provide a device in order to, if not dispense with, at least reduce their diversity. Furthermore, it is promising to note that these conclusions are also applicable to other languages due to the peculiar form third person singular often displays cross-linguistically (Benveniste, 1946) and due to the presence of the highly irregular verbs ‘be’ and ‘have’ in this reduced list. This work is thus consistent with Radford’s (2000) statement that children “seek perfection in the imperfect input they receive.”

Moreover, we can see how this hypothesis questions the status of nominative clitics within the child system since they surface in nearly all sentences, only preverbally, and their non-target omission is possible whenever the verb is morphologically rich enough to make the sentence interpretable as far as [Person] is concerned. Owing to these characteristics and others detailed in Palasis (2010a), it is therefore hypothesized that French children consider these clitics to be verbal prefixes. The morphological status of nominative clitics is further conceived to be one of the characteristics that distinguishes Spontaneous (colloquial) French –

acquired initially by all French natives– from Normed (normative) French –whose acquisition is much more unstable (Palasis, to appear). In order to account for this variation, the Pro-drop Parameter is split into two morphological parameters, i.e. the Verbal Prefixation Parameter and the Distinctive Suffixation Parameter. As shown in Table 15, this device does not lead to a binary partition as is the case for the Pro-drop Parameter. Rather, it aims to formalise microvariation, as illustrated with the pairs Spontaneous vs. Normed French and Normed Italian vs. Northern Italian dialects.¹³

Table 15: Splitting the Pro-drop Parameter

Parameters Grammars	Verbal Prefixation	Distinctive Verbal Suffixation	Examples	Gloss
Spontaneous French	+	–	<i>i-parle</i>	speak.3sg
Normed French	–	–	<i>il parle</i>	
Normed Italian	–	+	<i>parla</i>	
Trentino	+	+	<i>el-parla</i>	

In a nutshell: (i) all the French children acquire Spontaneous French as an L1; (ii) Spontaneous French displays verbal prefixation; (iii) verbal prefixation can be dropped under the morphological and computational economy principles described in this contribution. The next step would be to integrate this work into broader investigations on feature interpretability and economy principles since Gelderen (2008:297), for instance, states that interpretable features “are later reinterpreted as uninterpretable ones” and that this process applies to language acquisition as well as to language change.

Notes

¹ All the data come from Palasis (2005) and Palasis (2010a).

² The appellation of the parameter varies depending on the analysis (e.g. Rizzi, 1982:173,fn1). This matter is addressed in the course of the article.

³ Child Language Data Exchange System (MacWhinney, 2000 and the website at <http://childes.psy.cmu.edu/>).

⁴ In both corpora, the data are filtered in order to be representative, i.e. only the verbs uttered by at least three different children are included in the figures. See Palasis (2010b) with regard to generative grammar, data, and representativeness.

⁵ The data come from the South of France. This accounts for the transcription of the mid, closed, front vowel [e], e.g. [e] *est* ‘is’, instead of the standard [ɛ].

⁶ I would like to thank one of the reviewers for suggesting this way of looking at these figures.

⁷ In fact, “preverbal null clitic” would be the really neutral appellation. However, the term “nominative” is chosen here in order to distinguish these clitics from the other preverbal ones.

⁸ However, it has to be mentioned that the numerous [ja] forms in the corpus for *il y a* ‘there is/are’ are not included in the 803 total since there is no evidence that when a child utters [ja], he/she is handling a nominative clitic and the 3sg form of the verb *avoir*. This form was hence counted separately and amounts to an additional 285 occurrences.

⁹ Also see Tuller (2004), where these facts are presented and analyzed in adult French.

¹⁰ The analysis for *vouloir* is not detailed here since it presents a similar outcome.

¹¹ *Ça* can be clitic or strong, e.g. *ça ça va ici* ‘this it goes here’ (Palasis, 2010a:38-39).

¹² The same twenty children were recorded and videotaped between 4;0 and 6;0. These additional data, once coded and analysed, will tell us whether this prediction is borne out or not within this period.

¹³ The discrepancy between *i* and *il* is deliberate since only Spontaneous French shows a strict accommodation pattern of *il* to the phonological environment (Palasis, 2009).

References

- Auger, J. (1994). Pronominal clitics in Québec colloquial French: A morphological analysis. Doctoral dissertation, University of Pennsylvania.
- Benveniste, E. (1946). Structure des relations de personne dans le verbe. In *Problèmes de linguistique générale, 1*, pp. 225-236. Paris: Editions Gallimard, 1966.
- Blanche-Benveniste, C. (1994). Quelques caractéristiques grammaticales des “sujets” employés dans le français parlé des conversations. In M. Yaguello (Ed.), *Subjecthood and subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"*, pp. 77-107. Paris: Editions Ophrys.
- Brandi, L. & Cordin, P. (1989). Two Italian dialects and the null subject parameter. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter*, pp. 111-142. Dordrecht: Kluwer.
- Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- . (1993). A Minimalist Program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20: Essays in linguistics in Honor of Sylvain Bromberger*, pp. 1-52. Cambridge, MA: MIT Press.
- . (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- . (1998). Some observations on economy in Generative Grammar. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax*, pp. 115-127. Cambridge, MA: MIT Press.
- . (1999). Derivation by Phase. *MIT Working Papers in Linguistics* 18 ; reprinted in M. Kenstowicz (Ed.), *Ken Hale: A life in language* (2001), pp. 1-52. Cambridge, MA: MIT Press.
- . (2007). Approaching UG from below. In U. Sauerland & H.-M. Gartner (Eds.), *Interfaces + Recursion = Language?: Chomsky's Minimalism and the view from syntax-semantics*, pp. 1-29. Berlin: Mouton de Gruyter.
- Culbertson, J. (2010). Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86: 85-132.
- De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua* 115: 1195-1219.
- Gelderen, E.v. (2008). Where did late Merge go? Grammaticalization as feature economy. *Studia Linguistica* 62: 287-300.

- . (2009). Feature economy in the linguistic cycle. In P. Crisma & G. Longobardi (Eds.), *Historical syntax and linguistic theory*, pp. 93-109. Oxford: Oxford University Press.
- Hyams, N.M. (1986). *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- Kayne, R.S. (1975). *French syntax. The transformational cycle*. Cambridge, MA: MIT Press.
- MacWhinney, B. (2000). *The CHILDES project. Tools for analyzing talk (3rd edition)*. Mahwah, NJ: Lawrence Erlbaum.
- Martin, R. (1999). Case, the Extended Projection Principle, and Minimalism. In S. D. Epstein & N. Hornstein (Eds.), *Working Minimalism*, pp. 2-25. Cambridge, MA: MIT Press.
- Miller, J. & Weinert, R. (1998). *Spontaneous spoken language. Syntax and discourse*. Oxford: Oxford University Press.
- Muller, C. (1979). Les verbes les plus fréquents du français. In C. Muller (Ed.), *Langue française et linguistique quantitative. Recueil d'articles*, pp. 363-370. Genève: Editions Slatkine.
- Oliiviéri, M. (2010). From dialectology to diachrony. In C. Upton & B. Heselwood (Eds.), *Proceedings of Methods XIII: Papers from the 13th International Conference on Methods in Dialectology*, pp. 42-52. Frankfurt: Peter Lang.
- Palasis, K. (2005). Problèmes d'acquisition et le Paramètre du Sujet Nul. Mémoire de Master, Université de Nice-Sophia Antipolis.
- . (2009). From morphological affixes to syntactic arguments: What forces children to reinterpret nominative French clitics. Talk given at Going Romance XXIII, Nice, 3-5 December 2009.
- . (2010a). *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*. Villeneuve d'Ascq: ANRT Diffusion.
- . (2010b). Introducing new French child data: Thoughts on their gathering and coding. *Corpus 9* "La syntaxe de corpus": 33-51.
- . (to appear). Français spontané et français normé : bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In O. Bertrand & I. Schaffner (Eds.), *Variétés, variations & formes du français*, pp. 55-72. Palaiseau: Les Editions de l'Ecole Polytechnique.
- Palasis, K. & Oliiviéri, M. (2007). Beyond deficiency. Reconsidering the null subject phenomenon. In S. Baauw, J. Van Kampen & M. Pinto (Eds.), *The acquisition of Romance languages. Selected Papers from The Romance Turn II 2006*, pp. 127-143. Utrecht: LOT.

- Piaget, J. (1923). *Le langage et la pensée chez l'enfant. Etudes sur la logique de l'enfant*. Neuchâtel: Delachaux & Niestlé.
- Pierce, A. (1992). *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht: Kluwer.
- Prévost, P. (2009). *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam: John Benjamins.
- Radford, A. (2000). Children in search of perfection: Towards a Minimalist model of acquisition. *Essex Research Reports in Linguistics* 34.
- Rasetti, L. (1996). Null subjects and root infinitives in the child grammar of French. *Geneva Generative Papers* 4: 120-132.
- Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris.
- . (1986a). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17: 501-557.
- . (1986b). On the status of subject clitics in Romance. In O. Jaeggli & C. Silva-Corvalan (Eds.), *Studies in Romance linguistics*, pp. 391-419. Dordrecht: Foris.
- . (1992). Early null subjects and root null subjects. In L. Rizzi (Ed.) *Comparative syntax and language acquisition*, pp. 299-339. London: Routledge, 2000.
- . (1997). A parametric approach to comparative syntax: Properties of the pronominal system. In L. Haegeman (Ed.), *The new comparative syntax*, pp. 268-285. Harlow: Longman.
- Roberge, Y. (1990). *The syntactic recoverability of null arguments*. Kingston & Montreal: McGill-Queen's University Press.
- Roberts, I. & Holmberg, A. (2010). Introduction: Parameters in Minimalism Theory. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (Eds.), *Parametric variation: Null subjects in Minimalism Theory*, pp. 1-57. Cambridge: Cambridge University Press.
- Taraldsen, K.T. (1978). *On the nominative island condition, vacuous application, and the that-trace filter*. Bloomington, IN: Indiana University Linguistics Club.
- Tuller, L. (2004). Il 'impersonnel' et son absence en français parlé. In N. Rossi-Gensane (Ed.), *Mélanges en l'honneur de Nicole Gueunier*, pp. 229-242. Tours: PUFR.
- Wexler, K. (1998). Very early parameter setting and the Unique Checking Constraint: A new explanation of the optional infinitive stage. *Lingua* 106: 23-79.

4. Palasis, K. (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies*, 23(1), Numéro spécial 'L'hypothèse d'une diglossie en France', Massot, Benjamin & Paul Rowlett (Eds.), 17-35, <https://doi.org/10.1017/S0959269512000348>

*The case for diglossia: Describing the emergence of two grammars in the early acquisition of metropolitan French**

KATERINA PALASIS

BCL UMR 7320 CNRS-Université Nice de Sophia Antipolis

(Received January 2012; revised August 2012)

ABSTRACT

This article supports the diglossic approach to variation in metropolitan French by delving into the subject from the point of view of acquisition. Drawing on naturalistic data from 37 native French children between the ages of 2;3 and 4;0, the investigation exemplifies the existence of two cognate, but distinct grammars in the mind/brain of these children. The distinction between Spontaneous French (G_1 , all children) and Normed French (G_2 , 4 children by age 4) hinges upon two crucial characteristics, i.e. the morpho-syntactic status of nominative clitics and the emergence of the negative particle *ne*. Accusative clitics with imperatives and past-participle agreement are also examined in order to gain a comprehensive picture of the two grammars. Finally, the emergence of *ne* is interpreted as a trigger forcing a speaker to move from G_1 to G_2 due to the total unavailability of *ne* in G_1 .

INTRODUCTION

Variation is widely acknowledged in metropolitan French. For example, negation with or without the negative particle *ne*, and subjects with or without a resumptive clitic represent two emblematic and uncontroversial areas of variation in this language (Ashby, 1981; Lambrecht, 1981; Zribi-Hertz, 1994; Rowlett, 1998; Blanche-Benveniste, 2000; Coveney, 2002; Barra-Jover, 2004; Massot, 2008, among others). Although research agrees on the existence of variation, it is still a matter of lively debate as to how to formalise it. Two opposing stances are emerging, i.e. variationism and diglossia. The former considers French to be a single grammatical system with internal variation. The language is described as 'a conjunction of social and linguistic variableness' where 'variation [...] is better characterized

* I would like to thank Benjamin Massot, Paul Rowlett and two anonymous reviewers for their very constructive comments on earlier versions of this article. This research was supported by the French-German ANR-DFG grant awarded to the project 'Dialectal, acquisitional, and diachronic data and investigations on subject pronouns in Gallo-Romance' (DADDIPRO, 2012-2015, no. ANR 11 FRAL 007 01) led by Michèle Oliviéri (BCL UMR 7320 CNRS - Université de Nice Sophia Antipolis) and Georg Kaiser (University of Konstanz).

as style-shifting (rather than 'diglossia') (Gadet, 2003: 5 and Coveney, 2011: 78, respectively).¹ The latter follows Ferguson's (1959) original sociolinguistic formulation for Arabic, Greek, Swiss German and Haitian Creole, and suggests that what is traditionally labelled 'French' actually spreads over two distinct grammars (Barra-Jover, 2004, 2010; Rowlett, 2007; Massot, 2008, 2010; Zribi-Hertz, 2011). Ferguson's approach further implies that the two grammars are complementary. Schematically, one grammar – the 'high variety' (H) – is activated in formal situations, often (but not always) when writing, whereas the other grammar – the 'low variety' (L) – is used in informal situations (but see Buson and Billiez, this volume on some drawbacks of this dichotomy). Furthermore, variationism and diglossia make very different predictions about variation within each utterance. Indeed, the former 'predicts that all logically possible combinations of variants are in principle available, and [...] that speakers can freely combine marked H and marked L variants' whereas the latter predicts that 'each utterance 'chooses' one or the other grammar and remains consistent throughout' (Rowlett, 2011: 15 and Massot, 2010: 98, respectively).²

The distinction based on linguistic characteristics and distinct activation situations entails an additional dichotomy with regard to the speakers who are actually diglossic. Indeed, since 'the actual learning of H is chiefly accomplished by the means of formal education' (Ferguson, 1959: 331), the ability to activate H can be attributed to a subgroup of speakers only, i.e. the 'educated' ones (Massot, 2010: 89). Accordingly, it is further assumed that 'non-educated' adults and to an even higher extent preschool children should not be actively diglossic. This exclusion however is too strong since 'a 5-year-old French-speaking child might have started to actively learn SG [Standard Grammar] without schooling – in a family concerned with the transmission of SG' (Zribi-Hertz, 2011: 241).³ Nonetheless, it is generally expected that French-speaking children will initially produce utterances whose characteristics (almost) all belong to L and, only at a second stage closely identified with schooling, should these children generate an increasing number of utterances pertaining to H since 'adults invariably use L in speaking to children and children use L in speaking to one another' (Ferguson, 1959: 331). However, to the best of my knowledge, a clear picture of the emergence of this French diglossia is still missing.

This article therefore proposes to delve into the subject from the point of view of acquisition, by describing and formalising the emergence of diglossia in 37 speakers of metropolitan French between the ages of 2;3 and 4;0. The main aim is to exemplify the existence of two distinct grammars in the mind/brain

¹ The first part of the quotation is a free translation of 'une rencontre de variabilités sociales et langagières'.

² The latter is a free translation of 'chaque énoncé 'choisit' l'une ou l'autre grammaire et reste cohérent'.

³ Free translation of 'un enfant francophone de 5 ans peut avoir commencé son apprentissage actif de GS [grammaire standard] hors de l'institution scolaire – dans une famille soucieuse de lui transmettre GS'.

of these children by highlighting clustering as well as incompatibility effects for each grammar. The outline of the paper is as follows. Section 1 introduces the two sets of naturalistic data underlying this work. Section 2 then examines one of the major characteristics relevant to variation in French, i.e. subjects. The investigation is closely related to the long-standing and still lively debate surrounding the morpho-syntactic status of nominative clitics in colloquial French. I argue that early clitics are generated as verbal prefixes in an initial grammar of French common to all L1 speakers and termed Spontaneous French (G_1). Object clitics and past-participle agreement are also examined in this section in order to gain a comprehensive picture of this initial grammar. Section 3 then pinpoints the emergence of two key features incompatible with G_1 , i.e. subject DPs with no resumptive clitics (henceforth 'canonical subject DPs') and the negative particle *ne*. The emergence of a second grammar termed Normed French (G_2) is hence tentatively described and explained, and the G_1/G_2 distinction as well as the prediction made by diglossia about consistency within every utterance are discussed in the light of the very specific phonological accommodation pattern of the nominative clitic *il* in the data. Finally, I also briefly discuss learnability, theoretical and open issues relevant to diglossic acquisition.

1 THE CHILD CORPORA

This investigation is primarily based on two sets of spontaneous data collected in two different kindergarten classes of French-speaking children in Southern France. Corpus #1 stems from 5 sessions of audio recordings with 17 children between the ages of 2;3 and 3;1 (mean age 2;8). These data were collected during the first term of these children's first kindergarten year. Corpus #1 consists of 1,072 verbal utterances in total (full details in Palasis, 2005). Corpus #2 comes from audio and video recordings with 20 other children between 2;5 and 4;0 (mean age 3;3). It displays 9,748 verbal utterances which were gathered over the entire first kindergarten year of these children during 13 different sessions (full details in Palasis, 2010a, 2010b). Small groups of 3 to 5 children were constituted on a voluntary basis for both corpora, and the recordings took place whenever possible in a quiet room adjoining the children's usual classroom. One investigator – the same person in both studies – led each group. The data consist of spontaneous child-child and child-investigator interactions. Nevertheless, in order to encourage speech, the investigator suggested activities to each group, i.e. games or books, or asked the children to tell her about their activities in and out of school. The general atmosphere was intended to be informal and friendly.

Corpus #2 was further transcribed and coded along the CHILDES formats.⁴ Each utterance displays three tiers, i.e. a main orthographical transcript, a line of morpho-syntactic codes (%mor), and, whenever relevant, an 'error' tier (%err)

⁴ Child Language Data Exchange System; see MacWhinney (2000a, 2000b), and the website at <http://childes.talkbank.org/>. Corpus #2 is available at <http://childes.talkbank.org/browser/index.php?url=Romance/French/Palasis/>.

accounting for the children's non-target utterances. The utterances in (1) exemplify two possibilities with regard to subjects, i.e. a singleton nominative clitic in (1a), and associated strong and clitic forms in (1b). The same 20 children were further recorded and videotaped during their two subsequent kindergarten years, i.e. between the ages of 4;0 and 6;0. However, since these data have not yet been fully processed, they are referred to as Corpus #3 only very sporadically in this article.

- (1) Transcripts and codes from Corpus #2:⁵
- a. *il a fait un rêve* (Lucille, 2;9)
%mor: pro:cli:nôm|il&3S
 - b. *moi je fais la fourmi* (William, 2;9)
%mor: pro:ton:dg:nôm|moi&IS + pro:cli:d:nôm|je&IS

Let us now examine some of the characteristics of these children's utterances starting with a key aspect of variation in French, i.e. the morpho-syntactic status of nominative clitics.

2 DESCRIBING THE INITIAL GRAMMAR OF FRENCH (G₁)

2.1 *The morpho-syntactic status of nominative clitics*

The morpho-syntactic status of nominative clitics in French has been the matter of much attention for years (e.g. Heap & Roberge, 2001 for an overview ranging from traditional to generative grammar). As far as generative linguistics is concerned, ever since Kayne's (1975) seminal work on subject and object clitics, and Rizzi's (1986) comparison between French and some Northern Italian dialects, the following question has arisen: Should French nominative clitics be analysed as proper syntactic arguments (Kayne, 1975; Rizzi, 1986; De Cat, 2005) or should they be considered as preverbal morphological affixes (Auger, 1994; Kaiser, 1994; Zribi-Hertz, 1994; Culbertson, 2010)? The latter option is however envisaged only in colloquial French. Indeed, nominative clitics are analysed as full syntactic arguments in formal French, i.e. as 'pronouns' in the etymological sense, and various syntactic properties follow, as illustrated in (2) with Zribi-Hertz' (1994) examples. Firstly, the syntactic subject can surface as a noun or a clitic pronoun as in (2a). Sentence (2b) then illustrates the standard analysis of left dislocation, i.e. a base-generated topic in a peripheral projection (De Cat, 2007). Topics are 'preposed elements characteristically set off from the rest of the clause by a 'comma intonation' and normally expressing old information' (Rizzi, 2000: 245). Accordingly, topics are generally expected to be definite, hence the ungrammaticality with an indefinite or quantified expression, as illustrated in (2c).

- (2) Formal French (Zribi-Hertz, 1994: 131-4):
- a. *Jean/Il mangera ici*

⁵ nom = 'nominative/subject-related'; ton = 'strong'; dg = 'doubled item to the left of the clitic'; d = 'doubling item'.

The case for diglossia

Table 1. Different types of subjects between the ages of 2;3 and 4;0

Types	Corpus #1: 2;3-3;1		Corpus #2: 2;5-4;0		Examples
	#	%	#	%	
Clitic	459	50.9	5,831	64.6	<i>nan je veux mes épées</i>
Clitic +DP	203	22.5	2,804	31.1	<i>mon papa i vient me chercher elle est là ma photo moi j'ai fini ça on voit Lucille nous</i>
Null	240	26.6	348	3.9	<i>veut pas manger</i>
DP	0	0	39	0.4	<i>le train est là</i>
Total	902	100.0	9,022	100.0	

- b. *Pierre, il n'a rien dit*
 c. **Personne, il a rien dit*

Consequently, the debate arises in colloquial French due to attested utterances such as *personne il a rien dit* (Zribi-Hertz, 1994: 134), *quelqu'un il dansait* (Auger, 1994: 97), *tout le monde il se baignait* (Blanche-Benveniste, 2000: 42), or other utterances I have collected randomly over the last few months such as *on sait pas pourquoi un truc i marche, personne i veut ça* and *comme tout le monde i dit*. Let us now examine the relevant child data.

2.2 Nominative clitics in early speech

Children's first combinations of words emerge between the ages of 1;6 and 2;0. These early utterances can be either nominal, e.g. *la dame là*, or verbal with a finite or non-finite verb, e.g. *il mange, mangé le chien* (De Boysson-Bardies, 1996: 232 and Pierce, 1992: 84, respectively). Nominative clitics represent the most frequent type of subject whenever the utterance displays a finite verb (Prévost, 2009: 131-40 for an overview of the literature). Table 1, drawn from later stages (2;3-3;1 and 2;5-4;0), illustrates this persistent characteristic of child French. Indeed, nominative clitics surface in 73.4% and 95.7% of the children's verbal utterances in Corpus #1 and #2, respectively. Conversely, canonical subject DPs, i.e. DPs with no resumptive clitics, are virtually absent from these early data. These characteristics are also found in adult oral French, as suggested with the 80% clitic and 5.9% to 20% DP rates in Blanche-Benveniste (1994: 87, 90; 2003: 83). Nominative clitics are hence pervasive in child and adult oral French.

A second characteristic of nominative clitics in child speech is their position with regard to finite verbs. The observed data are clear-cut since Corpus #1 displays absolutely no postverbal clitics, and Corpus #2 contains just 6 given in (3). All the other nominative clitics in these data are hence preverbal, either strictly adjacent to the finite verb, as in Table 1, or separated from the verb by one or several other clitics, e.g. *i m l'a piqué*. Quite expectedly, all the postverbal clitics surface in

interrogative utterances. Indeed, in order to yield clitic inversion, the finite verb, which raises first from V to I to check inflectional features, needs to move further to C to satisfy the *Wh* Criterion (Rizzi, 1996), e.g. [CP *dort_i* [_{IP} *il_i dort_j* [_{VP} *t_i dort_j*]]]. It can be seen from (3d-f) however that 3 of the 6 child utterances displaying this kind of structure are non-target.

- (3) Clitic inversions in Corpus #2:
- a. *que se passe-t-il?* (Victor, 3;5)
 - b. *poisson où es-tu?* (Dylan, 2;11)
 - c. *où es-tu Maman Ours?* (Carla, 3;5)
 - d. *euh s(e) passe-t-il?* (Victor, 3;5)
 - e. *qu'est-c(e) qui passe-t-il?* (Kelsang, 3;7)
 - f. *qu'est-c(e) que s(e) passe-t-il?* (Kelsang, 3;7)

Interrogative utterances hence highlight the fact that nominative clitics are (nearly) exclusively preverbal at this stage. Three different analyses can account for this preverbal position, i.e. no I-to-C movement of the verb, prefixation of the clitic to the verb prior to movement, or a combination thereof.

Let us now consider the different types of DPs young children spontaneously utter with nominative clitics. Corpus #2 overwhelmingly displays definite DPs, as exemplified in (4a). However, in very occasional cases, the associate can be indefinite, as in (4b-d).⁶ The latter are hence reminiscent of the adult examples in Section 2.1, and comply with the morphological analysis of the clitic since *un garçon*, *personne*, *quelqu'un* and *tout le monde*, being indefinite or quantified, cannot be dislocated topics.

- (4) Different types of associates in child speech:
- a. *l'escargot i dort* (Dylan, 2;8)
 - b. *un garçon il attrape un œuf* (Mathilde, 3;0)
 - c. *personne i m l'a dit* (Matteo, 4;5, Corpus #3)
 - d. *oh quelqu'un i l'a déchiré!* (Dylan, 3;10, Corpus #3)

As far as nominative clitics are concerned, the observed child data show the following characteristics so far: (i) overwhelming co-occurrence with finite verbs only, (ii) preverbal position, (iii) possible co-occurrence with indefinite and quantified DPs. This set of properties suggests a morphological handling of these clitics in the child system. Arguments against the morphological analysis exist however. De Cat (2005) for instance critically discusses some of the consequences of this analysis, i.e. (i) possible redundancy of the pre- and post-verbal agreement markers, (ii) optionality of the preverbal clitics in some cases, (iii) unavailability of these clitics for syntactic operations due to their pre-syntactic insertion, (iv) morphological analysis of the preverbal clitics that appear between the subject clitic and the finite verb, and (v) the existence of doubling in spoken French, i.e. the co-occurrence of a morphological clitic and a DP in subject position.

⁶ And in the additional randomly collected utterance *tout le monde i s'assoit dessus* (Sephora, 8;0).

Nevertheless, it has also been claimed that these arguments do not invalidate the morphological analysis. Firstly, most of these arguments do not apply to early French, as illustrated in the near total absence of optionality and clitic inversion, and the possible co-occurrence with indefinite and quantified DPs. Secondly, whenever these arguments do apply, i.e. redundancy and morphological analysis of other clitics, they are construed as unproblematic (see Palasis, 2009, 2010b for a thorough discussion). The morphological status of nominative clitics is hence deemed to be a major characteristic of an initial grammar of French termed Spontaneous French (G₁) in this work. Let us now test two additional phenomena relevant to the morpho-syntactic debate, i.e. object enclisis and past-participle agreement with object proclitics.⁷

2.3 Accusative clitics in early speech

The morphological status assumed for nominative clitics in Spontaneous French implies that any constituent intervening between a nominative clitic and a finite verb is also morphological (in line with 'clitics can attach to material already containing clitics, but affixes cannot', Zwicky and Pullum, 1983: 504). This implication entails that: (i) accusative clitics are verbal prefixes too, and (ii) the co-occurrence of an accusative clitic and an associate DP instantiates clitic doubling, not clitic dislocation. Let us hence firstly turn to occurrences that could question the morphological analysis of clitics, i.e. object enclitics.

2.3.1 Object enclitics

Object enclitics occur in positive imperatives, e.g. *prends-le*. Enclisis takes place when the finite verb moves to the left periphery (Rizzi, 2000; Han, 1998). Thus, the order *v-cl* is possible only if the accusative clitic is syntactically independent from the verb. Indeed, if the clitic were morphological, viz. prefixed, the syntactic operation would move the verb together with its prefix to the left periphery yielding **le-prends*. Consequently, object enclisis can only surface if the child manipulates the clitic as a proper syntactic argument. Although in complementary distribution with nominative proclitics (e.g. imperative *prends-le* vs. indicative *je le prends*), object enclitics could hence cast doubt on the morphological analysis of nominative and accusative proclitics in the rest of the data.

On the one hand, young children regularly resort to imperative forms in their speech (Corpus #1: 15.9% of finite verbs; Corpus #2: 14.8%). On the other hand, both spontaneous and elicited data usually illustrate a lag of several months between the emergence of nominative and accusative clitics since the latter are uttered on a regular basis in obligatory contexts at circa 5;0 (Prévost, 2009: 140–143). Before this period, children usually resort to strong pronouns, full DPs, or object omission. Imperative forms in Corpus #2 comply with these early preferences since imperatives are primarily uttered intransitively, as in (5a), and when the verb

⁷ I am grateful to Michael Zimmermann (p.c.) for pointing out the former argument to me.

is transitive the object is overwhelmingly strong, as shown in (5b). Thus, only 7 different children utter a total of 15 object enclitics, as exemplified in (5c). Although always correctly positioned with regard to the finite verb, object enclisis is hence not productive at this stage. Consequently, except for the 15 above-mentioned exceptions which could belong to an emerging G_2 , no overt evidence invalidates the morphological handling assumed so far for G_1 . Moreover, Haverkort and Weissenborn (1991) noticed that younger children (2;0–2;4) initially produce positive imperatives with *CL-V* sequences, as illustrated in (5d).

(5) Positive imperatives in child speech:

- a. *regarde!* (Matteo, 2;11)
- b. *eh Lucille touche ça!* (William, 2;10)
- c. *mets le sur la table!* (Nina, 3;3)
- d. *le mets là-dedans!* (2;0, Haverkort and Weissenborn, 1991: 3)

Haverkort and Weissenborn (1991) analysed this non-target order as an absence of I-to-C movement. However, in line with the morphological hypothesis and further to what was said with regard to nominative clitic inversion in Section 2.2, it could also be envisaged that utterance (5d) displays a verb with a prefixed clitic in either I or C.

2.3.2 The Doubling/Agreement Correlation

The 'Doubling/Agreement Correlation' proposed by Tsakali and Anagnostopoulou (2008) is also of particular relevance to our morpho-syntactic debate. Indeed, according to this correlation object-clitic doubling and participle agreement with direct objects are in complementary distribution, i.e. 'if a language has participle agreement it lacks clitic doubling', and vice versa (Tsakali and Anagnostopoulou, 2008: 322). Formal French and Modern Greek illustrate these two possibilities, as shown in (6).

(6) The Doubling/Agreement Correlation:

- a. [-] Doubling, [+] Agreement (e.g. Formal French):
je l'_i'ai achetée (**la tarte_i*)
- b. [+] Doubling, [-] Agreement (e.g. Modern Greek):
tin_i eho agorasi tin tarta_i
it-ACC-FEM-SG have-1SG bought-PP-MASC the tart-ACC-FEM-SG
'I have bought the tart'

It is generally acknowledged however that 'for many speakers, this agreement [past-participle agreement with direct-object clitics] is not made in spoken French' (Kayne, 2000: 10). Following Tsakali and Anagnostopoulou's correlation and Kayne's statement, we might suggest that early French displays the [+] Doubling/[-] Agreement option, i.e. has morphological accusative clitics, and we can examine Corpus #2 in order to test child data for this correlation. As indicated earlier, however, accusative clitics are still scarce at this stage. Consequently, this investigation merely represents a first attempt to add the correlation to the debate, and Corpus #3 should be more informative on this matter. Indeed, utterances

displaying an accusative clitic together with an associate DP and a past participle are very rare in Corpus #2 with only two utterances meeting these requirements, as in (7a-b).

(7) Past-participle agreement in Corpus #2:

- a. *eh tu l'as pris le chien-chien?* (William, 3;0)
- b. *tu l'as pas mis bien ça?* (Matteo, 3;3)
- c. *y'a Matteo qui me l'a pris!* (Nina, 3;4)

(7a-b) display a masculine object which does not yield a modification of the default masculine past-participle form. Consequently, no overt evidence emerges in these two utterances with regard to [+] Agreement, and the morphological analysis of clitics, corresponding to the [+] Doubling/[-] Agreement option, can be maintained for G_1 . (7c) points in the same direction since it illustrates the [-] Agreement option with a previously-mentioned feminine object (*la pomme*).

2.4 From G_1 to G_2

Drawing on the observation of naturalistic child data, Section 2 aimed to describe the grammar of French that emerges spontaneously with young children in the initial stages of their linguistic development, hence its label Spontaneous French (G_1). One pivotal characteristic of this grammar is the morphological status of its nominative and accusative clitics. In order to buttress the hypothesis on diglossia, we now need to describe the emergence of a second grammar (G_2), by identifying properties incompatible with G_1 , and test the prediction made by diglossia about grammatical consistency within utterances.

3 WHY ADD G_2 ?

Diglossia posits that, at a particular point in time, speakers move from their initial single grammar to a choice between two cognate grammars. The immediate questions are why and when. Actually, we have already mentioned child utterances non-ascribable to G_1 . This section firstly details these occurrences, and then turns to a further crucial topic, viz. negation.

3.1 Canonical subject DPs

Canonical subject DPs, e.g. *le train est là* (see Table 1), represent the most emblematic incompatibility with G_1 due to the absence of the nominative prefix in this configuration.⁸ It is hence suggested that the corresponding 39 sentences in Corpus #2 pertain to a different grammar, i.e. G_2 . Two types of utterances emerge within the relevant series. Indeed, only 29 sound genuinely spontaneous since the other

⁸ The well-known 'null subject' phenomenon represents the other case of absence. This is accounted for within G_1 in Palasis (2012).

Table 2. *The emergence of canonical subject DPs in Corpus #2*

Names	Ages first DP	Session #	Spontaneous/All DPs
Mathilde	2;09.05	3	2/2
William	2;09.13	1	1/1
Dylan	2;09.23	3	2/2
Lucille	2;10.05	2	17/18
Nina	3;02.00	8	1/9
Kelsang	3;07.01	4	1/1
Enzo	3;07.15	11	1/1
Victor	3;07.27	9	0/1
Sara	3;10.13	13	4/4

10 correspond to previously heard nursery rhymes or stories, as in (8a-b) and (8c), respectively.

(8) Previously heard utterances in Corpus #2:

- a. *le facteur n'est pas passé* (Victor, 3;7)
- b. *vole mon petit chagrin* (Lucille, 3;4)
- c. *pas moi dit le chat* (Nina, 3;2)

Interestingly, these utterances display three main characteristics with respect to children, development and constituents. Firstly, as far as children are concerned, only 9 out of 20 utter canonical subject DPs, as detailed in Table 2. Furthermore, the vast majority are produced by one child only, Lucille, who utters 46.1% of the series. Secondly, the progression between the first and the last recordings is also of interest since 9 of the 29 spontaneous utterances belong to the last session in contrast with the other sessions that display a maximum of 3 DPs each.

Finally, these utterances stand out because of the presence of quantified elements which are scarce in the rest of the data. *Personne* for instance surfaces a total of 5 times in verbal utterances and only once as a subject, as in (9b).

(9) Quantified subjects in Corpus #2:

- a. *et quelqu'un regarde la main* (Sara, 3;11)
- b. *personne me le retourne comme ça* (Lucille, 3;4)

Drawing on these facts, it is surmised that these utterances illustrate the emergence of a second grammar, G_2 , and that contrary to G_1 , G_2 displays syntactic clitics. These are hence not obligatory, not limited to proclisis, and not compatible with indefinite and quantified DPs, as illustrated in (2), (3) and (9), respectively. Few children show signs of G_2 at this stage. The most obvious utterances come from Lucille as will be confirmed in the next section on the emergence of discontinuous negation.

3.2 *The emergence of the negative particle ne*

Sentential negation represents another uncontroversial area of variation in French. Formal French displays discontinuous negation, i.e. Neg1 (*ne*) + Verb + Neg2 (*pas*,

Table 3. Negative markers in Corpus #2

Negative markers	<i>pas</i>	<i>plus</i>	<i>jamais</i>	#	%
Simple	1,164	126	5	1,295	98.8
Discontinuous	7	9	0	16	1.2
Total	1,171	135	5	1,311	100.0

plus, etc.), whereas colloquial French shows simple negation, viz. the absence of the negative particle *ne* (Ashby, 1981, 2001; Lambrecht, 1981; Bernini & Ramat, 1996; Rowlett, 1998; Coveney, 2002; Barra-Jover, 2004; Massot, 2008, among others). Seminal work on early acquisition (1;8–2;2) shows the correct placement of *pas* with regard to finite and non-finite verbs, e.g. *veux pas lolo* vs. *pas manger la poupée* (Déprez & Pierce, 1993: 40). These facts are interpreted as evidence that young children project and master functional categories and syntactic movements pertaining to verbal inflection. No explicit comments are made however on markers other than *pas* and the preverbal particle *ne*. Let us hence turn to Corpus #2 in order to delve into these matters.

Children negate 13.5% of their verbal utterances in Corpus #2. This rate is consistent with previous analyses (e.g. Friedemann, 1993: 233). Table 3 shows that child negation displays three different markers at this stage, i.e. *pas*, *plus* and *jamais*, and that *pas* is by far the most frequent one. The table also highlights the rarity of discontinuous negation in these data since the negative particle *ne* emerges in 16 utterances only, i.e. 1.2% of the negated sentences.

The 16 occurrences of discontinuous negation are given in (10). They illustrate that *ne* is uttered by 4 children only (out of 20). Moreover, the type of utterance is of considerable significance: 11 of these occurrences are either songs (Mathilde) or nursery rhymes (Victor). Consequently, as for clitic inversion and canonical subject DPs, none of the utterances seems to pertain to genuine spontaneous speech.

(10) Discontinuous negation in Corpus #2:

- a. Mathilde (3;2), 10 occurrences in a row, same pattern, singing, e.g.:
il ne chante plus/il ne saute plus/il ne danse plus
- b. Victor (3;7), 1 occurrence, nursery rhyme:
le facteur n'est pas passé
- c. Lucille (2;10–2;11–3;2), 3 occurrences:
 - i. *ta photo n'est pas là?*
 - ii. *non non ce n'est pas moi*
 - iii. *maman elle n'est pas là*
- d. Sara (3;10), 2 occurrences:
 - i. *il n'a pas gagné pour les œufs*
 - ii. *ne te cache pas avec la photo!*

The 13 recording sessions over a period of 7 months exhibit precise information on the moment *ne* emerges in the corpus. Table 4 indicates the relevant ages and corresponding sessions. Lucille is the first child to utter *ne* at 2;10, Sara is the fourth

Table 4. *The emergence of discontinuous negation in Corpus #2*

Names	Ages first <i>ne</i>	Session #	Ages first session
Lucille	2;10.05	2	2;09.22
Mathilde	3;02.14	11	2;08.00
Victor	3;07.27	9	3;03.00
Sara	3;10.13	13	3;03.05

at 3;10, and the 16 remaining children – who are between 3;0 and 4;0 by the end of the investigation – only use simple negation. Unfortunately, the exact lag between the emergence of simple and discontinuous negation cannot be calculated within this corpus since the relevant 4 children were already producing simple negation from the outset of the study (in accordance with their ages reported in the last column of Table 4).

3.3 *A possible developmental trajectory*

The data reported in (10) present an additional peculiarity in comparison with the rest of the corpus. Thus, (10b) and (10c) exemplify the co-occurrence of two rare phenomena, viz. the emergence of *ne* together with the absence of the nominative clitic (1.2% and 0.4% of the utterances, respectively).⁹ A comparison of the names and ages in Tables 2 and 4 shows that 4 children produce canonical subject DPs and discontinuous negation in Corpus #2, and that both phenomena emerge at the same age for Lucille (2;10), Victor (3;7) and Sara (3;10). The fourth child, Mathilde, shows a lag, and hence a sequence, between the emergence of her first canonical DP (2;9) and discontinuous negation (3;2). This order is interestingly also displayed by the remaining children in Table 2 since William, Dylan, Nina, Kelsang and Enzo rarely produce canonical subject DPs in this corpus and no discontinuous negation. Canonical subject DPs hence seem to emerge prior to or no later than discontinuous negation within the course of acquisition.

Since discontinuous negation and canonical subject DPs are traditionally assumed to belong to the ‘high variety’ (H) of French, it is hypothesized that we might be starting to gain a picture of the actual emergence of G_2 through a second set of possibly correlated properties. The two distinct series of characteristics for G_1 and G_2 are recapitulated in Table 5.¹⁰

Table 5 sketches the possible existence of two cognitively distinct grammars in the mind/brain of French speakers. Drawing on adult data, Massot (2010: 103) suggests a clausal domain of application for each grammar. Applied to Corpus #2, this proposal predicts for instance that a morphological clitic will not co-occur with the negative particle *ne*, and that a syntactic clitic will not co-occur with simple

⁹ Culbertson (2010: 95) established the same type of correlation in child-directed speech showing that ‘*ne*-retention is clearly affected by the properties of the preceding subject’.

¹⁰ The table mentions a crucial distinction with regard to the Pro-drop Parameter which cannot be discussed here for reasons of space, but see Palasis (2010b, 2012) on this matter.

Table 5. *Two distinct series of properties (provisional version)*

Items	Properties	G ₁	G ₂
Clitics	Status	Morphological	Syntactic
	Presence	Obligatory	Optional
	Clisis	Proclisis only	Pro- and enclisis
	Association	Any type of DP	Constrained
	Pro-drop language	Yes	No
Negation	Type	Simple	Discontinuous

negation. The following section illustrates this particular point relying on the very specific behaviour of clitic *il* in the child data.

3.4 The strict accommodation pattern of *il*

A typical characteristic of the nominative clitic *il* in Corpus #2 is its very strict accommodation pattern according to the phonological environment. Indeed, the analysis of the 1,688 *il* clitics shows that this element surfaces as [il] before a vowel and [i] before a consonant, as illustrated in (11).

(11) Strict accommodation pattern of *il* in Corpus #2 (provisional version):

- a. *il/_V*:
il est à moi (William, 2;9)
- b. *i/_C*:
i veut manger (Lucille, 2;9)

On the one hand, these facts do not seem particularly enlightening since French-speaking children favour this CVCV pattern from their babbling (De Boysson-Bardies, 1996: 80). On the other hand, the pattern described in (11) admits of 11 exceptions in the corpus, and their analysis seems to point to an interesting phenomenon with regard to the G₁/G₂ distinction and the prediction made by diglossia about grammatical consistency within utterances. Indeed, as can be seen from the relevant sentences reported in (12), the exceptions are definitely not random. Thus, 10 of these utterances display the negative particle *ne* which we ascribed to G₂. Following Massot's (2010) proposal on consistency, it is hence further assumed that: (i) the non-elided *il* clitics in these utterances are also generated by G₂, and (ii) these clitics are syntactic arguments, contrary to the 1,677 other *il* clitics that are generated by G₁ as verbal prefixes. Corpus #2 also exemplifies consistency within G₁ since elided clitics co-occur with simple negation only, as illustrated in (13).¹¹ Rizzi's (1986: 401) rejection of **personne il ne mange* could then be reinterpreted in terms of G₁ and G₂ grammars as belonging to neither G₁ (which yields *personne i mange*) nor G₂ (due to the constraint on Topics).

¹¹ (13b) also demonstrates that consonant [n] and adjacency to the finite verb can be discarded as possible triggers to the accommodation pattern of *il*.

- (12) The 11 non-elided forms of *il/_C* (G_2):
- 9 occurrences, all on the same pattern (Mathilde, 3;2, singing):
il ne chante/saute/danse plus
 - il n'a pas gagné pour les œufs* (Sara, 3;10)
 - et après oh il dit qui est là* (Lucille, 3;2)
- (13) Elided clitics co-occur with simple negation only (G_1):
- i (*ne) colle pas* (Chloé, 2;7)
 - comme ça i (*ne) nous mangera pas* (Lucille, 3;1)

The correlations exemplified in (12a-b) and (13) between *ne* and syntactic clitics on the one hand, and simple negation and morphological clitics on the other hand call for the following additional comments. Firstly and similarly to negation, a lag appears between the emergence of the two different types of clitics. Section 2 reported the early appearance of morphological clitics at circa 1;8. The utterances in (12) now point to an emergence of syntactic clitics at 3;2 for Mathilde and Lucille, and various subsequent ages for the other children (still to be confirmed in Corpus #3). G_2 hence definitely emerges at different moments of the linguistic development, but always after G_1 .¹² The second comment relates to the morpho-syntactic status of *ne*. Since clusters such as *i-ne-chante pas* (CL-NEG-V) and *ne-i-chante pas* (NEG-CL-V) are not attested, the data suggest that *ne* is not available as a verbal affix in Spontaneous French, contrary to some Northern Italian dialects (Rizzi, 1986; Zanuttini, 1997). Thus, *ne* is available only in Normed French as the head of an independent negative projection, viz. NegP (following Pollock, 1989 and Rowlett, 1998, among others). Finally, as *ne* never co-occurs with a nominative prefix, a blocking effect is presumed between *ne* and morphological clitics. This blocking effect is further interpreted as a trigger that forces a speaker to move from G_1 to G_2 , since *ne* is totally unavailable in G_1 , whether as a verbal affix or as an independent syntactic head. The very existence of this blocking effect should also be accounted for. At this stage of the investigation, I merely tentatively question the status of *pas* in both grammars, suggesting possible different underlying structures altogether in line with the various existing analyses of *pas* in the literature, e.g. specifier of NegP (Pollock, 1989), initial VP adjunction (Rowlett, 1993, 1998), multiple specifiers in VP (Peters, 1999), or surmise a possible connection with the overall development of quantification, further to Schapansky's (2002) proposal that negation with *ne* requires negation and quantifying features.

4 CONCLUSION

This article aims to lend support to the French diglossia hypothesis (Barra-Jover, 2004, 2010; Rowlett, 2007; Massot, 2008, 2010; Zribi-Hertz, 2011). Drawing on naturalistic data from native French children under 4, two cognate but distinct grammars are hence described by hinging upon two crucial distinctions, i.e.

¹² See Barra-Jover (2010, this volume) and Palasis (2011) on G_2 triggering.

Table 6. Two distinct series of properties (final version to date)

Items	Properties	G ₁	G ₂
Clitics	Status	Morphological	Syntactic
	Presence	Obligatory	Optional
	Clisis	Proclisis only	Pro- and enclisis
	Association	Any type of DP	Constrained
	Accommodation	Yes	No
Negation	Pro-drop language	Yes	No
	Type	Simple	Discontinuous
	NegP	No	Yes
	Status of <i>pas</i>		Different?

the morpho-syntactic status of nominative clitics and the emergence of the negative particle *ne*. Thus, the initial Spontaneous French grammar (G₁) displays morphological proclitics and simple negation, whereas the subsequent Normed French grammar (G₂) features syntactic clitics and discontinuous negation. The detail of these characteristics is recapitulated in Table 6. G₁ is common to all native French speakers, whereas G₂ gradually emerges in one fifth of the observed children. It is also proposed that the existence of discontinuous negation in G₂ represents one of the triggers forcing a speaker to move from G₁ to G₂ due to the total unavailability of the particle *ne* in G₁. Finally, the data on canonical subject DPs, discontinuous negation, and the elision pattern of clitic *il* seem to support the prediction made by diglossia and argued for by Massot (2010) about grammatical consistency within each utterance. These observations call for further comments on learnability and theoretical issues, and provide directions for future research.

The early acquisition of two grammars inevitably raises questions concerning learnability. Firstly, I follow the overall generative rationale for language acquisition implying '[...] that the human language faculty predisposes the individual to become bilingual and that adequate theories of language and of grammar need to reflect this fact' (Meisel, 2001: 12). Accordingly, diglossia does not challenge our cognitive capabilities. Its formalisation however remains a matter of lively debate, especially within the Principles and Parameters framework (from Chomsky, 1981 to Yang and Roper, 2011). This article highlights clustering and blocking effects. Should these be construed as mere tendencies, macro-parameters, micro-parameters, learning biases (Newmeyer, 2005; Baker, 2008; Palasis, 2010b; Boeckx, 2011, respectively), or does Rowlett's 'bolt-on approach' account for 'the non-randomness of the location/nature of the variation' (Rowlett, this volume)?

Additional acquisition data are always welcome in order to further our understanding of linguistic competence. When fully processed, Corpus #3, i.e. naturalistic data from the same children between the ages of 4;0 and 6;0, should give us a wealth of detail about the emergence and development of French diglossia. Furthermore, elicitation tasks with children and Event-Related brain Potential measures with adults have already lent support to diglossia as far as Arabic is concerned (Khamis-Dakwar, Froud & Gordon, 2011 and Khamis-Dakwar & Froud,

2007, respectively). These paradigms could therefore inspire further research on French.

Last but not least, several important domains of acquisition had to be left aside in this article, notably interrogative and relative clauses. Nevertheless, diglossia could also well account for the 'degraded status' of the *wh*-in-situ question *Jean ne mange pas quoi?* reported in Boskovic (1998: 246) since the author interestingly highlights a blocking effect of NEG on *wh*-movement in covert syntax.¹³ As far as relative clauses are concerned, Guasti's (2004: 240) statement that 'children's deviations are likely due to their ignorance of relative pronouns, a lacuna that one must assume is filled in during the school years through explicit teaching [and] lexical learning' lends direct support to the developing aspect of G₂. Both these matters are, of course, already on the research agenda.

Address for correspondence:

Katerina Palasis

Laboratoire Bases, Corpus, Langage - (BCL) UMR 7320

Campus Saint Jean d'Angely - SJA3/MSH

24 avenue des Diables Bleus

F-06357 Nice Cedex 4

France

email: palasis@unice.fr

REFERENCES

- Adli, A. (2006). French *wh*-in-situ Questions and syntactic optionality: evidence from three data types. *Zeitschrift für Sprachwissenschaft*, 25: 163–203.
- Ashby, W. J. (1981). The loss of the negative particle *ne* in French: A syntactic change in progress. *Language*, 57: 647–87.
- Ashby, W. J. (2001). Un nouveau regard sur la chute du *ne* en français parlé tourangeau: s'agit-il d'un changement en cours? *Journal of French Language Studies*, 11.1: 1–22.
- Auger, J. (1994). Pronominal clitics in Québec colloquial French: A morphological analysis. PhD Dissertation, University of Pennsylvania.
- Baker, M. C. (2008). The macroparameter in a microparametric world. In: T. Biberauer (ed.), *The Limits of Syntactic Variation*. Amsterdam: John Benjamins, pp. 351–74.
- Barra-Jover, M. (2004). Interrogatives, négatives et évolution des traits formels du verbe en français parlé. *Langue française*, 141: 110–25.
- Barra-Jover, M. (2010). 'Le' français ou ce qui arrive lorsqu'un état de choses est observé comme une entité. *Langue française*, 168.4: 3–18.
- Barra-Jover, M. (This volume). Le français à l'école primaire : pourquoi l'approche diglossique est la meilleure façon d'apprendre le français académique.

¹³ The blocking effect of negation in French *wh*-in-situ questions is also examined by Chang (1997: 19–20) and Cheng and Rooryck (2000: 11), and is critically discussed by Adli (2006: 176–9) who investigates simple as well as discontinuous negation.

- Bernini, G. and Ramat, P. (1996). *Negative Sentences in the Languages of Europe: A Typological Approach*. Berlin: Mouton de Gruyter.
- Blanche-Benveniste, C. (1994). Quelques caractéristiques grammaticales des 'sujets' employés dans le français parlé des conversations. In: M. Yaguello (ed.), *Subjecthood and Subjectivity. Proceedings of the Colloquium 'The Status of the Subject in Linguistic Theory'*. London: Editions Ophrys, pp. 77-107.
- Blanche-Benveniste, C. (2000). *Approches de la langue parlée en français*, Paris: Editions Ophrys.
- Blanche-Benveniste, C. (2003). Les formes grammaticales de réalisation des sujets et leur inégale représentation en français contemporain. In: J.-M. Merle (ed.), *Le sujet*. Paris: Ophrys, pp. 73-90.
- Boeckx, C. (2011). Approaching parameters from below. In: A. M. Di Sciullo and C. Boeckx (eds), *The Bilingual Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: Oxford University Press, pp. 205-21.
- Boskovic, Z. (1998). LF Movement and the Minimalist Program. In: Z. Boskovic and H. Lasnik (eds), *Minimalist Syntax: The Essential Readings*. (2007). Oxford: Blackwell, pp. 243-57.
- Buson, L. and Billiez, J. (This volume). Perspectives diglossique et variationnelle – Complémentarité ou incompatibilité?
- Chang, L. (1997). *Wh-in-situ Phenomena in French*. MA thesis, University of British Columbia.
- Cheng, L. L.-S. and Rooryck, J. (2000). Licensing wh-in-situ. *Syntax*, 3.1: 1-19.
- Chomsky, N. (1981). Principles and parameters in syntactic theory. In: N. Hornstein and D. Lightfoot (eds), *Explanation in Linguistics: The Logical Problem of Language Acquisition*. London: Longman, pp. 32-75.
- Coveney, A. (2002). *Variability in Spoken French: A Sociolinguistic Study of Interrogation and Negation*. Bristol: Elm Bank.
- Coveney, A. (2011). A language divided against itself? Diglossia, code-switching and variation in French. In: F. Martineau and T. Nadasdi (eds), *Le français en contact : Hommages à Raymond Mougeon*. Québec: Presses de l'Université de Laval, pp. 51-85.
- Culbertson, J. (2010). Convergent evidence for categorial change in French: from subject clitic to agreement marker. *Language*, 86.1: 85-132.
- De Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants. De la naissance jusqu'à deux ans*. Paris: Editions Odile Jacob.
- De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua*, 115.9: 1195-219.
- De Cat, C. (2007). *French Dislocation. Interpretation, Syntax, Acquisition*. Oxford: Oxford University Press.
- Déprez, V. and Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry*, 24.1: 25-67.
- Ferguson, C. A. (1959). Diglossia. *Word*, 15.1: 325-40.
- Friedemann, M.-A. (1993). The underlying position of external arguments in French: a study in adult and child grammar. *Language Acquisition*, 3.3: 209-55.
- Gadet, F. (2003). *La Variation sociale en français*, Paris: Editions Ophrys.
- Guasti, M. T. (2004). *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.
- Han, C.-h. (1998). The structure and interpretation of imperatives: mood and force in Universal Grammar. PhD Dissertation, University of Pennsylvania.

- Haverkort, M. and Weissenborn, J. (1991). Clitic and affix interactions in early Romance. Paper presented at the 16th Annual Boston University Conference on Language Development.
- Heap, D. and Roberge, Y. (2001). Cliticisation et théorie syntaxique, 1971–2001. *Revue québécoise de linguistique*, 30.1: 63–90.
- Kaiser, G. A. (1994). More about INFL-lection and Agreement: The Acquisition of Clitic Pronouns in French. In: J. M. Meisel (ed.), *Bilingual First Language Acquisition: French and German Grammatical Development*. Amsterdam: John Benjamins, pp. 131–59.
- Kayne, R. S. (1975). *French Syntax. The Transformational Cycle*, Cambridge, MA: MIT Press.
- Kayne, R. S. (2000). Past-participle agreement in French and Italian. In: R. S. Kayne (ed.), *Parameters and Universals*. Oxford: Oxford University Press, pp. 10–24.
- Khamis-Dakwar, R. and Froud, K. (2007). Lexical processing in two language varieties: an event-related brain potential study of Arabic native speakers. In: M. A. Mughazy (ed.), *Perspectives on Arabic Linguistics XX*. Amsterdam: John Benjamins, pp. 153–66.
- Khamis-Dakwar, R., Froud, K. and Gordon, P. (2011). Acquiring diglossia: mutual influences of formal and colloquial Arabic on children's grammaticality judgments. *Journal of Child Language*, FirstView Article: 1–29, DOI:10.1017/S0305000910000784.
- Lambrecht, K. (1981). *Topic, Antitopic and Verb Agreement in Non-standard French*. Amsterdam: John Benjamins.
- MacWhinney, B. (2000a). *The CHILDES Project. Tools for Analyzing Talk: Transcription Format and Programs*, 3rd edn, vol. 1. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2000b). *The CHILDES Project. Tools for Analyzing Talk: The Database*, 3rd edn, vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.
- Massot, B. (2008). *Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie : arguments morphosyntaxiques*. PhD dissertation, Université Paris 8.
- Massot, B. (2010). Le patron diglossique de variation grammaticale en français. *Langue française*, 168.4: 87–106.
- Meisel, J. M. (2001). The simultaneous acquisition of two first languages: Early differentiation and subsequent development of grammars. In: J. Cenoz and F. Genesee (eds), *Trends in Bilingual Acquisition*. Amsterdam: John Benjamins, pp. 11–41.
- Newmeyer, F. J. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.
- Palasis, K. (2005). *Problèmes d'acquisition et le Paramètre du Sujet Null*. MA Thesis, Université de Nice-Sophia Antipolis.
- Palasis, K. (2009). On the variable morpho-syntactic status of the French subject clitics. In: G. A. Kaiser and E.-M. Remberger (eds), *Null-subjects, Expletives and Locatives in Romance*. Konstanz: Konstanz Working Papers of the Department of Linguistics, pp. 47–62, <http://ling.uni-konstanz.de/pages/publ/PDF/ap123.pdf>.
- Palasis, K. (2010a). Introducing new French child data: thoughts on their gathering and coding. *La syntaxe de corpus*. = *Corpus* 9: 33–51, <http://corpus.revues.org/index1801.html>.
- Palasis, K. (2010b). *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*, Villeneuve d'Ascq: ANRT Diffusion.
- Palasis, K. (2011). Français spontané et français normé: bien plus que deux variétés d'un même français. Implications acquisitionnelles et didactiques. In: O. Bertrand and I. Schaffner (eds), *Variétés, variations et formes du français*. Palaiseau: Les Editions de l'École Polytechnique, pp. 55–72.

- Palasis, K. (2012). Economy principles within child speech: when the nominative clitic does not surface. In: S. Ferré, P. Prévost, L. Tuller and R. Zebib (eds), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages*. Cambridge: Cambridge Scholars Publishing, pp. 261–83.
- Péters, H. M. (1999). An alternative proposal for French negation. *Linguistica Atlantica*, 21: 107–36.
- Pierce, A. (1992). *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*. Dordrecht: Kluwer Academic Publishers.
- Pollock, J.-Y. (1989). Verb movement, Universal Grammar and the structure of IP. *Linguistic Inquiry*, 20.3: 365–424.
- Prévost, P. (2009). *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. Amsterdam: John Benjamins.
- Rizzi, L. (1986). On the Status of Subject Clitics in Romance. In: O. Jaeggli and C. Silva-Corvalan (eds), *Studies in Romance Linguistics*. Dordrecht: Foris, pp. 391–419.
- Rizzi, L. (1996). Residual verb second and the *Wh* Criterion. In: A. Belletti and L. Rizzi (eds), *Parameters and Functional Heads*. Oxford: Oxford University Press, pp. 63–90.
- Rizzi, L. (2000). The fine structure of the left periphery. In: L. Rizzi (ed.), *Comparative Syntax and Language Acquisition*. London: Routledge, pp. 241–95.
- Rowlett, P. (1993). On the syntactic derivation of negative sentence adverbials. *Journal of French Language Studies*, 3.1: 39–69.
- Rowlett, P. (1998). *Sentential Negation in French*. Oxford: Oxford University Press.
- Rowlett, P. (2007). *The Syntax of French*. Cambridge: Cambridge University Press.
- Rowlett, P. (2011). Syntactic variation and diglossia in French. *Salford Working Papers in Linguistics and Applied Linguistics*, 1: 13–26.
- Rowlett, P. (This volume). Do French speakers really have two grammars?
- Schapansky, N. (2002). The syntax of negation in French: contrariety versus contradiction. *Lingua*, 112: 793–826.
- Tsakali, V. and Anagnostopoulou, E. (2008). Rethinking the clitic doubling parameter: the inverse correlation between clitic doubling and participle agreement. In: D. Kallulli and L. Tasmowski (eds), *Clitic Doubling in the Balkan Languages*. Amsterdam: John Benjamins, pp. 321–57.
- Yang, C. and Roeper, T. (2011). Minimalism and language acquisition. In: C. Boeckx (ed.), *The Oxford Handbook of Linguistic Minimalism*. Oxford: Oxford University Press, pp. 551–73.
- Zanuttini, R. (1997). *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford: Oxford University Press.
- Zribi-Hertz, A. (1994). La syntaxe des clitiques nominatifs en français standard et en français avancé. *Travaux de Linguistique et de Philologie*, 32: 131–48.
- Zribi-Hertz, A. (2011). Pour un modèle diglossique de description du français: quelques implications théoriques, didactiques et méthodologiques. *Journal of French Language Studies*, 21.2: 231–56.
- Zwicky, A. M. and Pullum, G. K. (1983). Cliticization vs. inflection: English N'T. *Language*, 59.3: 502–13.

5. Palasis, K. (2014). Étude morphosyntaxique des pronoms sujets à partir d'un corpus de maternelle. In R. Mir-Samii (Ed.), *Du sujet et de son absence dans les langues* (pp. 259-270). Le Mans: Université du Maine, <http://doi.org/10.13140/2.1.1586.9440>

Étude morphosyntaxique des pronoms sujets à partir d'un corpus de maternelle¹

Katerina Palasis

Université Nice Sophia Antipolis, CNRS, BCL, UMR 7320

palasis@unice.fr

Résumé : Les syntacticiens débattent depuis des décennies du statut morphosyntaxique des pronoms personnels sujets en français contemporain : ces pronoms sont-ils manipulés en syntaxe comme des morphèmes libres ou entrent-ils dans la morphologie du verbe sous la forme de marqueurs préverbaux de personne ? La seconde analyse implique que le français contemporain est, comme les autres langues romanes, une langue à « sujet nul ». La fréquence, la position, la forme phonologique et les contextes d'utilisation des pronoms sujets dans un corpus de maternelle (trente-sept enfants entre 2;3 et 4;11) apportent un éclairage développemental nouveau à ce débat. La très grande majorité des données indique un traitement morphologique de ces constituants. Néanmoins, l'analyse longitudinale met également en avant l'émergence progressive de plusieurs phénomènes cooccurrents, dont l'utilisation d'un paradigme de pronoms syntaxiques. Ces faits nous invitent à suivre l'hypothèse de la diglossie selon laquelle les locuteurs francophones manipulent deux grammaires, apparentées mais bien distinctes.

Abstract: This contribution addresses the long-standing question surrounding the morpho-syntactic status of subject clitics in contemporary French: Do French speakers process these items as free morphemes or agreement markers? The latter implies that contemporary French is a “null-subject” language, on a par with other Romance languages. Frequency, placement, phonological forms and contexts of subject clitics are examined in a kindergarten corpus (thirty-seven children between 2;3 and 4;11), and the results, considered in a developmental perspective, shed new light on the debate. The data indicate that children seem to process subject clitics as preverbal agreement markers. Nevertheless, the longitudinal perspective also highlights a progressive and concomitant emergence of several phenomena, one of which is the emergence of a paradigm of proper syntactic pronouns. These facts support the diglossic hypothesis, which poses that French speakers handle two cognate nevertheless distinct grammars.

La notion de sujet grammatical interpelle les grammairiens et les syntacticiens depuis l'Antiquité. En effet, celle-ci est extrêmement paradoxale. Dans l'histoire de la grammaire, la fonction sujet, à la différence d'autres fonctions telles que celle de complément d'objet, est ancienne et stable, et ce depuis la réflexion sur les parties du discours menée par les philosophes grecs dont Platon. Cette notion appartient également au fonds commun des connaissances de tout écolier dans la mesure où elle est décrite à des fins d'analyse grammaticale dès les premières années à l'école élémentaire. Néanmoins, le consensus et la simplicité ne sont qu'apparents car de nombreux questionnements et points de vue divergents existent également autour de cette notion. La prégnance et la stabilité de la notion contrastent alors avec son aspect complexe comme en témoigne la préconisation de Grevisse et Goose (2008) que « Le plus sage est sans doute de considérer la notion de sujet comme une espèce de postulat, et de fournir simplement des moyens de l'identifier ».

Comme d'autres chercheurs, nous souhaitons néanmoins continuer à développer la connaissance que la communauté scientifique a déjà de cette notion. Nous proposons donc de contribuer à la réflexion sur la notion de sujet syntaxique et de son absence dans les langues à travers l'étude de la nature des éléments potentiellement sujets en français contemporain oral

¹ Recherches effectuées dans le cadre du projet ANR-DFG « Données et analyses dialectologiques, acquisitionnelles et diachroniques des pronoms sujets en Gallo-Roman » (DADDIPRO, 2012-2015, n° ANR 11 FRAL 007 01) sous la direction de Michèle Olivieri (Université Nice Sophia Antipolis, CNRS, BCL, UMR 7320) et Georg Kaiser (Université de Constance, Allemagne).

enfantin. En Section 1, nous exposons la problématique sur la nature du sujet à l’oral et nous l’illustrons de données extraites de corpus adultes. La Section 2 détaille ensuite les données enfantines sur lesquelles nous nous appuyons dans nos travaux. La Section 3 rapporte nos analyses des pronoms clitiques enfantins, en termes de fréquence dans les énoncés, position par rapport au verbe fléchi, contextes syntaxiques et formes phonologiques. L’analyse de ces résultats nous permet alors en Section 4 de proposer une discussion sur le statut morphosyntaxique des pronoms sujets en français oral enfantin. Nous constatons que les clitiques des petits francophones possèdent de nombreuses caractéristiques communes avec les marqueurs d’accord présents dans les dialectes du nord de l’Italie. Nous terminons notre discussion en l’élargissant à plusieurs questions d’ordre plus général : le statut du français enfantin contemporain comparé au français standard, l’hypothèse d’une diglossie en France et les aspects diachroniques de ces deux questions.

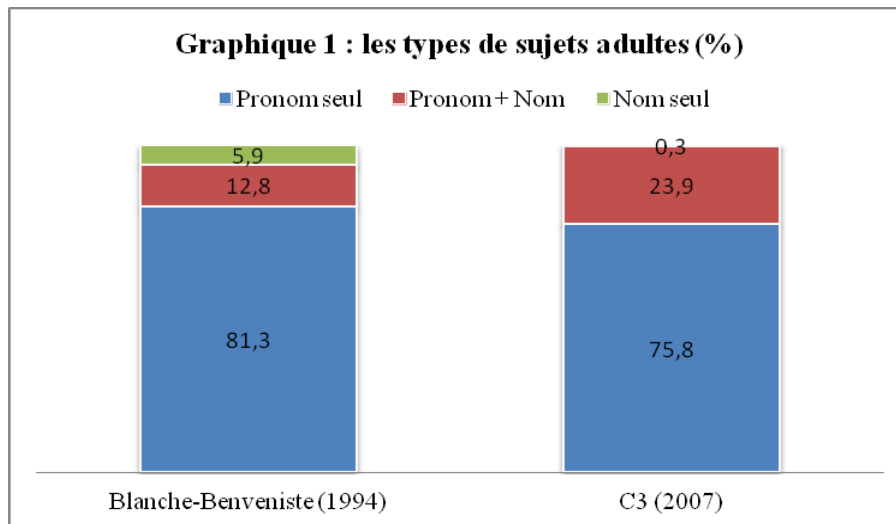
1. La problématique

Les constituants traditionnellement décrits comme sujets en français sont, la plupart du temps, soit nominaux soit pronominaux, comme illustré par Grevisse et Goosse (2008) dans des exemples tels que « le professeur écrit au tableau » ou « je travaille ». Or, les corpus oraux montrent que les locuteurs francophones associent souvent ces deux types de constituants au sein d’un même énoncé. Ainsi, Blanche-Benveniste (2000) rapporte l’inquiétude des grammairiens dès le XVII^e siècle face à ces cooccurrences avec l’exemple du grammairien Chifflet (1659) qui recommande de dire « mon père est malade » au lieu de « mon père il est malade ». Les corpus oraux actuels affichent également des taux de présence élevés des pronoms atones. Ainsi, une des caractéristiques qui distingue très nettement le français oral du français standard est la rareté des syntagmes nominaux seuls à l’oral dans la mesure où les locuteurs insèrent la plupart du temps un pronom atone devant chacun de leurs verbes fléchis, comme le montrent l’exemple (1) et les pourcentages du Graphique 1 ci-après.² L’étude de corpus oraux indique par ailleurs que les pronoms atones émergent associés non seulement à des syntagmes définis mais également indéfinis³, dans des propositions coordonnées, sans inversion dans les énoncés interrogatifs et que le pronom *il* est éliminé devant consonne (cf. exemples (2) et (3)). Ces caractéristiques sont généralement répertoriées comme distinguant le français oral du français standard (Kayne, 1975, Morin, 1979, Lambrecht, 1981, Ashby, 1984, Zribi-Hertz, 1994, Blanche-Benveniste, 2000, Fonseca-Greber et Waugh, 2003).

- (1) je crois que Lou elle va te montrer où il est le début. (C3, adulte)
- (2) tous les pirates i[z] ont des tatouages. (C3, adulte)
- (3) mais il est triste et i met les pieds dans l’eau ? (C3, adulte)

² Le contenu du corpus nommé C3 est détaillé en Section 2.

³ Sont considérés comme indéfinis : (i) les syntagmes dont le déterminant porte uniquement une valeur quantifiante (par exemple : *un, une, des, tous, chaque, aucun*) et (ii) les pronoms indéfinis (par exemple : *personne, rien*) (cf. Grevisse et Goosse, 2008 : 738).



Les cooccurrences « Pronom + Nom » portent un grand nombre de noms différents dans les grammaires. On trouve ainsi les termes de « reprise », « détachement », « redondance » (Grevisse et Goosse, 2008), « apposition » (Bonnard, 1997), « redoublement » et « dislocation » (Denis et Sancier-Chateau, 1994). La diversité des appellations traduit une diversité des approches et, très schématiquement, trois analyses sont envisagées : (i) le pronom est le sujet syntaxique et l'autre élément est relégué en dehors de la valence verbale (Kayne, 1975, De Cat, 2005) ; (ii) le pronom est un marqueur de personne et le véritable sujet grammatical est l'autre constituant (les pronoms portent alors le nom de « indices pronominaux » (Creissels, 1995) ou de « marqueurs d'accord affixaux » (Culbertson, 2010)) ; (iii) les deux éléments forment ensemble le sujet (Bauche, 1928). Cette absence de consensus provient en partie du statut phonologique des pronoms personnels du français moderne qui sont des pronoms sans accent propre (autrement dit atones ou clitiques), ce qui autorise les analyses mentionnées en (ii) et (iii) ci-dessus. Un tel débat n'existe pas dans les langues comme l'anglais par exemple dans lesquelles les pronoms personnels sont toniques. De plus, l'analyse traditionnelle du français (en (i) ci-dessus) implique que le français est une langue à part dans le groupe des langues romanes dans la mesure où les autres langues de ce groupe, telles que le portugais, l'espagnol ou l'italien, sont des langues dites « à sujet nul », c'est-à-dire des langues dans lesquelles la présence d'un sujet n'est pas obligatoire contrairement au français standard qui est dite « à sujet non-nul » (hors cas particuliers tels que les énoncés à l'impératif). *A contrario*, une analyse telle que celle proposée en (ii) ci-dessus implique de classer le français comme langue « à sujet nul ». De plus, les trois analyses mentionnées ci-dessus renvoient à un débat récurrent aussi bien en grammaire traditionnelle qu'en syntaxe moderne concernant le statut morphosyntaxique du pronom atone préverbal en français : pronom au sens étymologique du terme ou marqueur d'accord (Heap et Roberge, 2001 pour un historique de la question) ? Les partisans actuels des deux camps ont chacun des arguments de valeur (De Cat, 2005 vs Culbertson, 2010 par exemple). Nous proposons donc de reprendre ces arguments et de les mettre à l'épreuve de nouvelles données enfantines.

2. Le corpus

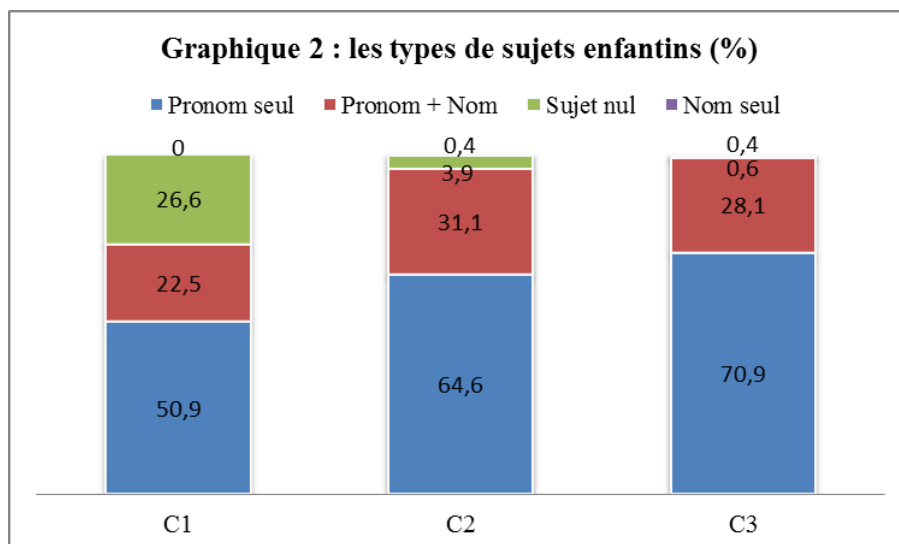
Nos recherches sur l'acquisition du français contemporain reposent sur un ensemble de données recueilli auprès de deux classes d'enfants de maternelle. Le premier groupe se compose de dix-sept enfants âgés de 2;3 à 3;1. Ils ont été enregistrés à l'occasion d'échanges spontanés avec un chercheur pendant cinq séances au cours de leur premier trimestre en toute petite section de maternelle. Ce corpus (dorénavant « C1 ») comporte 1.072 énoncés verbaux (Palasis, 2005). La seconde classe comprend vingt petits élèves. Ces enfants ont été

enregistrés à raison de dix à treize séances par an pendant leurs trois années de maternelle (soit un total de soixante-cinq heures d'enregistrements). Ces données audio et vidéo sont traitées par année scolaire. La première année (dorénavant « C2 ») comprend 15.992 énoncés provenant des vingt enfants alors âgés de 2;5 à 4;0 ainsi que 12.891 énoncés produits par le chercheur qui a conduit les séances (Palasis, 2010a, b).⁴ Le traitement des données de la seconde année (dorénavant « C3 ») est quasiment achevé (neuf séances sur dix ont été transcrites et codées). Nous disposons ainsi actuellement de 12.969 énoncés enfantins (dix-neuf enfants entre 3;6 et 4;11) et de 9.042 énoncés adultes. Nous espérons pouvoir traiter les données de la troisième année (« C4 » : 18 enfants entre 4;5 et 5;11) très prochainement. Nos analyses reposent donc sur des données enfantines à la fois transversales (avec un total de trente-sept enfants) et longitudinales (pour le moment, deux années scolaires consécutives), ainsi que sur des données produites par un adulte francophone au cours d'interactions informelles.

3. Analyse des sujets enfantins

3.1. Les pronoms atones

L'observation de C1-C3 confirme un certain nombre de faits qui avaient été mis en avant dans d'autres corpus francophones (cf. les travaux de Morgenstern, 2006 sur le développement du sujet et Prévost, 2009 pour une revue de quelques autres corpus). Les jeunes enfants n'associent quasiment jamais les pronoms atones à un verbe non-fléchi par exemple. Nos données permettent également de formuler des observations supplémentaires d'ordre développemental (entre 2;5 et 4;10 pour le moment). Le Graphique 2 illustre les faits suivants. Tout d'abord, nous constatons que les pronoms atones sont présents dans la quasi-totalité des énoncés verbaux enfantins. En effet, si nous ajoutons les catégories « Pronom seul » et « Pronom + Nom » (illustrées en (4)-(7) ci-dessous), nous constatons deux faits : les totaux sont très élevés dans les trois corpus (73,4 %, 95,7 %, 99 %) et ils augmentent avec l'âge des enfants. Cette omniprésence des pronoms atones implique alors que ceux-ci sont répétés lorsque plusieurs verbes fléchis sont juxtaposés et/ou coordonnés dans le même énoncé (cf. (7)).



⁴ Ces données (enregistrements audio et/ou vidéo alignés à une transcription orthographique, ainsi qu'un codage morphosyntaxique (%mor) et un codage des occurrences non-cibles (%err) ; Palasis, 2012b) sont disponibles sur le site de CHILDES (MacWhinney, 2000a,b) : <http://childes.talkbank.org/browser/index.php?url=Romance/French/Palasis/>.

- (4) et après il est content de retrouver ses amis. (C3, Carla, 4;7)
- (5) moi j'ai des feutres de Mickey à ma maison. (C3, Massimo, 3;6)
- (6) le nounours il arrive pas à mettre son manteau. (C3, Nina, 3;10)
- (7) elle joue elle danse elle met la musique et elle dort. (C3, Lucille, 3;10)

Nous constatons également que ces pronoms sont quasiment toujours préverbaux, quel que soit le corpus. En effet, C1 et C3 ne comportent aucun pronom postverbal et C2 n'en compte que six (cf. (8)). Les jeunes enfants n'utilisent donc globalement pas l'inversion lorsqu'ils forment des interrogations (cf. (9)). Pourtant, la structure de leurs énoncés interrogatifs se différencie de celle de leurs phrases déclaratives dans la mesure où le syntagme nominal n'apparaît pas au même endroit dans les deux types de structures. Nous illustrons cette différence par la paire d'exemples en (10) et (11) produite par le même enfant au cours de la même séance. Nous constatons ainsi une forte latéralisation du syntagme à gauche dans les déclaratives (91,6 %) et une tendance à la latéralisation du syntagme à droite dans les interrogatives (62,3 %).

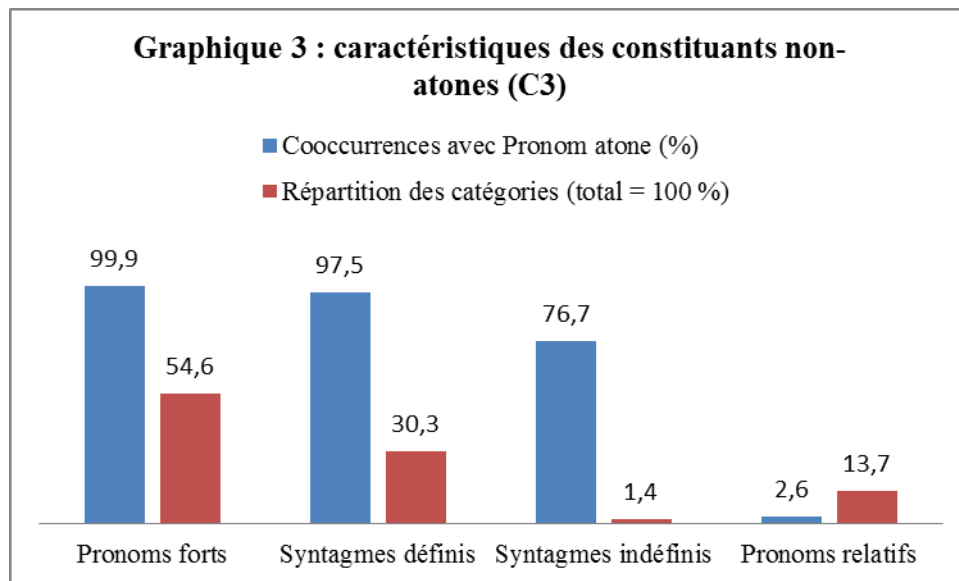
- (8) où es-tu Maman Ours ? (C2, Carla, 3;5)
- (9) où elle est ta maison ? (C3, Matteo, 3;11)
- (10) Sanbarbe il essaie de l'attraper. (C3, Matteo, 4;4)
- (11) et qu'est-ce qu'il fait Sanbarbe là ? (C3, Matteo, 4;4)

La catégorie non-cible « Sujet nul », illustrée en (12), diminue fortement et de façon attendue avec l'âge des enfants (26,6 %, 3,9 %, 0,6 %) (Palasis, 2012a). Finalement, la catégorie « Nom seul », illustrée en (13), est très faiblement représentée chez les jeunes enfants et n'augmente guère pendant la période étudiée (0 %, 0,4 %, 0,4 %). La baisse des Sujets nuls, l'augmentation concomitante des Pronoms et la stabilité basse des Noms semblent alors nous indiquer que ce sont des Pronoms et non des Noms qui prennent petit à petit la place des Sujets nuls dans le discours infantin. La nature des Sujets nuls enfantins semble donc pronominale, non nominale et le discours infantin est donc très loin de ressembler au français écrit qui, d'après les données de Blanche-Benveniste (1994), comporte entre 40 % et 80 % de syntagmes nominaux sujets. *A contrario* et en toute logique, ce discours s'apparente bien davantage aux productions adultes orales lorsque ceux-ci s'adressent à des enfants (cf. les corpus mentionnés en Section 1).

- (12) peux chanter ? (C1, Raphaël, 2;6)
- (13) Juliette est en vacances. (C3, Elena, 4;9)

3.2. Les autres constituants

Les exemples (5) et (6) ci-dessus illustrent deux types de cooccurrences possibles : le pronom atone émerge soit avec un pronom fort soit avec un syntagme nominal défini. Le Graphique 3 ci-dessous, réalisé à partir des données de C3, nous informe d'une part que, en effet, ces deux catégories émergent très rarement seules (dans 99,9 % et 97,5 % des cas, elles sont accompagnées d'un pronom atone) et, d'autre part, qu'elles couvrent 84,9 % des sujets du corpus (54,6 % + 30,3 %). Les dix-neuf enfants de C3 produisent ces deux types d'énoncés. Nous constatons ainsi que la plupart des cooccurrences est donc grammaticalement définie (*moi, le nounours*, etc.). L'étude des syntagmes nominaux réalisés seuls dans le même corpus (les 0,6 % dans le Graphique 2, soit $n=35$, cf. (13) ci-dessus) révèle la même caractéristique dans 71,4 % des cas, donc dans une proportion légèrement moins forte.



A *contrario*, les syntagmes nominaux indéfinis, qu'ils soient seuls ou accompagnés d'un pronom atone (cf. les exemples (14) et (15) ci-dessous), sont donc extrêmement peu présents en fonction sujet (1,4 %). Douze enfants sur les dix-neuf du groupe produisent ces syntagmes et, à la différence des constituants définis dont les taux d'association sont très homogènes d'un enfant à un autre (entre 88 et 100 %), les cooccurrences avec un pronom atone dans cette catégorie présentent des taux extrêmement hétérogènes (2 enfants : 0 %, 2 : 50 %, 1 : 66,7 %, 1 : 100 %). Notons également que l'un des deux enfants qui ne combine jamais un syntagme nominal indéfini avec un pronom atone dans ce corpus est le même et seul enfant à produire un pronom fort seul (cf. exemple (16)). Néanmoins, le Graphique 3 met en avant le fait que, lorsqu'ils émergent à cette place, les syntagmes nominaux indéfinis sont majoritairement accompagnés d'un pronom atone (76,7 %).

(14) des écureuils vit dans les arbres. (C3, Kelsang, 4;8)

(15) et même qu'un Pokemon il a des crocodiles. (C3, Massimo, 3;9)

(16) mais lui n'est pas d'accord. (C3, Kelsang, 4;7)

Terminons ce tour d'horizon par un type de cooccurrence particulièrement intéressant dans la mesure où il est considéré comme non-cible et donc inexistant dans les productions adultes (Labelle, 1988). Il s'agit de l'association d'un pronom relatif et d'un pronom atone, comme illustré en (17) et (18) ci-dessous. Bien qu'absentes du langage adressé aux enfants, ces cooccurrences représentent 2,6 % des pronoms relatifs et concernent quatre enfants dans C3 avec des taux relativement homogènes entre 9,1 % et 20 %. Nous nous sommes également intéressée à la forme du pronom relatif sujet *qui* dans les productions enfantines dans la mesure où la non-élision du [i] devant voyelle est interprétée comme un argument en faveur de l'analyse morphologique de ce [i] en français colloquial du Québec (Auger, 1994). Dix-sept enfants sur dix-neuf produisent le pronom relatif *qui* devant une voyelle dans C3 ($n=92$) et seuls 13 % de ces occurrences sont élidés. Globalement, les enfants n'élident donc pas *qui* devant voyelle et onze enfants ne l'élident jamais ($n=52$), comme illustré en (19).

(17) le petit renard qui il a faim. (C2, Mathilde, 3;2)

(18) ma mère qui il a dit qu'y aura des médailles. (C3, Massimo, 3;9)

(19) ah c'est le tee-shirt qui est trop grand ! (C3, Dylan, 3;8)

3.3. La forme du pronom *il*

De nombreux travaux ont montré que la forme phonologique des pronoms atones varie en fonction du contexte en français oral (Morin, 1979, Ashby, 1984). Ainsi, le pronom *il* est présent sous sa forme pleine devant voyelle et est élide devant consonne. D'après Pope (1934), cette réduction phonologique est attestée en français populaire depuis le XII^e siècle et Chervel (2011) nous informe que l'école primaire et la pression de la norme ont restauré la prononciation du [i] devant consonne au XIX^e siècle. Cette alternance dans les formes du pronom *il* est donc bien connue et attestée, que ce soit dans les corpus adultes ou enfantins (Grégoire, 1947, P. Guillaume, 1927). C2 et C3 ne dérogent pas à la règle puisque 99,4 % et 99,3 % des pronoms *il* dans ces deux corpus (totaux respectifs : $n=1688$ et $n=2208$) respectent l'alternance *i/il*, comme illustré en (20) et (21) ci-dessous.

(20) i veut sa maman parc'que il est tout seul. (C2, Nina, 3;3)

(21) et après i sort et après il habite ici. (C3, Dylan, 3;8)

L'élément nouveau que nous souhaitons mettre en avant est le contexte dans lequel l'alternance ne s'applique pas dans C2 (0,6 %, $n=11$) et C3 (0,7 %, $n=15$). Les données de C2 sont particulièrement tranchées à cet égard puisque dix énoncés sur onze présentent un contexte récurrent : la présence de la particule préverbale négative *ne*, comme illustré en (22). Notons que l'utilisation de la négation discontinue est très rare chez les enfants de cet âge puisqu'elle n'est présente que dans 1,2 % des énoncés verbaux négatifs de C2 (soit $n=16$). Dans C2, nous observons alors une corrélation systématique entre le type de négation (simple vs discontinue) et la forme de *il* (élide vs pleine) : C2 ne comporte ainsi aucun énoncé dans lequel la négation simple et la forme pleine de *il* ou la négation discontinue et la forme élide de *il* sont cooccurrents. Tous les énoncés négatifs respectent donc les associations suivantes : négation discontinue et forme pleine (cf. (22)) ou négation simple et forme élide (cf. (23)) (Palasis, 2013).

(22) il ne chante plus. (C2, Mathilde, 3;2)

(23) i nous mangera pas. (C2, Lucille, 3;1)

Qu'en est-il dans C3 ? Quinze énoncés sur 2208 (0,7 %) comportent une forme non-élide de *il* devant consonne dans ce corpus. L'étude des contextes dans ces énoncés révèle que sept d'entre eux sont négatifs et que la négation y est systématiquement représentée par le morphème discontinu *ne...pas*, comme dans C2. La corrélation décrite d'après les données de C2 entre le type de négation et la forme du pronom *il* existe donc également dans C3, comme illustré par la paire d'énoncés en (24) et (25) ci-dessous, produite par la même enfant, au cours de la même séance, avec le même verbe. La forme non-élide de *il* devant consonne peut donc être expliquée par la présence du morphème négatif *ne* dans 99,94 % des cas dans C2 et 99,64 % des cas dans C3. Nous reviendrons sur les 0,06 % ($n=1$) et 0,36 % ($n=8$) restants dans notre discussion.

(24) mais lui il ne voulait toujours pas. (C3, Lou, 4;9)

(25) mais i voulait pas que j' le nourrisse. (C3, Lou, 4;9)

4. Discussion

4.1. Sur le statut morphosyntaxique des clitiques

L'observation de nos données enfantines permet de confirmer et d'affiner les descriptions existantes sur les sujets oraux enfantins. Tout d'abord, nous avons noté que, non seulement les pronoms atones sont majoritairement présents dans les propositions indépendantes et principales, mais que leur présence augmente avec l'âge dans ces contextes (de 73,4 % dans C1 à 99 % dans C3). La proposition subordonnée relative représente alors le seul contexte

dans lequel le clitique est majoritairement absent (97,4 %). Cette dernière observation est intéressante à deux titres.

Tout d'abord, le taux d'absence très élevé n'est pas surprenant dans la mesure où le type de cooccurrence « Pronom relatif sujet + Clitique » n'existe pas dans les productions adultes (Labelle, 1988). Nous faisons alors l'hypothèse que ces cooccurrences chez les quatre enfants qui les produisent (total $n=11$) ne sont pas le résultat d'une répétition mais d'une computation morphosyntaxique⁵ au cours de laquelle l'enfant considère le clitique comme un marqueur morphologique, ce qui lui permet de l'associer à un pronom relatif sujet. Cette proposition nous renvoie alors à l'analyse de type (ii) mentionnée dans la Section 1 (Creissels, 1995, Culbertson, 2010). Par extension, nous pouvons alors supposer que ces enfants utilisent tous leurs clitics comme des marqueurs de personne, non seulement dans ce contexte avec un pronom relatif mais également dans les trois autres contextes étudiés (syntagmes nominaux indéfinis, syntagmes nominaux définis et pronoms toniques). Nous nous retrouvons alors dans une situation qui existe par ailleurs dans d'autres langues. En effet, les dialectes du nord de l'Italie sont bien connus dans la littérature pour deux raisons : ces dialectes doublent leurs pronoms toniques, leurs syntagmes nominaux définis et indéfinis et leurs pronoms relatifs avec un clitique et ce clitique est analysé de façon consensuelle comme un marqueur d'accord (Rizzi, 1986, Brandi et Cordin, 1989, Poletto, 2000), comme illustré en (26) ci-dessous. Malgré la présence constante du clitique, ces dialectes sont donc considérés comme des langues à sujet nul, au même titre que l'Italien standard par exemple. Néanmoins, le phénomène du doublement du sujet n'est pas un phénomène homogène dans cette aire dialectale. Poletto (1993, 2000) notamment a étudié cette variation. L'auteure a ainsi non seulement proposé un classement des dialectes en fonction du nombre de contextes d'application du doublement du sujet (de un à quatre) mais a également établi une liste d'implications entre les quatre contextes (cf. (27)-(29) ci-dessous).

(26) El Gianni el magna. « Jean il mange » (dialecte trentin, extrait de Rizzi, 1986)

(27) Si les syntagmes nominaux définis sont doublés par un clitique dans un dialecte, alors tous les pronoms toniques le sont également dans cette même langue (en trentin par exemple).

(28) Si les syntagmes nominaux indéfinis sont doublés par un clitique, alors tous les syntagmes nominaux définis et tous les pronoms toniques le sont également (en milanais par exemple).

(29) Si les variables dans les contextes *wh*- (c'est-à-dire les structures relatives, interrogatives et clivées) sont doublées par un clitique, alors le doublement est obligatoire avec tous les autres types de sujets (en ligurien par exemple).

Il est alors extrêmement intéressant de retrouver le même type d'implications dans nos corpus de français enfantin. En effet, dans C2, les trois enfants qui doublent les pronoms relatifs doublent également les syntagmes indéfinis, définis et les pronoms toniques. Dans C3, deux des quatre enfants qui doublent les pronoms relatifs doublent également les trois autres types de sujets (les deux autres enfants ne produisant aucun syntagme indéfini, l'implication ne peut être vérifiée à ce niveau). Ainsi, la présence quasi-obligatoire de ces clitics en position préverbale uniquement et les similitudes avec le doublement dans les dialectes du nord de l'Italie sont autant d'arguments qui nous conduisent à proposer une analyse de ces clitics en tant que marqueurs de personne dans les corpus enfantins étudiés. Nous proposons ainsi que dans un énoncé tel que *et après i sait* (Elena, 4;4) : (i) le clitique est utilisé comme un marqueur préverbal de personne ; (ii) il n'y a pas de sujet grammatical phonétiquement réalisé ; (iii) le français oral fait partie du même groupe que les autres langues romanes dites « à sujet nul » dans la mesure où la réalisation d'un sujet grammatical n'est pas obligatoire.

⁵ Nous suivons le type de raisonnement appliqué en psycholinguistique depuis les travaux fondateurs de Berko (1958) sur l'acquisition de la morphologie.

Néanmoins, nous avons également vu que certaines occurrences représenteraient alors des contrexemples à cette analyse morphologique. *Quid* par exemple des énoncés dans lesquels le sujet n'est pas doublé par un clitique (C1 : 0 %, C2 et C3 : 0,4 %, soit $n=39$ et $n=35$) ? La forme du clitique *il* lorsqu'il n'est pas élidé devant consonne nous semble alors apporter un éclairage supplémentaire à la discussion sur le statut morphosyntaxique des clitics en français et ouvre notre champ d'horizon à une discussion plus générale sur le statut du français étudié dans ces corpus enfants.

4.2. Sur le statut du français enfantin contemporain

L'hypothèse morphologique défendue ici fait donc les prédictions que le clitique est : obligatoire, préverbal et élidé devant consonne. Les énoncés avec un sujet non doublé par un clitique ou avec un clitique postverbal ou non-élidé devant consonne sont donc incompatibles avec cette analyse. Nous pouvons néanmoins rendre compte de leur existence et de leur fonctionnement en suivant les propositions de Massot (2008, 2010), Barra-Jover (2010, 2013), Zribi-Hertz (2011) et Massot et Rowlett (2013) sur l'hypothèse d'une diglossie en France (d'après les travaux de Ferguson, 1959). En effet, si, comme ces auteurs, nous postulons l'existence de deux grammaires du français, c'est-à-dire une grammaire acquise spontanément dès le plus jeune âge et généralement utilisée dans des contextes plutôt informels (appelons-la G_1) et une seconde grammaire apprise plus tard, souvent par le biais de l'école et utilisée dans des contextes plus formels (G_2), alors nous pouvons rendre compte de la variation entre les deux types de comportements du clitique *il* en postulant l'existence de deux paradigmes distincts de clitics. Les clitics de G_1 seraient des marqueurs de personne (obligatoires, préverbaux, élidés devant consonne) et les clitics de G_2 seraient des pronoms syntaxiques (pronoms au sens étymologique du terme, pré- ou postverbaux, de forme fixe quel que soit le contexte).

Les données sur la négation dans C2 et C3 étayaient cette hypothèse dans la mesure où elles montrent que les distributions de l'éllision et de la négation ne sont pas aléatoires mais que, au contraire, elles corréllent dans les énoncés, comme illustré avec les exemples (24) et (25), repris en (30) et (31) ci-dessous. La négation discontinue (en (31)), les clitics postverbaux ((8), repris en (32)), les clitics non-élidés devant consonne (en (31) et (33)), les pronoms forts et les syntagmes nominaux non-doublés (en (34) et (35)) relèvent alors de G_2 , les enfants utilisant les deux grammaires de façon différenciée mais inconsciente (un des principes de la diglossie). Le faible taux de représentation de G_2 dans nos corpus est cohérent avec la chronologie des acquisitions dans une situation diglossique et nous nous attendons à une augmentation des données de G_2 dans notre futur C4, en termes de proportion des énoncés par rapport à G_1 et du nombre d'enfants qui pratiquent G_2 (pour la négation discontinue par exemple, les 1,2 % des énoncés verbaux négatifs sont répartis sur quatre enfants dans C2 et les 1,8 % des énoncés verbaux négatifs sont répartis sur dix enfants dans C3).

(30) mais i voulait pas que j' le nourrisse. (C3, Lou, 4;9) → G_1

(31) mais lui il ne voulait toujours pas. (C3, Lou, 4;9) → G_2

(32) où es-tu Maman Ours ? (C2, Carla, 3;5) → G_2

(33) il va mettre les gens en prison (C3, Eliza, 4;5) → G_2

(34) mais lui n'est pas d'accord (C3, Kelsang, 4;7) → G_2

(35) mais des fois Lisa va à la cantine. (C3, Matteo, 4;2) → G_2

4.3. Sur l'aspect diachronique de la diglossie

La remarque de Hirschbühler et Labelle (2004), traduite ci-dessous, sur les données enfantines du futur roi Louis XIII (Foisil, 1989) nous incite à élargir notre champ

d'investigations à ce corpus enfantin du XVII^e siècle, décrit comme très précieux pour l'étude du français (Morgenstern, 2009).

Louis XIII a deux grammaires en compétition, une avec *ne* et une sans *ne*. Si ceci est correct, l'omission de *ne* à l'oral, qui est également une caractéristique du français oral contemporain, remonte aussi loin que le début du XVII^e siècle. (Hirschbühler et Labelle, 2004 : 196)

Nos investigations, actuellement en cours, portent pour le moment sur les énoncés verbaux négatifs de Louis entre 2;0 et 3;3 ($n=67$). Elles mettent en avant les résultats provisoires suivants. Tout d'abord, Louis, comme nos petits locuteurs contemporains, associe un clitique à quasiment chacun de ses verbes fléchis dans ce petit corpus (94 % des énoncés).⁶ Les quatre énoncés négatifs qui ne comportent pas de clitiques sont deux énoncés à sujet nul (exemple en (36)) et deux énoncés avec un syntagme nominal sujet seul (exemple en (37)). Nous constatons alors, dans un second temps, que les deux énoncés avec un syntagme nominal seul comportent la forme discontinue de la négation. Ces deux énoncés enfantins du XVII^e siècle combinent ainsi deux caractéristiques majeures qui ont été corrélées d'après l'observation de données du XXI^e siècle. Selon notre hypothèse, ces deux énoncés présentent alors deux caractéristiques cohérentes pour être classés comme relevant de G₂. Cette observation intéressante mais uniquement liminaire sera complétée prochainement par l'étude de l'alternance *i/il* dans ces mêmes énoncés puis par une observation étendue à l'ensemble des énoncés avant d'élargir l'investigation à une tranche d'âge plus importante.

(36) fai pa beau saboté. « on n'est pas beau en sabots » (Louis, 2;6)

(37) que le Roy ne veu pa descende.⁷ (Louis, 2;11)

4.4. Sur l'aspect diachronique du statut morphosyntaxique des clitiques

Nous avons mentionné en Section 1 que l'une des raisons pour lesquelles le débat sur le statut morphosyntaxique du pronom existe en français et non dans d'autres langues telles que l'anglais est son statut phonologique faible. En français, le pronom personnel est atone. Il ne peut donc pas être utilisé seul (en réponse à une question par exemple), ni être coordonné avec un autre pronom clitique, ni être séparé du verbe fléchi par autre chose qu'un autre clitique (Kayne, 1975), contrairement aux pronoms de l'ancien français (exemples extraits de Kaiser, 2009). En français moderne, les pronoms toniques sont utilisés dans ces trois contextes, comme illustré dans le Tableau 1.

	Ancien français	Français moderne
Isolation	jo	*je / moi
Coordination	tu é li	*tu et il / toi et lui
Séparation	je meismes preu i avrai ⁸	*il, souvent, mange / lui

Tableau 1 : évolution du statut des pronoms personnels en français

Nous avons donc dans un premier temps, entre le latin et l'ancien français, l'émergence d'un paradigme de pronoms préverbaux et indépendants du verbe. Le processus diachronique évoqué, lorsqu'un lien est postulé avec l'affaiblissement du caractère distinctif des suffixes verbaux depuis le latin, est la déflexivité, c'est-à-dire le processus par lequel un morphème flexionnel acquiert le statut de morphème libre, généralement par antéposition (cf. par exemple Moignet (1965), à la suite des travaux de G. Guillaume (2003 [1950-1959])). Puis, dans un deuxième temps, à partir du Moyen français, ces pronoms perdent de leur

⁶ Dufter et Stark (2007) ont noté la même proportion de clitiques dans les propositions principales négatives de Louis entre 4;0 et 5;0.

⁷ En réponse à la question : « Que dit-on ? ».

⁸ « je moi-même en tirerai un avantage ».

indépendance et deviennent clitiques. Ce processus d'affaiblissement est fréquemment qualifié de grammaticalisation (Hopper et Traugott, 1993) et est généralement distingué de la déflexivité (Bottineau, 2010). C'est à cette période qu'émergent les pronoms toniques et l'utilisation concomitante des deux séries de pronoms avec un même verbe fléchi.

La question qui était posée dans le cadre de cette étude était de savoir si les clitiques du français ont franchi la dernière étape de la grammaticalisation, celle qui consiste à passer du statut de clitique phonologique à celui d'affixe flexionnel. Nous avons mentionné que la question était largement débattue chez les syntacticiens et que les deux réponses possibles reposaient chacune sur des arguments de valeurs (De Cat, 2005 vs Culbertson, 2010). Nous suggérons que les partisans des deux camps ont finalement raison dans la mesure où ils décrivent chacun une langue différente : le français oral (G₁) dans lequel les clitiques sont manipulés comme des marqueurs préverbaux de personne et le français standard (G₂) dans lequel les clitiques sont manipulés en syntaxe au même titre que les syntagmes nominaux.

Bibliographie

- Ashby William J., 1984, The elision of /l/ in French clitic pronouns and articles, in Ernst Pulgram (éd.), *Romanitas: Studies in Romance Linguistics*, Ann Arbor, University of Michigan Press, p. 1-16.
- Auger Julie, 1994, *Pronominal Clitics in Québec Colloquial French: A Morphological Analysis*, PhD Dissertation, University of Pennsylvania.
- Barra-Jover Mario, 2010, « Le » français ou ce qui arrive lorsqu'un état de choses est observé comme une entité, *Langue française* N° 168, p. 3-18.
- Barra-Jover Mario, 2013, Linguistique et école primaire : en quoi l'approche diglossique est-elle la meilleure façon d'apprendre le français académique ?, *Journal of French Language Studies* N° 23, p. 87-108.
- Bauche Henri, 1928, *Le langage populaire*, Paris, Éditions Payot.
- Berko Jean, 1958, The Child's Learning of English Morphology, *Word* N° 14, p. 150-177.
- Blanche-Benveniste Claire, 1994, Quelques caractéristiques grammaticales des « sujets » employés dans le français parlé des conversations, in Marina Yaguello (éd.), *Subjecthood and Subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"*, London, Éditions Ophrys, p. 77-107.
- Blanche-Benveniste Claire, 2000, *Approches de la langue parlée en français*, Paris, Éditions Ophrys.
- Bonnard Henri, 1997, *Grammaire française à l'usage de tous*, Paris, Éditions Magnard.
- Bottineau Didier, 2010, Introduction, *Langages* N° 178, p. 3-10.
- Brandi Luciana et Patrizia Cordin, 1989, Two Italian Dialects and the Null Subject Parameter, in Osvaldo Jaeggli et Kenneth Safir (éds.), *The Null Subject Parameter*, Dordrecht, Kluwer Academic Publishers, p. 111-142.
- Chervel André, 2011, Histoire de l'orthographe française : réformes, enseignement, niveau, in Olivier Bertrand et Isabelle Schaffner (éds.), *Variétés, variations & formes du français*, Palaiseau, Les Éditions de l'École Polytechnique, p. 195-210.
- Chifflet Laurent, 1659, *Essay d'une parfaite grammaire de la langue française*, Anvers, Jacques van Meurs.
- Creissels Denis, 1995, *Éléments de syntaxe générale*, Paris, Presses Universitaires de France.
- Culbertson Jennifer, 2010, Convergent Evidence for Categorical Change in French: From Subject Clitic to Agreement Marker, *Language* N° 86, p. 85-132.
- De Cat Cécile, 2005, French Subject Clitics are not Agreement Markers, *Lingua* N° 115, p. 1195-1219.
- Denis Delphine et Anne Sancier-Chateau, 1994, *Grammaire du français*, Paris, Le Livre de Poche.
- Dufter Andreas et Elisabeth Stark, 2007, La linguistique variationnelle et les changements linguistiques 'mal compris' : le cas du *ne* de négation, in Bernard Combettes et Christiane Marchello-Nizia (éds.), *Études sur le changement linguistique en français*, Nancy, Presses Universitaires de Nancy, p. 115-128.
- Ferguson Charles A., 1959, Diglossia, *Word* N° 15, p. 325-340.
- Foisil Madeleine (éd.), 1989, *Journal de Jean Héroard Médecin de Louis XIII*, Paris, Fayard.
- Fonseca-Greber Bonnie et Linda R. Waugh, 2003, The Subject Clitics of Conversational European French: Morphologization, Grammatical Change, Semantic Change, and Change in Progress, in Rafael Núñez-Cedeño, Luis López et Richard Cameron (éds.), *A Romance Perspective on Language Knowledge and Use*, Amsterdam, John Benjamins, p. 99-117.
- Grégoire Antoine, 1947, *L'apprentissage du langage : la troisième année et les années suivantes*, Paris & Liège, E. Droz & Faculté de Philosophie et Lettres.
- Grevisse Maurice et André Goosse, 2008, *Le bon usage*, Bruxelles, De Boeck & Ducolot (14ème édition).
- Guillaume Gustave, 2003 [1950-1959], *Prolégomènes à la linguistique structurale I*, Québec, Les Presses de l'Université Laval.
- Guillaume Paul, 1927, Le développement des éléments formels dans le langage de l'enfant, *Journal de Psychologie* N° 24, p. 203-229.

- Heap David et Yves Roberge, 2001, Cliticisation et théorie syntaxique, 1971-2001, *Revue québécoise de linguistique* N° 30, p. 63-90.
- Hirschbühler Paul et Marie Labelle, 2004, Enclisis and proclisis of object pronouns at the turn of the 17th century: The speech of the future Louis XIIIth, in Randall S. Gess et Deborah Arteaga (éds.), *Historical Romance Linguistics: Retrospective and perspectives*, Amsterdam, John Benjamins, p. 187-208.
- Hopper Paul J. et Elizabeth Closs Traugott, 1993, *Grammaticalization*, Cambridge, Cambridge University Press.
- Kaiser Georg A., 2009, Losing the Null Subject. A Contrastive Study of (Brazilian) Portuguese and (Medieval) French, in Georg A. Kaiser et Eva-Maria Remberger (éds.), *Null-subjects, Expletives, and Locatives in Romance*, Konstanz, Konstanz Working Papers of the Department of Linguistics, p. 131-156.
- Kayne Richard S., 1975, *French Syntax. The Transformational Cycle*, Cambridge, MA, The MIT Press.
- Labelle Marie, 1988, *Prédication et mouvement : le développement de la construction relative chez les enfants francophones*, Thèse de Doctorat, Université d'Ottawa.
- Lambrecht Knud, 1981, *Topic, Antitopic and Verb Agreement in Non-standard French*, Amsterdam, John Benjamins.
- MacWhinney Brian, 2000a, *The CHILDES Project. Tools for Analyzing Talk: The Database*, Mahwah, NJ, Lawrence Erlbaum Associates, <http://childes.talkbank.org/>.
- MacWhinney Brian, 2000b, *The CHILDES Project. Tools for Analyzing Talk: Transcription Format and Programs*, Mahwah, NJ, Lawrence Erlbaum Associates, <http://childes.talkbank.org/>.
- Massot Benjamin, 2008, *Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie : arguments morphosyntaxiques*, Thèse, Université Paris 8.
- Massot Benjamin, 2010, Le patron diglossique de variation grammaticale en français, *Langue française* N° 168, p. 87-106.
- Massot Benjamin et Paul Rowlett (éds.), 2013, L'hypothèse d'une diglossie en France, *Journal of French Language Studies* N° 23.
- Moignet Gérard, 1965, *Le pronom personnel français. Essai de psycho-systématique historique*, Paris, Klincksieck.
- Morgenstern Aliyah, 2006, *Un « je » en construction : genèse de l'auto-désignation chez le jeune enfant*, Paris, Éditions Ophrys.
- Morgenstern Aliyah, 2009, *L'enfant dans la langue*, Paris, Presses Sorbonne Nouvelle.
- Morin Yves-Charles, 1979, La morphophonologie des pronoms clitiques en français populaire, *Cahier de linguistique* N° 9, p. 1-36.
- Palasis Katerina, 2005, *Problèmes d'acquisition et le Paramètre du Sujet Nul*, Mémoire de Master, Université Nice Sophia Antipolis.
- Palasis Katerina, 2010a, Introducing New French Child Data: Thoughts on their Gathering and Coding, *Corpus* N° 9 *La syntaxe de corpus*, p. 33-51, <http://corpus.revues.org/index1801.html>.
- Palasis Katerina, 2010b, *Syntaxe générative et acquisition : le sujet dans le développement du système linguistique du jeune enfant*, Villeneuve d'Ascq, ANRT Diffusion.
- Palasis Katerina, 2012a, Economy Principles within Child Speech: When the Nominative Clitic does not Surface, in Sandrine Ferré, Philippe Prévost, Laurice Tuller et Rasha Zebib (éds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages*, Newcastle, Cambridge Scholars Publishing, p. 261-283.
- Palasis Katerina, 2012b, *Romance French Child Corpus 1 (2;5-4;0)*, CHILDES, <http://childes.talkbank.org/browser/index.php?url=Romance/French/Palasis/>.
- Palasis Katerina, 2013, The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French, *Journal of French Language Studies* N° 23, p. 17-35.
- Platon, [1920-1964], *Le Sophiste, Oeuvres complètes*, Paris, Les Belles Lettres ; traduction de Diès.
- Poletto Cecilia, 1993, *La sintassi del soggetto nei dialetti italiani settentrionali*, Padova, Unipress.
- Poletto Cecilia, 2000, *The Higher Functional Field: Evidence from Northern Italian Dialects*, New York & Oxford, Oxford University Press.
- Pope Mildred K., 1934, *From Latin to Modern French*, Manchester, Manchester University Press.
- Prévost Philippe, 2009, *The Acquisition of French: the Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*, Amsterdam, John Benjamins.
- Rizzi Luigi, 1986, On the Status of Subject Clitics in Romance, in Osvaldo Jaeggli et Carmen Silva-Corvalan (éds.), *Studies in Romance Linguistics*, Dordrecht, Foris, p. 391-419.
- Zribi-Hertz Anne, 1994, La syntaxe des clitiques nominatifs en français standard et en français avancé, *Travaux de Linguistique et de Philologie* N° 32, p. 131-148.
- Zribi-Hertz Anne, 2011, Pour un modèle diglossique de description du français : quelques implications théoriques, didactiques et méthodologiques, *Journal of French Language Studies* N° 21, p. 231-256.

6. Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, 161, 125-143, <http://dx.doi.org/10.1016/j.lingua.2014.11.012>



Available online at www.sciencedirect.com

ScienceDirect

Lingua 161 (2015) 125–143

Lingua

www.elsevier.com/locate/lingua

Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony



Katerina Palasis*

Université Nice Sophia Antipolis, CNRS, BCL, UMR 7320, France

Received 7 January 2013; received in revised form 15 May 2014; accepted 30 November 2014

Available online 7 January 2015

Abstract

This contribution aims to propose a corpus-based analysis of variation and acquisition of subject clitics and preverbal negation in European French within a diglossic approach. The investigation collates previous and new, contemporary and diachronic, adult and child data from France and Belgium. The results point to an analysis of subject clitics as agreement markers in contemporary French. The negative particle *ne* is eliminated from the list of arguments against the morphological analysis of subject clitics, since negative utterances with agreement markers display postverbal negation only. A strong correspondence between two characteristics, i.e. morpho-syntactic status of subject clitics (agreement markers vs. arguments) and type of negation (simple vs. discontinuous), is established supporting the hypothesis on grammatical consistency and pointing to the existence of two different grammars of French (labeled chronologically G₁ and G₂). Diatopic data inform us that the correspondence seems to hold throughout France, but that diglossia does not appear to apply (or at least applies differently) when reaching the Belgian frontier. Finally, adult and child diachronic data (17th–19th century) also display an interesting co-occurrence in terms of clitics and negation, and invite us to further our understanding of the acquisition and processing of expletive clitics.

© 2014 Elsevier B.V. All rights reserved.

Keywords: Subject clitic; Negation; French; Diglossia; Acquisition; Diachrony

1. Introduction

The comparative study of preverbal clitics in colloquial and standard French represents a lively and inspiring area for research in a number of different perspectives. Indeed, the debate surrounding the distribution and forms of these preverbal elements, i.e. subject (*je* 'I', *tu* 'you-sg', *il* 'he', etc.), object (*me* 'me', *te* 'you-sg', *le* 'him', etc.), and adverbial clitics (*y* 'there', *en* 'from there'), together with the negative particle *ne*, is of interest to (at least) morpho-syntacticians studying French in synchrony, diachrony, diatopy, and/or cross-linguistically, language acquisitionists, sociolinguists, and cognitive scientists. This contribution brings these complementary points of view together by examining subject clitics and negation in new child data and by considering the results in the light of the other perspectives. The main goals of this paper are to contribute to (i) the debates surrounding the morpho-syntactic status of subject clitics and preverbal negation in European French, and (ii) the discussion on the formalization of variation in this language in synchrony, diachrony, and

* Correspondence to: BCL – Campus Saint Jean d'Angély – SJA3, 24 avenue des Diables Bleus, 06357 Nice Cedex 4, France.

Tel.: +33 (0)4 89 88 14 40; fax: +33 (0)4 89 88 14 50.

E-mail address: palasis@unice.fr.

<http://dx.doi.org/10.1016/j.lingua.2014.11.012>

0024-3841/© 2014 Elsevier B.V. All rights reserved.

diatopy. The investigation supports the following hypotheses: (i) subject clitics are agreement markers in colloquial French and syntactic subjects in standard French, and (ii) diglossia can account for this variation in France.

The article is organized as follows. Section 2 presents the theoretical debates on the morpho-syntactic status of subject clitics and preverbal negation, and on the formalization of variation in French. Section 3 examines subject clitics and negation in new child data from south-eastern France (12,969 utterances by 19 children). Section 4 turns to diatopy, and attempts to outline the geographical area where diglossia applies by examining fieldwork undertaken on child and adult central, Parisian, northern and Belgian French. Section 5 then assesses the plausibility of the current existence of diglossia in the brain/mind of French native speakers against adult and child data from the 17th century onwards. Finally, section 6 contains concluding remarks and leads for further research.

2. Variation in contemporary adult French

Preverbal clitics in French can be subject (*je* 'I', *tu* 'you-sg', *il* 'he', etc.), object (*me* 'me', *te* 'you-sg', *le* 'him', etc.), adverbial (*y* 'there', *en* 'from there'), or negative clitics (preverbal marker *ne* 'neg'). This contribution focuses on two of these categories, i.e. subject and negative clitics.

2.1. The morpho-syntactic status of subject clitics

A particularly long-standing debate surrounds the morpho-syntactic analysis of subject clitics in French, whether in traditional grammar or in more recent frameworks such as generative linguistics (overview in Heap and Roberge, 2001). In his classic work on standard French, Kayne (1975) described these elements as proper syntactic arguments occupying the canonical subject position, and cliticizing to the finite verb at the post-syntactic, phonological level (1a). Kayne's work followed by other influential investigations (Brandi and Cordin, 1989; Cardinaletti and Starke, 1999; De Cat, 2005; Rizzi, 1986; among others) hence analyzed these formatives as morphologically independent words fulfilling a syntactic role. In this configuration, when a subject clitic and a DP co-occur, the syntactic structure represents an instance of left-dislocation (also labeled 'Clitic Left Dislocation (CLLD)' after Cinque, 1990), and the DP occupies a topic position above the subject position (TopP, as in (1b); Rizzi, 1997).¹ Since definiteness characterizes topics, indefinite entities are not expected to occur in TopP, as shown in (1c–d)^{2,3}:

- (1) a. [_{IP} il parle]
 he speaks
 'he speaks'
- b. [_{TopP} l' enfant [_{IP} il parle]]
 the child he speaks
 'the child speaks'
- c. *_[TopP] un enfant [_{IP} il parle]]
 a child he speaks
 'a child speaks'
- d. *_[TopP] quelqu'un [_{IP} il parle]]
 someone he speaks
 'someone speaks'

In the 1970s–1980s, (socio-)linguists started to examine syntactic variation between written/standard French on the one hand and spoken/colloquial French on the other hand. Questions arose on the status of the latter with regard to the former (e.g. Blanche-Benveniste, 1983; Lambrecht, 1981), and structures that had previously been stigmatized were gradually quantified and described as belonging to non-standard varieties of French (e.g. Hulk, 1991; Zribi-Hertz, 1994). The co-occurrence of DPs and subject clitics in oral French is hence now well documented.⁴ First, it is established that the

¹ I leave aside the question of whether the DP is base-generated in the left periphery or copied to it (e.g. De Cat, 2007a vs. Grohmann, 2003).

² The term 'indefinite entities' (or 'QPs') refers to: (i) DPs whose determiners bear a quantifying value only, e.g. *un* 'a-masc', *une* 'a-fem', *des* 'some', *tous* 'all', *chaque* 'each', *aucun* 'no', etc., and (ii) indefinite pronouns, e.g. *personne* 'nobody', *rien* 'nothing', etc. (Grevisse and Goosse, 2008:738).

³ This definition and exclusions are controversial though. Reinhart (1981) for instance clearly distinguishes topichood from old information, and De Cat (2007b), among others, qualifies the exclusions reported in (c) and (d).

⁴ Note however that Baude (2006:26) reports that France is far behind with regard to oral databases. Gadet (2009:115) relates this lag to 'the influence of the "ideology of the standard" and, more specifically, the importance of normative attitudes'.

Table 1
Types of subjects in adult speech.

Corpus	Clitic	DP + Clitic	DP	Total <i>n</i>
Blanche-Benveniste (1994)	81.3%	12.8%	5.9%	203
C3	75.8%	23.9%	0.3%	6,357

phenomenon is common and long-standing in oral Metropolitan French, since 17th century grammarians were already concerned with the status and frequency of this co-occurrence: ‘Ne dites pas *Mon père, il est malade*, au lieu de *Mon père est malade*’ (Chifflet, 1659).⁵ Second, a general picture of the types of subjects speakers spontaneously use can be outlined. Thus, one major characteristic that distinguishes colloquial from standard French is the rarity of canonical subject DPs in the former. Indeed, speakers overwhelmingly introduce subject clitics before finite verbs, as shown in Table 1 with adult-directed and child-directed data (Blanche-Benveniste, 1994 and C3 exemplified in (2a)).⁶ Oral corpora also indicate that subject clitics emerge with indefinite entities, in coordinate VPs, preverbally in interrogative contexts, and that *il* ‘he/it’ is elided in colloquial French, contrary to standard French (Ashby, 1980, 1984; Blanche-Benveniste, 2000; Dye, 2011; Fonseca-Greber and Waugh, 2003; Kayne, 1975; Lambrecht, 1981; Morin, 1979; Zribi-Hertz, 1994; examples from C3 in (2b) and (2c)).

(2) Subjects in adult speech (C3):

- a. je crois que Lou elle va te montrer où il est le début.
I think that Lou she goes you show where it is the beginning
‘I think that Lou is going to show you where the beginning is’
- b. tous les pirates i[z] ont des tatouages.
all the pirates they have some tattoos
‘all pirates have tattoos’
- c. mais il est triste et i met les pieds dans l’ eau?
but he is sad and he puts the feet in the water
‘but he is sad and puts his feet in the water?’

Thus, the pervasive presence of preverbal subject clitics in many finite contexts has led some scholars to propose an alternative to the classic, syntactic analysis of these formatives by considering them as agreement markers not syntactic subjects in colloquial Metropolitan French as well as Quebec, Ontario, and Swiss varieties of French (Auger, 1994; Culbertson, 2010; Dye, 2011; Fonseca-Greber, 2000; Kaiser, 1994; Miller, 1992; Miller and Monachesi, 2003; Nadasdi, 1995; Pierce, 1992; Roberge, 1986, 1990; Zribi-Hertz, 1994).⁷ This analysis views co-occurring clitics and DPs as doubling, not dislocation. The DP occupies the subject position, and the emergence of the clitic is in principle less constrained.⁸ This alternative analysis of French subject clitics is hence reminiscent of the morphological analysis of subject clitics in some northern Italian dialects (henceforth NIDs) which have been shown to display clitics in all possible finite contexts, as illustrated in (3) with Poletto’s (2000) examples. Section 3.2 will point to interesting parallels between these dialects and French child speech.

(3) Doubling in NIDs:

- a. Pronominal DPs:
TI te magni sempre. (Venice)
YOU scl eat always
‘YOU always eat’
- b. Referential DPs:
el fio el mangia l pom. (Milan)
the boy scl eats the apple
‘the boy eats the apple’

⁵ ‘Do not say *My father, he is ill* for *My father is ill*’, cited in Blanche-Benveniste (2000:37).

⁶ Details on C3 are given in section 3.1.

⁷ King and Nadasdi (1997) illustrate that the analysis cannot be extended to all varieties of colloquial French though (Newfoundland French in Canada patterns like standard French with regard to subject clitics).

⁸ Doubling is not a unitary phenomenon though. See Poletto (2000) for qualifications on doubling with subject clitics, and Kallulli and Tasmowski (2008) with object clitics.

- c. QPs:
 un quidun el riverà in ritart. (Milan)
 a somebody scl will-arrive in late
 ‘somebody will arrive late’
- d. *Wh*-variables:
 le fomne che le neta le scale e endade via. (Friulian)
 the women who scl clean the stairs scl went away
 ‘the women who clean the stairs have left’

Section 3 will examine the properties of child subject clitics in French in terms of presence, position, repeatability, co-occurring elements, and // elision, between the ages of 3;6 and 4;10. These observations will be collated with previous results from the same children before 3;6 in order to (i) propose a longitudinal analysis of subject clitics in child French, and (ii) test the morphological analysis of these elements in this system compared to NIDs. A major counterargument to the analysis of subject clitics as agreement markers in French is the presence of intermediate clitics between the subject clitic and the finite verb, notably the preverbal negative marker *ne* (De Cat, 2005). The next section outlines the theoretical debate surrounding the status of this particular element.

2.2. The morpho-syntactic status of the negative particle *ne*

Zanutini (1997) established a threefold classification of Romance languages according to the distribution of their negative markers, as reported in (4).

- (4) Distribution of negative markers in Romance languages:
- a. Preverbal marker only, e.g. Italian:
 Gianni non legge articoli di sintassi.
 Gianni neg reads articles of syntax
 ‘Gianni does not read syntax articles’
 - b. Post-verbal marker only, e.g. Valdostain^{9,10}:
 lo film l’ èra pa dzen.
 the movie scl was neg beautiful
 ‘the movie was not good’
 - c. Co-occurring pre- and post-verbal markers, e.g. Standard French:
 je n’ ai pas parlé de toi.
 I neg have neg spoken of you
 ‘I have not spoken of you’

In Standard French, the preverbal clitic *ne* is traditionally analyzed as the head of an independent negative projection, viz. NegP (Ouhalla, 1991; Pollock, 1989; Rowlett, 1998; among others). Zanutini (1997), however, specifies that the case of French is complex due to sociological and geographical factors. Indeed, Ashby’s fieldwork (Ashby, 1976 onwards) has shown that *ne* can be frequently omitted in colloquial French, as in (5). The negative particle is hence generally absent from child-directed speech (Clark and de Marneffe, 2012; Dye, 2011).

- (5) Colloquial French:
 j’ ai pas parlé de toi.
 I have neg spoken of you
 ‘I have not spoken of you’

Variation between standard and colloquial French is thus widely documented in the literature, and different formalizations exist. Depending on the approach, the presence/absence of *ne* is deemed to be either sociolinguistically or grammatically constrained in adult speech (e.g. Ashby, 1981, 1982, 2001; Coveney, 1996, 2002 vs. Barra-Jover, 2004; Massot, 2010; further details in section 2.4), and its morpho-syntactic status is discussed. Ouhalla (1991:142–143) for instance, who

⁹ Valdostain is a variety of Franco-Provençal mainly spoken in Switzerland and eastern parts of France.

¹⁰ Markers come in different forms in this category, e.g. *nen* in Piedmontese, and different positions, e.g. Milanese *no* is lower than Valdostain *pa*. Their ‘defining property’ though is that the negative marker ‘always follows the main verb when it is finite and in a simple form’ (Zanutini, 1997:4).

applies the morphological analysis to subject clitics in colloquial French, suggests that the latter can co-occur with *ne*, and also attributes a morphological status to *ne* in the verbal complex since it appears ‘inside AGR [Agreement, i.e. the subject clitic] and outside TNS [Tense]’. To the contrary, De Cat (2005:1201–1203; 2007b:15–18), who mentions that *ne* is productive in spontaneous speech, argues that the negative particle cannot be morphological insofar as its value (negative or restrictive; referring to Godard, 2004), and its scope (clausal or longer-distance; referring to Godard, 2004; Milner, 1979; Rowlett, 1998) are syntactically determined. Culbertson (2010) also supports the syntactic analysis of *ne*. The morpho-syntactic status of this preverbal clitic is hence debated in adult data. Section 3.4 will examine *ne* in child data and attempt to shed different light on the matter.

2.3. Do both elements co-occur?

The presence of subject clitics and the absence of *ne* have been often related in adult- and child-directed speech. Drawing on adult data, Ashby (1977:62) for instance reports the existence of a correspondence between the absence of *ne* and the presence of a subject clitic (55% deletion with a clitic vs. 12.1% with a noun), and further suggests that: ‘The trend toward the synthesis of these clitics with the verb is likely the major reason for the advanced state of the loss of *ne* in this environment’.

As far as child-directed speech is concerned, Culbertson (2010:95) interestingly mentions that ‘*ne*-retention [average rate: 7.6%, 456/5990] is clearly affected by the properties of the preceding subject’ since *ne* is preferentially dropped after a subject clitic (retention rate in this context: 6.3%, 332/5229), whereas *ne* is largely maintained in the context of a DP with no clitic (83.3%, 20/24).¹¹ Crucially, Culbertson (2010:95–96) also mentions that this pattern is statistically reliable in the corpus, but does not know of any historical evidence which would correlate the emergence of the morphological status of subject clitics and the loss of the preverbal negative particle. This investigation will show that the same pattern exists in child data, will suggest a formalization within the diglossic approach to variation (Barra-Jover, 2004, 2010; Massot, 2008, 2010; Massot and Rowlett, 2013; Zribi-Hertz, 2011), and rely on diachronic data (Foisil, 1989; Martineau and Mougeon, 2003) in order to investigate the historical and cognitive aspects of the correspondence.

2.4. Formalizing variation: three possible approaches

The term diglossia stems from Ferguson’s (1959) seminal work on Arabic, Modern Greek, Swiss German and Haitian Creole.¹² The author defined diglossia as a language situation ‘where two varieties of a language exist side by side throughout the community, with each having a definite role to play’ (Ferguson, 1959:325). Thus, a sociolinguistically ‘high variety’ (H) of the language is activated in formal situations, whether written or spoken (poetry, newspaper editorials, university lectures, etc.), whereas a ‘low variety’ (L) is used for more ordinary purposes (folk literature, conversations with family, etc.). The speakers hence regard H either as prestigious compared to L or as the only language they use since they sometimes deny the existence of L (Ferguson, 1959:330). Another major characteristic of diglossia is the difference in the way children acquire the two varieties. Indeed, ‘L is invariably learned by children in what may be regarded as the “normal” way of learning one’s mother tongue’ whereas ‘H may be heard by children from time to time, but the actual learning of H is chiefly accomplished by the means of formal education’ (Ferguson, 1959:331). Thus, it is expected that non-educated speakers and preschool children should not be actively diglossic. As far as language acquisition is concerned, diglossia then makes the very strong prediction that there are no native speakers of H, and investigations in diglossic communities are expected to uncover the following sequence: (i) initial presence of one grammar only, and (ii) subsequent emergence and development of a second grammar.

French is a language that exhibits broad differences in various phonological, morphological, syntactic and lexical configurations, together with strong sociolinguistic constraints on the use of these different variants (e.g. Lodge, 2007 on the codification of French). The following question then naturally arose, as termed by Blanche-Benveniste (2003:317): ‘Les Français seraient-ils atteints de *diglossie*?’¹³ Some linguists answer the question negatively by adopting a variationist approach which accounts for variation within one unique grammar (e.g. Blanche-Benveniste and Martin, 2010; Coveney, 2011; Gadet, 2007). Other linguists answer the question positively, and argue that French actually spreads over two distinct grammars in the speakers’ mind/brain (Barra-Jover, 2004, 2010; Massot, 2008, 2010; Massot and Rowlett,

¹¹ Culbertson (2010) relies on the Lyon Corpus which includes 106,000 utterances from four French children and their caregivers (Demuth and Tremblay, 2008).

¹² The linguistic situation in Cyprus (with Cypriot Greek and Standard Modern Greek) has also recently been described as possibly diglossic, giving rise to very interesting acquisition and interface issues, and a new cover term for bilectalism, bidialectism and bilingualism, i.e. ‘bi-x’ (Grohmann and Leivada, 2012; Rowe and Grohmann, 2013).

¹³ ‘Have French people caught *diglossia*?’.

Table 2
The corpus.

Specifications	C1	C2	C3	C4
Date	2003/2004	2006/2007	2007/2008	2008/2009
Hours	13	25	20	20
Sessions	5	13	10	10
Recording	Audio	Audio & video	Audio & video	Audio & video
Adults	1	3	1	1
Utterances	N/A	12,891	9,042	N/A
Children	17	20	19	18
Age range	2;3–3;1	2;5–4;0	3;6–4;11	4;5–5;11
Utterances	1,072	15,992	12,969	N/A

2013; Palasis, 2013; Zribi-Hertz, 2011).¹⁴ Crucially, the variationist approach and diglossia then make very different predictions about combinations of variants. Indeed, the former does not formalize any particular constraint with regard to combinations, whereas the latter makes very strict predictions on this particular matter, and Massot (2010:90) even argues that: 'Pour chaque énoncé produit par un locuteur éduqué, celui-ci s'engage pour l'énoncé entier dans l'une de ses deux compétences [...]'.¹⁵ Finally, Villeneuve and Auger (2013) capture the differences and relationships between French and Picard with a model that aims to reconcile diglossia and variation. The next section will test the strongest thesis, i.e. diglossia, against contemporary child data.

3. Variation in contemporary child French

3.1. The data

The investigation relies on newly processed spontaneous data collected with 19 children in a kindergarten class in south-eastern France (henceforth C3, as detailed in Table 2). Table 2 shows that C3 is part of a broader, longitudinal and cross-sectional corpus which has already been partly examined (Palasis, 2005, 2010a,b, 2013).¹⁶

The same methodology was used throughout the corpus. The data consist of child–child and child–adult interactions. The investigator recorded small groups of 3 to 5 children in a room adjoining the children's usual classroom. The children were encouraged to tell the investigator about their activities in and out of school, to play games, or to look at books.

3.2. Assessing the status of subject clitics in contemporary child French

3.2.1. Presence and position with regard to the finite verb

Table 3 outlines the different types of subjects children use in the corpus. Observation of C3 corroborates previous findings: (i) preverbal clitics are pervasive in finite clauses, either appearing alone or co-occurring with a pronominal or referential DP (as illustrated in (6a), (6b) and (6b'), respectively)¹⁷; (ii) non-target null subjects (as in (6c)) decrease drastically during the period¹⁸; (iii) canonical subject DPs (as in (6d)) remain very rare throughout the whole period. In terms of development between C2 and C3, the concomitant decrease of null subjects and increase of clitics on the one hand, and the stable and low rate of DPs on the other hand suggest that the early null subjects are null clitics, not null DPs.

- (6) Child data (C3):
 a. et après i sait. (Elena, 4;4)
 and after he knows
 'and afterwards he knows'

¹⁴ The grammars are named differently though, depending on the background of the authors, e.g. 'demotic' and 'late classic' (Massot, 2008), 'standard' and 'dialectal' (Zribi-Hertz, 2011), 'G₁' and 'G₂' (Palasis, 2013).

¹⁵ 'Every time an educated speaker produces an utterance, he engages himself in one of his two grammars for the entire utterance [...]'.
¹⁶ C2 is available on CHILDES (Child Language Data Exchange System) at <http://childes.talkbank.org/browser/index.php?url=Romance/French/Palasis/>. The figures for C3 are provisional (9 sessions out of 10), and C4 has not been processed yet.

¹⁷ There are no postverbal clitics in C1 and C3, and only 6 in C2 (details and data from C2 in Palasis, 2013).

¹⁸ The null subject phenomenon is well-documented in French and other non-null subject languages. This matter cannot be addressed in this article for reasons of space but see Prévost (2009:150–160) for an overview of the literature on French, and Palasis (2012) for an analysis of the phenomenon in C1 and C2.

Table 3
Subjects in early French.

Types of subjects		C1	C2	C3
a	Clitic	50.9% (459)	64.6% (5,831)	70.9% (6,719)
b	Clitic + DP	22.5% (203)	31.1% (2,804)	28.1% (2,658)
c	Null	26.6% (240)	3.9% (348)	0.6% (60)
d	DP	0% (0)	0.4% (39)	0.4% (35)
Total		100% (902)	100% (9,022)	100% (9,472)

Table 4
Types and relative rates of doubling in C3.

Constituents	All	Doubled
Pronominal DPs	54.6% (1,697)	99.9% (1,696)
Referential DPs	30.3% (942)	97.5% (918)
Indefinite DPs	1.4% (43)	76.7% (33)
Relative pronouns	13.7% (426)	2.6% (11)
Total	100.0% (3,108)	85.5% (2,658)

- b. moi j' ai un livre de pirates. (Massimo, 3;6)
me I have a book of pirates
'I have a book of pirates'
- b'. mon papa il a une moto. (Massimo, 3;6)
my father he has a motorbike
'my father has a motorbike'
- c. veux pas ce livre! (Enzo, 4;1)
want not this book
'(I) do not want this book'
- d. le Père Noël m' a apporté une Barbie. (Mathilde, 3;9)
the Father Christmas me has brought a Barbie
'Father Christmas has brought me a Barbie doll'

Regular repetition of subject clitics in conjoined VPs follows directly from the overwhelming presence of these clitics in finite clauses, as illustrated in (7a), while (7b) presents the only occurrence of non-repetition.

(7) Conjoined VPs (C3):

- a. elle joue elle danse elle met la musique et elle dort. (Lucille, 3;10)
she plays she dances she puts the music and she sleeps
'she plays, dances, turns the music on, and sleeps'
- b. après elle soufflait et tout le monde goûtait et disait miam miam. (Elena, 4;10)
after she blew and all the world tasted and said yam yam
'afterwards she blew and everybody tasted (the cake) and said yam yam'

3.2.2. Co-occurring constituents

In addition to the co-occurrences with pronominal and definite DPs illustrated in (6b–b'), C3 displays clitic associations with indefinite DPs and relative pronouns, as reported in (8). Table 4 gives the breakdown and relative doubling rates for each type of co-occurrence.

(8) Doubling (C3):

- a. Indefinite DPs:
- i. et même qu' un Pokemon il a des crocodiles. (Massimo, 3;9)
and even that a Pokemon he has some crocodiles
'and even a Pokemon has crocodiles'
- ii. personne i m l' a dit. (Matteo, 4;5)
nobody he me it has told
'nobody told it to me'

- b. Relative pronouns:
- i. ma mère qui il a dit qu' y aura des médailles. (Massimo, 3;9)
my mother who he has said that there will-have some medals
'my mother who said that there will be medals'
 - ii. y a un sac qu' il est marron et noir.¹⁹ (Lina, 4;1)
there has a bag that he is brown and black
'there is a bag that is brown and black'

Table 4 shows that pronominal and referential DPs represent the vast majority of the subjects in child speech (84.9%), and that their average relative doubling rates are close to 100%. All 19 children use these two structures. These doubled DPs together with the non-doubled DPs ($n = 35$ in Table 3) indicate that definiteness is a general characteristic of subjects (71.4% of the latter are definite entities). Indefinite DPs are hence very rare in subject position (1.4%). They are nevertheless also mainly doubled (76.7%). 12 children out of 19 produce these indefinite DPs, but contrary to definite DPs the individual doubling rates are extremely heterogeneous (7 children: 100%; 1: 66.7%; 2: 50%; 2: 0%). Nonetheless, they indicate that most of the children predominantly apply the doubling strategy when introducing an indefinite DP in subject position. Interestingly, one of the two children who never double these indefinite DPs is the only child not to double a pronominal DP (occurrence reported in (14a)).

Finally, the average doubling rate for relative pronouns in C3 is 2.6%, with 4 children out of 19 who occasionally double these pronouns (individual rates: 9.1%, 15%, 15.4%, 20%). These occurrences are particularly relevant to the debate on the status of subject clitics since relative pronouns are never doubled in adult French (Labelle, 1988). Moreover, the form of the relative pronoun *qui* 'who/that' is used to test the status of [i] in [ki], since its non-deletion before a vowel is interpreted as supporting the analysis of [i] as an agreement marker in Quebec Colloquial French (henceforth QCF; Auger, 1994). In C3, 17 children out of 19 produce the relative pronoun *qui* before a vowel ($n = 92$), and only 13% of these occurrences are elided ($n = 12$). The children hence predominantly do not elide *qui* before a vowel (87%), and 11 children even never elide their forms ($n = 52$), as exemplified in (9a). The 4 children who occasionally double relative pronouns belong to this subgroup of 11. Auger (1994) also argues that the presence of two *qui*'s in *wh*-questions in QCF supports the morphological analysis of subject clitics in this language. Thus, QCF displays sequential *qui*'s (*qui qui est venu?* 'who who came'), as well as complex forms (e.g. *qui est-ce qui, qui c' qui, qui c'est qui*), but no structures with a single occurrence of the interrogative pronoun *qui* 'who' (Auger, 1994:91). C3 presents similar characteristics to QCF on this matter since three different combinations of two *qui*'s are attested in the data, i.e. *c'est qui qui, qui c'est qui*, and *qui est-ce qui*, as illustrated in (9b–d), and these combinations represent 79.5% ($n = 35$) of all the subject *wh*-questions. Simple *qui* pronouns are hence much less frequent than combinations thereof (20.5%, $n = 9$) and are uttered by 5 children only.

- (9) Relative and interrogative *qui* forms (C3):
- a. j' ai vu un cochon qui est parti. (Enzo, 4;1)
I have seen a pig that is gone
'I saw a pig that left'
 - b. c' est qui qui a la vache? (Elena, 4;9)
it is who who has the cow
'who has the cow?'
 - c. qui c' est qui était là? (Dylan, 3;8)
who it is who was there
'who was there?'
 - d. qui est c' qui m' a ouvert ça? (Enzo, 4;1)
who is it who me has opened that
'who opened that for me?'

In sum, relative pronouns can co-occur with subject clitics in child data despite the total absence of such combinations in child input. Second, interrogative pronouns show a strong tendency not to emerge as simple forms. In addition, let us remember that pronominal, referential and indefinite DPs also largely surface with subject clitics. Most of the different subject-related constituents in child French are hence doubled with a clitic. Doubling does not apply across-the-board though, contrary to NIDs. In order to test the strength of the doubling hypothesis in child French, the next section will report Poletto's (1993, 2000) description of doubling in the NIDs, and verify if the phenomena observed in C3 comply with Poletto's implications.

¹⁹ A reservation on this type of utterance is that *qu'* could be a complementizer not a relative pronoun.

Table 5
Doubling in NIDs.

Types	Tonic pronouns	DPs	QPs	Wh-variables	Dialects
1	+	–	–	–	Central Veneto
2	+	+	–	–	Trentino & North Lombard
3	+	+	+	–	Milanese & East Lombard
4	+	+	+	+	Friulian, Piedmontese & Ligurian

3.2.3. Cross-linguistic comparison with NIDs

Poletto (1993, 2000) proposed one-way implications in order to capture the variation found in NIDs with regard to doubling. Her implications are reported in (10), and Table 5 exemplifies the four possible combinations.

- (10) Poletto's (2000:140) one-way implications on doubling in NIDs:
- 'If [referential] DPs are doubled in a given dialect, tonic pronouns are also doubled.
 - If QPs are doubled, both DPs and tonic pronouns are doubled.
 - If variables in *wh*-contexts such as relative, interrogative, and cleft structures are doubled, doubling is always obligatory with all other types of subjects.'

With the exception of the Friulian dialects spoken in north-eastern Italy, Poletto's implications show a gradual geographical pattern in this area, moving west from the east coast (Type 1: Central Veneto dialects) to the French-Italian border (Type 4: Piedmontese and Ligurian dialects).²⁰ Our recordings of child French took place just a few kilometers away on the French side of this border. It can be seen from Table 5 that child data seem quite similar to Type 3 with some utterances compatible with Type 4. Following Roberge (1990), Culbertson (2010:121) integrated standard and colloquial French to Poletto's typology of subject doubling by: (i) distinguishing standard and colloquial French, (ii) placing standard French as a system with no doubling at all, and (iii) describing colloquial French as a Type 2 system. The data examined in this contribution suggest that standard and colloquial French differ even more deeply insofar as the latter seems to match Type 3, not Type 2.

Besides, the implications put forward by Poletto for entire linguistic systems also seem to apply at an individual level in C2 and C3. Indeed, in both sets of data, all the children double pronominal and referential DPs. In C2, the 3 children who double *wh*-variables belong to the small subgroup of 5 who double QPs. In C3, 2 of the 4 children who double *wh*-variables also double QPs. The other two children unfortunately never produce QPs in subject position, whether doubled or not, so the implication cannot be confirmed across-the-board in C3. The occurrences are nevertheless predominantly consistent with Poletto's implications at two levels: (i) by and large, the linguistic system we are describing seems to represent a geographical continuum with the contiguous NIDs with regard to doubling, and (ii) individually, the implications seem to hold. The connection between the latter and the former, and the underlying explanation of such implications could then represent interesting leads for further research in cognitive sciences.

3.2.4. // elision

Last but not least, previous literature on the matter has established that subject clitics vary in their form according to the phonological context in colloquial French, but not in standard French (e.g. Ashby, 1984; Morin, 1979; detailed account in Culbertson, 2010:90–93). These variations include fusion of *je* 'I' with certain verbs and elision of *il* 'he/it' before a consonant. Pope (1934:324) reported that the phonological reduction of *il* for instance is attested in French since the end of the 12th century 'with the common people', and according to Chervel (2011:198), it is 19th century schooling and its normative pressure that restored the pronunciation of the final // in pre-consonantal position.

These alternations have hence been extensively studied in adult data. // elision has also been highlighted in different child corpora (e.g. Grégoire, 1947; Guillaume, 1927; review of data in Clark, 1985), and C3 shows the same strict alternation pattern, i.e. full // before a vowel and elided // before a consonant (99.3% of the *il* clitics, total $n = 2208$), as exemplified in (11).²¹

- (11) // elision (C3):
 et après i sort et après il habite ici. (Dylan, 3;8)
 and after he goes out and after he lives here
 'and afterwards he goes out and afterwards he lives here'

²⁰ A gradual geographical variation pattern is also reported for object-clitic doubling in the Balkan languages (Friedman, 2008; Tomić, 2008).

²¹ The remainder 0.7% is analyzed in section 3.4.

Table 6
Negation in C3.

Forms	Simple	Discontinuous
<i>pas</i> 'not'	89.7% (967)	85.0% (17)
<i>plus</i> 'no more'	7.9% (85)	10.0% (2)
<i>jamais</i> 'never'	1.9% (21)	5.0% (1)
<i>que</i> 'only'	0.5% (5)	0.0% (0)
Total	100.0% (1,078)	100.0% (20)

3.3. Discussion on the morpho-syntactic status of subject clitics

The close examination of the finite clauses produced by a group of 19 children between 3;6 and 4;10 (C3) sheds light on the following characteristics of subject clitics at this age. First, their presence, which was already predominant in C1 and C2, becomes pervasive in C3 (73.4%, 95.7% and 99.0%, respectively). This increase co-occurs with a decrease of the null subjects (26.6%, 3.9% and 0.6%, respectively). DPs which represent the expected subject in standard French remain very low throughout the period (0%, 0% and 0.4%, respectively).²² Thus, subject clitics seem almost obligatory in this linguistic system. It follows that these formatives appear in complementary syntactic contexts, i.e. main and embedded clauses as well as coordinated structures, and that most of the pronominal, referential and indefinite subject DPs occur with a clitic (99.9%, 97.5% and 76.7%, respectively). Even *wh*-variables which never co-occur with a subject clitic in adult speech are sometimes doubled in these data (2.6%), yielding an average doubling-rate of 85.5% in C3. Second, the position of subject clitics with regard to their host is remarkably constant, i.e. always preverbal, even in interrogatives. Since the data display no clitic-verb inversions, two competing hypotheses can be drawn with regard to verb movement to the left periphery once the verb has checked its inflectional features in the agreement/tense area. In order to yield the correct order, the finite verb has to either remain in the agreement/tense area, or move higher with the preverbal clitic. The former hypothesis suggests that we should find no particular difference with regard to the order of the constituents in declarative and interrogative sentences, whereas the latter entails that the clitic-verb sequences should be found relatively higher in interrogative structures than in declarative sentences. An examination of the position of these sequences with regard to the subject DPs seems to favor the second hypothesis. Indeed, subject DPs mainly occur to the left of the clitic-verb sequences in declarative structures (91.6%, total $n = 1895$, as in (12a)), whereas they show a tendency to appear to their right in interrogative structures (62.3%, total $n = 162$, as in (12b)). We can hence hypothesize that clitic and verb move together past the DP in the latter only.

(12) Lateralization of the DP in declarative and interrogative structures (C3):

- a. Sanbarbe il essaie de l' attraper. (Matteo, 4;4)
Sanbarbe he tries to him catch
'Sanbarbe tries to catch him'
- b. et qu'est-c' qu' i fait Sanbarbe là? (Matteo, 4;4)
and what he does Sanbarbe there
'and what is Sanbarbe doing?'

Finally, we have seen that subject clitics and *wh*-elements display regular phonological patterns according to the context. On the one hand, subject clitics in Metropolitan child French hence seem to display a substantial number of characteristics that lead to an analysis in terms of agreement markers. On the other hand, doubling does not apply across-the-board, contrary to QCF and NIDs. Let us extend our investigation to negation in C3.

3.4. Negation in contemporary child French

Children overwhelmingly use simple rather than discontinuous negation in C3 (98.2% vs. 1.8%, total $n = 1098$). Table 6 lists the different negative adverbs found in these utterances and their relative occurrence rates. Examples are reported in (13).

(13) Simple negation (C3):

- a. mon papa i voulait pas. (Victor, 4;3)
my dad he wanted not
'my dad did not want'

²² In contrast, Blanche-Benveniste (1994:87) reports DP rates of 38.3% and 83.3% in two written corpora of French.

- b. on va plus toucher à ça. (Matteo, 3;11)
 one will no more touch at that
 ‘we will not touch that anymore’
- c. moi je mange jamais à la cantine. (Lucien, 4;2)
 me I eat never at the canteen
 ‘I never eat at the canteen’
- d. moi j’ ai vu que des dauphins. (Nina, 4;1)
 me I have seen only some dolphins
 ‘I only saw dolphins’

Discontinuous negation (1.8%, $n = 20$) hence barely extends compared to C2 (1.2%, $n = 16$; Palasis, 2013), but the contexts in which preverbal *ne* emerges and the form of *il* in these utterances call for comments. First, as far as contexts are concerned, one peculiarity comes into sight. Discontinuous negation appears in 6 different contexts, i.e. *je*, *il*, imperatives, *lui*, *elle* and relative *qui* ($n = 7, 6, 4, 1, 1$ and 1, respectively), and the *lui* context reported in (14a) is very singular. Indeed, this utterance combines two rare phenomena in the corpus which are both uncontroversially described as belonging to standard French, i.e. discontinuous negation ($n = 20$) and a non-doubled third-person strong pronoun ($n = 1$). Second, the form of *il* in these negative sentences ($n = 6$) stands out. When *il* surfaces before *ne*, it never obeys the strict alternation pattern described in section 3.2.4. The pair (14b–c) illustrates this contrast (same child, same session, same verb), which is due neither to */n/* nor to non-strict adjacency to the finite verb (as illustrated in (14d–e)).²³

(14) Interesting facts (C3):

- a. mais lui n’ est pas d’ accord. (Kelsang, 4;7)
 but him neg is not of agreement
 ‘but he disagrees’
- b. mais lui il ne voulait toujours pas. (Lou, 4;9)
 but him he neg wanted always not
 ‘but he still did not want to’
- c. mais i voulait pas que j’ le nourrisse. (Lou, 4;9)
 but he wanted not that I him feed
 ‘but he did not want me to feed him’
- d. i nage. (William, 4;2)
 he swims
 ‘he is swimming’
- e. Quentin i nous a fait une balançoire. (Lina, 4;3)
 Quentin he us has made a swing
 ‘Quentin made us a swing’

3.5. Discussion on negation

3.5.1. Implications for the diglossic hypothesis

The negative utterances in C3 represent an interesting test with regard to the diglossic approach mentioned in section 2.4. The strongest hypothesis defended by Massot (2010) predicts grammatical consistency throughout every utterance. Discontinuous negation is uncontroversially analyzed as belonging to the ‘high variety’ (H) of French. Consequently, any peculiarity emerging with discontinuous negation in child utterances should also be considered as pertaining to H. The examination of C2 highlighted a strong correspondence between the presence of *ne* and the non-elision of *il* (Palasis, 2013). Following Massot (2010), the entire sentences were considered as stemming from a unique grammar G_2 , which displays discontinuous negation and argument clitics, contrary to G_1 which features simple negation and agreement markers. Two series of predictions were made with regard to negative utterances. First: (i) simple negation (G_1) should never surface with a non-doubled subject DP (G_2), and (ii) discontinuous negation (G_2) should never surface with a doubled subject (G_1). Second: (i) all the subject clitics emerging with simple negation (G_1) should be elided before a consonant (G_1), and (ii) all the subject clitics emerging with discontinuous negation (G_2) should be full clitics (G_2). Both series of predictions on grammatical consistency are

²³ Note that both *il* and *je* are elided in (14c).

borne out in new C3, as exemplified in (13a) and (14a–c). Massot's (2010) strong and controversial proposal is hence supported by child data so far (up to 4;10).

Grammatical consistency raises questions however. Indeed, code-switching is a well-documented phenomenon in child bilingualism (at least since Leopold's (1949) investigations on his daughter). The diglossic hypothesis assumes that French speakers manipulate two grammars. Yet, C2 and C3 do not seem to show any signs of code-switching. Indeed, diglossia and bilingualism are different in the early stages insofar as the diglossic child starts his/her acquisition process with one grammar only, i.e. G_1 (there are no native speakers of G_2 , as detailed in section 2.4), and the very low rates of discontinuous negation and canonical subject DPs in the data illustrate that G_2 emerges subsequently and gradually. C4 will furnish us with further data on this particular matter, and code-switching is expected to emerge at some point of the development, perhaps when the two grammars get more balanced.

3.5.2. Implications for the status of subject clitics

The continuous absence of *ne* with elided subject clitics (G_1) in C3 is also of interest to the debate surrounding the status of subject clitics. *Ne* is often mentioned as problematic when analyzing subject clitics as agreement markers, since the analysis entails that *ne* would also have to be considered as morphological, which is controversial (De Cat, 2005; see section 2.2). The presence of *ne* in G_2 only (where clitics are analyzed as arguments) and its absence in G_1 (where the debate applies) seem to solve the problem, since *ne* and non-argument subject clitics belong to two different grammars. From this point of view, negation no longer represents an obstacle to the morphological hypothesis. Object and adverbial clitics still do though, and hence require further examination.²⁴

The twofold proposal, i.e. (i) capture variation within a diglossic approach and (ii) grant two different statuses to subject clitics according to the activated grammar, is primarily drawn from 21st century south-eastern child French. In order to further our understanding of diglossia and test the relevance of the approach, complementary questions are addressed in sections 4 and 5 with regard to diatopy and diachrony, i.e. Where does diglossia apply? and When did diglossia emerge in France?

4. Diatopy

On the one hand, researchers sometimes use the label 'European French' to refer to varieties of French found in France, Belgium and Switzerland (Culbertson, 2010; Ferdinand, 1996; Fonseca-Greber and Waugh, 2003; Martineau and Mougeon, 2003). On the other hand, it is well-known that these varieties display important lexical, phonological and morpho-syntactic divergences. The next section will examine oral adult and child data from different parts of France and Belgium in order to attempt to outline the geographical limits of G_1 .

4.1. The south-eastern boundary

Since French is spoken neither in Italy nor in Spain, the southern boundaries of G_1 follow. Nevertheless, interesting comments arise as far as cross-linguistic comparisons with neighboring linguistic systems are concerned. As mentioned in section 2.2, Zanuttini (1997) established a threefold classification of negation in Romance languages according to the distribution of their negative markers. Previous and new data on the acquisition of sentential negation in early south-eastern French clearly establish that children under five years old overwhelmingly use simple negation (see section 3.4). Their grammar hence descriptively complies with Zanuttini's Type b, as exemplified in (15a–d) with Valdotain and G_1 simple and complex verbal forms.²⁵

- (15) Comparing Valdotain and G_1 :
- a. lo film l' èra pa dzen. (Valdotain)
the movie scl was neg beautiful
'the movie was not good'
 - b. Sanbarbe il était pas content. (G_1 , Matteo, 4;5)
Sanbarbe he was neg happy
'Sanbarbe was not happy'

²⁴ Object and adverbial clitics were tentatively examined in child data in Palasis (2009, 2013). Note however that these items are relatively scarce in naturalistic data under the age of 5;0, e.g. 8729 subject clitics vs. 1105 object and 437 adverbial clitics in C2, and 9377 subject clitics vs. 1488 object and 159 adverbial clitics in C3.

²⁵ Note that Valdotain and G_1 also present very interesting similarities with regard to the status of their subject clitics. A detailed comparison is not possible in this article for reasons of space but see Diémoz (2007) for an investigation of subject clitics in Franco-Provençal.

- c. dz' i pa mindzà. (Valdotain)
 scl have neg eaten
 'I have not eaten'
- d. j' ai pas vu. (G₁, Nina, 4;2)
 I have not seen
 'I have not seen'

Interestingly, G₁ also geographically complies with Zanuttini's Type b since 'These languages are geographically contiguous, covering the area of southeastern France [referring to Occitan dialects, not French], the Romance part of Switzerland, and the western part and central part of northern Italy' (Zanuttini, 1997:4). Morphological subject clitics and simple postverbal negation then seem to co-occur in these systems too. Let us carry on our investigation on subject clitics and negation by moving further north.

4.2. The northern boundary

4.2.1. Central and Parisian French

Major fieldwork on simple negation has been carried out over the years with adult speakers in central and northern France (Armstrong and Smith, 2002; Ashby, 1976, 1981, 2001; Coveney, 1996, 2002; Hansen and Malderez, 2004; Pohl, 1975). The authors usually define the phenomenon as 'losing' or 'deleting' the preverbal negative particle *ne*, and research reports null or very low rates of *ne*-retention with children, teenagers and young adults (Garel, 1997; Hansen and Malderez, 2004).

Hansen and Malderez (2004) also examined the distribution of simple and discontinuous negation with regard to the nature of the subject. Interestingly, the authors report a correspondence between the presence of a subject clitic and the absence of the negative particle *ne* with an average 3.3% *ne*-retention rate with a clitic pronoun (e.g. *lui il (n') aime pas* 'him he does not like'), in contrast with an average 56.4% *ne*-retention rate with a DP subject (*mon père (n') aime pas* 'my father does not like'). Both facts, i.e. the absence of *ne* and the co-occurrence with the type of subject, are consistent with the diglossic approach advocated in this paper. They hence allow us to suggest that diglossia also applies to native speakers in these geographical areas.

4.2.2. Northern French

The characteristics of negation seem to change when moving further north. According to Hansen and Malderez (2004:20), who also refer to previous studies by Coveney (1996:64) and Armstrong and Smith (2002:28), speakers who lived in the Parisian area at the time of their investigations but who grew up in northern France display differences with the speakers who had spent their childhood in the Parisian area. The authors thus report that the former presented a statistically significant higher *ne*-retention rate than the latter. Interestingly, *ne*-retention is characteristic of Picard, a Gallo-Romance language of northern France and southern Belgium (61% in spoken Picard, Auger and Villeneuve, 2008), and monolingual speakers of French in this area and bilingual speakers of Picard and French display *ne*-retention rates of 31% and 33%, respectively when speaking French (Villeneuve and Auger, 2013:120). The authors nevertheless report a correlation between the type of negation and the type of subject in their northern French data, since doubling is not pervasive in this area (monolinguals: 42%; bilinguals: 25%).

4.2.3. Belgian French

Descriptions of Belgian French often indicate that this language is more conservative than the one spoken in France. Belgian French thus displays clitic-verb inversions in *wh*-questions (40% inversions in Belgium vs. 2% in France), singleton DP subjects (21% DPs in Belgium vs. 3% in France), as well as frequent *ne*-retention (De Cat, 2002, 2007b). The former led De Cat (2005) to argue for the syntactic analysis of subject clitics in 'spoken French'. In the face of these data, two hypotheses arise: (i) either Belgian speakers are not diglossic and their unique grammar resembles French G₂, (ii) or Belgian speakers are diglossic but their G₁ is 'one stage more conservative' than the G₁ spoken in France (Rowlett, 2013:54). Further data from early stages of acquisition would be necessary in order to disentangle this matter. The corpora examined so far nevertheless show that a term such as 'European French' actually covers broad linguistic diversity, and that scholars should be cautious about this heterogeneity.

5. Diachrony

The last perspective explored in this article on the acquisition of subject clitics and the negative particle *ne* in French is historical. We have hypothesized that French children first acquire G₁ that displays morphological subject clitics and

simple negation. The remainder of the contribution will illustrate that, contrary to common belief (see section 2.1), these characteristics seem to have been robustly present and interrelated in adult and child data for the last two centuries (section 5.1) or longer (section 5.2).

5.1. Adult data (17th to 19th century)

The ‘lack of agreement’ among linguists as to when *ne*-deletion became widespread in spoken French (somewhere between the 17th and the 20th century) drove Martineau and Mougeon (2003) to examine additional diachronic data in order to contribute to the resolution of this debate. Their approach is of special interest to us for two reasons. First, they analyzed sources assumed to reproduce ‘popular spoken French’, i.e. plays, diaries and dialectological surveys, which would be the closest we can get to G_1 . Second, they also investigated the evolution of the morpho-syntactic status of subject clitics over the centuries and the possible role of the latter in the progressive loss of the negative particle.

Ne-deletion turns out to be ‘extremely infrequent’ in their 17th century data, ‘still quite low’ in the 18th century, and ‘stratified along social class lines’ but spreading during the 19th century, initially confined to lower social classes in the Parisian area. The authors point to a strong correspondence with the type of subject as early as the 18th century with ‘the starting point’ of deletion being the context of subject clitics in contrast with high retention rates with DPs. Martineau and Mougeon (2003:139) hence conclude that *ne*-deletion is relatively recent in France and that the transformation of subject pronouns into affixes might have reinforced, but not necessarily triggered, its rise after the 18th century. Along with Hirschbühler and Labelle (1994), the authors also suggest that the diachronic evolution of the position of the non-finite verb with regard to the adverb *pas*, i.e. the loss of leftward movement of the verb rendering *ne* and *pas* adjacent (e.g. *ne vouloir point* vs. *ne point vouloir* ‘not want’), could represent another factor of reinforcement. Let us now collate these findings with 17th century child data.

5.2. Child data (17th century)

The *Journal de Jean Héroard* (Foisil, 1989) is a major source of information for any scientist investigating 17th century infancy and childhood in France. Indeed, Héroard was appointed personal physician to Louis XIII upon the birth of this future king in 1601, and spent the subsequent 26 years keeping a very detailed diary of the dauphin/king’s activities and behavior. Considered of prime importance for the study of child care and development in general, Héroard’s diary is also regarded as very informative as far as language acquisition is concerned since it displays a number of almost phonetically transcribed utterances (Ernst, 1989; Wirth Marvick, 1993). This corpus is thus referred to as ‘trustworthy’, ‘objective’, and ‘priceless’ for the study of French (Hirschbühler and Labelle, 2004:188; Ingram and Le Normand, 1996:357; Morgenstern, 2009:25; respectively), and has been used in various linguistic studies (e.g. Duffer and Stark, 2007; Ernst, 1985, 1989; Hirschbühler and Labelle, 2004; Ingram and Le Normand, 1996).

Hirschbühler and Labelle (2004:196) draw the following conclusions on negation from these data: ‘[. . .] Louis XIII has two grammars in competition, one with *ne* and one without. If this is right, the omission of *ne* in spoken speech, which is typical of contemporary spoken French, has its origin as early as the beginning of the 17th century [. . .].’ On the one hand, Martineau and Mougeon (2003) and Hirschbühler and Labelle (2004) reach opposing conclusions on the emergence of *ne*-deletion. On the other hand, the closeness of our hypothesis on diglossia with Martineau and Mougeon’s correspondence and Hirschbühler and Labelle’s statement urges us to examine the distribution of *ne* in this 17th century corpus, and test the conclusions drawn from contemporary child French. For reasons of time, 15 months have been investigated so far in the former, which correspond to the ages of 2;0 to 3;3 (from September 1603 to December 1604; Foisil, 1989:455–561), with the following outcome. First, the child’s negative utterances ($n = 67$) display variation as early as 2;3 since Louis produces four different types of negative structures during the period, as labeled in Table 7 and exemplified in (16a–d). Simple and discontinuous negation (types c and d, respectively) are dominant, and the latter increases sharply after 3;0.

Table 7
Louis’ different types of negative structures (2;0–3;3).

Types of negation		2;0–3;0	3;0–3;3	Total
a	Preverbal marker only	6.7% (2)	0.0% (0)	3% (2)
b	Preverbal marker + negative determiner	3.3% (1)	2.7% (1)	3% (2)
c	Postverbal marker only	36.7% (11)	13.5% (5)	23.9% (16)
d	Pre- and postverbal markers	53.3% (16)	83.8% (31)	70.1% (47)
Total		100.0% (30)	100.0% (37)	100.0% (67)

- (16) Louis' first negative verbal utterances (Foisil, 1989:468–482):
- a. a ne tare dire. (target: je ne le sçauroit dire) (2;3)
I neg know say
'I cannot say (it)'
 - b. je ne pante a nu ma. (je ne pense a nul mal) (2;4)
I neg think of no harm
'I think of no harm'
 - c. fai pa beau saboté. (on n'est pas beau en sabots) (2;6)
makes not nice with-clogs
'(one) does not look nice with clogs'
 - d. je né pas encore equivé. (je n'ai pas encore écrit) (2;8)
I neg-have not yet written
'I have not written (it) yet'

These data comply with the hypothesis on diglossia since it can be suggested that the child alternately applies G_1 and G_2 in his discourse. Let us now investigate the expected correlates, i.e. co-occurring subjects and the *i//* alternation.

A major characteristic of the child's speech is the overwhelming presence of subject clitics in his utterances between the ages of 4;0 and 9;0 (e.g. 94.7% of his negative matrix clauses between 4;0 and 5;0, Dufter and Stark, 2007:123). This characteristic is also prevalent in earlier data since all the negative sentences investigated in this contribution, except two, display a subject clitic (total $n = 63$).²⁶ The two exceptions show referential DPs. As far as the correspondence with negation is concerned, it is interesting to note that these extremely rare subject DPs co-occur with discontinuous negation (and subordination), as reported in (17a–b). These utterances thus exhibit properties of G_2 throughout, i.e. non-doubled subject DPs and discontinuous negation, and hence corroborate one of the predictions made for contemporary child French (illustrated in (14a)).

- (17) Louis' subject DPs (Foisil, 1989:516, 531):
- a. que le Roy ne veu pa descende.²⁷ (2;11)
that the King neg wants not get-down
'that the King does not want to get down'
 - b. alé faire bon guet pou Papa que les ennemi ne nou viene pa tué. (3;0)
go do good watch for Daddy that the enemies neg us come not kill
'go and do a good watch for Daddy so that the enemies do not come and kill us'

The remaining utterances with discontinuous negation exhibit clitics. Another characteristic of these clitics is the absence of co-occurring referential DPs or strong pronouns, contrary to what is observed in contemporary French. According to the diglossic approach on grammatical consistency, these clitics should be syntactic and not exhibit the *i//* alternation found with morphological clitics in contemporary child French.

On the one hand, the data show that the *i//* alternation is active in the child's speech with a clear phonological reduction of *il* before consonants (96.3%, total $n = 27$, as illustrated in (18a–c)). On the other hand and contrary to the predictions made by diglossia, some of these elided clitics emerge with discontinuous negation ($n = 10$, as in (18d)). These apparent counterexamples call for two comments though. First, they are alike in terms of clitics (80% display expletive *il* 'one') and verbs (60% display *faut* 'must-3sg'), as in (18d). Second, the rest of the corpus tells us that expletive *il* is systematically reduced before a consonant, even in postverbal position, as in (18e).

- (18) Louis' *i//* alternations (Foisil, 1989:484, 489, 500, 533, 541):
- a. i dit sa messe. (2;8)
he says his mass
'he celebrates mass'
 - b. il a de gan den. (2;9)
he has some big teeth
'he has big teeth'

²⁶ The null-subject utterances such as (16c) are not included in this amount.

²⁷ This utterance answers the question: 'What does one say?'

- c. i fau pa touné le cu a Moucheu le Daufin. (2;9)
 one must not turn the back to Mister the Dauphin
 'one must not turn his back to the Dauphin'
- d. i ne fau pas don le tué. (3;0)
 one neg must not so him kill
 'so one must not kill him'
- e. que faut i que je batte? (3;1)
 what must one that I hit
 'what must I hit'

Consequently, the tentative examination of Louis' data seems to show that the dauphin alternately produced simple and discontinuous negation and that the *i//* alternation was active in his speech at least for referential subject clitics.²⁸ However, the behavior of expletive *il* and the limited number of negative utterances with referential *il* studied so far do not allow us to propose reliable correlations at this stage. The very appealing hypothesis suggested by Hirschbühler and Labelle (2004) that 'Louis XIII has two grammars in competition' hence still requires further investigation.

6. Concluding remarks

This contribution aimed to propose a corpus-based analysis of variation and acquisition of subject clitics and preverbal *ne* in European French within the recent diglossic approach to variation in French (Massot and Rowlett, 2013). Contemporary as well as diachronic data were reviewed, stemming from previous and new adult and child corpora, from south-eastern, central, Parisian, northern France, Belgium, and northern Italy. The investigation aimed to test the morpho-syntactic status of subject clitics (agreement markers vs. arguments), and to formalize a possible correspondence with the types of negation in French (simple vs. discontinuous). The thesis on grammatical consistency defended by the proponents of diglossia was put to the test of new child data (C3), and the investigation attempted to outline the geographical and temporal limits of diglossia.

The results point to an analysis of subject clitics as agreement markers in contemporary south-eastern child French, and the negative particle *ne* has been eliminated from the list of arguments against this morphological analysis, since negative utterances with agreement markers display postverbal negation only. A strong correspondence between two characteristics, i.e. morpho-syntactic status of subject clitics (agreement markers vs. arguments) and type of negation (simple vs. discontinuous), is hence established from the data, which supports the hypothesis on grammatical consistency on the one hand, and points to the existence of two different grammars of French on the other hand (labeled chronologically G_1 and G_2). Indeed, developmental data also show that the emergence of discontinuous negation co-occurs with the emergence of other types of subjects, i.e. argument clitics and non-doubled subject DPs. Diatopic data complement the investigation, and inform us that the correspondences seem to hold moving north, but that diglossia does not appear to apply (or at least applies differently) when reaching the Belgian frontier. Finally, adult and child diachronic data (17th–19th century) also display interesting alternations in terms of full/elided clitics and simple/discontinuous negation, and invite us to further our understanding of the acquisition and processing of expletive clitics. Concomitantly, the analysis of C4 should provide us with additional information on the development of diglossia before the children reach primary school. Elicitation tasks with other children in kindergarten are also foreseen in order to test the hypotheses on subject clitics, negation and French diglossia.

Acknowledgements

This research was supported by the French-German ANR-DFG grant awarded to the project 'Dialectal, acquisitional, and diachronic data and investigations on subject pronouns in Gallo-Romance' (DADDIPRO, 2012-2015, no. ANR 11 FRAL 007 01) led by Michèle Oliviéri (Université Nice Sophia Antipolis, CNRS, BCL, UMR 7320) and Georg Kaiser (University of Konstanz).

I am very grateful to three anonymous reviewers for their very constructive comments on earlier versions of this paper. I also wish to thank Professor Ernst for his time and kind advice with regard to references in section 5.2, and Elisabetta

²⁸ The latter observation invites us to refine our investigations on the acquisition of expletive clitics in Louis' corpus as well as in contemporary data, in line with recent research on the status of expletive clitics in Old French, Modern French and dialects spoken in France (e.g. Kaiser et al., 2013; Oliviéri, 2011; Zimmermann, 2012).

Carpitelli, Richard Faure, Kleantes Grohmann, Jean-Pierre Lai, Frédéric Lavigne, Theoni Neokleous, and Michèle Oliiviéri for their helpful comments and support.

References

- Armstrong, N., Smith, A., 2002. The influence of linguistic and social factors on the recent decline of French *ne*. *J. French Lang. Stud.* 12, 23–41.
- Ashby, W.J., 1976. The loss of the negative morpheme *ne* in Parisian French. *Lingua* 39, 119–137.
- Ashby, W.J., 1977. Clitic Inflection in French: A Historical Perspective. Rodopi, Amsterdam.
- Ashby, W.J., 1980. Prefixed conjugation in Parisian French. In: Izzo, H. (Ed.), *Italic and Romance Linguistic Studies in Honor of Ernst Pulgram*. John Benjamins, Amsterdam, pp. 195–207.
- Ashby, W.J., 1981. The loss of the negative particle *ne* in French: a syntactic change in progress. *Language* 57, 647–687.
- Ashby, W.J., 1982. The drift of French syntax. *Lingua* 57, 29–46.
- Ashby, W.J., 1984. The elision of // in French clitic pronouns and articles. In: Pulgram, E. (Ed.), *Romanitas: Studies in Romance Linguistics*. University of Michigan Press, Ann Arbor, pp. 1–16.
- Ashby, W.J., 2001. Un nouveau regard sur la chute du *ne* en français parlé tourangeau: s'agit-il d'un changement en cours? *J. French Lang. Stud.* 11, 1–22.
- Auger, J., 1994. *Pronominal clitics in Québec colloquial French: a morphological analysis* (Doctoral dissertation). University of Pennsylvania.
- Auger, J., Villeneuve, A.-J., 2008. *Ne* deletion in Picard and in regional French: evidence for distinct grammars. In: Meyerhoff, M., Nagy, N. (Eds.), *Social Lives in Language – Sociolinguistics and multilingual speech communities*. John Benjamins, Amsterdam, pp. 223–247.
- Barra-Jover, M., 2004. Interrogatives, négatives et évolution des traits formels du verbe en français parlé. *Langue française* 141, 110–125.
- Barra-Jover, M., 2010. « Le » français ou ce qui arrive lorsqu'un état de choses est observé comme une entité. *Langue française* 168, 3–18.
- Baude, O., 2006. *Corpus Oraux. Guide des bonnes pratiques 2006*. Presses Universitaires d'Orléans & CNRS Editions, Orléans & Paris.
- Blanche-Benveniste, C., 1983. L'importance du 'français parlé' pour la description du 'français tout court'. *Recherches sur le français parlé* 5, 23–45.
- Blanche-Benveniste, C., 1994. Quelques caractéristiques grammaticales des "sujets" employés dans le français parlé des conversations. In: Yaguello, M. (Ed.), *Subjecthood and Subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"* Editions Ophrys, London, pp. 77–107.
- Blanche-Benveniste, C., 2000. *Approches de la langue parlée en français*. Editions Ophrys, Paris.
- Blanche-Benveniste, C., 2003. La langue parlée. In: Yaguello, M. (Ed.), *Le grand livre de la langue française*. Editions du Seuil, Paris, pp. 317–344.
- Blanche-Benveniste, C., Martin, P., 2010. *Le français: Usages de la langue parlée*. Peeters, Leuven.
- Brandi, L., Cordin, P., 1989. Two Italian dialects and the null subject parameter. In: Jaeggli, O., Safir, K. (Eds.), *The Null Subject Parameter*. Kluwer Academic Publishers, Dordrecht, pp. 111–142.
- Cardinaletti, A., Starke, M., 1999. The typology of structural deficiency: a case study of the three classes of pronouns. In: Riemsdijk, H.v. (Ed.), *Clitics in the Languages of Europe*. Mouton de Gruyter, Berlin & New York, pp. 145–233.
- Chervel, A., 2011. Histoire de l'orthographe française: réformes, enseignement, niveau. In: Bertrand, O., Schaffner, I. (Eds.), *Variétés, variations & formes du français*. Les Editions de l'Ecole Polytechnique, Palaiseau, pp. 195–210.
- Chifflet, L., 1659. *Essay d'une parfaite grammaire de la langue française*. Jacques van Meurs, Anvers.
- Cinque, G., 1990. *Types of A'-Dependencies*. MIT Press, Cambridge, MA.
- Clark, E.V., 1985. The acquisition of Romance with special reference to French. In: Slobin, D. (Ed.), *The Crosslinguistic Study of Language Acquisition*. Lawrence Erlbaum Associates, Hillsdale, pp. 688–782.
- Clark, E.V., de Marneffe, M.-C., 2012. Constructing verb paradigms in French: adult construals and emerging grammatical contrasts. *Morphology* 22, 89–120.
- Coveney, A., 1996. *Variability in Spoken French: Interrogation and Negation*. Elm Bank, Bristol.
- Coveney, A., 2002. *Variability in Spoken French: A Sociolinguistic Study of Interrogation and Negation*. Elm Bank, Bristol.
- Coveney, A., 2011. A language divided against itself? Diglossia, code-switching and variation in French. In: Martineau, F., Nadasdi, T. (Eds.), *Le français en contact: Hommages à Raymond Mougeon*. Presses de l'Université de Laval, Québec, pp. 51–85.
- Culbertson, J., 2010. Convergent evidence for categorial change in French: from subject clitic to agreement marker. *Language* 86, 85–132.
- De Cat, C., 2002. *French dislocation* (Doctoral dissertation). University of York.
- De Cat, C., 2005. French subject clitics are not agreement markers. *Lingua* 115, 1195–1219.
- De Cat, C., 2007a. French dislocation without movement. *Nat. Lang. Linguist. Theory* 25, 485–534.
- De Cat, C., 2007b. *French dislocation. Interpretation, syntax, acquisition*. Oxford University Press, Oxford.
- Demuth, K., Tremblay, A., 2008. Prosodically-conditioned variability in children's production of French determiners. *J. Child Lang.* 35, 99–127.
- Diémoz, F., 2007. Morphologie et syntaxe des pronoms personnels sujets dans les parlers francoprovençaux de la Vallée d'Aoste. A. Francke Verlag, Tübingen & Basel.
- Dufter, A., Stark, E., 2007. La linguistique variationnelle et les changements linguistiques 'mal compris': le cas du *ne* de négation. In: Combettes, B., Marchello-Nizia, C. (Eds.), *Etudes sur le changement linguistique en français*. Presses Universitaires de Nancy, Nancy, pp. 115–128.
- Dye, C.D., 2011. Reduced auxiliaries in early child language: converging observational and experimental evidence from French. *J. Linguist.* 47, 301–339.
- Ernst, G., 1985. *Gesprochenes Französisch zu Beginn des 17 Jahrhunderts: Histoire particulière de Louis XIII (1605–1610)*. Niemeyer, Tübingen.
- Ernst, G., 1989. Le langage du prince. In: Foisil, M. (Ed.), *Journal de Jean Héroard, Médecin de Louis XIII*. Fayard, Paris, pp. 189–214.
- Ferdinand, R.A., 1996. *The Development of Functional Categories: The Acquisition of the Subject in French*. Holland Academic Graphics, La Hague.
- Ferguson, C.A., 1959. Diglossia. *Word* 15, 325–340.
- Foisil, M. (Ed.), 1989. *Journal de Jean Héroard, Médecin de Louis XIII*. Fayard, Paris.

- Fonseca-Greber, B., 2000. *The change from pronoun to clitic to prefix and the rise of null subjects in spoken Swiss French* (Doctoral dissertation). University of Arizona.
- Fonseca-Greber, B., Waugh, L.R., 2003. The subject clitics of conversational European French: morphologization, grammatical change, semantic change, and change in progress. In: Núñez-Cedeño, R., López, L., Cameron, R. (Eds.), *A Romance Perspective on Language Knowledge and Use*. John Benjamins, Amsterdam, pp. 99–117.
- Friedman, V.A., 2008. *Balkan object reduplication in areal and dialectological perspective*. In: Kallulli, D., Tasmowski, L. (Eds.), *Clitic Doubling in the Balkan Languages*. John Benjamins, Amsterdam, pp. 35–63.
- Gadet, F., 2007. *La variation de tous les français*. Linx 57, 155–164.
- Gadet, F., 2009. Stylistic and syntactic variation: introduction. In: Beeching, K., Armstrong, N., Gadet, F. (Eds.), *Sociolinguistic Variation in Contemporary French*. John Benjamins, Amsterdam, pp. 115–120.
- Garel, C., 1997. *Etude de la négation dans la langue française à travers le langage des jeunes de 17-19 ans*. BA. Université de Paris X Nanterre.
- Godard, D., 2004. French negative dependency. In: Corblin, F., De Swart, H. (Eds.), *Handbook of French Semantics*. CSLI Publications, Stanford, CA, pp. 351–389.
- Grégoire, A., 1947. *L'apprentissage du langage: la troisième année et les années suivantes*. E. Droz & Faculté de Philosophie et Lettres, Paris & Liège.
- Grevisse, M., Goosse, A., 2008. *Le bon usage*, 14th ed. De Boeck & Duculot, Bruxelles.
- Grohmann, K.K., 2003. *Prolific Domains: On the Anti-locality of Movement Dependencies*. John Benjamins, Amsterdam.
- Grohmann, K.K., Leivada, E., 2012. Interface ingredients of dialect design: Bi-X, socio-syntax of development, and the grammar of Cypriot Greek. In: Di Sciullo, A.M. (Ed.), *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces*. John Benjamins, Amsterdam, pp. 239–262.
- Guillaume, P., 1927. *Le développement des éléments formels dans le langage de l'enfant*. J. Psychol. 24, 203–229.
- Hansen, A.B., Malderez, I., 2004. *Le ne de négation en région parisienne: une étude en temps réel*. Langage et société 107, 5–30.
- Heap, D., Roberge, Y., 2001. Cliticisation et théorie syntaxique, 1971–2001. *Revue québécoise de linguistique* 30, 63–90.
- Hirschbühler, P., Labelle, M., 1994. Changes in verb position in French negative infinitival clauses. *Lang. Var. Change* 6, 149–178.
- Hirschbühler, P., Labelle, M., 2004. Enclisis and proclisis of object pronouns at the turn of the 17th century: the speech of the future Louis XIIIth. In: Gess, R.S., Arteaga, D. (Eds.), *Historical Romance Linguistics: Retrospective and Perspectives*. John Benjamins, Amsterdam, pp. 187–208.
- Hulk, A., 1991. *Les pronoms clitiques sujets et la théorie linguistique*. Actes du XVIIIe Congrès International de Linguistique et de Philologie Romanes. Niemeyer, Tübingen, pp. 504–514.
- Ingram, D., Le Normand, M.-T., 1996. A diary study on the acquisition of Middle French: a preliminary report on the early language acquisition of Louis XIII. In: Stringfellow, A., Cahama-Amitay, D., Hughes, E., Zukowski, A. (Eds.), *BUCLD 20: Proceedings of the 20th Annual Boston University Conference on Language Development*. Cascadilla Press, Somerville, MA, pp. 352–363.
- Kaiser, G.A., 1994. More about INFL-lection and agreement: the acquisition of clitic pronouns in French. In: Meisel, J.M. (Ed.), *Bilingual First Language Acquisition: French and German Grammatical Development*. John Benjamins, Amsterdam, pp. 131–159.
- Kaiser, G.A., Olivieri, M., Palasis, K., 2013. Impersonal constructions in Northern Occitan. In: Carrilho, E., Magro, C., Álvarez, X. (Eds.), *Current Approaches to Limits and Areas in Dialectology*. Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 345–366.
- Kallulli, D., Tasmowski, L. (Eds.), 2008. *Clitic Doubling in the Balkan Languages*. John Benjamins, Amsterdam.
- Kayne, R.S., 1975. *French Syntax. The Transformational Cycle*. MIT Press, Cambridge, MA.
- King, R., Nadasdi, T., 1997. Left dislocation, number marking, and (non-)standard French. *Probus* 9, 267–284.
- Labelle, M., 1988. *Prédication et mouvement: le développement de la construction relative chez les enfants francophones* (Doctoral dissertation). Université d'Ottawa.
- Lambrecht, K., 1981. *Topic, Antitopic and Verb Agreement in Non-standard French*. John Benjamins, Amsterdam.
- Leopold, W.F., 1949. *Speech Development of a Bilingual Child*. Northwestern University Press, Evanston, IL.
- Lodge, A., 2007. The history of French. In: Ayoun, D. (Ed.), *French Applied Linguistics*. John Benjamins, Amsterdam, pp. 21–48.
- Martineau, F., Mougeon, R., 2003. A sociolinguistic study of the origins of *ne* deletion in European and Quebec French. *Language* 79, 118–152.
- Massot, B., 2008. *Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie: arguments morpho-syntaxiques* (Doctoral dissertation). Université Paris 8.
- Massot, B., 2010. *Le patron diglossique de variation grammaticale en français*. *Langue française* 168, 87–106.
- Massot, B., Rowlett, P. (Eds.), 2013. *L'hypothèse d'une diglossie en France*. J. French Lang. Stud. 23.
- Miller, P.H., 1992. *Clitics and constituents in phrase structure grammar* (PhD Thesis, 1991). Utrecht University, Garland, New York.
- Miller, P.H., Monachesi, P., 2003. Les pronoms clitiques dans les langues romanes. In: Godard, D. (Ed.), *Langues romanes. Problèmes de la phrase simple*. CNRS Editions, Paris, pp. 67–123.
- Milner, J.-C., 1979. Le système de la négation en français et l'opacité du sujet. *Langue française* 44, 80–105.
- Morgenstern, A., 2009. *L'enfant dans la langue*. Presses Sorbonne Nouvelle, Paris.
- Morin, Y.-C., 1979. *La morphophonologie des pronoms clitiques en français populaire*. *Cahier de linguistique* 9, 1–36.
- Nadasdi, T., 1995. *Variation morphosyntaxique et langue minoritaire: Le cas du français ontarien* (Doctoral dissertation). University of Toronto.
- Olivieri, M., 2011. *Typology or reconstruction: the benefits of dialectology for diachronic analysis*. In: Berns, J., Jacobs, H., Scheer, T. (Eds.), *Romance Languages and Linguistic Theory 2009. Selected Papers from 'Going Romance' Nice 2009*. John Benjamins, Amsterdam, pp. 239–253.
- Ouhalla, J., 1991. *Functional Categories and Parametric Variation*. Routledge, London.
- Palasis, K., 2005. *Problèmes d'acquisition et le Paramètre du Sujet Nul* (Master's thesis). Université Nice Sophia Antipolis.
- Palasis, K., 2009. On the variable morpho-syntactic status of the French subject clitics. In: Kaiser, G.A., Remberger, E.-M. (Eds.), *Null-subjects, Expletives and Locatives in Romance*, Arbeitspapier 123. Konstanz Working Papers of the Department of Linguistics, Konstanz, pp. 47–62. <http://ling.uni-konstanz.de/pages/publ/PDF/ap123.pdf>
- Palasis, K., 2010a. *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant* (Doctoral dissertation). ANRT Diffusion, Villeneuve d'Ascq.

- Palasis, K., 2010b. Introducing new French child data: thoughts on their gathering and coding. *Corpus* 9, 33–51. <http://corpus.revues.org/index1801.html>
- Palasis, K., 2012. Economy principles within child speech: when the nominative clitic does not surface. In: Ferré, S., Prévost, Ph., Tuller, L., Zebib, R. (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages*. Cambridge Scholars Publishing, Newcastle, pp. 261–283.
- Palasis, K., 2013. The case for diglossia: describing the emergence of two grammars in the early acquisition of Metropolitan French. *J. French Lang. Stud.* 23, 17–35.
- Pierce, A., 1992. *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*. Kluwer Academic Publishers, Dordrecht.
- Pohl, J., 1975. L'omission de *ne* dans le français contemporain. *Le français dans le monde* 111, 17–23.
- Poletto, C., 1993. *La sintassi del soggetto nei dialetti italiani settentrionali*. Unipress, Padova.
- Poletto, C., 2000. *The Higher Functional Field: Evidence from Northern Italian Dialects*. Oxford University Press, Oxford.
- Pollock, J.-Y., 1989. Verb movement, universal grammar and the structure of IP. *Linguist. Inq.* 20, 365–424.
- Pope, M.K., 1934. *From Latin to Modern French*. Manchester University Press, Manchester.
- Prévost, P., 2009. *The Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. John Benjamins, Amsterdam.
- Reinhart, T., 1981. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27, 53–94.
- Rizzi, L., 1986. On the status of subject clitics in Romance. In: Jaeggli, O., Silva-Corvalan, C. (Eds.), *Studies in Romance Linguistics*. Foris, Dordrecht, pp. 391–419.
- Rizzi, L., 1997. The fine structure of the left periphery. In: Rizzi, L. (Ed.), *Comparative Syntax and Language Acquisition*. Routledge, London & New York, pp. 241–295.
- Roberge, Y., 1986. Subject doubling, free inversion, and null argument languages. *Can. J. Linguist.* 31, 55–79.
- Roberge, Y., 1990. *The Syntactic Recoverability of Null Arguments*. McGill-Queen's University Press, Kingston & Montreal.
- Rowe, C., Grohmann, K.K., 2013. Discrete bilingualism: towards co-overt prestige and diglossic shift in Cyprus. *Int. J. Sociol. Lang.* 224, 119–142.
- Rowlett, P., 1998. *Sentential Negation in French*. Oxford University Press, Oxford.
- Rowlett, P., 2013. Do French speakers really have two grammars? *J. French Lang. Stud.* 23, 37–57.
- Tomić, O.M., 2008. Towards grammaticalization of clitic doubling: clitic doubling in Macedonian and neighbouring languages. In: Kallulli, D., Tasmowski, L. (Eds.), *Clitic Doubling in the Balkan Languages*. John Benjamins, Amsterdam, pp. 65–87.
- Villeneuve, A.-J., Auger, J., 2013. 'chtileu qu'i m'freumereu m'bouque i n'est point coér au monne': grammatical variation and diglossia in Picardie. *J. French Lang. Stud.* 23, 109–133.
- Wirth Marvick, E., 1993. Louis XIII and his doctor: on the shifting fortunes of Jean Héroard's Journal. *French Hist. Stud.* 18, 279–300.
- Zanutini, R., 1997. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford University Press, Oxford.
- Zimmermann, M., 2012. *The evolution of expletive subject pronouns in French* (Doctoral dissertation). University of Konstanz.
- Zribi-Hertz, A., 1994. La syntaxe des clitiques nominatifs en français standard et en français avancé. *Travaux de Linguistique et de Philologie* 32, 131–148.
- Zribi-Hertz, A., 2011. Pour un modèle diglossique de description du français: quelques implications théoriques, didactiques et méthodologiques. *J. French Lang. Stud.* 21, 231–256.



Language Acquisition

7. Palasis, K., Faure, R. & Lavigne, F. (2019). Explaining variation in *wh*-position in child French: A statistical analysis of new seminaturalistic data. *Language Acquisition*, 26(2), 210-234, <https://doi.org/10.1080/10489223.2018.1513004>



ISSN: 1048-9223 (Print) 1532-7817 (Online) Journal homepage: <https://www.tandfonline.com/loi/hlac20>

Explaining variation in *wh*-position in child French: A statistical analysis of new seminaturalistic data

Katerina Palasis, Richard Faure & Frédéric Lavigne

To cite this article: Katerina Palasis, Richard Faure & Frédéric Lavigne (2019) Explaining variation in *wh*-position in child French: A statistical analysis of new seminaturalistic data, *Language Acquisition*, 26:2, 210-234, DOI: [10.1080/10489223.2018.1513004](https://doi.org/10.1080/10489223.2018.1513004)

To link to this article: <https://doi.org/10.1080/10489223.2018.1513004>



Published online: 05 Sep 2018.



Submit your article to this journal [↗](#)



Article views: 90



View Crossmark data [↗](#)



Explaining variation in *wh*-position in child French: A statistical analysis of new seminaturalistic data

Katerina Palasis, Richard Faure, and Frédéric Lavigne

Université Côte d'Azur, CNRS, BCL, Provence-Alpes-Côte d'Azur, France

ABSTRACT

The two possible positions for *wh*-words (i.e., in situ or preposed) represent a long-standing area of research in French. The present study reports on statistical analyses of a new seminaturalistic corpus of child L1 French. The distribution of the *wh*-words is examined in relation to a new verb tripartition: Free *be* forms, the Fixed *be* form *c'est* 'it is', and Other Verbs. Results indicate that a discriminating variable is verb form (i.e., Free vs. Fixed), regardless of verb type (i.e., *be* vs. Other Verbs), and that there is a correlation between the *wh*-in-situ position and the Fixed *be* form. The Fixed *be* form is thus identified as the component that leads to *wh*-in-situ utterances, in contrast to other languages such as English. Overuse of the Fixed *be* form in child speech could also account for the predominance of *wh*-in-situ in child object questions compared to adjunct questions and child *wh*-questions in general compared to adult questions.

ARTICLE HISTORY

Received 31 July 2017
Accepted 09 August 2018

1. Introduction

Variation in *wh*-position is well documented in both adult and child Metropolitan French (i.e., the variety of French spoken in France). To control for geographic variation, we leave aside other varieties of French spoken throughout the world (in Belgium, Switzerland, Quebec, etc.), whose *wh*-questions are also the matter of much investigation (Boeckx, Stateva & Stepanov 2001; Starke 2001; Mathieu 2009; Tailleux 2014, among others). Sociolinguistic studies on the variability of the *wh*-position in adult Metropolitan French show very different distributions of the two potential positions (45.9%, 10.5%, and 16.1% for the in-situ position in Ashby 1977; Coveney 1995, 2002 respectively; see example (1a) in Section 1.1). In the same vein, naturalistic, child-directed data also display variable in-situ rates (38.1%, 51%, and 16.6% in Strik & Pérez-Leroux 2011; Geveler & Müller 2015; Becker & Gotowski 2015 respectively). Proportions in child corpora also differ. However, it is reported that French preschool children systematically produce more *wh*-in-situ questions than adults (90.9%, 62.5%, and 66.7% for three children in Hamann 2006). In sum, there is variation in both the children's input and output, and the two patterns do not seem to overlap neatly. These facts lead to crucial questions on what constrains *wh*-position in adult and child French, and why their utterances are not alike.

Based on statistical analyses of a new seminaturalistic child corpus, the goal of the present article is to seek a variable that correlates with the position of the *wh*-words in child French and thus may explain the variation previously identified. Ideally, this variable could also account for the discrepancy between child and adult *wh*-in-situ questions and show coherence with languages with no *wh*-in-situ in matrix information-seeking questions, such as English (*Where does he live?* vs. **He lives where?*).¹

CONTACT Katerina Palasis ✉ palasis@unice.fr 📍 Université Côte d'Azur, BCL, Campus Saint Jean d'Angély 3, 24 avenue des Diables Bleus, Nice 06300 France.

¹We exclude echo questions, which are not information-seeking questions but rather requests for repetition of material that was not perceived or not understood. In the following example, B utters an echo question because either s/he is astonished by A's assertion or s/he did not hear what A said. His/her aim is not to know where X lives.

A: He lives on Mars.

B: He lives WHERE? (capitals denote emphasis)

Since different, nonmutually exclusive constraints might apply, Section 1 first reviews the existing literature on variation in adult and child French to explore possible connections with previous research. Section 2 details our method. It provides the verb tripartition we will statistically examine, the related Verb-form hypothesis we propose to account for the variation in the *wh*-position in child data and our corpus. Section 3 displays the statistical analyses and results using the Cochran-Mantel-Haenszel Test for repeated 2 x 2 tests of independence (Cochran 1954; Mantel & Haenszel 1959). Finally, Section 4 covers our conclusions on the Verb-form hypothesis in child French and highlights the compatibility of our hypothesis with adult French and English.

1.1. Variation in adult French

To form matrix information-seeking questions like *Where does he live?*, most languages allow either initial, preposed *wh*-words or final, in-situ *wh*-words. More rarely, languages like French allow both possibilities, as exemplified in (1).² Much investigation has focused on uncovering the reasons for this unusual variation in adult French using both specific and interfaced approaches. Generative grammar posits that the position to the right of the verb, as in (1a), is the underlying, “in-situ” position and that (1b) is obtained by moving the *wh*-word from this underlying position to a “preposed/fronted” position in the left periphery of the sentence. There is a wide consensus on the claim that *wh*-movement is constrained, but there still is a lively debate on the constraints that actually operate when a *wh*-word is in-situ or preposed.

- (1) a. In-situ *wh*-word: *Il vit où?*
 he live-PRS.3SG where
 ‘Where does he live?’
- b. Preposed *wh*-word: *Où il vit?*³
 where he live-PRS.3SG
 ‘Where does he live?’

Many approaches have focused on syntactic explanations based on: (i) overt vs. covert movement (Aoun, Hornstein & Sportiche 1981); (ii) strong vs. weak features, with a possible late insertion (Rizzi 1991/1996; Bošković 2000); (iii) phrasal vs. feature movement (Cheng & Rooryck 2000); (iv) inflectional-phrase (IP) movement (Munaro, Poletto & Pollock 2001; Munaro & Pollock 2005; Déprez, Syrett & Kawahara 2012, 2013); (v) a split determiner phrase (DP) in a (silent) left-peripheral operator and a *wh*-in-situ acting as a variable (Mathieu 2004); or (vi) default-positioning of the *wh*-phrase at the right edge of the clause to mark focalization (Hamlaoui 2010).

Since the 1980s, most linguists have tried to tie these syntactic accounts to interpretive differences, with the exception of Adli (2006), who argues for syntactic optionality between the two possible *wh*-positions in French. Semantic and pragmatic aspects of *wh*-questions, notably D(iscourse)-linking, have been the focus of particular attention. A *wh*-word is D(iscourse)-linked when the answer to the question belongs to a limited set of references shared by the speaker and hearer. As Pesetsky (1987:107–108) notes, the *wh*-phrase *which books* in the question *Which books did you read?* is D-linked because the speaker and hearer have the same, limited set of books in mind, whereas *who* or *what* are normally not D-linked since questions like *Who came?* or *What happened?* can refer to any person or event. The author concedes that D-linking does not interfere with the *wh*-position in

²Dryer (2013) reports that 97.5% of 902 languages studied worldwide display either “initial” (29.3%) or “not initial” (68.2%) interrogative phrases. French and Hausa (a West African language) are examples of “mixed” languages (2.5%).

³Accepted in spoken French (Gadet 1989; Grevisse 1993; Riegel, Pellat & Rioul 1994), whereas standard French requires clitic-verb inversion (*où vit-il ?*).

simple questions in English, since English has both types of *wh*-words but only one *wh*-position. With respect to French, the notion of D-linking has developed in two directions.

First, the discourse constraint is loosened from the limited set of felicitous answers suggested in Pesetsky (1987) to a more general notion of “presupposed context” (Coveney 1989; Obenauer 1994; Chang 1997; Boeckx 2000; Cheng & Rooryck 2000). Within this framework, in-situ *wh*-questions in French, but not preposed *wh*-questions, are construed as seeking information on a strongly presupposed context. For instance, *Tu as mangé quoi?* ‘What have you eaten?’ presupposes that the hearer has eaten something, whereas the preposed equivalent question *Qu’est-ce que tu as mangé?* does not make such a presupposition and can be answered *Rien* ‘Nothing’. This presupposition triggers movement of the IP to a discourse position in the left periphery (Munaro, Poletto & Pollock 2001; Munaro & Pollock 2005), or a need for repair by prosodic means (Cheng & Rooryck 2000), or both (Déprez, Syrett & Kawahara 2012, 2013). The grammatical category of the *wh*-word—that is, whether it is an adverb or a pronoun—is also claimed to interfere with its position in connection with its interpretation. There is even debate as to the possibility of interpreting adverbs when in-situ (Mathieu 1999; Reinhart 2006). The analyses that depend on presupposition have, however, been challenged. A second approach in spoken French claims that in-situ *wh*-questions are not necessarily presuppositional. For example, *Tu fais quoi dans la vie?* ‘What do you do for a living?’ can be answered *Rien, je suis au chômage* ‘Nothing, I am unemployed’ (Mathieu 2004; Baunaz 2011). This approach suggests that an in-situ *wh*-phrase is not necessarily presuppositional in the variety of French under scrutiny.

Finally, Coveney (1995) investigated the orthogonal, “socio-stylistic value” of different *wh*-structures by examining pragmatic, semantic, syntactic, and social factors, to include communication function, the type of verb, and the length of the *wh*-phrase. The work discovered a strong preference for both low informational verbs (e.g., forms of the copula *être* ‘be’) and *wh*-phrases of three or more syllables (e.g., *combien de* ‘how many’) in in-situ rather than preposed *wh*-structures.

In sum, different specific and interfaced approaches have been used to account for the variation in the *wh*-position in adult Metropolitan French, but no consensus has yet been reached. One main difficulty is the variability in the proportion of in-situ versus preposed *wh*-words in the different data sets used in previous studies. This variability suggests that there is no constant or straightforward pattern for *wh*-questions in the French children are exposed to in their first years of life and raises the question of the (possibly multiple) constraints children apply to their own productions.

1.2. Variation in child French

Variation with respect to *wh*-position is also reported in matrix *wh*-questions in both naturalistic and elicited L1 French data from monolingual preschool children (Crisma 1992; Déprez & Pierce 1993; Hulk 1995, 1996; Plunkett 1999, 2000a, 2000b, 2001; Hamann 2000, 2006; Zuckerman & Hulk 2001; De Cat 2002; Oiry & Demirdache 2006; Strik 2007, 2008, 2012; Jakubowicz 2011; Oiry 2011; Becker & Gotowski 2015; Gotowski & Becker 2016). The present article focuses on matrix *wh*-questions produced by typically developing French L1 children. We leave aside the abundant literature on embedded questions, long-distance questions, multilingualism, and language impairment. In addition, since it is widely acknowledged that children do not produce inversion or other features specific to standard (written) French at this stage of development (Hamann 2006; De Cat 2007; Palasis 2015), we focus only on the spoken variety of the language.

Wh-structures emerge around the age of 2;00 and mostly feature the copula *est* ‘is’ and the *wh*-word *où* ‘where’ (Prévost 2009). Studies report that preschool children usually produce both in-situ and preposed *wh*-words from the outset but that the proportions vary a good deal from one child to another (Crisma 1992; Déprez & Pierce 1993; Plunkett 1999; Hamann 2006; Strik 2007). Table 1 presents the *wh*-in-situ rates and raw numbers in four children between 1;08 and 2;09 in naturalistic data, thus giving an overview of this variability. Note that the data also suggest an absence of a straightforward, common developmental pattern since Marie’s rates decrease over time while the

Table 1. In-situ *wh*-words in naturalistic child speech (% of all *wh*-questions and raw numbers).⁴

	1;08	1;09	1;10	1;11	2;00	2;01	2;02	2;03	2;04	2;05	2;06	2;07	2;08	2;09
Marie	90.9 (10)						81.2 (26)		80.6 (25)					
Louis			62.5 (5)				84.2 (16)	88.6 (31)						
Augustin					66.7 (2)				94.4 (67)					85 (17)
Philippe					0			1.3 (1)			11.5/40.7 (23/81) ⁵			

other three increase. It has also been shown that children produce higher rates of preposed *wh*-words in elicited contexts than in naturalistic settings, although this difference may be due to methodological biases (Zuckerman & Hulk 2001).

Another characteristic of children's questions concerns the syntactic status of the *wh*-phrase. Children distinguish between arguments and adjuncts (Roeper & de Villiers 1992). Indeed, arguments are required by the verb and appear in a fixed position, whereas adjuncts usually freely relate to the sentence and can appear in various positions. Child studies generally report more in-situ positions in argument questions than adjunct questions, as illustrated in (2) (Strik 2007; Jakubowicz 2011).⁶

(2) a. Argument *wh*-word: *Tu vois qui?*

you see-PRS.2SG who

'Who do you see?'

b. Adjunct *wh*-word: *Comment on joue?*

how one play-PRS.3SG

'How do we play?'

Moreover, in a comparative study of children and adults, Zuckerman & Hulk (2001) show that children behave differently from adults only in argument questions. Indeed, children favor the in-situ position and adults favor the preposed position in argument questions, whereas both groups produce proportionally more preposed *wh*-phrases in adjunct questions. Strik (2007) hypothesized a possible relationship between the verbal-phrase (VP) internal position of objects and their in-situ position in child data. Interestingly then, there does not seem to be a correlation between the input children receive from their caregivers and the children's subsequent productions (Zuckerman & Hulk 2001; Strik & Pérez-Leroux 2011; Becker & Gotowski 2015). Overall, children are reported to produce more in-situ questions than adults, both in naturalistic and elicited corpora, before they enter kindergarten and during their kindergarten years (Becker & Gotowski 2015).⁷

In summary, variation in the position of the *wh*-words is well documented in preschool child Metropolitan French, but no particular correlation between the child and adult productions of *wh*-questions has been established. A number of explanations have, however, been put forward to account for child-adult differences in use. These are examined in the following section.

⁴Marie, Louis, and Augustin's data are part of the Geneva corpus (Rasetti 2003). Rates are from Hamann (2006). Philippe's data come from Suppes, Smith & Léveillé's (1973) corpus. Rates are from Crisma (1992).

⁵Excluding/including the questions of the type *C'est qui NP ?* 'It is who NP?' The author mentions that these forms could be considered as "fixed forms" and could therefore not be counted as instances of *wh*-in-situ questions (Crisma 1992:120).

⁶Authors also address this distinction by using the "adverb" vs. "pronoun" terminology (e.g., Mathieu 1999 mentioned in Section 1.1). We will use these terms when referring to our work. Indeed, the adverb/pronoun distinction is more relevant in our data than the adjunct/argument distinction, since our corpus mostly features sentences like *Où il va?* 'Where is he going?' and *C'est qui 'Who is it?'*, where *où* can be viewed as an argument of the verb despite its adverbial status, and the pronoun *qui* is not an argument per se since it is not selected for by the presentational *c'est* 'it is'.

⁷Twenty children in first kindergarten year, 2;09–3;10 (corpus from Palasis 2010). Rates for *qu'est-ce que/quoi* 'what': children in-situ: 88.2% vs. adult: 16.6% (Becker & Gotowski 2015).

1.3. Variation in child French compared to adult French

Variation in child-adult data is usually accounted for in terms of differential abilities. However, many of the purely syntactic accounts given for adults (Section 1.1) cannot be tested in child preschool data, since the utterances are short at this stage (no long-distance effects), with rare or no use of quantifiers (no intervention effects). Consequently, we find only one syntactic account but several interfaced approaches that include semantics and pragmatics in addition to syntax. In what follows we summarize these studies and discuss their strengths and weaknesses.

An argument for derivational economy has been frequently advanced to explain the prevalence of *wh*-in-situ questions in child speech compared to adults (Zuckerman & Hulk 2001; Hamann 2006; Strik 2007, 2008; Jakubowicz 2011). Indeed, in-situ structures result from covert movement of the *wh*-word, which is described as computationally more economical than overt movement, since the former affects features whereas the latter moves whole categories (Chomsky 1995). The authors thus ascribe the decrease in *wh*-in-situ questions from children to adults to the development of the children's syntactic abilities with regard to Merge and Move. However, this approach, taken alone, does not seem to account for: (i) the presence of some preposed *wh*-words from the outset (all children in Table 1); (ii) an absence of decrease in early *wh*-in-situ questions in some children (Louis, Augustin, and Philippe in Table 1); (iii) the fact that the final state adult language is not restricted to preposed *wh*-words exclusively; and (iv) the total absence of matrix *wh*-in-situ questions in the early productions of children acquiring languages with preposed *wh*-words in information-seeking questions (e.g., English, Dutch, Italian, German, Swedish), but in-situ *wh*-words in echo-questions (Stromswold 1995; Van Kampen 1997; Guasti 2000; Roeper & De Villiers 2011; Strik 2012).

Indeed, if derivational economy were the sole explanation for the prevalence of in-situ questions in child speech, one would expect at least some early errors in these languages, due to the existence of in-situ echo questions in the child's input—for example, 16% of the *what* questions in Brown's (1973) corpus (Eve's files examined in Gotowski & Becker 2016). It is also doubtful whether overt movement is really more complex than covert movement, since some kind of movement is necessary in both cases (Becker & Gotowski 2015).

Moving from core syntax to the interface between syntax and semantics, Palasis and Faure (2014) examined the verbs involved in *wh*-questions in kindergarten speech between 2;06 and 4;11 and observed that in-situ questions feature significantly more copulas than lexical verbs (Fisher exact test, $p < .0001$). This led the authors to propose a "Lexical hypothesis"—that is, the existence of a correlation between the movement of the *wh*-word to the left periphery and the semantic information contained in lexical verbs but not copulas. Palasis and Faure (2014) suggested that the preposed *wh*-phrases are endowed with a quantificational, Distributive-Selective feature. When the *wh*-phrase is preposed, lexical verbs feed this type of operator. A similar relationship based on the informational value of lexical versus copula-type verbs was put forward in adult speech (Coveney 1995). We find it interesting that the same constraint has been invoked in different studies addressing the same phenomenon in child and adult speech. The discrepancy in in-situ rates between child and adult speech could thus be related to a difference in the use of copulas. It has been established that copulas are more frequent in child than adult speech (Guillaume 1927; Bassano 1998).

Nevertheless, there are two reasons why the Lexical hypothesis requires closer examination. First, Coveney's (1995) and Palasis and Faure's (2014) analyses are not completely congruent. What Coveney (1995) labeled as "copulas" are forms of the verb *être* 'be' and other verbs such as *il y a* 'there is', whereas Palasis and Faure (2014) considered only forms of *être* 'be', excluding other verbs. Second, the lexical/copula partition also calls for further investigation of the nature of the sentences—that is, predicational versus nonpredicational (Higgins 1973/1979; Roy 2013). Lexical verbs are always predicational, since they always attribute a property denoted by the verbal phrase to the subject (e.g., *Paul likes bananas*). Copulas, on the other hand, can be either predicational (e.g., *Paul is nice*) or, when they establish an equation between two referents, nonpredicational (e.g., *Paul is my English teacher*). Thus, the predicational/nonpredicational division does not correspond exactly to

the lexical/copula distinction, and it might be worth exploring this more closely in connection with the position of the *wh*-words. This question is addressed in [Section 2.1](#).

In addition to these open questions in syntax and semantics, the pragmatic/discourse lead has also been investigated to account for the discrepancy between child and adult *wh*-in-situ structures (Becker & Gotowski 2015; Gotowski & Becker 2016). These authors followed: (i) Mathieu's (2004) proposal that the *wh*-in-situ position corresponds to background information, that is, knowledge shared by the speaker and hearer, also called "common ground"; and (ii) Schaeffer & Matthewson's (2005) hypothesis that children under 4;00 lack the pragmatic ability to distinguish between their own knowledge and that of their interlocutors. Becker & Gotowski (2015) and Gotowski & Becker (2016) examined *wh*-words expressing 'what' (i.e., *quoi* and *qu'est-ce que*) between 2;09 and 3;10 and suggested that children under 4;00 overuse the *wh*-in-situ position because they overattribute referents to the common ground. This pragmatic/discourse hypothesis predicts that the in-situ rate should decrease after 4;00. The open question here concerns the actual knowledge young children have of common ground, and recent changes in experimental settings to have children participate rather than observe tasks offer opportunities to investigate this knowledge (Clark 2015). New experimental studies show that, as participants in the task, infants under 2 years of age gesture appropriately with respect to the amount of information they share with the hearer, and 3-year-olds also behave relevantly in ignorance compared to common-ground conditions (Liszkowski, Carpenter & Tomasello 2008; Liebal et al. 2009; Liebal, Carpenter & Tomasello 2013). These findings suggest that young children may in fact have greater awareness of common ground than previously attested (Clark 2015).

This brief review of some hypotheses to account for variation in *wh*-position in child French and its discrepancy with adult speech shows that possible explanations are numerous, appealing, and probably not mutually exclusive. In the present article, we focus on research taking a syntactic/semantic approach to the lexical/copula partition in child and adult French (Coveney 1995; Palasis and Faure 2014) and test the influence of the verbal form on the position of *wh*-words using statistical analyses of a large corpus of child productions. This method complies with the first two aims mentioned in [Section 1](#): (i) to test for a variable that correlates with *wh*-word position in child French, and (ii) to reconsider the discrepancy between child and adult *wh*-in-situ questions. In [Section 2](#), we present the Verb-form hypothesis, our subjects and data, and we address the third aim: (iii) to relate the hypothesis to languages with no *wh*-in-situ in matrix information-seeking questions, like English.

2. Method

We aim to extend previous findings on the position of *wh*-words in child French at several levels by: (i) elaborating on the lexical/copula distinction proposed in adult and child French ([Section 2.1](#)); (ii) proposing a fine-grained, statistical examination of *wh*-positions in relation to the form of co-occurring verbs ([Section 2.2](#)); and (iii) supplying new, seminaturalistic data to the field ([Sections 2.3–2.6](#)).

2.1. A new verb tripartition: "Free *be* forms", "Fixed *be* form" and "Other Verbs"

The partition between lexical verbs and copulas emerged in both adult and child studies on the *wh*-in-situ position (Coveney 1995; Palasis and Faure 2014). It requires finer examination, though, since the label "copula" was applied differently in the two studies and, as noted in [Section 1.3](#), does not coincide precisely with the (non-)predicational division, another dichotomy at the interface of syntax and semantics (Higgins 1973/1979; Roy 2013). These considerations lead us to abandon the standard lexical-copula dichotomy in favor of a new verb tripartition based on the conjunction of two characteristics: (i) the type of verb (lexical vs. copula), and (ii) its co-occurring subject pronoun (the pronoun *ce* 'that' vs. other pronouns). This new approach also has the advantage of addressing cross-linguistic variation between English and French. [Table 2](#) displays declarative sentences in English and French,

Table 2. Declarative sentences in English and French.

Examples	Specifications		
	Verb	Sentence	Form
(3) Paul likes bananas	(4): lexical	predicational	free
(4) <i>Paul (il) aime les bananes</i>			
(5) Paul is nice	(6): copula	predicational	free
(6) <i>Paul (il)/(*c') est gentil</i>			
(7) Paul is my English teacher	(8): copula	nonpred/equative	free
(8) <i>Paul *il/*(c') est mon professeur d'anglais</i>			fixed
(9) This man is Paul	(10): copula	nonpred/identificational	free
(10) <i>Cet homme *il/*(c') est Paul</i>			fixed
(11) The cook is Paul	(12): copula	nonpred/specificational	free
(12) <i>Le cuisinier *il/*(c') est Paul</i>			fixed

distinguished according to three specifications that we will discuss in this section: “Verb” (lexical vs. copula), “Sentence” (predicational vs. nonpredicational), and “Form” (free vs. fixed).

Predication shows that the copula *be* is semantically manifold (Higgins 1973/1979; Roy 2013). A declarative sentence is predicational when a property denoted by the verbal phrase is attributed to the subject. The copula *be* and lexical verbs can both convey predication. The syntactic relationship between the verb and the subject is the same in English and French in these predicative structures—that is, the subject can be freely chosen among any singular or plural, masculine, neutral, or feminine noun or pronoun. We shall henceforth call these predicative *be* forms “Free *be* forms” and these predicative, lexical verbs “Other Verbs”. Table 2 labels these two specifications “Verb” (lexical vs. copula) and “Sentence” (predicational vs. nonpredicational) and provides English and French counterparts for each case (examples 3–4 and 5–6 respectively).

On the other hand, a declarative sentence is nonpredicational when it features an equation between two referents. The copula *be* is the only verb found in these nonpredicational structures. Higgins (1973/1979) further divided these nonpredicational, copular sentences into three subtypes: equative, identificational, and specificational. Table 2 provides English and French counterparts for each subtype (examples 7–8, 9–10, and 11–12 respectively). An interesting finding is that English and French do not behave the same way in these nonpredicational sentences. Indeed, only English can freely use any pronoun in these contexts, whereas spoken French displays a unique and constrained form—that is, the presentational form *c'est* ‘it is’, which obligatorily features the elided, neutral, third-singular pronoun *c’* ‘it’.⁸ We shall henceforth call the former “Free *be* forms”, the latter “Fixed *be* form”, and label the specification “Form”. Coveney (1995) included the fixed form *il y a* ‘there is/are’ in his “copulas”. Since English and French are not completely identical in this respect—because while the choice of the pronoun is constrained in both languages (*il/there*), the conjugation of the verb varies only in English (*is/are*)—we choose not to follow Coveney (1995) and categorize *il y a* ‘there is/are’ as “Other Verbs”.

⁸The question is abundantly discussed in the literature. Here we build on recent research from Amary-Coudreau (2010, 2014) and Roy (2013). Note that the third-singular pronoun also takes the form *c’* ‘it’ when the sentence has a generic interpretation, as in (i):

- (i) La glace, *(c') est bon.
the ice-cream it be-PRS.3SG good
‘Ice-cream is good’.

However, this *c'est* is free: It is analyzable as *ça + est* ‘it is’ (*ça* shows up when the verb is lexical, as in (ii) and (iii)), contrary to *c'est* in Table 2, which is an unanalyzable, fixed form.

- (ii) La glace, j' aime ça.
the ice-cream I like-PRS.1SG that
‘I like ice-cream’.
- (iii) La glace, ça me plaît.
the ice-cream that to.me appeal-PRS.3SG
‘I like ice-cream’.

When the declarative sentences displayed in Table 2 are transformed into *wh*-questions, English only allows the preposed *wh*-position, whereas French potentially accepts preposed and in-situ positions (Huddleston 1988 and Grevisse 1993 respectively). Table 3 displays preposed *wh*-structures in English (examples 13–17), and Table 4 provides their preposed equivalents in French (examples 18, 20, 22, 24, 26) along with their *wh*-in-situ counterparts (examples 19, 21, 23, 25, 27). The shaded area in each table shows the boundaries with respect to our three specifications—that is, Verb (lexical vs. copula), Sentence (predicational vs. nonpredicational), and Form (free vs. fixed).

Moreover, the gray shade in Table 4 shows that the Sentence and Form partitions are redundant, since they establish the same dichotomy in French (examples 18–21 vs. 22–27). We choose to retain the Form partition only, for two reasons. First, Table 5 illustrates that combining the features of the Verb and Form partitions provides a new verb tripartition that has the advantage of increased clarity. Second, the Form partition is the only distinction that corresponds to a difference between English and French: English has no Fixed *be* form. English thus displays two types of verbs—“(Free) *be* forms” (14–17) and “Other Verbs” (13), whereas French has three types—“Free *be* forms” (20–21), the “Fixed *be* form” (22–27), and “Other Verbs” (18–19).

2.2. Hypotheses and predictions on the *wh*-position in English and French

Using the analysis proposed in Table 5, Section 2.2 spells out two hypotheses and their predictions about English and French *wh*-questions.

Table 3. English *wh*-structures.

Examples	Specifications		
	Verb	Sentence	Form
(13) What does Paul like?	lexical	predicational	free
(14) How is Paul?	copula	predicational	free
(15) Who is Paul?	copula	nonpred/equative	free
(16) Who is this man?	copula	nonpred/identificational	free
(17) Who is the cook?	copula	nonpred/specificational	free

Table 4. French *wh*-structures.

Examples	Specifications		
	Verb	Sentence	Form
(18) <i>Qu'est-ce qu'il aime Paul?</i>	(19): lexical	predicational	free
(19) <i>Il aime quoi Paul ?</i>			
(20) <i>Comment il est Paul?</i>	(21): copula	predicational	free
(21) <i>Il est comment Paul ?</i>			
(22) <i>Qui *il/(c) est Paul?</i>	(23): copula	nonpred/equative	fixed
(23) <i>*Il/(c) est qui Paul ?</i>			
(24) <i>Qui *il/(c) est cet homme?</i>	(25): copula	nonpred/identificational	fixed
(25) <i>*Il/(c) est qui cet homme ?</i>			
(26) <i>Qui *Il/(c) est le cuisinier?</i>	(27): copula	nonpred/specificational	fixed
(27) <i>*Il/(c) est qui le cuisinier ?</i>			

Table 5. Verb tripartition.

Categories	Features		
	Free <i>be</i> forms	Fixed <i>be</i> form	Other Verbs
Verb (± copula)	+	+	–
Form (± fixed)	–	+	–

2.2.1. *The Lexical hypothesis (H1)*

The dichotomy labeled “Verb” focuses on the opposition between “Free *be* forms” plus the “Fixed *be* form” on the one hand and “Other Verbs” on the other hand (Table 5). The corresponding hypothesis with respect to *wh*-position is the Lexical hypothesis (Palasis and Faure 2014, henceforth H1, introduced in Section 1.3). H1 makes the prediction (P1) that questions with a *be* form, whether free or fixed, should display an in-situ *wh*-word, whereas questions with another verb should exhibit a preposed *wh*-word because of the semantic information carried by verbs other than *be*. P1 will be statistically examined for child French in Section 3.1 (Table 12, Result 1). H1, however, requires further examination, since it relies in part on a category that blends two cases: “Free *be* forms” plus the “Fixed *be* form”. If H1 is correct, it makes the additional prediction (P1’) that the position of the *wh*-word should be in-situ with all *be* forms, whether free or fixed. P1’ will be statistically examined by comparing “Free *be* forms” versus the “Fixed *be* form” in Section 3.2 (Table 13, Result 2) and “Free *be* forms” versus “Other Verbs” in Section 3.3 (Table 14, Result 3). At this stage, we cannot yet test “Other Verbs” against the “Fixed *be* form”, since the two modalities differ in two features, as shown in Table 5. We will be able to conduct the test in Section 3.4 (Table 15, Result 4). As mentioned in Section 1.3, H1 might be too broad to account for the variation found in French. Moreover, Table 3 illustrates that H1 cannot be validated in English, which displays the two verbal categories but allows no variation in matrix information-seeking questions. Section 2.2.2 therefore spells out a second hypothesis by focusing on the French “Fixed *be* form” *c’est* ‘it is’.

2.2.2. *The Verb-form hypothesis (H2)*

The dichotomy labeled “Form” focuses on the opposition between the “Fixed *be* form” on the one hand and “Free *be* forms” plus “Other Verbs” on the other hand (Table 5). The corresponding hypothesis on the *wh*-position is the new “Verb-form hypothesis” (H2). Building on H1, H2 makes the prediction (P2) that questions with the “Fixed *be* form” should display *wh*-in-situ words, whereas questions with “Free *be* forms” and “Other Verbs” should display a preposed *wh*-word. P2 will be statistically examined by comparing the “Fixed *be* form” versus “Other Verbs” in Section 3.4 (Table 15, Result 4) and the “Fixed *be* form” versus a new category “All Free forms” in Section 3.5 (Table 16, Result 5).

This complies with our first aim—to clearly identify a variable that would correlate with the position of the *wh*-words in child French. We find it interesting that H2 could also satisfy our other two aims by addressing the discrepancy between child and adult *wh*-in-situ questions and showing coherence with languages like English with no *wh*-in-situ matrix information-seeking questions, as will be discussed in Section 4. Section 2.3 gives details of the children and the utterances we examined.

2.3. *Subjects and data*

The subjects were 17 native speakers of L1 French in a kindergarten class in the south of France (eight male, nine female). The children were audio- and video-recorded between 2;06 and 4;11 during 23 sessions over a 20-month period (two school years from November to June). Transcripts and codes follow CHILDES guidelines (MacWhinney 2000).⁹ The data involve 40 hours of child-child and child-adult seminaturalistic interactions. The children were offered activities during the recording sessions to create situations for language production and interaction, but these were chosen to resemble ordinary everyday conversations as closely as possible (Eisenbeiss 2010). No attempt was made to elicit specific linguistic structures, and the children were free to interact spontaneously. For each session, the investigator and three to five children were seated around a small table with books and games. The investigator encouraged the children to narrate their activities in and out of school, “read” books, and play games. An animal matching game proved particularly

⁹The first year is available online at <http://childes.talkbank.org/access/French/Palasis.html>

effective for collecting child questions. Each child chose a board to fill with six different animal cards by taking turns picking a card from the pack. If a card did not fit a child's board, the player was encouraged to ask whether another child could use that animal, as shown in (28). Such semistructured tasks obviously require the investigator to interact with the children, unlike naturalistic studies where the researcher interferes as little as possible. To minimize potential elicitation artifacts, the investigator (INV) was careful not to prime any particular *wh*-words or structures when asking questions to the children and hence produced *yes/no* questions rather than *wh*-questions, as illustrated in (28). Nevertheless, the independent status of adult and child questions in the corpus is crucial, and the matter is therefore addressed statistically in Section 2.6.

(28) Transcript excerpt (MAT, 3;03,25; brackets indicate unpronounced segments):

- INV: *Tu as pioché la vache.*
 you have-PRS.2SG pick-PTCP the.F cow
 'You have picked the cow.'
- INV: *Est-c(e) qu' elle est à toi la vache?*
 Q she be-PRS.3SG to you the.F cow
 'Is the cow yours?'
- MAT: *Nan.*
 no
 'No.'
- INV: *Alors i(l) faut demander ...*
 so one must-PRS.3SG ask-INF
 'So you must ask ...'
- MAT: *C' est à qui?*
 it be-PRS.3SG to who
 'Whose is it?'
- INV: *Voilà.*
 there
 'Good.'
- MAT: *C' est à qui la vache?*
 it be-PRS.3SG to who the.F cow
 'Whose is the cow?'

The child corpus comprises a total of 27,464 utterances, including 2,527 questions (9.2%). Table 6 shows that these questions can be divided across six different types with very different frequencies, as

Table 6. The different types of questions in the child corpus.

	Nonverbal	Yes/No	<i>Est-ce que</i>	Finite matrix <i>wh</i> -	Finite embedded <i>wh</i> -	Nonfinite <i>wh</i> -	Total
<i>n</i>	721	757	45	913	84	7	2,527
%	28.5	30.0	1.8	36.1	3.3	0.3	100.0

illustrated in (29). This research investigated the most frequent type at this stage of the children's development: finite matrix *wh*-questions (36.1% of the total questions). No echo questions were found in the corpus.

(29) Examples of different question types:

a. Nonverbal:

Et moi? (LUS, 2;10,27)

and me

'And me?'

b. Yes/No:

Tu viens? (MAI, 2;08,00)

you come-PRS.2SG

'Are you coming?'

c. *Est-ce que* 'is it that':

Est-c(e) qu' il était cassé ton biberon? (LUS, 2;10,27)

Q he be-PST.3SG break-PTCP your bottle

'Was your bottle broken?'

d. Finite matrix *wh*-:

Où elle est ta maison? (MAT, 3;11,18)

where she be-PRS.3SG your house

'Where is your house?'

e. Finite embedded *wh*-:

Tu sais ce que c' est hein un TGV?

(LSN, 3;11,17)

you know-PRS.2SG what that it be-PST.3SG hm a.M TGV

'Do you know what a TGV (high-speed train) is?'

f. Nonfinite *wh*-:

J' arrive pas comment l' enlever. (LUS, 3;10,00)

I manage-PRS.1SG NEG how it-OBJ take.off

'I can't figure out how to take it off.'

2.4. Distributing the data

First, repetitions were excluded from the analysis (self-repetitions and repetitions of the previous speaker, $n = 9$). The remaining occurrences ($n = 904$) were classified according to the following criteria:

- *Wh*-word (12 modalities + potential combinations with the interrogative marker *est-ce que* ‘is it that’):
 - *que* ‘what’, *quoi* ‘what’, *à quoi* ‘to what’, *qui* ‘who’, *à qui* ‘to whom’, *quel(les)* ‘which’, *le(s)quel(les)* ‘which one(s)’, *où* ‘where’, *pourquoi* ‘why’, *comment* ‘how’, *quand* ‘when’, *combien* ‘how many/much’.
 - Any combination of the preceding *wh*-words with the interrogative marker *est-ce que*—for example, *qu’est-ce que* ‘what is it that’, *quand est-ce que* ‘when is it that’, etc.
- Type of verb (three modalities, as described in Table 5):
 - Free *be* forms: the copula *être* ‘be’ used freely, that is, with any pronoun except within the fixed *c’est* form—for example, *Comment il est Paul?* ‘How is Paul?’ (20 in Table 4; with the caveat in Fn. 8).
 - Fixed *be* form: the copula *être* ‘be’ used within the fixed, presentational structure *c’est* ‘it is’—for example, *C’est qui Paul?* ‘Who is Paul?’ (23 in Table 4).
 - Other Verbs: all verbs except the copula *être* ‘be’, used with any pronoun—for example, *Il aime quoi Paul?* ‘What does Paul like?’ (19 in Table 4).
- Position of the *wh*-word in the sentence (two modalities):
 - Beginning: at the beginning of the sentence—that is, to the left of the verb (what generative grammar calls “preposed/fronted *wh*–” under the assumption that the *wh*-word moves from an initial, lower position to a final, left-periphery position).
 - End: at the end of the sentence—that is, to the right of the verb (what generative grammar calls “*wh*-in-situ”).
- Child: the data from each child were tagged separately to control for quantitative and qualitative individual variation. Table 11 (in Section 3) shows the quantitative distribution of occurrences in the final corpus across the 17 children, displayed from the oldest to the youngest child. To maintain sufficient numbers of occurrences for statistical analysis no longitudinal analysis was conducted.

Table 7 displays the distribution of the total of 904 *wh*-questions according to the *wh*-word (in order of decreasing frequency) and the syntactic structure. First, the children used 11 of 12 possible *wh*-words (*combien* ‘how many/much’ never appeared in the data). The most frequent *wh*-questions in this corpus feature *que/quoi* ‘what’ (33%), *où* ‘where’ (22.1%), *qui* ‘who’ (20.7%), and *pourquoi* ‘why’ (10.6%). The less frequent *wh*-words are *comment* ‘how’ (5.8%), *à qui* ‘to whom’ (5.5%), *quand* ‘when’ (1.5%), *lequel/quel* ‘which (one)’ (0.3% and 0.2%, respectively), and *à quoi* ‘to what’ (0.2%). This frequency distribution is consistent with previous studies (as reported in Prévost 2009:346–347). Second, Table 7 exhibits 12 different syntactic structures, three of which are exemplified in (30). The structures vary depending on:

Table 7. Distribution of questions according to *wh*-word and syntactic structure.

Structures*	<i>quoi/que</i>	<i>où</i>	<i>qui</i>	<i>pourquoi</i>	<i>comment</i>	<i>à qui</i>	<i>quand</i>	<i>lequel</i>	<i>à quoi</i>	<i>quel</i>	Total
1a S V _o <i>wh</i> -	50/2	15	12	0	4	0	0	0	2	1	86
1b S V _{be} <i>wh</i> -	0	55	0	0	1	1	0	0	0	0	57
2a <i>wh</i> - (S) V _o	1	9	31	79	41/6	0	0	0	0	1	168
2b <i>wh</i> - (S) V _{be}	0	112/1	3	8	0	0	0	0	0	0	124
3a <i>wh</i> - V _o S	2	0	0	0	0	0	0	0	0	0	2
3b <i>wh</i> - V _{be} S	0	6	0	0	0	0	0	0	0	0	6
4 <i>c’est wh</i> -	170/2	1	30/2	0	0	45	0	3	0	0	253
5 <i>wh</i> - <i>c’est</i>	14	0	2	9	0	4	0	0	0	0	29
6 <i>c’est wh</i> - <i>que</i> S V	0	1	53	0	0	0	12	0	0	0	66
7 <i>wh</i> - <i>c’est que</i> S V	0	0	54	0	0	0	2	0	0	0	56
8 <i>wh</i> - ESK S V _o	55	0	0	0	0	0	0	0	0	0	55
9 <i>wh</i> - ESK V _o S	2	0	0	0	0	0	0	0	0	0	2
Total	298	200	187	96	52	50	14	3	2	2	904

*S = Subject, V_o = Other Verbs, V_{be} = Free *be* forms, *c’est* = Fixed *be* form, (x) = x is optional, ESK = *est-ce que* ‘is it that’, gray shade = only one grammatical position for the *wh*-word in the sentence (see Section 2.5), y/z = z were excluded.

(i) the verb form: “Free *be* forms” (V_{be} ; 1b, 2b, 3b), the “Fixed *be* form” (*c’est* ‘it is’; 4, 5, 6, 7), and “Other Verbs” (V_o ; 1a, 2a, 3a, 8, 9); (ii) syntactic criteria: Subject-Verb inversion (3a, 3b), insertion of the *c’est ... que* ‘it is ... that’ cleft (6, 7), and insertion of the *est-ce que* ‘is it that’ interrogative marker (henceforth ESK; 8, 9); and (iii) the position of the *wh*-word: Beginning vs. End (2, 3, 5, 7, 8, 9 vs. 1, 4, 6). Note that since Subject-Verb inversions, clefts, and ESK are mutually exclusive, we did not expect to find examples of all 48 theoretically possible structures. It is also important to understand that each *wh*-word was not realized in every structure.

(30) The three most frequent structures in the child corpus:

a. Structure 4 (*c’est wh*-):

C’ est quoi ça? (LIN, 3;04,02)
 it be-PRS.3SG what that
 ‘What’s that?’

b. Structure 2a (*wh*- S V_o):

Comment on joue à ce jeu? (LUS, 3;04,05)
 how one play-PRS.3SG to this.M game
 ‘How do we play this game?’

c. Structure 2b (*wh*- S V_{be}):

Où il est l’aspirateur? (DYL, 4;02,21)
 where he be-PRS.3SG the.M vacuum cleaner
 ‘Where is the vacuum cleaner?’

2.5. Filtering out the data

Before searching the corpus for a possible discriminating variable, it was necessary to exclude some occurrences from further analysis. There were no errors of *wh*-position (such as a preposed *quoi* ‘what’) in the utterances retained for analysis, but to comply with the constraints of French grammar, some *wh*-words can only appear in one of the two potential positions. Since these instances are therefore not subject to the alternation of interest in our study, we discarded those utterances where the *wh*-words could appear in only one of the two positions. The exclusions are as follows (shaded cells in Table 7).

- *Pourquoi* ‘why’ ($n = 96$; Structures 2a, 2b, and 5):

The *wh*-word *pourquoi* ‘why’ always appears to the left of the verb (e.g., MAI, 2;10,23: *Pourquoi i(l) veut pas le prêter?* ‘Why doesn’t he want to lend it?’). This *wh*-word is assumed to be directly generated in the left periphery, higher than the other *wh*-words (Rizzi 2001; Hamann 2006; Myers & Pellet 2014).

- *Qui* ‘who’ ($n = 34$; Structures 2a and 2b):

The *wh*-word *qui* ‘who’ is always preposed to the verb when used as a subject (e.g., MAT, 3;08,09: *Qui a le mien alors?* ‘So who has mine?’). These utterances were also discarded from our analysis. The occurrences of *qui* when *qui* is the object of the verb were retained ($n = 12$; Structure 1a). Note, however, that the latter are all

ungrammatical, since they always display the reflexive verb *s'appeler* 'to be called', with which children used *qui* instead of *comment* 'how' to ask *Tu t'appelles comment?* 'What's your name?'.

- The *wh*-words in Verb-Subject inversion contexts ($n = 10$; Structures 3a, 3b, and 9):

In-situ *wh*-words are ungrammatical in inversion contexts (e.g., LSN, 4;03,15: *Elmer où es-tu?* vs. **Elmer es-tu où?* 'Elmer where are you?'; VIK, 3;05,08: *Que se passe-t-il?* vs. **Se passe-t-il quoi?* 'What's happening?'). The "Wh-criterion," which assumes obligatory structural proximity between the *wh*-word (in the specifier position of the complementizer phrase CP) and the *wh*-feature (picked up by the verb in its I-to-C movement), accounts for this phenomenon (Rizzi 1991/1996).

- Clefts of the type *c'est ... que S V* 'it is that S V' ($n = 66 + 56$; Structures 6 and 7):

The *wh*-word is always to the left of V. We discarded both structures and hence leave discussion of movement of the *wh*-word around the cleft for future research (but see Labelle 1990, 1996; Oiry 2011).

- D-linked *wh*-words ($n = 18$):

The literature reports a possible interaction between D-linking and the in-situ position in French (see Section 1). The corpus displays 18 D-linked *wh*-words (four *quoi* 'what', one *où* 'where', two *qui* 'who', six *comment* 'how', two *quel* 'which', and three *lequel* 'which one'). Since D-linking is a possible factor that is likely to influence the position of the *wh*-word, but the number of D-linked utterances in our corpus was too low for statistical investigation, we discarded these 18 items to prevent them from biasing the analyses.¹⁰

Note that *wh*-ESK $S V_o$ items were not discarded ($n = 55$; Structure 8) because all the *wh*-ESK combinations in the corpus are occurrences of *qu(e)+est-ce que* 'what is it that', which have an in-situ equivalent in Structure 1a—that is, $S V_o$ *quoi*, as illustrated in (31a) vs. (31b). The items in Structure 8 were retained for analysis.

- (31) a. Structure 8 (*wh*-ESK $S V_o$):

Qu' est-ce que t(u) as Elena? (LOU, 4;04,24)
 what Q you have-PRS.2SG Elena
 'What's wrong Elena?'

- b. Structure 1a ($S V_o$ *wh*-):

J' ai mangé quoi? (LOU, 4;04,24)
 I have-PRS.1SG eat-PTCP what
 'What did I eat?'

After discarding *wh*-words that can appear in only one of the two *wh*-positions in French, we were left with 624 *wh*-questions. Table 8 displays these utterances, which constitute the final corpus on which statistical analyses were performed. Table 9 shows the distribution of these 624 occurrences with respect to the variables of interest: the two potential *wh*-positions and the three verb types. The predominance of in-situ position (61.9%) and *be* forms (69.7%) shown in Table 9 indicates that this corpus is similar to previous sets of child data (as described in Sections 1.2 and 1.3).

Before moving to the statistical examination of these data, Section 2.6 first compares the child distribution in Table 9 with the language used by the investigator during interactions to exclude imitation as a source of bias.

¹⁰We thank an anonymous reviewer for drawing our attention to this factor in our data.

Table 8. Final corpus for analysis.

		<i>quoi</i> 'what'	<i>où</i> 'where'	<i>qui</i> 'who'	<i>comment</i> 'how'	<i>à qui</i> 'to whom'	<i>à quoi</i> 'to what'	Total
1a	S V _o <i>wh-</i>	50	15	12	4	0	2	83
1b	S V _{be} <i>wh-</i>	0	55	0	1	1	0	57
2a	<i>wh-</i> S V _o	1	9	0	41	0	0	51
2b	<i>wh-</i> S V _{be}	0	112	0	0	0	0	112
4	<i>c'est wh-</i>	170	1	30	0	45	0	246
5	<i>wh- c'est</i>	14	0	2	0	4	0	20
8	<i>wh-</i> ESK S V _o	55	0	0	0	0	0	55
	Total	290	192	44	46	50	2	624

Table 9. Distribution of occurrences according to variables of interest.

> <i>Wh</i> -position Category	Beginning	End	Total <i>n</i> (%)
Free <i>be</i>	112	57	169 (27.1)
Fixed <i>be</i>	20	246	266 (42.6)
Other Verbs	106	83	189 (30.3)
Total <i>n</i> (%)	238 (38.1)	386 (61.9)	624 (100)

2.6. Addressing the independence of the children's and investigator's *wh*-questions

The investigator's *wh*-questions were extracted and filtered using the same criteria as those described in Section 2.5 for the child data. Table 10 displays the distribution of the adult *wh*-questions ($n = 1,377$) with respect to the same variables as in Table 9.

Two comments are in order with regard to previous adult corpora reported in Section 1. First, the investigator's rate for *wh*-in-situ questions (58.5%) is higher than in other data sets, the highest previously reported in-situ rate in child-directed speech being 51% (Geveler & Müller 2015). Second, the investigator's rate for *be* forms is lower than the children's rate (60.6% vs. 69.7%), which is consistent with studies reporting that copulas are less frequent in adult than child speech (Guillaume 1927; Bassano 1998). Nevertheless, it is crucial to examine the independence of the distribution of the investigator's and children's *wh*-questions in this corpus through statistical measures.

The independence of the distribution of *wh*-positions as a function of the verbal variable (displayed in Table 5) was compared in adult and child data. First, we examined the distribution of frequencies in the adult data according to the same 2 x 2 tables of frequencies as detailed in Section 3 for the child data. Second, we compared adult and child distributions for each table, leading to three-level 2 x 2 x 2 contingency tables, with frequencies for the *wh*-position, the verbal variable under test, and whether the data are child or adult utterances. This series of tests statistically addresses the question: Does the relationship between the *wh*-position and the verbal variable depend on the speaker (adult vs. child)?

We used the *lme* package in *R* (Pinheiro & Bates 2000) to first run a model of mutual independence, which tests for the effect of each of the three variables without interactions between the variables. Second, we ran a model of conditional independence, which tests for the effect of each variable with two-way interactions between the variables. Model comparison by the Akaike's

Table 10. Distribution of occurrences in investigator's data according to variables of interest.

> <i>Wh</i> -position Category	Beginning	End	Total <i>n</i> (%)
Free <i>be</i>	108	50	158 (11.5)
Fixed <i>be</i>	106	570	676 (49.1)
Other Verbs	357	186	543 (39.4)
Total <i>n</i> (%)	571 (41.5)	806 (58.5)	1,377 (100)

Information Criterion (henceforth AIC) allowed us to conclude that, for each verbal variable under study, including the two-way interactions in the model better accounts for the data (lower AICs) than not including the interactions (higher AICs). In other words, the distribution of *wh*-positions as a function of the verbal variable is statistically different in the adult and child data. This outcome enables us to conclude that the children's distributions of frequencies examined in Tables 12–19 in Section 3 are different from the investigator's distributions.

3. Results and discussion

The 624 child utterances remaining after data filtering (Table 9) were submitted to the Cochran–Mantel–Haenszel Test (henceforth CMH, Cochran 1954; Mantel & Haenszel 1959) for repeated 2 x 2 tests of independence. CMH allowed us to take three variables into consideration. First, we could test the independence of the variables of interest: the verb form in the sentence (Fixed *be* form vs. Free *be* forms vs. Other Verbs) and the position of the *wh*-word in the sentence (Beginning vs. End). Second, it enabled us to submit the data in as many 2 x 2 tables as there were participants with available data, hence taking into account a third variable corresponding to repeated measures in different children, each with a different amount and distribution of data (McDonald 2011). Table 11 displays the quantitative distribution of occurrences in the final corpus across the 17 children, ranging from the oldest to the youngest child. All the children contributed to the data in the tests, with occasional exceptions indicated in the relevant sections.

3.1. Testing H1: *Wh*-position according to verb type (Other Verbs vs. All *be* forms)

Table 12 investigated whether *wh*-words are more likely to be preposed with Other Verbs or more in-situ with the copula (Free *be* forms and the Fixed *be* form), as suggested in Palasis and Faure (2014) (H1, described in Section 2.2.1). Our first result thus counted Free *be* forms and the Fixed *be* form together to uncover a possible overall pattern for All *be* forms.

Analyzing the independence between *wh*-position and verb type—that is, Other Verbs vs. All *be* forms (Free and Fixed)—using CMH, the result was $\chi^2_{\text{CMH}} = 39.11$, 1 *df*, $p < .00001$ (Result 1). Overall, the *wh*-position is significantly more frequent at the beginning of the questions with Other Verbs and at the end with *be* forms. Cramér's V, which determines the degree of association between variables, indicates an intermediate effect size for Result 1 (H1, $V = 0.25$, Cramér 1946). In Section 3.5, we will determine the effect size for Result 5 (H2), and compare the two results.

At this stage, the statistical analysis leads us to reject the null hypothesis and validate H1. Consequently, despite the fact that French allows the speaker to place the *wh*-word either preposed or in-situ with both types of verbs, Result 1 indicates that the placement of the *wh*-word is related to the type of verb, linking the in-situ position with *be* forms, in line with Coveney's (1995) findings on adult data. However, Result 1 is not fully satisfactory, since the category All *be* forms combines two cases: Free *be* forms plus the Fixed *be* form. Sections 3.2 and 3.3 will address this issue to discover whether this finding can be further refined.

Table 11. Quantitative distribution of occurrences in final corpus across 17 children.

	KEL	LOU	ELE	VIC	CAR	ENZ	LIZ	MAT	LSN	LIN	LUS	NIN	WIL	LAN	DYL	MAI	MAS	<i>n</i>
Free <i>be</i>	6	2	3	5	0	4	1	29	5	6	8	5	22	0	31	0	42	169
Fixed <i>be</i>	1	6	6	24	5	14	12	52	13	5	23	14	5	17	16	20	33	266
Other Verbs	0	8	10	7	0	3	5	22	13	2	15	15	12	5	9	43	20	189
<i>n</i>	7	16	19	36	5	21	18	103	31	13	46	34	39	22	56	63	95	624
%	1.1	2.6	3.0	5.8	0.8	3.4	2.9	16.5	5.0	2.1	7.4	5.4	6.3	3.5	9.0	10.1	15.2	100

Table 12. *Wh*-position according to verb type ($n = 624$, all children).

	Beginning	End
Other Verbs	106	83
All <i>be</i> forms	132	303

Table 13. *Wh*-position according to copula form ($n = 435$, all children).

	Beginning	End
Free <i>be</i> forms	112	57
Fixed <i>be</i> form	20	246

3.2. Refining H1: *Wh*-position according to copula form (Free *be* forms vs. the Fixed *be* form)

To test the impact of each *be* form on the position of the *wh*-word separately, Table 13 examined each form independently: Free *be* forms (any person) versus the Fixed *be* form (*c'est* 'it is').

Analyzing the relationship between *wh*-position and the Free *be* forms vs. the Fixed *be* form using CMH, the result was $\chi^2_{CMH} = 138.61$, 1 *df*, $p < .00001$ (Result 2). Overall, the *wh*-position is significantly more likely to occur at the beginning with the Free *be* forms and more at the end with the Fixed *be* form. Consequently, despite the fact that French grammar permits both in-situ and preposed *wh*-positions with both types of *be* forms, Result 2 indicates that the in-situ position of the *wh*-word correlates with only one form of the copula, the Fixed form *c'est* 'it is'.

Result 2 thus refines H1, by examining the Free and Fixed *be* forms separately. Their significant impact on the *wh*-position indicates that the partition suggested by H1 is insufficiently fine-grained. Result 2, however, also limits the scope of our findings, since it is based on only a subset of the data, the *be* forms. In Section 3.3, we will connect Result 2 to the rest of the data on the basis of this new variable, the form of the verb (Free vs. Fixed).

3.3. Abandoning H1: *Wh*-position according to the type of Free verb forms (Free *be* forms vs. Other Verbs)

To connect Result 2 to the rest of the data, Table 14 examined Free *be* forms in relation to Other Verbs, which are also free verb forms. Since one child (CAR) had no data for this test, data were submitted in only sixteen 2 x 2 tables.

Analyzing the relationship between *wh*-position and Free *be* forms vs. Other Verbs using CMH, the result was $\chi^2_{CMH} = 2.48$, 1 *df*, $p = 0.12$ (Result 3). Overall, the *wh*-position is not significantly different when comparing Free *be* forms and Other Verbs, which are also free verb forms. Result 3 shows that the type of free verb (i.e., *be* vs. Other Verbs) is not the discriminating variable for the position of the *wh*-words in the corpus. We therefore abandon the Lexical hypothesis (H1) and test the Verb-form hypothesis (H2), which suggests that it is the form of the verb that correlates with the position of the *wh*-word in the sentence (Section 2.2.2). H2 makes the predictions that the position of the *wh*-word is related to: (i) the Fixed *be* form vs. Other Verbs, and (ii) the Fixed *be* form vs. Other Verbs together with Free *be* forms (Sections 3.4 and 3.5 respectively).

Table 14. *Wh*-position according to the type of Free verb forms ($n = 358$, 16 children).

	Beginning	End
Free <i>be</i> forms	112	57
Other Verbs	106	83

Table 15. *Wh*-position with Fixed *be* form vs. Other Verbs ($n = 454$, 16 children).

	Beginning	End
Fixed <i>be</i> form	20	245
Other Verbs	106	83

3.4. First step toward H2: *Wh*-position according to verb form (Fixed *be* form vs. Other Verbs)

To test the impact of the Fixed *be* form on the position of the *wh*-word, Table 15 compared the Fixed *be* form vs. Other Verbs, which are free verb forms. This test excludes Free *be* forms. As in the previous test, one child (KEL) had insufficient data for this test ($n = 1$), so only sixteen 2 x 2 tables were calculated.

Analyzing the relationship between *wh*-position and Fixed *be* form vs. Other Verbs using CMH, the result was $\chi^2_{CMH} = 116.87$, 1 *df*, $p < .00001$ (Result 4). *Wh*-position is significantly more frequent at the beginning with Other Verbs and more common at the end with the Fixed *be* form. Result 4 thus supports a twofold conclusion: (i) verb form is indeed a discriminating variable with respect to *wh*-position, and (ii) the in-situ position seems strongly related to the Fixed *be* form.

3.5. Confirming H2: *Wh*-position according to verb form (Fixed *be* form vs. All Free forms)

Table 16 further investigated the impact of the Fixed *be* form on the position of the *wh*-words by testing the Fixed *be* form against All Free forms, regardless of verb type. The new subset All Free forms was created by merging Other Verbs with Free *be* forms.

Analyzing the relationship between *wh*-position and the Fixed *be* form vs. All Free forms using CMH, the result was $\chi^2_{CMH} = 170.97$, 1 *df*, $p < .00001$ (Result 5). The *wh*-position is significantly more common at the beginning with All Free forms and at the end with the Fixed *be* form. Result 5 therefore confirms Result 4 and suggests that a discriminating variable for *wh*-word position is verb form (Fixed vs. Free), regardless of verb type (*be* vs. Other Verbs). Cramér's *V*, which determines the degree of association between variables, indicates a large effect size for Result 5 (H2, $V = 0.52$), which is twice the effect size for Result 1 (H1, $V = 0.25$, Section 3.1).

Taken together, Result 5 and Cramér's *V* provide strong support for the Verb-form hypothesis (H2, Section 2.2.2), which predicts that questions with the Fixed *be* form should display *wh*-in-situ words. Our investigation therefore highlights a correlation between the form of the verb and the position of the *wh*-word. Although this covariation does not constitute proof of a causal relation between the two variables, it is compatible with the hypothesis of a causal effect of the form of the verb on the position of the *wh*-word. In any case, these results allow us to accomplish our primary objective, to identify a variable that correlates with the position of *wh*-words in child French. We discuss the predictions made by H2 with respect to our other aims—that is, the clarification of the relationship between child and adult French, and the compatibility with English, in Section 4.

Table 16. *Wh*-position with Fixed *be* form vs. All Free forms ($n = 624$, all children).

	Beginning	End
Fixed <i>be</i> form	20	246
All Free forms	218	140

3.6. *Wh*-position according to grammatical category of the *wh*-word (adverb vs. pronoun)¹¹

Previous studies have shown that *wh*-in-situ questions are more frequent with object *wh*-words than adjuncts in child data (Zuckerman & Hulk 2001; Strik 2007; Jakubowicz 2011). A possible relationship between the VP-internal position of objects and their in-situ position in child *wh*-questions has been proposed (Strik 2007). We address this matter by statistically examining the position of the adverbial *wh*-words in the corpus (*où* ‘where’ and *comment* ‘how’) and the pronominal *wh*-words (*quoi* ‘what’, *qui* ‘who’, *à qui* ‘to whom’, *à quoi* ‘to what’) as a function of verb form (Free *be* forms, the Fixed *be* form, and Other Verbs).

Table 17 shows the distribution of all the *wh*-words retained for analysis according to grammatical category and syntactic structure (as in Table 8). Table 17 shows that children use more pronominal *wh*-words than adverbial *wh*-words, which is consistent with previous descriptions (Clark 1985; Prévost 2009; Plunkett 2000a).

3.6.1. Adverbial *wh*-words

The distribution of adverbial *wh*-words according to verb form in the corpus is quite clear-cut. Adverbs only occur in sentences with free verbs (Other Verbs, Structures 1a and 2a in Table 17; Free *be* forms, Structures 1b and 2b). Only one adverb appears with the Fixed *be* form (Structure 4: LSN, 3;05;08: *ben c'est où les trucs?* ‘Where are the things?’). The Verb-form hypothesis (H2, tested and validated in Sections 3.4–3.5), which opposes All Free forms to the Fixed *be* form, predicts that we should not obtain significant results when testing the position of adverbial *wh*-words with Free *be* forms against Other Verbs. Table 18 examined the distribution of the two adverbs retained for analysis (*où* ‘where’ and *comment* ‘how’) in data from the 14 children who produced sufficient examples for this test ($n = 224$).

Analyzing the relationship between the position of *wh*-adverbs and the two types of verbs (Free *be* forms vs. Other Verbs) using CMH, the result was $\chi^2_{CMH} = 0.124$, 1 *df*, $p = 0.72$ (Result 6), indicating that the *wh*-position is not significantly different with Free *be* forms and Other Verbs. Result 6 therefore confirms H2 for this subset of the corpus, similar to Result 3 for the entire corpus (Section 3.3, Table 14). As expected, (i) adverb position is not significantly different when comparing

Table 17. *Wh*-words according to grammatical category and syntactic structure.

		Adverbial		Pronominal		Total	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
1a	S V _o <i>wh</i> -	19	8.0	64	16.6	83	13.3
1b	S V _{be} <i>wh</i> -	56	23.5	1	0.3	57	9.1
2a	<i>wh</i> - S V _o	50	21.0	1	0.3	51	8.2
2b	<i>wh</i> - S V _{be}	112	47.1	0	0.0	112	17.9
4	<i>c'est wh</i> -	1	0.4	245	63.5	246	39.4
5	<i>wh- c'est</i>	0	0.0	20	5.2	20	3.2
8	<i>wh</i> - ESK S V _o	0	0.0	55	14.2	55	8.8
	Total	238	38.1	386	61.9	624	100.0

Table 18. Position of adverbial *wh*-words ($n = 224$, 14 children).

	Beginning	End
Free <i>be</i> forms	112	56
Other Verbs	39	17

¹¹See Footnote 6 on the distinction between adverbs vs. pronouns on the one hand and adjuncts vs. arguments on the other hand.

the two types of verbs, since both are free forms; and (ii) adverbial *wh*-words are mainly preposed in these sentences, in relation with the free form of the verbs.

Result 6 is consistent with previous descriptions of a higher incidence of preposing adverbial *wh*-words compared to object *wh*-words (Zuckerman & Hulk 2001; Strik 2007; Jakubowicz 2011). Result 6, however, points to a different explanation, since it suggests that it is the absence of the Fixed *be* form in adverbial *wh*-questions that underlies the observed asymmetry. It would be interesting to reexamine previous corpora to test our hypothesis.¹²

3.6.2. Pronominal *wh*-words

Unlike adverbial *wh*-words, pronominal *wh*-words occur with both free and fixed verb forms—that is, Other Verbs and Free *be* forms on the one hand (Structures 1a–1b, 2a–2b, and 8 in Table 17) and the Fixed *be* form on the other hand (Structures 4 and 5). The Verb-form hypothesis makes the prediction that we should obtain significant results when testing the position of pronominal *wh*-words in the corpus. Table 19 examined the distribution of the pronominal *wh*-words retained for analysis (*quoi* ‘what’, *qui* ‘who’, *à qui* ‘to whom’, and *à quoi* ‘to what’) in data from the 16 children who produced sufficient examples for this test ($n = 385$).

As in Section 3.5 (Table 16, Result 5), we analyzed the relationship between the position of the *wh*-words and verb form (the Fixed *be* form vs. All Free forms) using CMH. The result was $\chi^2_{CMH} = 68.01$, 1 *df*, $p < .00001$ (Result 7), indicating that the position of the pronominal *wh*-words is significantly more frequent at the end of the sentence with the Fixed *be* form than with All Free forms. Result 7 confirms H2 for this subset of the corpus, that the form of the verb (free vs. fixed) is related to the position of the *wh*-word.

Once again, these findings are consistent with previous descriptions of a preference for keeping object *wh*-words in-situ rather than adjuncts (Zuckerman & Hulk 2001; Strik 2007; Jakubowicz 2011). Result 7, however, just like Result 6, suggests that it is the massive presence of the Fixed *be* form in these pronominal *wh*-questions that leads to the observed asymmetry, not the grammatical category of the *wh*-words.

4. Conclusions

Based on statistical analyses of a new seminaturalistic child French corpus, the present article set out to clearly identify a variable that correlates with the position of *wh*-words in child language. Ideally, this variable would also explain the incidence of *wh*-in-situ questions in child corpora in comparison with adult French and show coherence with languages with no *wh*-in-situ in matrix information-seeking questions, such as English.

Using an analysis combining syntactic and semantic considerations (following Coveney 1995 and Palasis and Faure 2014), we examined two features, verb type (copula *be* vs. Other Verbs) and verb form (Free vs. Fixed), and suggested a new verb tripartition: Free *be* forms, the Fixed *be* form *c’est* ‘it is’, and Other Verbs (Table 5). Using this classification, the Verb-form hypothesis makes two predictions: (i) a discriminating variable for *wh*-position is verb form (Free vs. Fixed), regardless of verb type (*be* vs. Other Verbs); and (ii) the *wh*-in-situ position is related to the Fixed *be* form (Section 2.2.2).

Table 19. Position of pronominal *wh*-words ($n = 385$, 16 children).

	Beginning	End
Fixed <i>be</i> form	20	244
All Free forms	56	65

¹²We thank an anonymous reviewer for drawing our attention to the importance of this asymmetry.

We statistically examined the distribution of the *wh*-position (in-situ vs. preposed) as a function of verb form (Free *be* forms, the Fixed *be* form, and Other Verbs) in 624 matrix *wh*-questions by 17 French-speaking children between 2;06 and 4;11 (Table 9). Results suggest that the distribution of the *wh*-position is not random in this child corpus but rather that there is a correlation between the Fixed *be* form and the in-situ position of *wh*-words (e.g., *c'est quoi ça?* 'what is that'; Table 16, Result 5 in Section 3.5, $p < .00001$). Although this finding does not constitute proof of a causal relation between the two variables, it is consistent with a hypothesis of a causal effect of verb form on *wh*-position. Further work is needed to investigate this relationship.

Our results also shed new light on a previously discussed partition in child French: the differential behavior of adverbial and pronominal *wh*-words. Our findings are consistent with previous results: object *wh*-words tend to remain in-situ while adverbial *wh*-words are more likely to be preposed (Zuckerman & Hulk 2001; Strik 2007; Jakubowicz 2011). Results 6 and 7 (Section 3.6) suggest that it is the absence of the Fixed *be* form in adverbial *wh*-questions compared with its strong predominance in pronominal *wh*-questions that may underlie this observed asymmetry. It would be interesting to reexamine previous corpora for similar correlations.

On the basis of these results, we are now in a position to reassess previous approaches and suggest an explanation for the *wh*-in-situ/preposed alternation in child French. Approaches reviewed in Section 1.3 hypothesized that *wh*-in-situ questions are derivationally more economical than *wh*-preposed questions (Zuckerman & Hulk 2001; Hamann 2006; Strik 2007, 2008; Jakubowicz 2011). We questioned the idea of derivational cost as the sole explanation for the high frequency of in-situ questions in French child speech compared to adults, and this for two reasons. First, adults also produce *wh*-in-situ structures. Second, in *wh*-preposed-only languages (such as English), *wh*-in-situ is absent in the very earliest child productions. Nevertheless, we claim that cost can explain our data if we change perspective. A notable difference with the Fixed *be* form *c'est* 'it is' compared to All Free forms (Free *be* forms and Other Verbs) is that Free forms are always semantically heavier than the Fixed *be* form (see also Coveney's 1995 conclusions on adult speech, as reported in Section 1.1). Indeed, even Free *be* forms involve an operation of predication that does not apply in the case of the Fixed *be* form. We follow Amary-Coudreau (2011), for whom French predicational *be* (our Free *be* forms) takes a small clause whose structure depends on how the predicate organizes the relations between its arguments. In contrast, the Fixed *be* form *c'est* 'it is' only plays a role at the discourse level, by bringing together topical and focal elements. This would mean that French generates *wh*-phrases in-situ by default and so requires an additional factor to trigger preposing. This is in line with the generativist assumption that movement must be motivated and Palasis and Faure's 2014 suggestion reported in Section 1.3. This is also a step toward discarding the full optionality hypothesis (Adli 2006). In our child recordings, preposed *wh*-questions are especially frequent when there is a contrast between two or more possibilities, as with a set of cards, for example, as in (32a), or when there is an element of surprise (32b). We tentatively suggest that a feature related to the notion of contrast could be the trigger for *wh*-preposing in child French.

- (32) a. *Où il est le papa oiseau?* (MAS, 2;07,05)
 where he be-PRS,3SG the daddy bird
 'Where is the daddy bird (card)?'
- b. *Qu' est-ce qu' i(l) raconte euh?* (MAI, 4;00,24)
 what Q he say-PRS,3SG er
 'What is he talking about er?'

Our final point concerns the secondary objectives of the present study: The Fixed *be* form might also provide the beginnings of an answer to these questions. First, the variable could account for the

discrepancy between child and adult *wh*-in-situ questions (reported in Section 1). The Verb-form hypothesis makes the prediction that children should produce more *wh*-in-situ questions than adults if they produce more *wh*-questions with the Fixed *be* form than adults. The difference could be at least partly developmental, in line with Guillaume's (1927) and Bassano's (1998) findings that French children produce more copulas than adults and Clark's (1978) observations that children initially lack specific verbs and use more general-purpose verbs instead (in English, *do* and *make*). The phenomenon is well known in language acquisition as "overextension". This developmental hypothesis makes the prediction that all the children's in-situ rates should be higher than all adults' rates. Idiolectal and contextual factors could also be envisaged to explain the variation in children on the one hand and adults on the other. It would be interesting to study adult corpora in light of this new variable.

Regarding the compatibility of the Verb-form hypothesis with languages with no *wh*-in-situ in matrix information-seeking questions, our prediction is that there should be no *wh*-in-situ questions in languages with no Fixed *be* form. The prediction is borne out in English, since it is admitted that English displays no Fixed *be* form: *Le problème *(c') est la cravate* vs. *The problem *(it) is the tie* (Roy 2013). Additional work remains to be done to test the Verb-form hypothesis in other, non-in-situ languages. Finally, the Verb-form hypothesis addresses the matter of variation in the *wh*-position in French with an interfaced syntactic-semantic approach, which could probably be enriched with fine-grained syntactic, semantic, and pragmatic considerations.

Acknowledgments

We thank Carlos Aguilar for his assistance on the test of conditional independence. This work has also benefited from helpful discussions and feedback from colleagues at BCL as well as audiences at LSRL44 and The Romance Turn VI Conferences.

References

- Adli, Aria. 2006. French *wh*-in-situ questions and syntactic optionality: Evidence *wh* from three data types. *Zeitschrift für Sprachwissenschaft* 25. 163–203.
- Amary-Coudreau, Valérie. 2010. Qu'est-ce que "c'" est? *Le français moderne* 2. 259–277.
- Amary-Coudreau, Valérie. 2011. Copular sentences and Binding Theory: The case of French and Principle C. Ms. Université de Caen.
- Amary-Coudreau, Valérie. 2014. Identité stricte ou partielle et identification dans les phrases à copule. Comment les identifier? *SHS Web of Conferences* 8. 2295–2312.
- Aoun, Joseph, Norbert Hornstein & Dominique Sportiche. 1981. Some aspects of wide scope quantification. *Journal of Linguistic Research* 1(3). 69–95.
- Ashby, William J. 1977. Interrogative forms in Parisian French. *Semasia* 4. 35–52.
- Bassano, Dominique. 1998. Sémantique et syntaxe dans l'acquisition des classes de mots: l'exemple des noms et des verbes en français. *Langue française* 118(1). 26–48.
- Baunaz, Lena. 2011. *The grammar of French quantification*. Dordrecht: Springer.
- Becker, Misha & Megan Gotowski. 2015. Explaining children's *wh*-in situ questions: Against economy. In Elizabeth Grillo & Kyle Jepson (eds.), *Proceedings of the 39th annual Boston University Conference on Language Development [BUCLD 39]*, 88–100. Somerville, MA: Cascadilla Press.
- Boeckx, Cedric. 2000. Decomposing French questions. In Jim Alexander, Na-Rae Han & Michelle Minnick (eds.), *Proceedings of the 23rd Annual Penn Linguistics Colloquium*, 69–80. Philadelphia: University of Pennsylvania Working Papers in Linguistics.
- Boeckx, Cedric, Penka Stateva & Arthur Stepanov. 2001. Optionality, presupposition, and *wh*-in situ in French. In Joaquim Camps & Caroline R. Wiltshire (eds.), *Romance syntax, semantics and L2 acquisition*, 57–72. Amsterdam: John Benjamins.
- Bošković, Željko. 2000. Sometimes in [Spec, CP], sometimes in-situ. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalism in honor of Howard Lasnik*, 53–87. Cambridge, MA: MIT Press.
- Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Chang, Lisa. 1997. *Wh-in-situ phenomena in French*. Vancouver: University of British Columbia master's thesis.

- Cheng, Lisa Lai-Shen & Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 3(1). 1–19.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: The MIT Press.
- Clark, Eve V. 1978. Strategies for communicating. *Child Development* 49(4). 953–959.
- Clark, Eve V. 1985. The acquisition of Romance with special reference to French. In Dan Slobin (ed.), *The cross-linguistic study of language acquisition*, 688–782. Hillsdale, NJ: Lawrence Erlbaum.
- Clark, Eve V. 2015. Common ground. In Brian MacWhinney & William O’Grady (eds.), *The handbook of language emergence*, 328–353. Malden, MA: Wiley-Blackwell.
- Cochran, William G. 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* 10(4). 417–451.
- Coveney, Aidan. 1989. Pragmatic constraints on interrogatives in spoken French. *York Papers in Linguistics* 13. 89–99.
- Coveney, Aidan. 1995. The use of the QU-final interrogative structure in spoken French. *Journal of French Language Studies* 5(2). 143–171.
- Coveney, Aidan. 2002. *Variability in spoken French: A sociolinguistic study of interrogation and negation*. Bristol: Elm Bank.
- Cramér, Harald. 1946. *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.
- Crisma, Paola. 1992. On the acquisition of wh-questions in French. *Geneva Generative Papers* 0(1–2). 115–122.
- De Cat, Cécile. 2002. *French dislocation*. York, UK: University of York dissertation.
- De Cat, Cécile. 2007. *French dislocation. Interpretation, syntax, acquisition*. Oxford/New York: Oxford University Press.
- Déprez, Viviane, Kristen Syrett & Shigeto Kawahara. 2012. Interfacing information and prosody: French wh-in-situ questions. In Irene Franco, Sara Lusini & Andrés Saab (eds.), *Romance languages and linguistic theory 2010: Selected papers from Going Romance Leiden 2010*, 135–154. Amsterdam: John Benjamins.
- Déprez, Viviane & Amy Pierce. 1993. Negation and functional projections in early grammar. *Linguistic Inquiry* 24(1). 25–67.
- Déprez, Viviane, Kristen Syrett & Shigeto Kawahara. 2013. The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions. *Lingua* 124. 4–19.
- Dryer, Matthew S. 2013. Position of interrogative phrases in content questions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, <http://wals.info/chapter/93>. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Eisenbeiss, Sonja. 2010. Production methods in language acquisition research. In Elma Blom & Sharon Unsworth (eds.), *Experimental methods in language acquisition research*, 11–34. Amsterdam: John Benjamins.
- Gadet, Françoise. 1989. *Le français ordinaire*. Paris: Armand Colin.
- Geveler, Jasmin & Natascha Müller. 2015. Wh-fronting and Wh-in-situ in the acquisition of French: Really variants? In Pedro Guijarro Fuentes, Katrin Schmitz & Natascha Müller (eds.), *The acquisition of French in multilingual contexts*, 43–65. Bristol: Multilingual Matters.
- Gotowski, Megan & Misha Becker. 2016. An information-structural account of children’s wh-in situ questions in French. In Laurel Perkins, Rachel Dudley, Juliana Gerard & Kasia Hitczenko (eds.), *Proceedings of the 6th Conference on Generative Approaches to Language Acquisition North America [GALANA 2015]*, 24–35. Somerville, MA: Cascadilla Press.
- Grevisse, Maurice. 1993. *Le bon usage*. Louvain-la-Neuve, Belgium: Editions Duculot.
- Guasti, Maria Teresa. 2000. An excursion into interrogatives in Early English and Italian. In Marc-Ariel Friedemann & Luigi Rizzi (eds.), *The acquisition of syntax: Studies in comparative developmental linguistics*, 105–128. Harlow: Pearson Education.
- Guillaume, Paul. 1927. Le développement des éléments formels dans le langage de l’enfant. *Journal de Psychologie* 24. 203–229.
- Hamann, Cornelia. 2000. The acquisition of constituent questions and the requirements of interpretation. In Marc-Ariel Friedemann & Luigi Rizzi (eds.), *The acquisition of syntax: Studies in comparative developmental linguistics*, 170–201. London: Longman.
- Hamann, Cornelia. 2006. Speculations about early syntax: The production of wh-questions by normally developing French children and French children with SLI. *Catalan Journal of Linguistics* 5. 143–189.
- Hamlaoui, Fatima. 2010. Anti-givenness, prosodic structure and “intervention effects.” *The Linguistic Review* 27. 347–364.
- Higgins, Francis Roger. 1973/1979. *The pseudo-cleft construction in English*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Huddleston, Rodney. 1988. *English grammar: An outline*. Cambridge: Cambridge University Press.
- Hulk, Aafke. 1995. L’acquisition du sujet en français. *Recherches linguistiques de Vincennes* 24. 33–53.
- Hulk, Aafke. 1996. The syntax of wh-questions in child French. *Amsterdam Series in Child Language Development* 5. 129–172.
- Jakubowicz, Céilia. 2011. Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua* 121(3). 339–351.
- Kampen, Jacqueline van. 1997. *First steps in wh-movement*. Delft: Eburon.
- Labelle, Marie. 1990. Predication, WH-movement and the development of relative clauses. *Language Acquisition* 1(1). 95–119.

- Labelle, Marie. 1996. The acquisition of relative clauses: Movement or no movement? *Language Acquisition* 5(2). 65–82.
- Liebal, Kristin, Malinda Carpenter & Michael Tomasello. 2013. Young children's understanding of cultural common ground. *British Journal of Developmental Psychology* 31. 88–96.
- Liebal, Kristin, Tanya Behne, Malinda Carpenter & Michael Tomasello. 2009. Infants use shared experience to interpret pointing gestures. *Developmental Science* 12(2). 264–271.
- Liszkowski, Ulf, Malinda Carpenter & Michael Tomasello. 2008. Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition* 108. 732–739.
- MacWhinney, Brian. 2000. *The CHILDES Project. Tools for analyzing talk: Transcription format and programs*. Mahwah, NJ: Lawrence Erlbaum. <http://childes.talkbank.org/>.
- Mantel, Nathan & William Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22(4). 719–748.
- Mathieu, Eric. 1999. WH in situ and the intervention effect. *UCL Working Papers in Linguistics* 11. 441–472.
- Mathieu, Eric. 2004. The mapping of form and interpretation: The case of optional WH-movement in French. *Lingua* 114(9–10). 1090–1132.
- Mathieu, Eric. 2009. Les questions en Français: Micro et macro-variation. In France Martineau, Raymond Mougeon, Terry Nadasdi & Annie Tremblay (eds.), *Le français d'ici: Etudes linguistiques et sociolinguistiques de la variation*, 37–66. Toronto: GREF.
- McDonald, John H. 2011. *Myths of human genetics*. Baltimore: Sparky House.
- Munaro, Nicola, Cecilia Poletto & Jean-Yves Pollock. 2001. Eppure si muove! On comparing French and Bellunese wh-movement. In Pierre Pica & Johan Rooryck (eds.), *Linguistic variation yearbook*, 167–180. Amsterdam: John Benjamins.
- Munaro, Nicola & Jean-Yves Pollock. 2005. Qu'est-ce-que (*qu'est-ce que*)? A case study in comparative Romance interrogative syntax. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford handbook of comparative syntax*, 542–606. New York: Oxford University Press.
- Myers, Lindsay Lee & Stéphanie Pellet. 2014. *Pourquoi* in spoken French. In Stacey Katz Bourns & Lindsay Lee Myers (eds.), *Perspectives on linguistic structure and context*, 157–182. Amsterdam: John Benjamins.
- Obenauer, Hans-Georg. 1994. *Aspects de la syntaxe A-barre. Effets d'intervention et mouvements des quantifieurs*. Paris: University of Paris VIII dissertation.
- Oiry, Magda & Hamida Demirdache. 2006. Evidence from L1 acquisition for the syntax of *wh*-scope marking in French. In Vicenc Torrens & Linda Escobar (eds.), *The acquisition of syntax in Romance languages*, 289–315. Amsterdam: John Benjamins.
- Oiry, Magda. 2011. *L'acquisition des questions à longue-distance par des enfants français. Stratégie à dépendance directe versus indirecte et questions alternatives*. Editions Universitaires Européennes.
- Palasis, Katerina. 2010. *Syntaxe générative et acquisition: le sujet dans le développement du système linguistique du jeune enfant*. Villeneuve d'Ascq, France: ANRT Diffusion.
- Palasis, Katerina. 2015. Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua* 161. 125–143.
- Palasis, Katerina & Richard Faure. 2014. Grammatical and interpretive constraints on the development of *wh*-questions in L1 French. Poster presented at The Romance Turn VI Conference, September 18–20, Universitat de les Illes Balears, Palma de Mallorca.
- Pesetsky, David. 1987. *Wh-in-situ: Movement and unselective binding*. In Eric Reuland & Alice Ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.
- Pinheiro, José C. & Douglas M. Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Plunkett, Bernadette. 1999. Targeting complex structure in French questions. In Annabel Greenhill, Heather Littlefield & Cheryl Tano (eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development [BUCLD 23]*, 764–775. Somerville, MA: Cascadilla Press.
- Plunkett, Bernadette. 2000a. The emergence of periphrastic questions in child French. In Michael Perkins & Sara Howard (eds.), *New directions in language development and disorders*, 105–117. New York: Springer.
- Plunkett, Bernadette. 2000b. What's "what" in French questions. *Journal of Linguistics* 36(3). 511–530.
- Plunkett, Bernadette. 2001. Attract and covert merge: Predicting interrogative variation. In Galina M. Alexandrova & Olga Arnaudova (eds.), *The minimalist parameter*, 159–174. Amsterdam: John Benjamins.
- Prévost, Philippe. 2009. *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam: John Benjamins.
- Rasetti, Lucienne. 2003. *Optional categories in early French syntax: A developmental study of root infinitives and null arguments*. Geneva: Université de Genève dissertation.
- Reinhart, Tanya. 2006. *Interface strategies: Optimal and costly computations*. Cambridge, MA: The MIT Press.
- Riegel, Martin, Jean-Christophe Pellat & René Rioul. 1994. *Grammaire méthodique du français*. Paris: Presses Universitaires de France.
- Rizzi, Luigi. 1991/1996. Residual verb second and the *wh* criterion. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and functional heads*, 63–90. Oxford: Oxford University Press.

- Rizzi, Luigi. 2001. On the position “Int(errogative)” in the left periphery of the clause. In Guglielmo Cinque & Giampaolo Salvi (eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 287–296. Oxford: Elsevier.
- Roeper, Thomas & Jill G. de Villiers. 1992. ordered decisions in the acquisition of wh-questions. In Jürgen Weissenborn, Helen Goodluck & Thomas Roeper (eds.), *Theoretical issues in language acquisition*, 191–236. Hillsdale, NJ: Lawrence Erlbaum.
- Roeper, Tom & Jill G. de Villiers. 2011. The acquisition path for wh-questions. In Jill G. de Villiers & Tom Roeper (eds.), *Handbook of generative approaches to language acquisition*, 189–246. Dordrecht: Springer.
- Roy, Isabelle. 2013. *Nonverbal predication: Copular sentences at the syntax-semantics interface*. Oxford: Oxford University Press.
- Schaeffer, Jeannette & Lisa Matthewson. 2005. Grammar and pragmatics in the acquisition of article systems. *Natural Language & Linguistic Theory* 23. 53–101.
- Starke, Michal. 2001. *Move dissolves into merge: A theory of locality*. Geneva: University of Geneva dissertation.
- Strik, Nelleke. 2007. L’acquisition des phrases interrogatives chez les enfants francophones. *Psychologie française* 52(1). 27–39.
- Strik, Nelleke. 2008. *Syntaxe et acquisition des phrases interrogatives en français et en néerlandais: Une étude contrastive*. Paris: Université de Paris 8 dissertation.
- Strik, Nelleke. 2012. Wh-questions in child bilingual acquisition of French: Derivational complexity and cross-linguistic influence. *Canadian Journal of Linguistics* 57(1). 133–151.
- Strik, Nelleke & Ana Teresa Pérez-Leroux. 2011. Jij doe wat girafe?: Wh-movement and inversion in Dutch-French bilingual children. *Linguistic Approaches to Bilingualism* 1(2). 175–205.
- Stromswold, Karin. 1995. The acquisition of subject and object wh-questions. *Language Acquisition* 4(1–2). 5–48.
- Suppes, Patrick, Robert Smith & Madeleine Léveillé. 1973. The French syntax of a child’s noun phrases. *Archives de Psychologie* 42. 207–269.
- Tailleur, Sandrine. 2014. The French wh interrogative system: Evolution and clefting. In Marie-Hélène Côté & Eric Mathieu (eds.), *Variation within and across Romance languages: Selected papers from the 41st Linguistic Symposium on Romance Languages [LSRL 41], Ottawa, 5–7 May 2011*, 315–330. Amsterdam: John Benjamins.
- Van Kampen, Jacqueline (see Kampen)
- Zuckerman, Shalom & Aafke Hulk. 2001. Acquiring optionality in French wh-questions: An experimental study. *Revue québécoise de linguistique* 30(2). 71–97.



Exclusivity! *Wh*-fronting is not optional *wh*-movement in Colloquial French

Richard Faure¹ · Katerina Palasis¹

Received: 3 June 2019 / Accepted: 23 May 2020 / Published online: 23 June 2020
© The Author(s) 2020

Abstract This article revisits the long-standing issue of the alternation between *wh-in-situ* and *wh-ex-situ* questions in French in the light of diglossia and cross-linguistic data. A careful preliminary examination of the numerous *wh*-structures in Metropolitan French leads us to focus on Colloquial French, which undoubtedly displays both *wh-in-situ* and *wh-ex-situ* questions. Within this dataset, *wh-ex-situ* questions without the *est-ce que* ‘is it that’ marker are more permissive than *in-situ* regarding weak-islandhood and superiority. In a Relativized Minimality framework, we suggest that *wh-ex-situ* items bear an additional feature, which permits them to bypass these constraints. Colloquial French is thus a *wh-in-situ* language that allows for *wh-ex-situ* under specific conditions, like other *wh-in-situ* languages. Hence we argue against free variation and claim that *wh*-fronting is not driven by a *wh*-feature, but by another feature. Exploring the contexts where *wh-ex-situ* is licensed, we highlight a type of non-exhaustive contrast specific to questions, namely Exclusivity, and provide a formalization. The article therefore also contributes to the larger debate on information structure in questions.

Keywords *Wh*-questions · *In-situ* · *Ex-situ* · Contrast · Diglossia · French

1 Preliminaries: Optionality?

The present article discusses the common idea that *wh-ex-situ* is the normal/default way to form a *wh*-question in Colloquial Metropolitan French and argues that *wh-ex-situ* is actually more marked than *wh-in-situ*.¹ In tackling this question, we shall

¹Metropolitan French refers to the language used in European France (so-called ‘Metropole’), as opposed to other varieties of French in other parts of the world (e.g., Overseas France, Belgium, Switzerland,

✉ R. Faure
richard.faure@univ-cotedazur.fr

¹ Université Côte d’Azur, CNRS, BCL, Nice, France

contribute to the more general problem of information structure in questions, which is notably thorny and yet understudied (Beyssade 2006; Constant 2014; Engdahl 2006).

Metropolitan French is a ‘mixed language’ in the typological literature on *wh*-positions.^{2,3} This language hence features both *in-situ* and *ex-situ wh*-phrases (*wh*Ps), as illustrated in (1) and (2) with an argument and an adjunct, respectively

- (1) a. **Qui** il a vu?
who he has seen
b. Il a vu **qui**?
he has seen who
‘Who has he seen?’
- (2) a. **Où** il a travaillé?
where he has worked
b. Il a travaillé **où**?
he has worked where
‘Where has he worked?’

The received view is that speakers set a binary parameter in relation with the checking of the inherent *wh*-feature on the *wh*-item. This parameter can be described as overt vs. covert *wh*-movement (e.g., English in 3a vs. Chinese in 3b; Huang 1982) or unobligatory (English) vs. obligatory (Chinese) formation of a prosodic chunk between the *wh*-item and the verb (Richards 2010, 2016; among other analyses). In both approaches, Chinese Logical Form (in 3c) is identical to English surface form (in 3a) with regard to *wh*-placement.

- (3) a. [**Who**_i do] you like t_i?
b. Ni xihuan **shei**?
you like who
c. [shei_i [ni xihuan e_i]]

In this binary perspective, Colloquial French is deemed to follow the same pattern as English, and *wh-in-situ* questions are seen as marked, which entails that they have received most of the attention.⁴ However, works like Baunaz (2011, 2016) and

Quebec, etc.). We elaborate on Colloquial Metropolitan French in Sect. 2 and do not take a stand on the other varieties of French. However, we wish to thank an anonymous reviewer for pointing out that the results achieved here carry over to other varieties of French.

²‘Mixed languages’ are rare (i.e., 2.5% of 902 languages per Dryer 2013), and include Bahasa Indonesian, Egyptian Arabic, Igbo, Malagasy, among others. Cheng (1991) is one of the first important contributions to this typology.

³We shall leave aside echo questions, which strongly favor *wh-in-situ* across languages (Wachowicz 1974a,b). We shall not be examining embedded questions either because they mostly appear with *ex-situ wh*Ps for reasons that might be different from what we have in direct questions (but see Sect. 6 for a few remarks). Finally, we shall omit questions with subject *wh*Ps since they always appear preverbally in French and it cannot be shown whether they are *in-situ* or fronted.

⁴See Baunaz (2011, 2016); Boeckx (2000); Boeckx et al. (2001); Bošković (2000); Chang (1997); Cheng and Rooryck (2000); Coveney (1989, 1995); Denham (2000); Déprez (2018); Déprez et al. (2012, 2013); Larrivé (2016); Lasnik and Saito (1992); Mathieu (2004, 2009); Munaro et al. (2001); Munaro and Pol-

Vergnaud and Zubizarreta (2005) have shown that *wh-in-situ* questions actually come in several guises, one of which is only very slightly marked.⁵ *Wh-ex-situ* questions must then be either an optional, freely available variant (of at least one of the varieties of *wh-in-situ* questions) or a different type of *wh*-question. Surprisingly, *wh-ex-situ* are understudied and still await more careful analysis, which we intend to provide here (but see Baunaz 2011:223, fn. 229).

We shall be restricting our attention to the syntactic structures displayed in (1) and (2), a stand that calls for explanation, given the many other options available in French (Mathieu 2009). First, in order to be relevant, the *ex-* and *in-situ* counterparts examined in this article need to be strict prosodic minimal pairs, which leads us to exclude some types of *wh-in-situ* questions. According to Martin (1975) and Vergnaud and Zubizarreta (2005), *wh-in-situ* questions come in two prosodically distinct guises, as in (4a-b) (their 3).⁶ Both can be divided into two prosodic phrases $\varphi 1$ and $\varphi 2$. However, the first guise (a) has a rising accent on the subject DP ($\varphi 1$) and a falling accent on the VP ($\varphi 2$), whereas the second one (b) has a rising accent on the subject DP+V ($\varphi 1$) and a falling accent on the PP containing the *wh* ($\varphi 2$). The latter has an emphatic flavor.

- (4) a. [La jeune artiste] $_{\varphi 1}$ [a dansé avec qui] $_{\varphi 2}$?
 the young artist has danced with whom
 α . [La jeune artiste] $_{\varphi 1}$ [a dansé avec qui] $_{\varphi 2}$?
 1. Non-presuppositional
 2. Presuppositional: Partitive
 β . [La jeune artiste] $_{\varphi 1}$ [a dansé avec /qui\ \setminus] $_{\varphi 2}$
 3. Presuppositional: Specific
 b. [La jeune artiste a dansé] $_{\varphi 1}$ [avec qui] $_{\varphi 2}$?

Moreover, type (4a) is also reported to come in two prosodic types, α and β , the latter displaying an additional rise-fall accent on the *wh* (Baunaz and Patin 2011). Finally, α and β correspond to three different interpretations (1, 2, 3; Baunaz 2011, 2016).⁷ In order to control for the prosodic facts, we shall ignore the subtle difference between 1. and 2. (as in Baunaz and Patin 2011) and use the least-marked type of *wh-in-situ* questions ($\alpha\alpha$) to contrast it with *wh-ex-situ*. The exceptional types ($\alpha\beta$) and (b) will

lock (2005); Myers (2007); Quillard (2000); Shlonsky (2012); Tailleux (2014); Vergnaud and Zubizarreta (2005); Zubizarreta (2003). See also GENWH 2018, *The Geneva WH-orkshop on Optional Insituness* at <https://genwh2018.wordpress.com/>, last accessed 18 May 2020.

⁵This is also assumed more or less explicitly in Adli (2006), Aoun et al. (1981), Baunaz and Patin (2011), Hamlaoui (2010, 2011), and Rizzi (1996).

⁶Beysade (2006) also proposes different prosodic patterns for *wh-ex-situ* in Standard French, a variety of French we shall not be concerned with here (see Sect. 2).

⁷Here are Baunaz's (2016) definitions (The reader is referred to the original article for details):

"A partitive *wh*-phrase is an object, which belongs to a presupposed set containing more objects. Each of the objects of the set can potentially be referents to the answer of the *wh*-phrase, i.e., all are alternatives." (p. 134).

"Specificity narrows down the context to familiar individuals, excluding alternatives. A constituent question involving specificity entails an answer referring to a familiar individual that the interlocutor has in mind." (p. 137)

be mentioned only when needed in order to avoid confusion.⁸ We shall informally refer to (a β) and (b) as ‘stressed *wh-in-situ*.’ Finally, we shall not elaborate on the final, interrogative intonation in *wh-in-situ* questions, which was argued to be like *yes/no* questions along with a presuppositional analysis of these questions (Cheng and Rooryck 2000; Déprez et al. 2013). Indeed, there is variation among speakers (Déprez et al. 2012; Tual 2017), and this type of intonation seems to concern only a variety of French in which *wh-in-situ* is excluded from subordinate clauses, which is not the case of the one under study here (see also Baunaz 2011:43). Second, French allows variation in its *wh-ex-situ* questions with the possible, additional insertion of *est-ce que* ‘is it that’. In this contribution, in order to work from minimal pairs, we shall only consider the *ex-situ qu’est-ce que* ‘what is it that’ in contrast with the *in-situ quoi* ‘what’. The other combinations (e.g., *où est-ce que* ‘where is it that’), which are quite rare in adult, oral corpora, are left for further investigation.⁹

This said, the article will argue against optionality between *wh-ex-situ* and *in-situ* questions, and will claim that French is actually a *wh-in-situ* language that sometimes allows for *wh-ex-situ*, under specific conditions, much like other *wh-in-situ* languages. Such an alternation is well documented in languages across the world and is illustrated in (5) with Eastern Armenian (Megerdooonian and Ganjavi 2000, glosses as in original).

- (5) a. Ara-n **vor girk-en** e k’artatsel? [Eastern Armenian]
 Ara-NOM which book-ACC is read
 ‘Which book did Ara read?’
 b. **Vor girk-en** e Ara-n k’artatsel?
 which book-ACC is Ara-NOM read
 ‘Which book did Ara read?’

The unmarked way to form a *wh*-question is (5a). (5b) is an instance of fronting or scrambling. This type of alternation is widespread and is also attested in American Sign Language (Abner 2011), Korean (Beck 2006; Beck and Kim 1997), Mandarin Chinese (Hoh and Chiang 1990, Cheung 2008), some Northern Italian Dialects, including Trevigiano (Bonan 2019), Persian (Megerdooonian and Ganjavi 2000), and Turkish (Özsoy 2009), to name but a few.

A reasonable claim is that this additional movement of a *wh*-item in *wh-in-situ* languages does not come for free. However, no or little motivation has been given for these frontings and scramblings in the past literature. Based on child and adult data, we shall aim to show that French behaves like a *wh-in-situ* language, and to provide

⁸(4b) is arguably a case of ‘Question with Declarative Syntax,’ in which the focus is replaced *in situ* by a *wh*, and the interrogative meaning is acquired at the pragmatic level and does not directly follow from clause-typing (Bobaljik and Wurmbrand 2015), which solves Baunaz’s (2011) issue on a subset of D-linked multiple questions in English. It is beyond the scope of this article to compare the syntax of the various types of *wh-in-situ* questions. Poletto and Pollock (to appear) (and in earlier work) argue that there are at least two syntactic types of *wh-in-situ*, respectively displaying overt remnant-IP and remnant-vP movement. We claim that there is at least one type with covert *wh*-movement and no IP-movement (see Sect. 3), in line with Bonan’s (2019) cross-Romance perspective.

⁹Ratios of other combinations/all *wh-est-ce que* forms: 2.3% ($N = 15/651$; Dekhissi and Coveney 2018), 3% ($N = 73/2391$; Palasis et al. 2019), 3.7% ($N = 5/134$; Reinhardt 2019). Details on *c’est wh- que* and *wh- c’est que* with (69).

a motivation for the *wh*P not to be *in-situ*, but *ex-situ*. In our view, the movement is triggered by a focus feature that can be defined as Exclusive, and for which we propose a semantics.

The article is organized as follows: In Sect. 2, we define the contours of Colloquial French, as opposed to Standard French, by relying on the Diglossic Hypothesis. This point is crucial because not all patterns of *wh*-questions belong to the same variety. Section 3 then examines *in-situ* and *ex-situ*, simple and multiple questions in order to test and compare their sensitivity to constraints on *wh*-movement. Since (at least some) *wh-in-situ* questions appear as the unmarked way to form *wh*-questions in Colloquial French, the semantic properties of *wh-ex-situ* are further investigated in Sect. 4 in order to uncover a possible and common trigger to *wh*-fronting in this language. Section 5 then formalizes the suggested trigger, namely Exclusivity, a contrastive operator. Finally, we draw some conclusions and suggest leads to further research in Sect. 6.

2 Defining Colloquial French

The study must consider only one variety of Metropolitan French in order to explain the *in-situ/ex-situ* alternation. Thus, it is crucial to start with a definition of Colloquial French (the subject of our study), in contrast with Standard French (which we shall leave aside). We claim that the diglossic hypothesis provides touchstones for determining which *wh*-structures belong to one or the other variety, Subject-Clitic inversion (henceforth SCLI) being particularly relevant to the issue. Note that no variety possesses all the patterns of *wh*-questions.

2.1 The diglossic hypothesis

Syntactic variability has been widely acknowledged in Metropolitan French. Three main areas of variation have been thoroughly examined in adult speech: subject clitics, negation and interrogative structures (Ashby 1977a,b, 1981; Coveney 2002, 2003; Lambrecht 1981; Pohl 1965, 1975, among others). Two different hypotheses currently account for this variability, that is, Variationism and Diglossia.¹⁰ The debate mainly hinges upon the number of grammars that a French native speaker actually handles. The variationist approach suggests that the different variants of French belong to the same grammar, and that variation is due to social and stylistic factors (Beeching et al. 2009; Blanche-Benveniste 1997; Coveney 2011; Gadet 2003, among others). There are no grammatical constraints on the combinations of variants in this unique grammar (Rowlett 2011). In contrast, the diglossic hypothesis, which builds on Ferguson's (1959) work, suggests that a variant of French belongs to one of two cognate, but nevertheless distinct, grammars (Barra-Jover 2004, 2010; Massot 2008, 2010; Massot and Rowlett 2013; Zribi-Hertz 2011). Variation is explained by the co-existence of two grammars in the minds/brains of the French native speakers. Let us call these two grammars G1 and G2.

¹⁰Villeneuve and Auger (2013) have aimed to reconcile both approaches based on French and Picard data.

Table 1 Formalizing Diglossia in child data

Items	Diverging statuses	
	G1	G2
Nominative clitics (<i>je</i> 'I', <i>tu</i> 'you', <i>il</i> 'he', etc.)	Agreement markers	Subject pronouns
	Proclitics only	Pro- or enclitics
	Obligatory	Optional
	Elided (<i>i</i> instead of <i>il</i> _C)	Non-elided (always <i>il</i>)
Nominative clitic and co-occurring DP	Doubling	Dislocation
Negation ((<i>ne</i>)... <i>pas</i> 'not')	Simple	Discontinuous

In the framework of Diglossia, Massot (2010) examined the combination of six binary variables in spontaneous adult data, for example, simple negation *pas* 'not' (G1) vs. discontinuous negation *ne...pas* (G2), and the use of the third-singular pronoun *on* 'one' (G1) vs. first-plural pronoun *nous* 'we' (G2). The author reported that his informant never code-switched within the boundaries of his clauses, and hence argued in favor of the strong grammatical constraint that a characteristic of one grammar (e.g., G1 *pas* or *on*) cannot co-occur with a characteristic of the other grammar (e.g., G2 *ne...pas* or *nous*) within the same clause. Thus, a sentence with code-switching (e.g., **on ne gagnait pas énormément* 'one did not earn much') is not expected within the diglossic hypothesis (Massot 2010:100).

Other variants in adult French are also well documented and the most emblematic topic is probably the morpho-syntactic status of subject clitics, notably *il* 'he' (Ashby 1984; Culbertson 2010; De Cat 2005; Kayne 1975; Legendre et al. 2010; Morin 1979; Rizzi 1986; Roberge 1990; Zribi-Hertz 1994, among others). In G1, *il* takes the elided form *i*- before consonant-initial constituents (e.g., *i-travaille* 'he works'), is analyzed as a preverbal agreement marker, and hence cannot undergo SCLI. In contrast, in G2, *il* is never elided and, as a proper pronoun, can appear pre- or postverbally (e.g., *il travaille, travaille-t-il*).

More evidence in favor of the diglossic approach has been adduced in studies in acquisition. Indeed, the two grammars are assumed to emerge subsequently, the first one being acquired at home and the second being mainly learned 'by the means of formal education' (Ferguson 1959:331). Preschool children are therefore expected to initially handle G1 only and develop G2 later. Palasis (2013, 2015) highlighted this asynchrony in kindergarten data, and showed that children consistently used the items of the same grammar within the boundaries of their utterances, as detailed in Table 1 and instantiated in (6).

- (6) Consistent use of G1 or G2 within the boundaries of an utterance (Palasis 2015:135):
- a. Mais *i* voulait *pas* que j(e) le nourrisse. (LOU, 4;9, G1)
but he wanted not that I him feed
'But he didn't want me to feed him.'
 - b. Mais lui *il ne* voulait toujours *pas*. (LOU, 4;9, G2)
but him he neg wanted always not
'But he still didn't want.'

Table 2 Distribution of *wh*-position in relation with SCLI

	SCLI	Elided <i>il</i> 'he'	Grammar	<i>Wh</i> -position
a. <i>i travaille où?</i> 'he works where'	No	Yes	G1	in-situ
b. <i>où i travaille?</i> 'where he works'	No	Yes	G1	ex-situ
c. * <i>travaille-t-il où?</i> 'works he where'	Yes	No	G2	*in-situ
d. <i>où travaille-t-il?</i> 'where works he'	Yes	No	G2	ex-situ

SCLI is rare in child questions (0.1% of the interrogatives with a subject clitic) and, when instantiated, not always fully mastered, as seen in (7c).¹¹

(7) SCLI in child data:

- a. *Où es- tu Maman Ours?* (CAR, 3;5, G2)
where are you Mummy Bear
'Where are you Mummy Bear?'
- b. *Que se passe-t- il?* (VIC, 3;5, G2, clearly imitating adult intonation)
what REFL happens it
'What is going on?'
- c. **Qu' est- c(e) qui passe-t- il?* (KEL, 3;7, G2)
what is it that happens it
'What is going on?' (intended)

These facts are consistent with the diglossic hypothesis that French children start with G1 (no SCLI) and develop G2 (SCLI) only later.

2.2 Diglossia and *wh*-questions

Let us now place *wh*-questions in the picture by determining which *wh*-interrogative structures belong to G1 and G2. SCLI, which is found only in G2, is particularly relevant to the matter, as illustrated in Table 2.

Table 2 highlights that G1 displays both *in-situ* (in a) and *ex-situ wh*-positions (in b). In contrast, we have no positive evidence in our child dataset that G2 also features *wh-in-situ*, that is, there are no occurrences combining a non-elided clitic and an *in-situ wh*-word (e.g., *il travaille où?* 'he works where'). Therefore, no conclusion can be drawn with regard to variation of the *wh*-position in G2, which at this stage may or may not be described as an exclusively *wh-ex-situ* language, on a par with English.¹² We shall assume that G1 is Colloquial French and G2 Standard French.

¹¹Examples are from our corpus of seminaturalistic data, which contains a total of 913 finite, matrix *wh*-questions produced by 17 children between 2;06 and 4;11, all native speakers of L1 Metropolitan French (Palasis et al. 2019).

¹²For an insightful account of the behavior of *wh*P's in SCLI G2 questions, see Poletto and Pollock (to appear) (and previous work).

In this section, we observed that Colloquial French is the only variety of French that undoubtedly features an *in-situlex-situ* alternation and that *wh-ex-situ* questions with SCLI belong exclusively to Standard French (Table 2). Consequently, we shall concentrate on Colloquial French in the remainder of the paper and leave Standard French for further research.¹³

3 Evidence that *wh-in-situ* questions are unmarked questions

In this section, we compare *wh-in-situ* and *wh-ex-situ* questions, provide evidence that *wh-in-situ* is the unmarked counterpart of *wh-ex-situ*, and conclude that Colloquial French displays only covert *wh*-movement. In Sect. 3.1, we examine simple and multiple questions with regard to strong and weak islands. This leads us to adopt the theoretical framework of Relativized Minimality in Sect. 3.2. Section 3.3 then widens the investigation to Superiority. These tests highlight when *in-situ* is either unacceptable or degraded compared to *ex-situ*, and hence enable us to pinpoint the feature(s) *ex-situ wh*-items carry, contrary to their *in-situ* counterparts.

Before we proceed however, a note on D-linking and the French counterpart of *which* is in order. The demonstration will rely on bare *wh*Ps only and thus discard *quel* ‘which’ *wh*Ps for three reasons. First, French *quel wh*Ps are ambiguous between ‘which (book)’ and ‘what (book)’ and, therefore, cannot be used straightforwardly to test D-linking. Second, French has an unambiguously D-linked *wh*-item, namely *lequel* ‘which one.’ But French *lequel* needs to be more strongly related to the context than English *which*. Pending more research on this topic, we decide to leave it aside. Finally, the best reason is that bare *wh*Ps can be contextually D-linked, as already noted in Pesetsky’s (1987) seminal paper. This means that when comparing a ‘which’ phrase with a bare *wh*P, we can never be sure that we are comparing a D-linked and a non-D-linked *wh*P. Consequently, in order to exclude possible D-linking effects, we shall contrast only bare *wh*Ps in strongly controlled contexts.

3.1 Islands

In *wh-ex-situ* languages, some configurations are degraded when the *wh*P is in the (matrix) CP-domain and in a relation with certain types of embedded clauses or after a negative operator. This degradation is assessed when the movement from the lower to the higher position is blocked. These configurations thus provide good tests to check whether movement takes place overtly or covertly. They are metaphorically named ‘islands,’ and can be either ‘strong’ (i.e., they block all types of extractions) or ‘weak’ (i.e., some extractions are possible; Cinque 1990; Ross 1967; Szabolcsi 2006). Adjuncts are well-known strong islands, as in (9), and negative operators are examples of weak islands, as in (17) and (18). In each configuration, *wh-in-situ* is provided in (a) and *wh-ex-situ* in (b). As mentioned earlier, context is crucial to the

¹³ Partitioning the data has seemed indispensable to other authors before us (Baunaz 2011, 2016; Hamlaoui 2010, 2011; Mathieu 2004). We strengthen the dichotomy by resorting to the diglossic hypothesis.

analysis. (8) sets an overall context for the examples in Sect. 3 and is fleshed out when necessary.¹⁴

- (8) Overall context: *The university library is closing and gives out most of its books and stationery. Before getting an item, a possible recipient has to be interviewed by Pierre, one of the librarians, in order for him to know what the new destination of the given item will be. Then, another librarian, Julien, is allowed to give the item to the interviewee. Pierre and Julien regularly exchange positions. Importantly, due to the size of the library, nobody really knows the exact list of books and other items given out.*

3.1.1 Strong islands

In this section, we shall see two crucial points: *wh-in-situ* involves covert movement; *wh-ex-situ* is more constrained, a point that we shall verify in the next section as well.

Examples (9) display strong, adjunct islands (i.e., *avant que* ‘before that’ clauses):

- (9) Simple questions’ obedience to strong islandhood:
We are at the very beginning of the giving-out process. The library director finds out that Julien gave out a book to someone before Pierre had interviewed him/her. She asks Julien:
- a. *T’ as donné un livre [avant que Pierre il ait vu **qui**]?
 you have given a book before that Pierre he has seen who
- b. ***Qui**_i t’ as donné un livre [avant que Pierre il ait vu t_i]?
 who you have given a book before that Pierre he has seen

In these simple questions, both *in-situ* and *ex-situ wh*Ps are unacceptable. The reason for b’s unacceptability is obviously syntactic: *qui* is prevented from moving out of the *avant que* clause. Crucially, we also take a’s unacceptability as indication that *qui* normally covertly moves to check the *wh*-feature, which it cannot do here because there is no possible escape from the island. In fact, contrary to weak islands (see fn. 24 and Sect. 3.1.2), all accounts of adjunct, strong islands we are aware of are dependent on the *wh*P moving out of the clause, be it in terms of subadjacency/barriers (Huang 1982; Chomsky 1976), phases (Müller 2010), or the eventive structure of the whole sentence (Truswell 2007).¹⁵ The relevant LF of both a and b is schematized in (10).

¹⁴In all the examples in Sect. 3, it is important for the *wh*P not to range among a given set, i.e., not to be D(iscourse)-linked (Pesetsky 1987). Otherwise, another type of *wh-in-situ* question is triggered (partitive, or exclusive with stressed *wh-*), whose behavior regarding islandhood is different, as shown in Baunaz (2011, 2016).

¹⁵No good representational account in terms of Minimality is available, as acknowledged by Luigi Rizzi himself, who also posits a subadjacency constraint to explain strong islands (Rizzi 2001). The reason for that is that no feature has been found so far that could be responsible for the Minimality effect. More on Relativized Minimality in Sect. 3.2. Another family of explanations accounts for sentences like (9b) in terms of processing difficulties involving the interaction between the necessity of holding a term in working memory until the gap (its interpretation position) is reached, and the lexical semantic processing factors at the embedded clause boundary (Kluender and Kutas 1993). But they do not address the degradation of cases like (9a), where supposedly nothing has to be held in memory.

(10) LF: Q WH... t

It is time to pause here to consider another very influential explanation, namely the hypothesis of an overt movement of the *wh*P followed by a remnant-IP movement (Munaro et al. 2001, and followers). A simple sentence like (11a) is derived through the steps described in b and c.

- (11) a. Tu vois qui?
 b. Step1: qui_i tu vois t_i
 c. Step2: [tu vois t_i]_j qui_i t_j

According to this hypothesis, (9a) is ruled out because the strong island blocks the *wh*-movement required in Step1. As appealing as it may be, this theory will not be retained here, mainly because it does not seem to extend to Colloquial French¹⁶ (see Bonan 2019; Etxepare and Uribe-Etxebarria 2005; Manzini and Savoia 2011; Polletto and Pollock 2015; Uribe-Etxebarria 2003, for more details and tests on Spanish and Italian dialects; and Bonan 2019; Baunaz 2011; Cheng and Rooryck 2000; Mathieu 2002, for a stand similar to ours on Colloquial French). Indeed, the theory predicts that the *wh*-term should always be rightmost (i.e., the “sentence-finality requirement”), whereas Colloquial French displays sentences like *T’as vu quoi hier?* ‘what did you see yesterday?’, *T’as donné quoi à Paul?* ‘what did you give to Paul,’ in which the prosodic pause signaling deaccenting of *hier* and *à Paul* is optional after *quoi*. Another option would be to assume several derivational steps topicalizing *hier* or *à Paul*, then fronting the *wh*-items before the remnant movement,¹⁷ but these steps would be loosely motivated.

Moreover, Bonan (2019) entertains the idea that patterns like (14) could be evidence for a French *wh*-position of a third type, namely intermediate, IP-internal, above vP, as in Trevigiano.¹⁸

- (12) Tu vois Pierre **ici/lundi**.
 you see Pierre here/on Monday
- (13) Tu vois Pierre **où/quand?**
 you see Pierre where/when
- (14) ??Tu vois **où/quand** Pierre?
 you see where/when Pierre
 ‘Where/when do you see Pierre?’
- (15) *Tu as **où** vu Pierre?
 you have where seen Pierre
 ‘Where did you see Pierre?’

¹⁶It crucially rests on varieties of French that feature SCLI (G2), which we excluded (see Sect. 2), but are well accounted for in the remnant-movement approach. This means that Bonan’s (2019) position is right, according to which both remnant-IP movement and *bona fide wh*-insitiveness are necessary to account for the variety of *wh*-insitiveness across (and sometimes within) languages.

¹⁷Jean-Yves Pollock (p.c.).

¹⁸We thank an anonymous reviewer for pointing out this fact to us.

Time and location adjuncts appear clause-finally in French as evidenced in (12), which is also possible for the corresponding *wh*-terms in (13), but not obligatorily if we consider (14). In (14), *où/quand* have moved to an IP-internal position. However, Baunaz (2011) points out the optionality and degradation of this kind of example, and the specific semantic (strongly presuppositional) and prosodic conditions under which this kind of sentence is marginally acceptable. Bonan (2019) concludes that (14) is better seen as an instance of non-feature-driven scrambling. This is all the more plausible given that an IP-internal movement would also derive a sentence like (15), where *où* would ungrammatically surface between the T (auxiliary *as*) and v/V (past participle *vu*). Be that as it may, its presuppositional character and its specific prosodic conditions exclude this intermediate *wh*-position from our study (see 4 and fn. 7).

Coming back to the hypothesis that both overt and covert movement must be considered if we want to explain the patterns under examination, multiple questions¹⁹ confirm the previous results and take us a step further in hinting at the possibility that *wh-ex-situ* structures are more constrained than *wh-in-situ* ones:²⁰

(16) Strong islandhood in multiple questions:

One morning, the library director asks Pierre:

a. *In situ* (1)

*Hier, t' es parti [avant que Julien il ait donné **quoi** à **qui**?]
yesterday you have left before that Julien he has given what to whom

a'. *In situ* (2)

T' as donné **quoi** [avant que Julien il ait vu **qui**?]
you have given what before that Julien he has seen who

b. *Ex situ*

***Qu'est-ce que**; t' as donné t_i [avant que Julien il ait vu **qui**?]
what you have given before that Julien he has seen who

While unacceptable when both *wh*-arguments are in the island in (16a), the sentence becomes better when the matrix verb displays a *wh*-argument as in (16a'). A possible explanation is that one *wh*-item in the matrix clause is required (if the other one is trapped in an embedded island) and suffices to covertly check the *wh*-feature on matrix C, hence freezing the movement of the second *wh*-item (here *qui* in the embedded clause).²¹

¹⁹We limit ourselves to two *wh*-items here.

²⁰(16a') sounds echoic to an anonymous reviewer (native speaker of French), whereas our informants found it acceptable in the given context as a request for information. This discrepancy may be due to an idiolectal difference.

²¹Alternatively, one could propose an explanation along the lines of Richards's (1998:604–608) *Principle of Minimal Compliance*. Simplifying the framework, constraints apply only once. In our example (16a'), both *wh* move, but the second one is not submitted to subjacency, since this constraint was already checked and found non-violated by the first *wh*. "It appears to be true quite generally that in cases involving multiple *wh*-movement to a single [+wh] complementizer, only the first-moved *wh*-word will have to obey Subjacency; the other *wh*-movements are free from Subjacency. This is true regardless of the levels at which the *wh*-words move [SS or LF]." (p. 608)

In contrast with *wh-in-situ*, strong islandhood blocks *ex-situ* multiple questions (16b), albeit one of the *wh*-items is in the matrix. Following what we just saw, this indicates that it is not sufficient that the *wh*-feature on C is checked by the moved *wh*. Our hypothesis is that the degradation is due to the necessity for the second *wh* to move as well, which it cannot do because it is trapped in the adjunct island.²² Weak island facts also point towards *wh-ex-situ* structures involving an additional operation.

3.1.2 Weak islands

In this section, we show that both syntax (more precisely, Minimality but not Intervention) and semantics (Contradiction) are at play in our weak-island facts. As in the previous section, we give simple and multiple question examples. We illustrate weak-island effects with a negative operator in (17) and (18).²³

- (17) Negative operator *pas* ‘not’ (argument condition):
The library director finds out that Julien refused to hand out an item. She asks Pierre:
- a. *Julien il a pas donné **quoi**?
 Julien he has not given what
 - b. **Qu’est-ce que**_i Julien il a pas donné t_i?
 what Julien he has not given
- (18) Negative operator *pas* ‘not’ (adjunct condition):
The library director knows that one day instead of leaving at 5 pm Julien stayed. She asks Pierre:
- a. *Julien il est pas parti à l’heure **quand**?
 Julien he has not left on time when
 - b. ?***Quand**_i Julien il est pas parti à l’heure t_i?
 when Julien he has not left on time

Weak islands are taken to display an acceptability asymmetry between argument and adjunct extractions, which is also reported here for *wh-ex-situ* questions (17b vs. 18b).²⁴ However, the examples are of interest here because *wh-in-situ* questions do not feature the asymmetry, and are rated as badly for arguments as for adjuncts (sentences in 17 and 18). Even if this is not exactly in line with the judgments reported in Baunaz (2011), whose informants tend to accept the configurations repre-

²²Once again, additional stress on the *wh-in-situ* repairs the island violation and triggers a pair-list reading (Hirschbühler 1979).

²³We shall not use factive islands here because the judgments of our informants are less clear-cut and no clear contrast appears when the context changes.

²⁴We follow Abrusán’s (2014) semantic theory of weak islands based on the idea that the interaction between the island that contains certain types of predicates and some question words (but crucially not all) yields a contradiction (i.e., configurations in which there is no maximally informative answer contrary to what is required in the act of questioning). This account best explains the argument/adjunct asymmetry and why the unacceptability disappears in certain contexts.

sented in (17a) and (18a), Baunaz (2011:44, 60–61) nevertheless notes that they are constantly felt to be degraded with respect to their *ex-situ* counterparts and that the argument/adjunct asymmetry exists in *ex-situ* questions only.²⁵

This means that the *in-situ* argument questions in (17a) are probably degraded because of a factor independent from Abrusán's contradiction theory, which explains why the adjunct questions in (18a-b) are degraded. Importantly, note that the degradation CANNOT be due to an intervention effect, because under current accounts (Beck 2006; Haida 2008; Mayr 2014, and works based on them) intervention effects arise when the LF is as in (19) (intervention between Q and the *wh*), that is when the *wh*-item does not move. This is different from what we saw in (10) in Sect. 3.1.1 with covert movement and no possible intervener between Q and the *wh* (see Beck 2006 building on a reasoning on D-linked multiple questions in Pesetsky 2000).

(19) LF: *Q ... intervener ... WH

The multiple-question data in (20) and (21) confirm these results:

(20) Weak-islands in multiple questions: Negative operator *pas* 'not' (argument condition):

After a few days, Pierre and Julien have been alternatively in charge of handing out the items. The library director knows that they both gave books, but that Julien refused some items to some people. She asks Pierre:

- a. *Julien il a pas donné **quoi à qui**?
Julien he has not given what to whom
- b. **Qu'est-ce que** Julien il a pas donné **à qui**?
what Julien he has not given to whom
- b'. ?**À qui** Julien il a pas donné **quoi**?
to whom Julien he has not given what

(21) Weak islands in multiple questions: Negative operator *pas* 'not' (adjunct condition):

After a few days of vacation, the director comes back to the library and asks about the employees. She finds out that Julien worked from all the desks every day except for one day. She further asks:

- a. *Julien il a pas travaillé **où quand**?
Julien he has not worked where when
- a'. *Julien il a pas travaillé **quand où**?
Julien he has not worked when where
- b. ***Quand** Julien il a pas travaillé **où**?
when Julien he has not worked where
- b'. ***Où** Julien il a pas travaillé **quand**?
where Julien he has not worked when

²⁵Note that the asymmetry resurfaces and (only) the argument questions become better when the *in-situ* questions are set in a context that makes them partitive or specific (in which case, the *wh*-item is stressed, a pattern that we do not examine in detail here; see the discussion around 4 in Sect. 1).

(22) Q WH ... Foc/Neg ... WH

Multiple *ex-situ* questions in weak-island contexts are noteworthy in three respects: 1) Like simple questions, they display the argument/adjunct asymmetry, as illustrated in (20b/b') vs. (21b/b'); 2) Multiple *ex-situ* questions are better than their *in-situ* counterparts; 3) They are degraded, however, with respect to simple *ex-situ* questions in the same contexts (see 17b). Aspect 1) is expected in weak-island situations, where adjunct *wh*-questions are ungrammatical for independent semantic reasons (see fn. 24). Aspect 3) is unexpected. However, the marginal acceptability of (20b/b') suggests that the contrast between simple and multiple questions may be due to processing difficulties.²⁶

Finally Aspect 2) confirms the observation made on simple questions: French *wh-in-situ*, but not *wh-ex-situ* questions are degraded in contexts where Abrusán's (2014) principle (fn. 24) is not at play, namely when the question bears on an argument. Consequently, the degradation must be attributed to another factor. We follow Baunaz (2016) in positing that a Relativized Minimality effect applies here à la Starke (2001) and Rizzi (2004).

3.2 Relativized minimality

In this section, we explain what Relativized Minimality is and how it can account for the marginality of *in-situ* questions, but also for the acceptability of *ex-situ* questions, provided that we posit that the latter are endowed with an additional feature.

In a configuration such as (23) or (24),²⁷ Y intervenes between X and Z because they share a feature (or a bundle of features) α . Note that if Y has a richer feature structure (for example bearing both α and β features), the same effect arises, as in (24). In derivational terms (i.e., Minimal Link or Attract Closest Constraint; Chomsky 1995), X probes for an item that bears the feature α , but it meets Y "before" its actual goal Z, Z is therefore left behind, and the derivation crashes. In contrast, (25) displays no effect because the feature structure of X is richer than that of the possible intervener Y. Thus X is allowed to probe for Z past Y to check its feature β .

(23) *X α ... Y α ... Z α ...

(24) *X α ... Y $\alpha\beta$... Z α ...

(25) X $\alpha\beta$... Y α ... Z $\alpha\beta$...

Following Rizzi (2004, 2014) and Baunaz (2011, 2016), we assume that Minimality effects arise between features of the same family. In particular, *Wh*, *Neg*, and *Focus* features belong to the same group of Quantificational features. In this framework, the

²⁶Unless speakers construe the second *wh* as remaining *in situ*, in which case we are in configuration (i) hereafter and an intervention effect could arise between Q and WH2. This is unlikely, however, given our analysis of (16b) and Dayal's (1996, 2003) account for multiple questions, in which both *wh* must be in CP (i.e., configuration (i) does not arise for multiple *ex-situ* questions).

(i) Q WH1 ... intervener ... WH2

²⁷Where X asymmetrically c-commands Y and Y asymmetrically c-commands Z, X and Y being of the same structural type, and where X and Z form a chain.

in-situ sentences in (17a, 20a) are degraded because the C[+wh] head, the *wh-in-situ*, and the intervening operator (e.g., *pas* ‘not’) all share the same quantificational feature. Thus, the *wh* is not attracted to CP at LF and the derivation crashes (see also Beck 1996; Kratzer and Shimoyama 2017 [2001]:139–140).²⁸

Put otherwise, *wh-in-situ* questions in (17) and (20) are ruled out because the *wh*-items are not endowed with an extra feature that would allow them to escape, namely they only carry a *wh*-feature. Conversely, the acceptability of (17b) shows that the *ex-situ wh*-item carries an extra feature. *Ex-situ* questions like (17b) provide evidence that *wh*-items can sometimes be extracted from weak islands (e.g., with a negative operator). Under the previous analysis, the extraction is possible only if *ex-situ*, but not *in-situ wh*-items bear an additional β feature, as schematized in (25).²⁹ The remainder of Sect. 3 will test these assumptions against the Superiority Condition.

3.3 Superiority

In addition to strong- and weak-island effects, multiple questions allow us to test Superiority, another hallmark of *wh*-movement (Chomsky 1973). We shall examine *in-situ* and *ex-situ* facts separately and show that the former, but not the latter exhibit such an effect.

3.3.1 Superiority and *in-situ*

Let us observe the following multiple question, which features two *in situ wh*Ps:

- (26) Julien a donné **quoi à qui**?
Julien has given what to whom
- (27) Possible LFs for (26):
- a. **quoi** Julien a donné **à qui**?
 - a'. **quoi à qui** Julien a donné?
 - b. **à qui** Julien a donné **quoi**?
 - b'. **à qui quoi** Julien a donné?

First, it is noteworthy that a single-pair reading for a multiple question like (26) seems to be freely available. However, some speakers also allow for a pair-list reading. Second, two competing LFs are possible in this case: (27a),³⁰ in which *quoi* checks the *wh*-feature on C and is the sorting key (i.e., the answer is (28a), and (27b), in which *à qui* checks the *wh*-feature on C and is the sorting key (i.e., the answer is (28b)).

²⁸Crucially, as already noted in fn. 25, according to Baunaz (2016), the questions become acceptable when the *wh*-item is endowed with a feature [+specific] (and sometimes [+partitive]).

²⁹Likewise, some speakers of French, including an anonymous reviewer, find (18b) slightly better than (18a) (hence the rating ?*) because the former violates only Abrusán’s island constraint, while the latter additionally violates Relativized Minimality.

³⁰Or (27a’) if both *wh* must be in CP to obtain a pair-list reading as proposed in Dayal (1996, 2003).

Crucially, (with no *wh* stressed), (28a) is highly preferred over (b), which suggests that the LF of (26) is (27a).

- (28) a. Julien a donné un livre à Mme A. et un stylo à Mme B.
 Julien has given a book to Mrs A. and a pen to Mrs B
- b. #? À Mme A. Julien a donné un livre et à Mme B. un stylo.
 to Mrs A. Julien has given a book and to Mrs B. a pen

This is reminiscent of the Superiority Condition (29), whose application to *wh*-movement predicts the ungrammaticality in (30c).

- (29) Superiority Condition (Chomsky 1973):
 No rule can involve X,Y in the structure:
 ... X ... [... Z ... – WYV ...] ...,
 where the rule applies ambiguously to Z and Y, and Z is superior to Y.
- (30) Application to *wh*-movement (Kuno and Robinson 1972):
- a. [C_{+WH} [IP Who saw what?
 b. Who_i C_{+WH} [IP t_i saw what?
 c. *What_i C_{+WH} did [IP who see t_i?

In (30), C_[+wh] can attract the closest *wh* as in (b), but not the farthest, as in (c); (23) illustrates this configuration. Crucially, (c) dramatically improves if *what* is D-linked (Pesetsky 1987), which could correspond to the configuration in (25).³¹ For Colloquial French, there is further complication though. The fact that D-linking rescues Superiority violation as just mentioned for English can be seen from the embedded-question patterns in (31) (still in the library scenario).

- (31) a. ??/✓ Le directeur se demande à qui t' as donné quoi.
 the director himself asks to whom you have given what
- b. *Le directeur se demande ce que qui a lu.
 the director himself asks what who has read

As expected, in (31a), movement of the indirect object *à qui* over the direct object *quoi* gives rise to a Superiority effect. However, as in English, (a) improves a lot in a context where *à qui* and *quoi* are D-linked, for example if the library director knows that there are only two items to give out (a book and a pen) and only two visitors (say, Maria and Samantha). Crucially, (b), with the object that has moved past the subject, is unacceptable and does not improve in any context. Whatever the reason for that, Superiority tests involving subjects do not apply in French and we shall test Superiority based only on direct and indirect objects.³² Thus, (26) and (28) show that Colloquial French *wh-in-situ* questions are actually sensitive to Superiority.

³¹Despite much resemblance, Superiority is not easily amenable to Relativized Minimality, as pointed out in Rizzi (2011).

³²There has also been discussion around the order of direct and indirect objects (Larson 1988). Note however that (i), which features the reverse order of (26), is strongly unacceptable, no matter the prosody. We take it to be proof that (26) features the basic word order.

Table 3 Properties of *wh-in-situ* and *wh-ex-situ* questions in Colloquial French

	Properties	In-situ	Ex-situ
A	Sensitivity to strong islandhood	✓	✓
B	Sensitivity to weak islandhood	✓	*
C	Sensitivity to Superiority	✓	*
D	Pair-list reading	✓	✓
E	Single-pair reading	✓	??

3.3.2 Superiority and *ex-situ*

Multiple questions with *wh-ex-situ* lack the Superiority effect, as shown in (32).³³ Note that speakers manifest a strong preference for a pair-list reading of these questions, even if some of them do not exclude a single-pair reading.

- (32) a. **Qu'est-ce que t' as donné à qui?**
 what you have given to whom
- b. **À qui t' as donné quoi?**
 to whom you have given what

3.4 Intermediate summary

Table 3 summarizes the facts observed in Sect. 3 with regard to islands, Superiority and interpretation. It shows that *wh-ex-situ* questions differ from *in-situ* ones in their weaker sensitivity to weak islands and Superiority, properties we attributed to a specific, additional property, which we shall discuss in Sect. 5 (the difference between single-pair and pair-list readings is touched upon in conclusion).

Property A shows that both *wh-in-situ* and *wh-ex-situ* questions involve movement. Property B shows that covert movement in *wh-in-situ* questions is more restricted than overt movement in *wh-ex-situ* questions. Property C shows that (covert) movement in *wh-in-situ* questions is in fact *wh*-movement. On the other hand, weaker sensitivity to weak islandhood and lack of Superiority effect point to a movement of a different nature for *wh-ex-situ* (see configuration 25). Economy and parsimony also dictate such a conclusion. Since covert movement checks the *wh*-feature, we conclude that Colloquial French is a *wh-in-situ* language and that another type of feature is necessary to trigger overt movement.

This is reminiscent of the alternation mentioned in Sect. 1 with *wh-ex-situ* questions analyzed as more marked than *wh-in-situ* questions based on the cross-linguistic observation that when a language has the parameter '(unmarked) *wh-in-situ*,' it also displays marked, overtly-moved counterparts via fronting, as in Armenian in (5b) (Megerdooian and Ganjavi 2000), or via scrambling, as in Korean in (33) (Beck 2006, glosses as in original).

(i) *T' as donné à qui quoi?
 you have given to whom what

³³(a) is sometimes rated as slightly better than (b), but both are deemed acceptable.

- (33) a. *Minsu-man **nuku-lül** po-ass-ni?
 Minsu-only who-Acc see-Past-Q
 ‘Who did only Minsu see?’
 b. **nuku-lül** Minsu-man po-ass-ni?
 who-Acc Minsu-only see-Past-Q
 ‘Who did only Minsu see?’

Uncovering the content of the markedness assumed for *wh-ex-situ* in Colloquial French will be the goal of the remainder of the article. Our task is to clarify the properties of the additional feature advocated in Sect. 3.2 in the frame of Relativized Minimality. *Wh*-fronting without *wh*-movement is not a new idea though. Bošković (2002) attributed movement to focus in multiple *wh*-fronting languages, such as Bulgarian, and Hamlaoui (2010, 2011) tied *wh*-fronting and focus together in French in an OT framework. We shall agree with these proposals in Sect. 4, and elaborate on Focus in Sect. 5.

4 Semantic properties of *wh-ex-situ* questions: Acquaintance with contrast

This section aims to clarify which contexts license *wh-ex-situ* questions in order to uncover their semantics in Colloquial French. We shall examine three contexts where *wh-ex-situ* are possible and (unstressed) *in-situ* impossible. First, *wh-ex-situ* is acquainted with contrast in exclusive-pairing contexts in child speech (Sect. 4.1) and explicitly contrastive scenarios in adult speech (Sect. 4.2). Note that, in these contexts, contrast on other elements in the question is just a hint at the contrast on the *wh*P. Second, an exclusive selection in a set also triggers *wh*-fronting (Sect. 4.3). It should be emphasized that stressed *wh-in-situ* often (though probably not always) seems to appear in free alternation with *wh-ex-situ* in these situations. The emphasis that goes along with stress also points towards focus. Stress must then be carefully controlled for when evaluating the *in-situ* examples.

4.1 Exclusive pairing in child speech

Let us start with child speech, which offers the clearest patterns. In our corpus (detailed in fn. 11), children often ask questions out of the blue, as in (34). The situation is as follows: The child (WIL, 2;10.18) is playing, stops, turns to the adult, and asks the question. These questions feature *in-situ wh*-words.

- (34) Ma photo elle est où?
 my photo she is where
 ‘Where is my photo?’

Nevertheless, the corpus also displays *ex-situ wh*-questions, as in (35), uttered in a row by the same child (MAS, 2;7.5):

- (35) a. **Où** il est le poisson?
 where he is the fish

- b. **Où** il est le perroquet?
where he is the parrot
- c. **Où** elle est la tortue?
where she is the tortoise
'Where is the fish? Where is the parrot? Where is the tortoise?'

In (35), the child was given a board with six animal pictures to match with six individual cards, which means that we have two sets of items, each member of which has a unique correspondent in the other set. In other terms, we have a mapping of the member of a set onto the member of another set. Moreover, the position of an item is only understandable with respect to another one, namely in contrast with another one. In our view, this contrast triggers the fronting of the *wh*-word *où* 'where' because it requires information about the relative position of the card on the board.³⁴

4.2 Explicitly contrastive contexts

Hamlaoui (2011:fn. 22) noted that explicit contrast triggers obligatory *wh-ex-situ* questions in adult French. We reproduce the author's minimal pair A1/A1'³⁵ in (36).

- (36) A0: Tu vas faire **quoi**, pendant les vacances?
you will do what during the break
'What will you be doing during the break?'
- A1: (En)fin, **qu'est-ce (que)** t(u) AIMERAIS faire?
actually what you would.like do
'Actually, what would you like to do?'
- A1': #/*(En)fin, t(u) AIMERAIS faire **quoi**?
actually you would.like do what
'Actually, what would you like to do?'

Hamlaoui (2011) considered the relationship between modal verbs and the *inlex-situ* alternation and convincingly argued that modal verbs such as *tu aimerais* 'you would like' do not favor *wh*-fronting per se. They favor fronting only when they are *contrastively focused*, as in A1, where *aimerais* 'would like (to do)' is focused in contrast with the preceding verb *vas* 'will (do)' in A0.³⁶ We observe that the speaker wants to identify an object (what the addressee would like to do), whose identity *may* be different from that of another object (what the addressee will actually do). Thus, the

³⁴This is in line with (Palasis et al. 2019): In child speech, *wh-ex-situ* questions are favored when the sentence has more content, namely when it contains elements that can be contrasted with others in the context. On the other hand, presentational sentences with the fixed (and semantically empty) *be* form *c'est* 'it is' strongly favor *wh-in-situ*.

³⁵Numbering as in the original document, glosses and translations adapted. Once again, in A1' the *wh-in-situ* question with a modal is rescued if *quoi* is stressed.

³⁶*Aimerais* bears a C accent (Marandin et al. 2004), the French accent corresponding to the (rise-fall) contrastive topic B-accent in English (Büring 2003; Jackendoff 1972).

contrast between *aimerais* and *vas* goes along with a contrast between the content of *qu'est-ce que* in A1 and *quoi* in A0.³⁷

4.3 Teasing apart Contrast on the *wh*-item and Contrast on the non-*wh*-part

The previous two sections have shown that when a non *wh*-element is contrasted in a question (i.e., when it is a contrastive topic), the contrast is not limited to the non-*wh* part of the question and *wh*-fronting occurs. The contrasted non-*wh* part plus the *whP* form a pair that is in turn contrasted with another alternative pair (e.g., 35: <the fish card, the fish square> vs. <the parrot card, the parrot square>, etc.).³⁸ These examples show that fronting requires the *whP* to be an active part of the contrast, otherwise fronting does not occur. The following scenario and examples buttress the argument:

- (37) Dinner scenario: *Three guests: Marie, Paul and Guillaume. The host cooks the main course and asks the guests to bring three items: wine, dessert and cheese.*
- (38) Question: *qui a apporté quoi?* ‘who brought what?’. If we take the guests to be the sorting key, this question is a “strategy” (in Roberts’s 1996 and Büring’s 2003 terms) that contains the following set of subquestions:
- a. What did Marie bring?
 - b. What did Paul bring?
 - c. What did Guillaume bring?

Let us now consider the next two subscenarios:

- (39) Subscenario 1: *The host said: “I need wine, dessert and cheese. Bring what you want.”*

The key point is that the three required items might not distribute complementarily across the guests, that is they might all bring wine for example. In this scenario, the subquestions (38a–c) can translate as (40), where *Marie, Paul* and *Guillaume* all bear a C-accent, that is, they are contrastive topics (CT).³⁹

- (40) a. Elle a apporté **quoi**, Marie?
 b. Il a apporté **quoi**, Paul?
 c. Il a apporté **quoi**, Guillaume?

In these (sub)questions, note that the *wh*- is *in-situ* despite the CT. Examples (40) thus illustrate that CT per se does not trigger *wh*-fronting. Let us now consider Subscenario 2:

³⁷On the basis of expressive contexts, Obenauer (1994:357) already suggested that in *ex-situ wh*-questions the referent is outside the domain Δ examined by the speaker, whereas *in-situ wh*-questions locate the variable in a non-empty domain Δ .

³⁸Actually, these questions can be described as subquestions belonging to the “strategy” *what is where?*, in Roberts’s (1996) and Büring’s (2003) terms.

³⁹See fn. 36. This fact contrasts with English, which does not allow for B-accent in questions, even on contrastive topics (Constant 2014).

- (41) Subscenario 2: *The host said: "I need wine, dessert and cheese. May each of you choose an item, so that we have everything for dinner."*

In this scenario, the subquestions (38a-c) can translate as (42), where *Marie*, *Paul* and *Guillaume* are again contrastive topics.

- (42) a. **Qu'est-ce qu'**elle a apporté, Marie ?
 b. **Qu'est-ce qu'**il a apporté, Paul ?
 c. **Qu'est-ce qu'**il a apporté, Guillaume ?

The difference between Subscenario 1 (*wh-in-situ*) and Subscenario 2 (*wh-ex-situ*) is that the three items (wine, dessert and cheese) are mutually exclusive in the latter only. They are in complementary distribution: If Marie brings wine, Paul cannot bring wine and he has to bring cheese or dessert. The contexts for (40) and (42) thus disentangle Contrast on the non-*wh*-part of the question and Contrast on the *wh*-item. Although Contrast on the non-*wh*-part favors Contrast on the *wh*-item, it does not necessarily trigger *wh*-fronting (40), whereas Contrast on the *wh*-item does (42) because the items underlying the *wh*- (wine, dessert and cheese) are in mutually exclusive distribution.

To make it clearer, here is another situation where *wh*-fronting is available in simple questions with no contrast on the *non-wh* part. Consider context (43) and questions (44) and (45):

- (43) A: "At work, I had a computer issue. I had to go to Marie, Paul or Guillaume to solve it."
 (44) B: "Ah oui? Et **qui** t' as vu, finalement?"
 really and who you have seen eventually
 (45) B: ??#"Ah oui? Et t' as vu **qui**, finalement?"
 really and you have seen who eventually

B has to select one of three individuals and does not know which one. S/he asks the question in order for A to perform this operation for him/her. In this context, (45)⁴⁰ with the prosody described for (4aα) is felt to be degraded. In contrast, (44) is perfectly natural, which shows that *wh-ex-situ* questions are optimal when there is a selection in a set, EXCLUDING the rest of the set.

4.4 Intermediate summary

In Sect. 4, we observed that *wh*-fronting occurs when:

- 1) There is a one-to-one mapping from one set onto another set (4.1, 4.3).
- 2) There is a potential contrast between two (sets of the) possible referents of the *whP* (4.2).
- 3) The speaker's question implies that the addressee can select one item only from a set, and hence has to exclude the rest of the set in order to answer the question (4.3, ex. 44).

⁴⁰As repeatedly said above, (45) is also possible if *qui* is stressed, since we are in a partitive context in the sense of Baunaz (2011, 2016) (see fn. 7).

Section 5 will aim to provide a formalization that captures the above observations.

5 Formalizing an exclusivity operator

In the previous section, we highlighted that contrast seems to be the hallmark of *wh-ex-situ* in Colloquial French and that it is thus absent in (unstressed) *wh-in-situ*. Section 5 will address the matter of what it means for a *wh*-item to be contrasted and will attempt to formalize this contrastive feature. We shall review two hypotheses: 1) The feature is a Contrastive Topic feature, 2) The feature is an idiosyncratic feature, and elaborate on the latter.

5.1 Hypothesis 1: Contrastive Topic

A brief presentation of Rooth's (1985, 1992) semantics for focus and Constant's (2014) semantics for contrastive topics (henceforth CT) is in order before we concentrate on CT in *wh*-questions.

5.1.1 Rooth's semantics for focus and Constant's semantics for Contrastive Topics

The crucial idea in Rooth (1985, 1992) is that focus marking on a phrase yields a set of alternatives (focus is indicated with the subscript F). To see how these alternatives work, let us consider (46).

- (46) Marie [est venue]_F.
Marie has come

A sentence like (46) has two semantic values corresponding to two levels of interpretation. For its ordinary semantic value, the meaning of the sentence obtains via the usual rules of composition. Thus, (46) has the meaning (47) (ignoring tense and intensions). The focus semantic value (noted F) is obtained by generating a set of propositions that includes the asserted proposition and all the propositions that can be obtained by substituting the possible alternatives for the focused item, thus yielding (48).

- (47) $[[\text{Marie [est venue]}_F]]^\circ = \text{come}(m)$
 (48) $[[\text{Marie [est venue]}_F]]^F = \{\lambda w.P(m)(w)|P_{\langle e, \langle s, t \rangle \rangle}\} = \{\text{Marie came; Marie left; Marie smiled}\}$

Let us now see what happens if some phrase in the sentence is additionally CT-marked. According to Constant (2014), sentences with a CT contain an operator responsible for the movement of the phrase that bears the CT feature, movement which triggers a λ -abstraction (much like Quantifier Raising). (49) gives the semantics of this operator and (50) a simple example.

- (49) a. $[[\text{CT-}\lambda i \varphi]]^\circ_g = \lambda x. [[\varphi]]^\circ_{g[i \rightarrow x]}$ (ordinary semantic value)
 b. $[[\text{CT-}\lambda i \varphi]]^F_g = \{\lambda x. [[\varphi]]^F_{g[i \rightarrow x]}\}$ (focus semantic value)

(50) [Marie₄]_{CT} CT-λ₄ t₄ [est venue]_F.

Like Rooth (1985, 1992), Constant (2014) assumes a two-tier semantics. The first tier is the ordinary semantics in (49a). At this level, (50) means *come(m)*, much like (47). The second tier is the level where the foci (including CT) are evaluated (49b). In the example, two phrases bear a focus feature: the DP *Marie* and the VP *est venue*. Combining the focus semantic value of *est venue* (see 48) with the operator gives (51b), which can then combine with the focus semantic value of *Marie* (say, the set {*Marie*, *Pierrette*}) by pointwise functional application, yielding the set of sets of propositions (51c).⁴¹

- (51) a. $[[[(50)]]^\circ = come(m)$
 b. $[[[CT-\lambda_4 t_4 [est venue]_F]]^F = \{\lambda x. \{x \text{ came; } x \text{ left; } x \text{ smiled}\}\}$
 c. $[[[(50)]]^F = \{\{Marie \text{ came; } Marie \text{ left; } Marie \text{ smiled}\}, \{Pierrette \text{ came; } Pierrette \text{ left; } Pierrette \text{ smiled}\}\}$

5.1.2 A CT-feature on *wh*?

Let us now move on to *wh*-questions. According to Rooth (1992) and Vallduví and Vilkuna (1998), *wh*-items are intrinsically endowed with a focus feature that “generates a ‘*wh*-set,’ a domain over which they ‘quantify’” (Vallduví and Vilkuna 1998:86). Consequently, the meaning of a question is the set of propositions that constitutes the possible answers (Hamblin 1973). In order to understand its mechanics, let us consider question (52a) and its LF (52b). Let us moreover assume that the denotation of *who* in *w* is the set of individuals given in (53) and that the denotation of *you saw* (ignoring tense) is as (54). By pointwise functional application of (54) to (53), we obtain (55), the meaning of (52). Note that (55) is a set of propositions, much like (48) the focus semantic value of (46).

- (52) a. Who did you see?
 b. Q who_i did you see t_i ?
 (53) $[[who]]^w = \{Guillaume; Paul; Marie\}$
 (54) $[[you \text{ saw}]]^w = \lambda x_e. \text{you saw } x (w)$
 (55) $[[Who \text{ did you see}]]^w = \{\text{you saw Guillaume; you saw Marie; you saw Paul}\}$

Now, we saw abundant evidence in the previous sections that some *wh*-items are contrastive while others are not.⁴² Consequently, we propose that the *wh* may be both focused and contrastive in some questions. If, as we saw, focus is responsible for questions denoting sets of propositions, contrast (another type of focus) would make them a set of sets of propositions. This idea was tentatively entertained as a theoretical possibility in Constant (2014:112–113), who hypothesized that *wh*-items could be endowed with a CT feature on top of their focus feature. Under this hypothesis, (52)

⁴¹The result is close to Buring’s (2003), but is syntactically anchored and achieved compositionally.

⁴²For independent evidence on a distinction between contrast and focus, on both the prosodic and the interpretive sides, see Erteschik-Shir (2007), Katz and Selkirk (2011), É. Kiss (1998); Kratzer and Selkirk (2010), Lee (1999), Molnár (2002), Rochemont (1986), Selkirk (2008) and Vallduví and Vilkuna (1998).

has two possible LFs. The first one is (52b) with the meaning (55) and the second one is (56a) with the meaning (56b):

- (56) a. Q who_1 CT- λ_1 you saw t_1 .
 b. $[[\text{Q who}_1 \text{ CT-}\lambda_1 \text{ you saw } t_1]]^F = \{\{\text{you saw Guillaume}\}, \{\text{you saw Marie}\}, \{\text{you saw Paul}\} \dots\}$
 \approx For each person, did you see him/her?

Under interpretation (55), (52) will translate in French as the *wh-in-situ* question (57), whereas under interpretation (56), it will take the form of the *wh-ex-situ* question (58) (for a context in which such a question can be used, see the discussion around 44).

- (57) T' as vu **qui**?
 you have seen who
- (58) **Qui** t' as vu?
 who you have seen

One advantage of this hypothesis is that the CT feature exists independently from questions. French CT occupies a position above IP, arguably the same as in English in (59).

- (59) a. What did you give to Marie and Paul?
 b. [À Marie]_{CT}, j' ai donné un livre, [à Paul]_{CT} j' ai donné un CD.
 to Marie I have given a book to Paul I have given a CD

Nevertheless, though appealing, there are several reasons to be suspicious about this idea of a *wh*-item marked as CT. First, we have no independent evidence that a CT feature can be assigned to a *wh*-item in a question in French, although we know that an element of the non-*wh* part of the question can receive such a feature (see 35, 36, 40 and fn. 36).

Second, CT can be assigned to multiple elements in a French assertive clause, as shown in (60), which is not the case for questions. Otherwise, we would expect multiple *wh*-fronting, even if it can be banned for independent reasons.

- (60) a. When did you give the book and the CD to Marie and Paul?
 b. [À Marie]_{CT}, [le livre]_{CT} je l' ai donné lundi,
 to Marie, the book I it have given Monday
 [à Paul]_{CT} [le CD]_{CT} je l' ai donné mercredi.
 to Paul the CD I it have given Wednesday

The third (related) objection is that the instantiation of the *wh* in the answer cannot be a CT, but must be focus. (58) cannot be felicitously answered with (61) (which we indicate with #). Note that a sentence like (61) is felicitous in certain circumstances, notably when it is an instance of 'Lone CT' (Constant 2014 contra Wagner 2012).

- (61) # [Marie]_{CT} je l' ai vue.
 Marie I her have seen

To conclude this section, although the idea of imposing a CT feature on the *wh*-item in French *wh-ex-situ* is appealing, it cannot be fully probed, and runs into too many objections. Instead, we shall pursue the second hypothesis, namely that there is an idiosyncratic feature on Colloquial French *wh-ex-situ* items.

5.2 Hypothesis 2: Exclusivity

In the previous section, we tackled the idea that *wh*-items are intrinsically focused and can be contrastive on top of that, in a Hamblin semantics for questions, in which *wh*-items are variables. However, our findings in Sect. 3 showed that Colloquial French displays covert *wh*-movement, which points more towards *wh*-items being quantifiers than variables. This is why we adopt Karttunen's (1977) semantics for questions, in which *wh*-words are treated as existential quantifiers and *wh*Ps as generalized quantifiers of type $\langle\langle e, t \rangle, t \rangle$. In this framework, a question also denotes a set of propositions, but this set is created by the question operator and not by the *wh*-variable.

Before we proceed, it is important to note that the operator we are going to discuss is not a question operator. It is a separate, focus operator that feeds the question operator in the same way as the operator *que* 'only' applies in (62). This does not mean that *que* or our operator interact in a neutral way with the problem of whether the question is weakly exhaustive, strongly exhaustive, or non-exhaustive,⁴³ but they are clearly distinct (see the discussions in 5.3.2 and 5.3.3).

- (62) T' as été que où?
 you have been only where
 'Where have you only been?'

Building on the results in Sect. 4, (58) roughly means "Which x , y and $z \in \{\text{Marie; Paul; Guillaume}\}$ are such that you saw x , but not y and z ." We propose that this meaning can arise through composition with a contrastive operator that overtly attracts the *wh*. Since the role of this operator is to exclude alternatives, we dub it Exclusivity (henceforth [Excls]). It is also more precise and has the advantage of avoiding the overused term *contrast*. (63) provides a formalization for this Exclusivity operator:⁴⁴

- (63) Exclusivity operator:
 $[[\text{Excls}]] = \lambda P_{\langle \tau, \langle s, t \rangle \rangle} . \lambda z_{\tau} . \lambda w'_{s'} . \exists y \in D_{\tau}^c [y \neq z \wedge \neg P(y)(w')]. P(z)(w')$.
 Where D_{τ}^c represents the set of the contextually relevant items of type τ .

The operator is polymorphic since questions can bear on items of any type τ . It says that the property obtained once we have abstracted over the IP is true of a referent to the exclusion of some other(s). The first, presupposed part of the formula ($\exists y \in D_{\tau}^c [y \neq z \wedge \neg P(y)(w')]$) requires further discussion.

⁴³Unless we transfer to an operator below the question operator part of the burden attributed to it in the discussions around weak/strong/absence of exhaustivity in questions. This research program is beyond the scope of this article, but see the commentary to the answer-set (74).

⁴⁴We follow Heim and Kratzer's (1998) convention in placing the presuppositional part of the formula between ':' and '.' (to be discussed below).

5.3 Presuppositions in questions

This section explores the presuppositional status of the subpart $\exists y \in D_{\tau}^c[y \neq z \wedge \neg P(y)(w')]$. First, we unfold our approach in an answer-based theory of presuppositions in questions (5.3.1). Second, we motivate the existential quantifier and show that it implies that Exclusivity is distinct from Exhaustivity (5.3.2). Third, we develop the answer-set approach and the predictions [Excls] makes on a specific example to better illustrate its behavior (5.3.3).

5.3.1 Favoring an answer-based approach

We formalized presupposition as a definedness condition on questions, making use of Heim and Kratzer's (1998) convention. Presuppositions in questions are far less studied than in declarative sentences. There are two families of approaches, either answer-based or question-based (phrasing as in Fitzpatrick 2005):

- (64) Answer-based approach:
A presupposition of a question is something that is entailed by every possible answer to it. (Keenan and Hull 1973)
- (65) Question-based approach:
A presupposition of a question is a necessary condition for a successful interrogative speech act. (Katz 1972)

While the latter posits that a question can inherit the presupposition from its constituents, the former assumes that no presupposition projects in questions, but that a question is infelicitous only if none of its answers is defined. Here is Guerzoni's (2003) Question Bridge Principle (based on Stalnaker 1978):

- (66) Question Bridge Principle:
A question is felicitous in c ONLY IF it can be felicitously answered in c (i.e., if at least one of its answers is defined in context c).

Although both could be necessary,⁴⁵ the answer-based approach is less redundant because in most cases the question-based approach predicts two reasons for the infelicity of the question: The question is undefined AND its possible answers are undefined (Guerzoni 2003, here simplified a lot). For this reason, we follow answer-based approaches, although our results are in principle harmlessly translatable to the other framework.

5.3.2 Existential quantification and Exhaustivity

Regarding the quantifier in $\exists y \in D_{\tau}^c[y \neq z \wedge \neg P(y)(w')]$, existentiality must be posited for both logical and empirical reasons. First positing a universal quantifier

⁴⁵See Fitzpatrick (2005) on *how-come* and *why*-questions. To our knowledge, a detailed study of this phenomenon and whether the various presupposition triggers have the same effect is still to be done.

would result in a contradiction.⁴⁶ Assuming the set *D* of three individuals {Guillaume; Paul; Marie}, if each proposition about one *x* in *D* in the question denotation of *Who did you see?* is defined only if some $y \neq x$ in *D* was not seen, the projection must be existential. Indeed, if it were universal we would end up presupposing for each *x* in *D* that there is a $y \neq x$ in *D* such that *y* was not seen, which is contradictory. Second, an answer like (67) is perfectly felicitous for (58) because (67) displays an exclusion, but not of all the alternatives in the set (55), as shown by the possibility of remaining agnostic about Guillaume.⁴⁷ Consequently, [Excls] implies that *at least one* element must be excluded.⁴⁸

- (67) J'ai vu Marie, j'ai pas vu Paul, mais je me souviens pas pour Guillaume.
'I saw Marie, I didn't see Paul, but regarding Guillaume, I cannot remember.'

A note is in order here on how Exclusivity relates to Exhaustivity because the latter can also involve exclusion. On the one hand, a question like (52) *Who did you see?* can be answered exhaustively in two ways, assuming that the speaker saw Marie: (68a) is the weakly exhaustive answer and (68b) is the strongly exhaustive one.⁴⁹

- (68) a. I saw Marie.
b. I saw Marie, but not Paul and Guillaume.
c. I saw Marie, but not Paul.
d. I saw Marie, but not Guillaume.
e. I saw Marie, Paul and Guillaume.

On the other hand, (68c) and (d) are exclusive but partial, non-exhaustive answers. Finally, if the correct answer is (68e), it is exhaustive but not exclusive (all three accessible individuals were seen, none being excluded). Consequently, Exclusivity does not imply Exhaustivity and Exhaustivity does not imply Exclusivity. They are two distinct operations. We argue that this applies to declarative sentences, as illustrated in (68), and questions. Put otherwise, the structure under examination here, *wh-ex-situ*, is not a mark of exhaustivity in Colloquial French questions.

A confirmation comes from a close, but distinct construction, namely clefts. One (frequent) means to mark exhaustivity in French questions is to cleft the *wh*-item, as illustrated in (69a) (Rouquier 2014). (69b) then shows that Exhaustivity and [Excls] are separate features and that both can aggregate to form an interrogative

⁴⁶This was pointed out to us by an anonymous reviewer.

⁴⁷Part of the discussion here was inspired by Yabushita's (2017) discussion on contrastive operators. Note that (55), which serves as a baseline for the present discussion, is provisional and is revised in (72)/(74).

⁴⁸See Hara (2006), Lee (1999, 2017), and Molnár (2002) on the 'at least' condition.

⁴⁹A weakly exhaustive answer contains all the positive answers to a question, whereas a strongly exhaustive answer contains both the positive and the negative answers to a question (Beck and Rullman 1999; George 2011; Groenendijk and Stokhof 1984; Heim 1994; Sharvit 2002, among others). To account for exhaustive answers, Heim (1994) and Dayal (2016) designed specific answerhood operators. Alternatively, one can imagine a covert *wh-only* operator that is below $C_{[+WH]}$ and accounts for exhaustivity (Nicolae 2015).

cleft with additional *wh*-fronting.⁵⁰ The combination of Exhaustivity and Exclusivity yields (70), the answer-set to (69b), to be compared to (74) (without Exhaustivity) and (72) (without Exhaustivity and Exclusivity).

- (69) a. [C' est **qui** que] t' as vu?
it is who that you have seen
- b. [**Qui** c' est que] t' as vu?
who it is that you have seen
- (70) [[Who Excls did you see]]^w = {you saw Guillaume but not Marie and Paul; you saw Marie but not Guillaume and Paul; you saw Paul but not Marie and Guillaume; you saw Guillaume and Marie, but not Paul; you saw Guillaume and Paul but not Marie; you saw Marie and Paul but not Guillaume}

5.3.3 *In-situ* and *ex-situ* Answer-sets

To illustrate the effect of the presupposition $\exists y \in D_r^c [y \neq z \wedge \neg P(y)(w')]$, let us reuse scenario (43), and questions (44) and (45), repeated here and modified to (43'), (44') and (45') (underlined part added, judgments reassessed):

- (43) A: "At work, I had a computer issue. I had to go to Marie, Paul or Guillaume to solve it."
- (44) B: "Ah oui? Et **qui** t' as vu, finalement?"
really and who you have seen eventually
- (45) B: ??#"Ah oui? Et t' as vu **qui**, finalement?"
really and you have seen who eventually
- (43') A: "At work, I had a computer issue. I had to go to Marie, Paul or Guillaume, or the three of them to solve it."
- (44') B: ??#"Ah oui? Et **qui** t' as vu, finalement?"
really and who you have seen eventually
- (45') B: "Ah oui? Et t' as vu **qui**, finalement?"
really and you have seen who eventually

We have to clarify why *in-situ* (45') is better than *ex-situ* (44') in the latter scenario, whereas *ex-situ* (44) is better than *in-situ* (45) in (43). To explain this observation, let us consider the answer-sets corresponding to (44)/(44') and (45)/(45'), starting with (45)/(45'). (55) was a first approximation, not taking into account that *qui* 'who' is ambiguous between singular and plural. First, we assume here

⁵⁰Zumwald Küster (2018:105) proposes that the third form with additional *est-ce* inversion (e.g., *qui est-ce que*) is not used as a clefting device but as an unanalyzed chunk that allows the speaker to dispense with subject-verb inversion (*qui est-ce que t'as vu?* vs *qui as-tu vu?*).

that the correct answer is selected by Dayal's (1996) answerhood operator (71), which picks up the maximally informative answer in the answer-set, that is (72) for (45)/(45').

- (71) $\text{Ans}(Q) = \lambda w. \iota p [p_w \wedge p \in Q \wedge \forall p' [[p'_w \wedge p' \in Q] \rightarrow p \subseteq p']]$
 (72) $[[\text{Who did you see}]]^w = \{\text{you saw Guillaume; you saw Marie; you saw Paul; you saw Guillaume and Marie; you saw Guillaume and Paul; you saw Marie and Paul; **you saw Guillaume, Marie and Paul**}\}$

Second, to construe the answer-set to (44)/(44'), let us consider answers like (73a).

- (73) a. Excls J' ai vu Marie.
 I have seen Marie
 b. Marie [Excls [IP j'ai vu t.

To be felicitous, (73a) must mean 'I saw Marie and there is someone else (either Guillaume or Paul) that I did not see.' Put otherwise, *Marie* is attracted to the specifier of Excls°. (73b) is (73a)'s LF. Note that *Marie* is *in situ* in (73a) whereas *qui* is *ex-situ* in the corresponding question. But in the context of (44) *Marie* carries the rise-fall accent marking specificity that the out-of-the-blue utterance would not display (see the analysis of example 4b and fn. 7). The reason why questions have both the options of fronting and accenting, whereas only accenting is available to the answer is left to future research. The answer-set corresponding to (44)/(44') is thus (taking the exclusion "but not" into account, boldface and underlining explained below):

- (74) $[[\text{Who Excls did you see}]]^w = \{\text{**you saw Guillaume but not Marie; you saw Guillaume but not Paul**; you saw Guillaume but not Marie and Paul; you saw Marie but not Guillaume; you saw Marie but not Paul; you saw Marie but not Guillaume and Paul; you saw Paul but not Marie; you saw Paul but not Guillaume; you saw Paul but not Marie and Guillaume; you saw Guillaume and Marie, but not Paul; you saw Guillaume and Paul but not Marie; you saw Marie and Paul but not Guillaume}\}$

The two answer-sets being in place (72 to the *in-situ* question and 74 to the *ex-situ* question), let us come back to Scenarios (43) and (43'), starting with (43).

Scenario (43) implies that the speaker did not meet all three individuals, and hence cannot entail the proposition *you saw Guillaume, Marie and Paul*. This latter proposition being part of the answer-set (72) to the *in-situ* question (45), this question is not optimal in this context because it includes an answer that is undefined. Conversely, a question like (44) featuring [Excls] will be felicitous. All the propositions in its answer-set (74) are defined in context (43). Suppose now that Guillaume is the only person that the speaker saw. In this frame, the bolded and underlined answers in (74) are the correct answers to *ex-situ* (44). Note that if the question is also exhaustive, the correct answer will be the underlined one only.

In contrast, in the frame of (43'), A's utterance entails that the proposition 'you saw Guillaume, Marie and Paul' is a contextually possible answer, that is, there is

a possible answer that is in the denotation of the *in-situ* question (45'), but not of the *ex-situ* question (44'). Consequently *in-situ* (45') comes out as more optimal than *ex-situ* (44').

Note that in the frame of (43), uttering *in-situ* (45) is not impossible but would sound like a presupposition cancellation; that is, the pre-construed answer-set, which is narrower and more specific than *in-situ* (45), will be enriched with the proposition 'you saw Guillaume, Marie and Paul.' Conversely, *ex-situ* (44') is not impossible, but sounds bizarre because the pre-construed answer-set of possibilities in the frame of (43') is larger than what (44') denotes (it includes 'you saw Guillaume, Marie and Paul'). Uttering (44') then has the effect of narrowing down the pre-construed answer-set. In a nutshell, depending on the contexts, either *in-situ* or *ex-situ* questions will sound degraded, because a pragmatic adjustment will be required. Nevertheless, because this operation is available, the judgments are never as clear-cut as in the case of a presupposition cancellation like (75), where there is a contradiction between the presupposition 'Jean smoked' and b's utterance.

- (75) a. Jean il a arrêté de fumer. b. #En fait, il a jamais fumé.
'Jean stopped smoking. Actually, he has never smoked.'

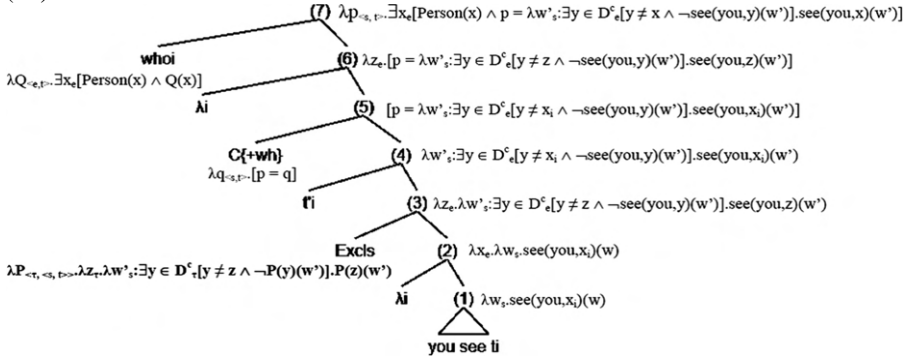
Finally, imagine that *ex-situ* (44) is answered with (76a) (which does not belong to its answer-set). Here again an adjustment is necessary and (76a) sounds like (76b), with the corrective *même* 'even' in the context of (43). Likewise, this adjustment is not as sharp as a presupposition cancellation. This comes as natural in the answer-based approach to questions presuppositions (64), since in this frame (44)'s presupposition is only pragmatic, i.e., construed on the pre-construed answer-set (74). More research is needed however to better understand how this transfer between the answers and the question works exactly.

- (76) a. J' ai vu Marie, Paul et Guillaume.
I have seen Marie Paul and Guillaume
b. J' ai vu Marie, Paul et même Guillaume.
I have seen Marie Paul and even Guillaume

5.4 Representing the Exclusivity operator

Figure (77) shows how this operator works in example (44). [Excls] is placed above IP to account for the landing site of the *ex-situ whP*. If it were not for this surface position, other locations could have been envisaged, like above vP. Let us note that $\exists y \in D_c^c[y \neq x \wedge \neg \text{see}(\text{you}, y)(w')]$ is embedded as expected in an answer-based approach to presuppositions in questions (it does not project and is therefore not a definedness condition for the question itself, but rather for its answers).

(77)



Remarks:

- We ignore tense.
- We set up the operator to type e .
- We treat the traces as free variables of type e , which we represent with an x bearing the index of their binder.
- We follow Karttunen’s (1977) semantics for questions revised as in Dayal (2016). In particular, we do not claim that questions denote the set of their true answers but only of their possible answers, as in Hamblin (1973).
- The *wh*-item undergoes two movements, first to the [Excls] operator’s specifier, second (covertly) to the interrogative complementizer $C[+wh]$ ’s specifier.
- The appended λ (under nodes 2 and 6) represent the abstraction triggered by these movements (see coindexation).

The steps of the composition are as follows (numbers refer to 77):

- (2) $\lambda x_e.\lambda w_s.see(you,x_i)(w)$ (predicate abstraction)
- (3) $[[[Excls]]] ([[2]]) = \lambda z_e.\lambda w'_s.\exists y \in D^c_e[y \neq z \wedge \neg see(you,y)(w')].see(you,z)(w')$ (functional application)
- (4) $[[[3]]] ([[t_i]]) = \lambda w'_s.\exists y \in D^c_e[y \neq x_i \wedge \neg see(you,y)(w')].see(you,x_i)(w')$ (functional application)
- (5) $[[[C[+wh]]]] ([[4]]) = [p = \lambda w'_s.\exists y \in D^c_e[y \neq x_i \wedge \neg see(you,y)(w')].see(you,x_i)(w')]$ (functional application)
- (6) $\Rightarrow \lambda z_e.[p = \lambda w'_s.\exists y \in D^c_e[y \neq z \wedge \neg see(you,y)(w')].see(you,z)(w')]$ (predicate abstraction)
- (7) $[[[whoi]]] ([[6]]) = \exists x_e[Person(x) \wedge p = \lambda w'_s.\exists y \in D^c_e[y \neq x \wedge \neg see(you,y)(w')].see(you,x)(w')]$ (functional application)

The outcome is in (78) (abstraction over the free variable p and creation of a set of propositions/a question). In words, (44) means that the speaker is asking the hearer which person the hearer has seen, presuming that there is also at least one person that the hearer could have seen but has not seen.

(78) $\lambda p_{<s,t>}. \exists x_e[Person(x) \wedge p = \lambda w'_s.\exists y \in D^c_e[y \neq x \wedge \neg see(you,y)(w')].see(you,x)(w')]$

In this section, we proposed a formalization of the observations presented in Sect. 4 on the contrast that drives overt *wh*-fronting in Colloquial French. Trying to import a CT feature, we found out that contrast in questions is better viewed as a presupposition triggered by an Exclusivity operator that excludes at least one of the possible answers.

6 Conclusions

This article examined *in-situ* and *ex-situ wh*-questions in Colloquial French, a language that undoubtedly displays both structures, as shown in Sect. 2. The aims were twofold: establish which *wh*-position is the unmarked, default one, and describe the characteristics of the other, marked *wh*-position.

In Sect. 3, we showed that (a subset of) *wh-in-situ* questions are sensitive to constraints on *wh*-movement (i.e., weak-islandhood and Superiority), which indicates that *in-situ* questions are cases of unmarked questions with covert *wh*-movement. We also examined *wh-ex-situ* questions and showed that they are insensitive to these constraints, which makes them instances of a special type of question, where *wh*-fronting is not driven by a *wh*-feature, but by another feature. Consequently, we suggested that Colloquial French is a *wh-in-situ* language and that *wh-ex-situ* is *wh*-fronting, not *wh*-movement. These results provide arguments against the optionality stance.

In Sect. 4, we examined which feature could be triggering *wh*-fronting in Colloquial French, and we showed that it is a type of contrast, namely Exclusivity, which implies that at least one of the possible answers is excluded, thus putting the focus on the correct ones. Interestingly, this exclusive/contrastive feature, although cognate to well-known features like Contrastive Topic, is not reducible to them. This way, we have contributed to the debate on information structure in questions (Beysade 2006; Constant 2014; Engdahl 2006) by showing that phenomena like focus or contrast need to be adjusted to be able to fit in with the semantics of questions, sometimes giving rise to specific pragmatic effects. In Sect. 5, we formalized the Exclusivity operator [[Excls]] as a definedness condition (63).

Finally and more generally, this article opens a program of research on the varieties of questions across languages. First, it might be interesting to consider other *wh-in-situ* languages that also display *wh-ex-situ* questions (see Sect. 1) in light of the Exclusivity operator defined in Sect. 5. It is promising to note that (Cheung 2008) argues that *wh*-fronting marks contrastive focus in Chinese, a result close to ours.

Another question concerns pair-list readings, which most speakers highly prefer for *wh-ex-situ* multiple questions (Sect. 3.2), whereas single-pair readings are freely accessible for *wh-in-situ* ones. Moreover, the strong-island facts in (16) showed that in LF both *wh*-words have to be in CP in *ex-situ* multiple *wh*-questions much like what is required to have a pair-list reading (Dayal 2003, 2006; Kitagawa et al. 2004; Kotek 2017). Yet, it is difficult to derive this reading from our Exclusivity operator, unless the contrast the operator involves just favors it. More work is definitely needed to assess this hypothesis.

Embedded questions also challenge the present analysis because they supposedly never feature *wh-in-situ* (Bobaljik and Wurmbrand 2015), although they can have

the semantics of *wh-in-situ* direct questions (see Baunaz 2011:227, fn. 213). Yet, recently-collected data show that *wh-in-situ* is gaining ground in embedded questions (Gardner-Chloros and Secova 2018; Poletto and Pollock to appear). Further research could establish whether embedded *in-situ* introduces an alternation equivalent to what is found in direct questions.

Finally, the *ex-situ/in-situ* alternation is not limited to *wh-in-situ* languages. For example, English has at least three varieties of *wh*-questions: *ex-situ* with and without pied-piping, the latter being more expressive (79, 80, from Obenauer 1994); and *wh-in-situ* questions, especially when the *wh* is focused (81, from Bobaljik and Wurmbrand 2015).

(79) (*Well,) **on whose** help did you count?

(80) Well, **whose** help did you count **on**?

(81) A: All the creative people—our R&D, marketing, in-house ad staff—that’s all done here in Jersey.

B: But the sneaks are made **where**? China? Malaysia? [From Stabler, *Law and Order*]

More research is definitely needed in order to broaden the empirical coverage of the proposal on Exclusivity in Colloquial French and see whether it can also capture cross-linguistic facts.

Acknowledgements We would like to thank Diti Bhadra, Jonathan Bobaljik, Gennaro Chierchia, Cecilia Poletto, Jean-Yves Pollock, the audiences at the MIT ling/lunches and the Cambridge Romance Seminar, and the members of the Department of Linguistics at the Geneva seminar for useful comments and discussions. We are also grateful to three anonymous reviewers, two of which are native speakers of French and confirmed our data. All the remaining errors are ours.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abner, Natasha. 2011. WH-words that go bump in the right. In *West Coast Conference on Formal Linguistics (WCCFL) 28*, eds. Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz, 24–32. Somerville: Cascadia Proceedings Project.
- Abrusán, Marta. 2014. *Weak island semantics*. Oxford: Oxford University Press.
- Adli, Aria. 2006. French *wh-in-situ* questions and syntactic optionality: Evidence from three data types. *Zeitschrift für Sprachwissenschaft* 25: 163–203.
- Aoun, Joseph, Norbert Hornstein, and Dominique Sportiche. 1981. Some aspects of wide scope quantification. *Journal of Linguistic Research* 1(3): 69–95.

- Ashby, William J. 1977a. *Clitic inflection in French: A historical perspective*. Amsterdam: Rodopi.
- Ashby, William J. 1977b. Interrogative forms in Parisian French. *Semasia* 4: 35–52.
- Ashby, William J. 1981. The loss of the negative particle *ne* in French: A syntactic change in progress. *Language* 57: 647–687.
- Ashby, William J. 1984. The elision of // in French clitic pronouns and articles. In *Romanitas: Studies in Romance linguistics*, ed. Ernst Pulgram, 1–16. Ann Arbor: University of Michigan Press.
- Barra-Jover, Mario. 2004. Interrogatives, négatives et évolution des traits formels du verbe en français parlé. *Langue Française* 141: 110–125.
- Barra-Jover, Mario. 2010. «Le» français ou ce qui arrive lorsqu'un état de choses est observé comme une entité. *Langue Française* 168(4): 3–18.
- Baunaz, Lena. 2011. *The grammar of French quantification*. Dordrecht: Springer.
- Baunaz, Lena. 2016. French “quantifiers” in questions: Interface strategies. *Acta Linguistica Hungarica* 63(2): 125–168.
- Baunaz, Lena, and Cédric Patin. 2011. Prosody refers to semantic factors: Evidence from French *wh*-words. In *Actes de interface discours et prosodie 2009*, eds. Hiyon Yoo and Elisabeth Delais-Roussarie, 96–107.
- Beck, Sigrid. 1996. Quantified structures as barriers for LF movement. *Natural Language Semantics* 4: 1–56.
- Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14: 1–56.
- Beck, Sigrid, and Shin-Sook Kim. 1997. On *Wh*- and operator scope in Korean. *Journal of East Asian Linguistics* 6: 339–384.
- Beck, Sigrid, and Hotze Rullman. 1999. A flexible approach to exhaustivity in questions. *Natural Language Semantics* 7(3): 249–298.
- Beeching, Kate, Nigel Armstrong, and Françoise Gadet. 2009. *Sociolinguistic variation in contemporary French*. Amsterdam: Benjamins.
- Beysade, Claire. 2006. La structure de l'information dans les questions: quelques remarques sur la diversité des formes interrogatives en français. *Linx* 55: 173–193.
- Blanche-Benveniste, Claire. 1997. La notion de variation syntaxique dans la langue parlée. *Langue Française* 115: 19–29.
- Bobaljik, Jonathan D., and Susi Wurmbrand. 2015. Questions with Declarative Syntax tell us what about selection? In *50 years later: Reflections on Chomsky's aspects, MIT working papers in linguistics*, eds. Ángel J. Gallego and Dennis Ott, 13–31. Cambridge: MIT Press.
- Boeckx, Cedric. 2000. Decomposing French questions. In *23rd Annual Penn Linguistics Colloquium (PLC)*, eds. Jim Alexander, Na-Rae Han, and Michelle Minnick. Vol. 1 of *Working papers in linguistics*, 69–80. Philadelphia: University of Pennsylvania.
- Boeckx, Cedric, Penka Stateva, and Arthur Stepanov. 2001. Optionality, presupposition, and *wh*-in situ in French. In *Romance syntax, semantics and L2 acquisition*, eds. Joaquim Camps and Caroline R. Wiltshire, 57–72. Amsterdam: Benjamins.
- Bonan, Caterina. 2019. On clause-internally moved *wh*-phrases: *Wh*-to-foc, nominative clitics, and the theory of Northern Italian *wh*-in situ. PhD diss., Université de Genève.
- Bošković, Željko. 2000. Sometimes in [Spec, CP], sometimes in-situ. In *Step by step: Essays on minimalism in honor of Howard Lasnik*, eds. Roger Martin, David Michaels, and Juan Uriagereka, 53–87. Cambridge: MIT Press.
- Bošković, Željko. 2002. On multiple *wh*-fronting. *Linguistic Inquiry* 33(3): 351–383.
- Büring, Daniel. 2003. On D-trees, beans, and B-accent. *Linguistics and Philosophy* 26(5): 511–545.
- Chang, Lisa. 1997. *Wh*-in-situ phenomena in French, MA thesis, University of British Columbia.
- Cheng, Lisa Lai-Shen. 1991. On the typology of *Wh*-questions. PhD diss., MIT.
- Cheng, Lisa Lai-Shen, and Johan Rooryck. 2000. Licensing *wh*-in-situ. *Syntax* 3(1): 1–19.
- Cheung, Candice Chi-Hang. 2008. *Wh*-fronting in Chinese. PhD diss., University of Southern California, Los Angeles.
- Chomsky, Noam. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, eds. Stephen R. Anderson and Paul Kiparsky, 232–286. New York: Holt, Rinehart & Winston.
- Chomsky, Noam. 1976. Conditions on rules of grammar. *Linguistic Analysis* 2(4): 303–351.
- Chomsky, Noam. 1995. *The minimalist program. Current studies in linguistics*. Cambridge: MIT Press.
- Cinque, Guglielmo. 1990. *Types of A'-dependencies*. Cambridge: MIT Press.
- Constant, Noah. 2014. Contrastive topic: Meanings and realizations, PhD diss., University of Massachusetts.

- Coveney, Aidan. 1989. Pragmatic constraints on interrogatives in spoken French. *York Papers in Linguistics* 13: 89–99.
- Coveney, Aidan. 1995. The use of the QU-final interrogative structure in spoken French. *Journal of French Language Studies* 5(2): 143–171.
- Coveney, Aidan. 2002. *Variability in spoken French: A sociolinguistic study of interrogation and negation*. Bristol: Elm Bank.
- Coveney, Aidan. 2003. Le redoublement du sujet en français parlé: Une approche variationniste. In *Structures linguistiques et interactionnelles du français parlé*, eds. Anita Berit Hansen and Maj-Britt Mosegaard, 111–143. Copenhagen: Museum Tusulanum Press.
- Coveney, Aidan. 2011. A language divided against itself? Diglossia, code-switching and variation in French. In *Le français en contact: Hommages à Raymond Mougéon*, eds. France Martineau and Terry Nadasdi, 51–85. Québec: Presses de l'Université de Laval.
- Culbertson, Jennifer. 2010. Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86(1): 85–132.
- Dayal, Veneeta. 1996. *Locality in WH quantification: Questions and relative clauses in Hindi*. Dordrecht: Kluwer.
- Dayal, Veneeta. 2003. Two types of universal terms in questions. In *Semantics: Critical concepts in linguistics*, ed. Javier Gutiérrez-Rexach, 373–385. London: Routledge.
- Dayal, Veneeta. 2006. Multiple Wh-questions. In *The Blackwell companion to syntax*, eds. Martin Everaert and Henk Van Riemsdijk, 275–327. Malden: Blackwell.
- Dayal, Veneeta. 2016. *Questions*. Oxford: Oxford University Press.
- De Cat, Cécile. 2005. French subject clitics are not agreement markers. *Lingua* 115(9): 1195–1219.
- Dekhissi, Laurie, and Aidan Coveney. 2018. La variation dans l'emploi des interrogatives partielles dans le cinéma de banlieue. In *L'interrogative en français*, eds. Marie-José Béguélin, Aidan Coveney, and Alexander Guryev, 119–151. Berne: Peter Lang.
- Denham, Kristin. 2000. Optional wh-movement in Babine-Witsuwit'en. *Natural Language & Linguistic Theory* 18(2): 199–251.
- Déprez, Viviane. 2018. On en est où? Intonation, information structure and the Embedding Problem. Paper presented at the GEN-*wh*2018, Geneva, 16–18 January 2018.
- Déprez, Viviane, Kristen Syrett, and Shigeto Kawahara. 2012. Interfacing information and prosody: French *wh-in-situ* questions. In *Romance languages and linguistic theory 2010. Selected papers from Going Romance Leiden 2010*, eds. Irene Franco, Sara Lusini, and Andrés Saab, 135–154. Amsterdam: Benjamins.
- Déprez, Viviane, Kristen Syrett, and Shigeto Kawahara. 2013. The interaction of syntax, prosody, and discourse in licensing French *wh-in-situ* questions. *Lingua* 124: 4–19.
- Dryer, Matthew S. 2013. Position of interrogative phrases in content questions. In *The world atlas of language structures online*, eds. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/93>, last accessed 18 May 2020.
- É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2): 245–273.
- Engdahl, Elisabet. 2006. Information packaging in questions. *Empirical Issues in Syntax and Semantics* 6: 93–111.
- Erteschik-Shir, Nomi. 2007. *Information structure: the syntax-discourse interface*. New York: Oxford University Press.
- Ettxepare, Ricardo, and Myriam Uribe-Etxebarria. 2005. In-situ wh-phrases in Spanish: Locality and quantification. *Recherches linguistiques de Vincennes* 33: 9–34.
- Ferguson, Charles A. 1959. Diglossia. *Word* 15(1): 325–340.
- Fitzpatrick, Justin. 2005. The whys and how comes of presupposition and NPI licensing in questions. In *West Coast Conference on Formal Linguistics (WCCFL) 24*, eds. John Alderete, Chung-hye Han, and Alexei Kochetov, 138–145. Somerville: Cascadilla Proceedings Project.
- Gadet, Françoise. 2003. *La variation sociale en français*. Paris: Ophrys.
- Gardner-Chloros, Penelope, and Maria Secova. 2018. Grammatical change in Paris French: In situ question words in embedded contexts. *Journal of French Language Studies* 28: 181–207.
- George, Benjamin R. 2011. Question embedding and the semantics of answers. PhD diss., UCLA.
- Groenendijk, Jeroen, and Martin Stokhof. 1984. Studies on the semantics of questions and the pragmatics of answers. PhD diss., Amsterdam University.
- Guerzoni, Elena. 2003. Why even ask? On the pragmatics of questions and the semantics of answers, PhD diss., MIT.

- Haida, Andreas. 2008. The indefiniteness and focusing of question words. In *Semantic and Linguistic Theory (SALT) 18*, eds. Tova Friedman and Satoshi Ito, 376–393.
- Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of language* 10: 41–53.
- Hamlaoui, Fatima. 2010. Anti-givenness, prosodic structure and “intervention effects”. *The Linguistic Review* 27: 347–364.
- Hamlaoui, Fatima. 2011. On the role of phonology and discourse in Francilian French *wh*-questions. *Journal of Linguistics* 47(1): 129–162.
- Hara, Yurie. 2006. Implicature unsuspendable: Japanese contrastive *wa*. In *Proceedings of the 2004 Texas Linguistics Society Conference*, eds. Pascal Denis, Eric McCready, Alexis Palmer, and Brian Reese, 35–45. Somerville: Cascadilla Press.
- Heim, Irene. 1994. Interrogative semantics and Karttunen’s semantics for “know”. *Israeli Association for Theoretical Linguistics (IATL)* 9: 128–144.
- Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden: Blackwell.
- Hirschbühler, Paul. 1979. The syntax and semantics of *and WH*-constructions. PhD diss., Indiana University.
- Hoh, Pau-San, and Wen-yu Chiang. 1990. A focus account of moved *wh*-phrases at *s*-structure in Chinese. *Lingua* 81(1): 47–73.
- Huang C. T. James. 1982. Logical relations in Chinese and the theory of grammar. PhD diss., MIT.
- Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA-London: MIT Press.
- Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1): 3–44.
- Katz, Jerrold J. 1972. *Semantic theory*. New York: Harper & Row.
- Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4): 771–816.
- Kayne, Richard S. 1975. *French syntax. The transformational cycle*. Cambridge: MIT Press.
- Keenan, Edward L., and Robert D. Hull. 1973. The logical presuppositions of questions and answers. In *Präsuppositionen in Philosophie und Linguistik*, eds. Sándor János Petőfi and Dorothea Franck, 441–466. Frankfurt: Athenäum.
- Kitagawa, Yoshihisa, Dorian Roehrs, and Satoshi Tomioka. 2004. Multiple *Wh*-interpretations. In *Generative grammar in a broader perspective. 4th GLOW in Asia 2003*, ed. Huang-Jin Yoon, 209–234. Seoul: The Korean Generative Grammar Circle & Institute for Cognitive Science.
- Kluender, Robert, and Marta Kutas. 1993. Subadjacency as a processing phenomenon. *Language and cognitive processes* 8(4): 573–633.
- Kotek, Hadas. 2017. Dissociating intervention effects from superiority in English *wh*-questions. *The Linguistic Review* 34(2): 397–417.
- Kratzer, Angelika, and Elisabeth Selkirk. 2010. Distinguishing contrastive, new and given information. Paper presented at the Georgetown Linguistics Society (GLS) Conference, February 2010.
- Kratzer, Angelika, and Junko Shimoyama. 2017 [2001]. Indeterminate pronouns: The view from Japanese. In *Contrastiveness in information structure, alternatives and scalar implicatures*, eds. Chungmin Lee, Ference Kiefer, and Manfred Krifka, 123–143. Dordrecht: Springer.
- Kuno, Susumu, and Jane J. Robinson. 1972. Multiple *wh* questions. *Linguistic Inquiry* 3: 463–487.
- Lambrecht, Knud. 1981. *Topic, antitopic and verb agreement in non-standard French*. Amsterdam: Benjamins.
- Larrivé, Pierre. 2016. Les interrogatives in-situ sont-elles pragmatiquement marquées en français vernaculaire ? Investigation synchronique et historique. Paper presented at the Kolloquium Konstanz, June 2016.
- Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3): 335–391.
- Lasnik, Howard, and Mamoru Saito. 1992. *Move alpha: Conditions on its application and output*. Cambridge: MIT Press.
- Lee, Chungmin. 1999. Contrastive topic: A locus of the interface—evidence from Korean and English. In *The semantics/pragmatics interface from different points of view 1*, ed. Ken Turner, 317–342. Oxford-Amsterdam: Elsevier.
- Lee, Chungmin. 2017. Contrastive topic, contrastive focus, alternatives, and scalar implicatures. In *Contrastiveness in information structure, alternatives and scalar implicatures*, eds. Chungmin Lee, Ference Kiefer, and Manfred Krifka, 3–21. Dordrecht: Springer.
- Legendre, Géraldine, Jennifer Culbertson, Isabelle Barrière, Thierry Nazzi, and Louise Goyet. 2010. Experimental and empirical evidence for the status and acquisition of subject clitics and agreement

- marking in adult and child spoken French. In *Movement and clitics*, eds. Vincenc Torrens, Linda Escobar, Anna Gavarró, and Juncal Gutiérrez, 333–360. Newcastle: Cambridge Scholars Publishing.
- Manzini, M. Rita, and Leonardo M. Savoia. 2011. *Wh*-in situ and *wh*-doubling in Northern Italian varieties: Against remnant movement. *Linguistic Analysis* 37(1–2): 79–114.
- Marandin, Jean-Marie, Claire Beyssade, Elisabeth Delais-Roussarie, and Annie Rialland. 2004. Discourse marking in French: C accents and discourse moves. In *Handbook of French semantics*, eds. Francis Corblin and Henriëtte de Swart, 455–476. Stanford: CSLI.
- Martin, Philippe. 1975. Analyse phonologique de la phrase française. *Linguistics* 146: 35–68.
- Massot, Benjamin. 2008. Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie: arguments morphosyntaxiques. PhD diss., Université Paris 8, Saint-Denis.
- Massot, Benjamin. 2010. Le patron diglossique de variation grammaticale en français. *Langue Française* 168(4): 87–106.
- Massot, Benjamin, and Paul Rowlett. 2013. L'hypothèse d'une diglossie en France. *Journal of French Language Studies* 23(1).
- Mathieu, Eric. 2002. The Syntax of non-canonical quantification: a comparative study. PhD diss., University College London.
- Mathieu, Eric. 2004. The mapping of form and interpretation: The case of optional WH-movement in French. *Lingua* 114(9–10): 1090–1132.
- Mathieu, Eric. 2009. Les questions en Français: Micro et macro-variation. In *Le français d'Ici: Etudes linguistiques et sociolinguistiques de la variation*, eds. France Martineau, Raymond Mougeon, Terry Nadasdi, and Annie Tremblay, 37–66. Toronto: GREF.
- Mayr, Clemens. 2014. Intervention effects and additivity. *Journal of Semantics* 31: 513–554.
- Megerdooomian, Karine, and Shadi Ganjavi. 2000. Against optional *Wh*-movement. In *Proceedings of Western Conference on Linguistics (WECOL)* 12, 358–369.
- Molnár, Valéria. 2002. Contrast from a contrastive perspective. In *Information structure, discourse structure and discourse semantics*, eds. Ivana Kruijff-Korbyová and Mark Steedman, 99–114. Helsinki: Brill-Rodopi.
- Morin, Yves-Charles. 1979. La morphophonologie des pronoms clitiques en français populaire. *Cahier de linguistique* 9: 1–36.
- Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41(1): 35–82.
- Munaro, Nicola, Cecilia Poletto, and Jean-Yves Pollock. 2001. Eppur si muove! On comparing French and Bellunese *Wh*-movement. In *Linguistic Variation Yearbook 1*, eds. Pierre Pica and Johan Rooryck, 167–180. Amsterdam: Benjamins.
- Munaro, Nicola, and Jean-Yves Pollock. 2005. Qu'est-ce-que (*qu*)-est-ce que? A case study in comparative romance interrogative syntax. In *The Oxford handbook of comparative syntax*, eds. Guglielmo Cinque and Richard S. Kayne, 542–606. New York: Oxford University Press.
- Myers Lindsay Lee. 2007. WH-interrogatives in spoken French: A corpus-based analysis of their form and function. PhD diss., University of Texas, Austin.
- Nicolae, Andreea C. 2015. Questions with NPIs. *Natural Language Semantics* 23(1): 21–76.
- Obenauer, Hans Georg. 1994. Aspects de la syntaxe A-barre: effets d'intervention et mouvements des quantificateurs. Doctorat d'Etat, Université Paris 8, Saint-Denis.
- Özsoy, Sumru A. 2009. Turkish as a (non)-*wh*-movement language. In *Turcological letters to Bernt Brendemoen*, eds. Eva. A. Csató, Gunvald Ims, Joakim Parslow, Finn Thiesen, and Emel Türker, 221–232. Oslo: Novus Forlag.
- Palasis, Katerina. 2013. The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies* 23(1): 17–35. Special Issue 'L'hypothèse d'une diglossie en France', Massot, Benjamin & Paul Rowlett.
- Palasis, Katerina. 2015. Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua* 161: 125–143.
- Palasis, Katerina, Richard Faure, and Frédéric Lavigne. 2019. Explaining variation in *wh*-position in child French: A statistical analysis of new seminaturalistic data. *Language Acquisition* 26(2): 210–234.
- Pesetsky, David. 1987. WH-in-situ: Movement and unselective binding. In *The representation of (in)definiteness*, eds. Eric Reuland and Alice ter Meulen, 98–129. Cambridge: MIT Press.
- Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge: MIT Press.
- Pohl, Jacques. 1965. Observations sur les formes d'interrogation dans la langue parlée et dans la langue écrite non-littéraire. *Actes du Xe Congrès International de Linguistique et de Philologie Romanes* 2: 501–512.

- Pohl, Jacques. 1975. L'omission de *ne* dans le français contemporain. *Le français dans le monde* 11: 17–23.
- Poletto, Cecilia, and Jean-Yves Pollock. 2015. Arguing for remnant movement in Romance. In *Remnant movement*, ed. Günther Grewendorf, 135–178. Berlin: De Gruyter.
- Poletto, Cecilia, and Jean-Yves Pollock. to appear. Remnant movement and smuggling in some Romance interrogative clauses. In *Smuggling in syntax*, eds. Adriana Belletti and Chris Collins. Oxford: Oxford University Press.
- Quillard, Virginie. 2000. Interroger en français parlé: Etudes syntaxique, pragmatique et sociolinguistique. PhD diss., Université de Tours.
- Reinhardt, Janina. 2019. Regularity and variation in French direct interrogatives: The morphosyntax and intonation of question forms in reality TV shows, audio books and teaching materials, PhD diss., University of Konstanz, Konstanz.
- Richards, Norvin. 1998. The principle of minimal compliance. *Linguistic Inquiry* 29(4): 599–629.
- Richards, Norvin. 2010. *Uttering trees*. Cambridge: MIT Press.
- Richards, Norvin. 2016. *Contiguity theory*. Cambridge: MIT Press.
- Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In *Studies in Romance Linguistics*, eds. Jaeggli Osvaldo and Carmen Silva-Corvalan, 391–419. Dordrecht: Foris.
- Rizzi, Luigi. 1996 [1991]. Residual verb second and the *Wh* criterion. In *Parameters and functional heads*, eds. Adriana Belletti and Luigi Rizzi, 63–90. Oxford: Oxford University Press.
- Rizzi, Luigi. 2001. Relativized minimality effects. In *The handbook of contemporary syntactic theory*, eds. Mark Reuben Baltin and Chris Collins, 89–110. Oxford: Blackwell.
- Rizzi, Luigi. 2004. Locality and left periphery. In *Structures and beyond: The cartography of syntactic structures (3)*, ed. Adriana Belletti, 223–251. Oxford: Oxford University Press.
- Rizzi, Luigi. 2011. Minimality. In *The Oxford handbook of linguistic minimalism*, ed. Cedric Boeckx. Oxford-New York: Oxford University Press.
- Rizzi, Luigi. 2014. Syntactic cartography and the syntacticisation of scope-discourse semantics. In *Mind, values, and metaphysics - Philosophical essays in honor of Kevin Mulligan*, ed. Anne Reboul, 517–533. Cham: Springer.
- Roberge, Yves. 1990. *The syntactic recoverability of null arguments*. Kingston & Montreal: McGill-Queen's University Press.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Ohio State University Working Papers in Linguistics* 49: 91–136.
- Rochemont, Michael S. 1986. *Focus in generative grammar*. Amsterdam: Benjamins.
- Rooth, Mats. 1985. Association with focus. PhD diss., University of Massachusetts.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1): 75–116.
- Ross, John R. 1967. Constraints on variables in syntax. PhD diss., MIT (published as *Infinite Syntax!* Norwood: Ablex, 1986).
- Rouquier, Magali. 2014. *L'émergence des constructions clivées, pseudo-clivées et liées en français*. Paris: Classiques Garnier.
- Rowlett, Paul. 2011. Syntactic variation and diglossia in French. *Salford Working Papers in Linguistics and Applied Linguistics* 1: 13–26.
- Selkirk, Elisabeth. 2008. Contrastive focus, givenness and the unmarked status of “discourse-new”. *Acta Linguistica Hungarica* 55(3): 331–346.
- Sharvit, Yael. 2002. Embedded questions and ‘De Dicto’ readings. *Natural Language Semantics* 10(2): 97–123.
- Shlonsky, Ur. 2012. Notes on *wh* in situ in French. In *Functional heads: The cartography of syntactic structures*, eds. Laura Brugè, Anna Cardinaletti, G. Giusti, Nicola Munaro, and Cecilia Poletto, 242–252. New York: Oxford University Press.
- Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9: 315–332.
- Starke, Michal. 2001. Move dissolves into merge: A theory of locality. Ph.D. diss., University of Geneva.
- Szabolcsi, Anna. 2006. Strong vs. Weak Islands. In *The Blackwell Companion to Syntax 4*, eds. Martin Everaert, Henk van Riemsdijk, 479–531. Malden: Blackwell.
- Tailleur, Sandrine. 2014. The French *wh* interrogative system: Evolution and clefting. In *Variation within and across Romance languages. Selected papers from the 41st Linguistic Symposium on Romance Languages [LSRL 41], Ottawa, 5–7 May 2011*, eds. Marie-Hélène Côté and Eric Mathieu, 315–330. Amsterdam: Benjamins.
- Truswell, Robert. 2007. Locality of *Wh*-movement and the individuation of events. PhD diss., University College London.

- Tual, Lucas. 2017. Prosodic and syntactic properties of French interrogative sentences. Paper presented at the 43rd *Incontro di Grammatica Generativa*, Pavia, 17 February 2017.
- Uribe-Etxebarria, Myriam. 2003. In situ questions and masked movement. *Linguistic Variation Yearbook* 2(1): 217–257.
- Vallduví, Enric, and Maria Vilkkuna. 1998. On rheme and kontrast. In *Syntax and semantics: The limits of syntax*, eds. Peter W. Culicover and Louise McNally, 79–108. San Diego: Academic Press.
- Vergnaud, Jean-Roger, and Maria Luisa Zubizarreta. 2005. The representation of focus and its implications: towards an alternative account of some ‘intervention effects’. In *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*, eds. Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinhenz, and Jan Koster, 641–660. Berlin: de Gruyter.
- Villeneuve, Anne-José, and Julie Auger. 2013. ‘Chtileu qu’i m’freumereu m’bouque i n’est point coér au monne’: Grammatical variation and diglossia in Picardie. *Journal of French Language Studies* 23(1): 109–133.
- Wachowicz, Krystyna A. 1974a. Multiple questions. *Linguistica Silesiana* 1: 155–166.
- Wachowicz, Krystyna A. 1974b. On the syntax and semantics of multiple questions. PhD diss., University of Texas, Austin.
- Wagner, Michael. 2012. Contrastive topics decomposed. *Semantics and Pragmatics* 5(8): 1–54.
- Yabushita, Katsuhiko. 2017. Partition semantics and pragmatics of contrastive topic. In *Contrastiveness in information structure, alternatives and scalar implicatures*, eds. Chungmin Lee, Ference Kiefer, and Manfred Krifka, 23–45. Dordrecht: Springer.
- Zribi-Hertz, Anne. 1994. La syntaxe des clitiques nominatifs en français standard et en français avancé. *Travaux de Linguistique et de Philologie* 32: 131–148.
- Zribi-Hertz, Anne. 2011. Pour un modèle diglossique de description du français: quelques implications théoriques, didactiques et méthodologiques. *Journal of French Language Studies* 21(2): 231–256.
- Zubizarreta, Maria Luisa. 2003. Intervention effects in the French *wh*-in-situ construction: Syntax or interpretation? In *Selected papers from the 31st symposium on Romance languages*, eds. Rafael Nunes-Cedeno, Luis Lopez, and Richard Cameron, 359–380. Amsterdam: Benjamins.
- Zumwald Küster, Géraldine. 2018. *Est-ce que* et ses concurrents. In *L’interrogative en français*, eds. Marie-José Béguelin, Aidan Coveney, and Alexander Guryev, 95–118. Berne: Peter Lang.

ARTICLE

Wh-in-situ in child French: Deictic triggers at the syntax-semantics interface

Katerina Palasis , Richard Faure and Fanny Meunier

Université Côte d'Azur, CNRS, BCL, France
Email: Katerina.Palasis@univ-cotedazur.fr

(Received 01 July 2022; revised 13 January 2023; accepted 02 February 2023; first published online 17 March 2023)

Abstract

This article investigates the development of *wh*-in-situ questions in French by examining a three-year kindergarten dataset of spontaneous productions with 16 children between 2;5 and 5;11. The distribution of the *wh*-phrases is statistically examined in relation to age, verb form (Fixed *be* form *c'est* 'it is' vs. Free *be* forms vs. Free lexical verbs), and grammatical category of the *wh*-word (Pronoun vs. Adverb). Results show that *wh*-in-situ remains prevalent throughout the period despite a steady increase in *wh*-ex-situ. Verb form (Fixed vs. All free forms) is a discriminating variable for the *wh*-position in all three years, and it interacts with the category of the *wh*-word. The Fixed *be* form *c'est* favours in-situ *wh*-pronouns (*c'est qui Taz ?*), whereas the Free forms favour *wh*-ex-situ questions, and massively co-occur with *wh*-adverbs (*combien ça coûte ?*). The emergence of the ex-situ *qu'est-ce que* 'what is it that', as opposed to the in-situ *quoi* 'what', is identified as a factor accounting for the gradual increase in *wh*-ex-situ. Finally, most outliers (*wh*-in-situ with Free forms) are shown to belong to the same paradigm as *c'est* in-situ questions: non-presuppositional questions, which are visible from the frequent use of *là* 'there', like *c'est*, a deictic item.

Keywords: *wh*-questions; French; acquisition; copula; lexical verbs; *quoi/que* alternation; deixis

1. INTRODUCTION

The position of *wh*-phrases in French *wh*-questions is the centre of much attention for at least two reasons. First, Colloquial French is usually described as a 'mixed' language, showing both in-situ and ex-situ *wh*-phrases in matrix information-seeking questions (e.g., Dryer, 2013) as compared to other languages that show either only *wh*-in-situ items (e.g., Chinese; Huang, 1982) or only *wh*-ex-situ items (e.g., English; Nguyen and Legendre, 2022).¹ Second, French-speaking children and adults differ in their distribution of *wh*-positions. It is usually

¹We focus on matrix information-seeking questions, hence leaving aside echo-questions, probe-questions and embedded structures (but see Ledegen (this issue) on the latter).



reported that children utter more *wh*-in-situ questions than adults (Zuckerman and Hulk, 2001; Strik, 2007; Strik and Pérez-Leroux, 2011; Becker and Gotowski, 2015; Thiberge, 2020). This article aims to explain this early prevalence of the *wh*-in-situ position and the difference with adult productions. It will statistically investigate the production of matrix *wh*-questions in 16 typically developing French L1 children during their three kindergarten years. The investigation will pinpoint two factors at the syntax-semantics interface that account for the prevalence of the *wh*-in-situ position in child but not adult French, namely the Fixed *be* structure *c'est* ‘it is’ and the adverb *là* ‘there’. An attempt at the unification of these in-situ question triggers under an umbrella category (deixis) will be made as well as an examination of the counter-examples. The data will also draw our attention to a third factor: the morpho-syntactic difference between in-situ *quoi* ‘what’ and ex-situ *qu’est-ce que* ‘what is it that’, and the gradual emergence of the latter throughout the three-year kindergarten period.

1.1 The target language

In this contribution, we will leave Standard French aside, because we assume that preschool children are mainly exposed to and hence initially acquire Colloquial French (e.g., Hamlaoui, 2011). Colloquial French is often but not always described as a mixed language. Baunaz (2011, 2016) compared ‘Non-Standard Colloquial French’ (NSCF) with Chinese and described both as *wh*-in-situ languages with covert *wh*-movement. The author also proposed that NSCF displays three types of in-situ *wh*-phrases: one non-marked and two marked types (Types 1, 2 and 3, respectively). Type 1 in-situ *wh*-phrases are analysed as covert Split-DPs (i.e., variables with a phonologically null interrogative operator). They are non-presuppositional and have a rising intonation. Types 2 and 3 in-situ *wh*-phrases have specific and partitive interpretations, and fall-rise and falling intonations, respectively. Both entail covert movement of the entire *wh*-phrase. Furthermore, Baunaz (2011: 75) described in-situ and ex-situ structures in NSCF as syntactically ‘similar’, with identical processing loads (following Adli, 2006), and she related the difference in overt versus covert movement to ‘interpretation, as witnessed by intonation’. We will elaborate on Baunaz’s (2011, 2016) proposal in Section 4.

Faure and Palasis (2021) also described ‘Colloquial French’ as a *wh*-in-situ language with covert *wh*-movement; however, contrary to Baunaz (2011), they suggested that covert and overt movement are different. They showed that movement is less restricted in *wh*-ex-situ questions, which is evidenced by lower sensitivity to weak islands (e.g., negation; see Examples 1a vs. 1b) and Superiority (i.e., order of arguments in multiple questions; see Example 1c, its alternate *c’* vs. Example 1d and the answers they allow). The authors thus attributed overt movement of the *wh*-phrases to the presence of a different trigger, namely Exclusivity. A *wh*-question endowed with this feature presupposes that not all of the possible answers, say from a set of three, are correct but that two at most are (see Examples 1e vs. 1f).

- (1) *Wh*-in-situ and *wh*-ex-situ structures differ in adult data (Faure and Palasis, 2021: 64–74):
Context: Julien runs a stand at a car boot sale. Nadia checks out his sales at the end of the day:
- a. **T(u) as pas vendu quoi* ?² (no stress on *quoi*, sensitivity to weak islandhood)
 you have not sold what
 ‘What haven’t you sold?’
 - b. *Qu’est-ce que t(u) as pas vendu* ? (no sensitivity to weak islandhood)
 what is it that you have not sold
 ‘What haven’t you sold?’
 - c. *Qu’est-ce que t(u) as vendu à qui* ? (answer: *a book to Sam, a pen to John*)
 what is it that you have sold to whom
 ‘What have you sold to whom?’
 - c’. *À qui t(u) as vendu quoi* ? (answer: *to Sam a book, to John a pen*)
 to whom you have sold what
 ‘What have you sold to whom?’
 - d. *T(u) as vendu quoi à qui* ? (no stress on *à qui*, answer: *a book to Sam, a pen to John, #to Sam a book, to John a pen*)
 you have sold what to whom
 ‘What have you sold to whom?’
Context: Marie enquires which friend(s) among a group of three Jane saw last week-end:
 - e. *T(u) as vu qui* ? (in-situ *qui* = anyone and all of them)
 you have seen who
 ‘Who did you see?’
 - f. *Qui t(u) as vu* ? (ex-situ *qui* = at least one person from the presupposed set cannot be the answer)³
 who you have seen
 ‘Who did you see?’

Baunaz’s (2011, 2016) and Faure and Palasis’ (2021) proposals both hinge upon the syntax-semantics interface and account for adult Colloquial French.⁴ This language will be considered here as the target language for children.

1.2 The child system

Research has shown that French preschoolers usually produce more in-situ *wh*-questions than adults (Zuckerman and Hulk, 2001; Strik, 2007; Strik and Pérez-Leroux, 2011; Becker and Gotowski, 2015; Thiberge, 2020). Young children nevertheless also produce ex-situ *wh*-items from the onset around the

²Incomplete words are transcribed with the omitted material in parentheses.

³An anonymous reviewer wonders whether the answer ‘everyone’ is totally excluded, as predicted. According to our informants, it is not but it has a corrective flavour.

⁴Hypotheses on Chinese and English also assume that the default *wh*-position only is related to the *wh*-feature and that the other *wh*-position is triggered by another feature (i.e., contrastive focus for Chinese ex-situ and presupposition for English in-situ; see Cheung, 2008 and Biezma, 2020, respectively).

age of 2;0 (Crisma, 1992; Déprez and Pierce, 1993; Hamann, 2006; Strik, 2007; Prévost, 2009), and the in-situ rates vary widely from one child to another. A small review of children between 1;8 and 2;9 (Palasis, Faure and Lavigne, 2019) reported in-situ rates ranging from 1.3% (Philippe at 2;3 in Crisma, 1992) to 94.4% (Augustin at 2;04 in Rasetti, 2003) in naturalistic speech, which seems to display more in-situ *wh*-items than experimental contexts (Zuckerman and Hulk, 2001). The in-situ ratio also seems to be different with pronouns (e.g., *quoi* ‘what’) compared to adverbs (e.g., *où* ‘where’), and there are more in-situ *wh*-phrases with the former than with the latter (Strik, 2007; Jakubowicz, 2011; Strik and Pérez-Leroux, 2011).

The discrepancy between the ratios of in-situ and ex-situ *wh*-phrases in adult and child speech has often been ascribed to differential abilities between adults and children. Generativists posit that in-situ *wh*-questions converge by checking a *wh*-feature covertly in Logical Form, whereas ex-situ *wh*-questions converge by checking the *wh*-feature in LF and overt syntax (Bayer and Cheng, 2017 for a review). The latter is described as cognitively more costly, and derivational economy is assumed in children compared to adults (Hamann, 2006; Jakubowicz, 2011; Strik, 2012). Differing pragmatics have also been evoked in the literature, such as overuse of common ground in children entailing more *wh*-in-situ questions (Gotowski and Becker, 2016). However, hypotheses based on differential abilities make the prediction that all children, whatever their language, should initially produce more in-situ *wh*-phrases, which is not the case in target *wh*-ex-situ languages (e.g., English, Dutch, Italian, German, Swedish; Stromswold, 1995; Van Kampen, 1997; Guasti, 2000; Roeper and De Villiers, 2011; Strik, 2012).

Resorting to another type of difference between adults and children, Palasis *et al.* (2019) showed that Verb form (i.e., Fixed *be* form *c'est* ‘it is’⁵ vs. Free *be* forms and Lexical verbs) was a discriminating variable with regard to *wh*-position, and that there was a correlation between the *wh*-in-situ position and the Fixed *be* form in their child data (see Examples 2a vs. 2b and 2c). The difference between adults and children was then ascribed to different stages of the diversification of the verbal system with more *c'est* forms in children than in adults (following Guillaume, 1927; Clark, 1978; Bassano, Eme and Champaud, 2005). The hypothesis also accounted for the well-documented asymmetry between adults and children with regard to *wh*-pronouns and *wh*-adverbs: Children utter more in-situ *wh*-pronouns than adults (e.g., *quoi* ‘what’, *qui* ‘who’), whereas both groups favour the ex-situ position for *wh*-adverbs (e.g., *où* ‘where’, *comment* ‘how’; Zuckerman and Hulk, 2001; Strik, 2007; Strik and Pérez-Leroux, 2011; Becker and Gotowski, 2015). Palasis *et al.* (2019) reported that most pronouns were uttered with the Fixed *be* form *c'est* ‘it is’ that favours in-situ (see 2a); whereas most adverbs were uttered with the Free verb forms that favour the ex-situ position (see 2b and 2c).

⁵This *c'est* is not to be confused with the *c'est* that is used in clefts (*c'est X que* ‘it is X that’). Fixed *be c'est* is a variant of the copula that appears in non-predicational copular sentences that posit an equation between two terms: *Bob c'est mon meilleur ami* ‘Bob is my best friend’ (Higgins, 1973; Roy, 2013 for French).

- (2) Correlation between the Fixed form *c'est* 'it is' and *wh*-in-situ in child data (Palasis et al., 2019):
- a. *C'est quoi ça ?* (LI, 3;4)
it is what that
'What is that?'
 - b. *Où elle est ta maison ?* (MT, 3;11)
where she is your house
'Where is your house?'
 - c. *Comment on joue à ce jeu ?* (LU, 3;4)
how one plays at this game
'How do we play this game?'

In this study, we will examine a new dataset for 16 children in their third kindergarten year and connect these new data with the previously examined data for the same children in their first two kindergarten years (Palasis et al., 2019), hence providing a new three-year longitudinal investigation into the *wh*-questions for 16 children. More specifically, we will examine the distribution of *wh*-in-situ questions in Year 3 and compare our results with Years 1 and 2 in order to shed light on new developmental aspects of the position of *wh*-words in preschoolers. Section 2 will describe the entire dataset and the predictions we make building on previous hypotheses. Section 3 will statistically examine the data as a function of the *wh*-position (in-situ vs. ex-situ), the verb form (fixed vs. free), and the grammatical category of the *wh*-phrase (pronoun vs. adverb). Section 4 will examine the counter-examples to the correlation between the Fixed *be* form *c'est* 'it is' and the *wh*-in-situ position, provide a comparison with the target language described in Section 1.1, and elaborate on deixis. We will end with concluding remarks and future research in Section 5.

2. METHOD

This research will statistically investigate the production of matrix *wh*-questions in typically developing preschool French L1 children. In Section 2, we will describe the initial kindergarten dataset (2.1), the participants (2.2), the data cleaning process leading to the final dataset (2.3), and the predictions we make building on previous hypotheses (2.4).

2.1 The initial kindergarten dataset

The data were collected with a class of preschoolers during their three kindergarten years in the South of France (total $n = 41,845$ utterances; see Table 1). Small groups of two to four children were gathered in a quiet room close to their usual classroom and were encouraged by the researcher to narrate their activities in and out of school, 'read' books and play games. There were child-child and child-researcher interactions, and the latter were kept as natural as possible. The children were video- and audio-recorded during 20 to 25-minute sessions, 13 times during their first year, and 10 times during their second and third kindergarten years

Table 1. The kindergarten corpus

Specifications	Year 1	Year 2	New Year 3 (5/10 sessions)	Total (provisional)
Dates	2006/2007	2007/2008	2008/2009	
Hours	25	20	20 (10)	
Sessions	13	10	10 (5)	
Recording types	Audio & video	Audio & video	Audio & video	
Adults	3	1	1	
Adult utterances	12,891	9,291	4,947	27,129
Children (all/ L1 French)	20/17	19/17	18/ 16	
Age range	2;5-4;0	3;6-4;11	4;5-5;11	
Child utterances (all/ 16)	15,992/13,320	14,348/12,242	11,505/ 10,501	41,845/36,063

(which amounted to an average interval between two sessions of three to four weeks, except for holidays). The first two years were examined with regard to *wh*-questions in Palasis *et al.* (2019). This article will focus on the *wh*-in-situ questions in the newly processed data for the third year (5 sessions out of 10 are available for analysis, transcribed and coded with the CHILDES tools; MacWhinney, 2000) and on a new longitudinal analysis of the three years.

2.2 Participants

The participants were the native speakers of Hexagonal French in the kindergarten class described in Section 2.1. All the children in the class were asked to participate ($n = 20$), but four children were excluded in the final dataset: two children did not have a native French background at home, and two children moved during the period. The set thus comprises the data for 16 children (9 female) during a three-year period ($n = 36,063$ utterances). Figure 1 shows the distribution of the data according to each child over the three years. The children were between 2;5 (MS: youngest child at the beginning of Year 1) and 5;11 (KE: oldest child by the end of Year 3).

2.3 The final dataset

The children produced six types of questions that distributed in frequent and infrequent types (see Examples 3 and Figure 2). Finite matrix *wh*-questions (initial $n = 1,084$) represented one of the frequent types, with an average of 34.5% of the children's questions during the three-year period.

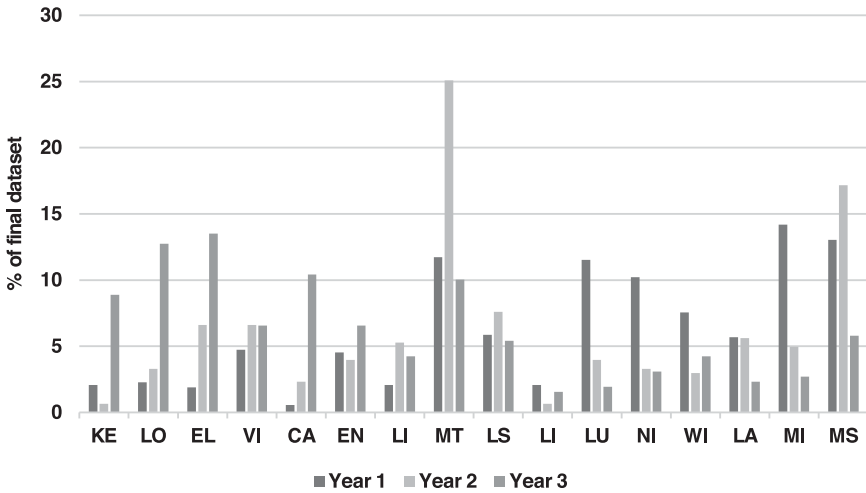


Figure 1. Participation per year per child.

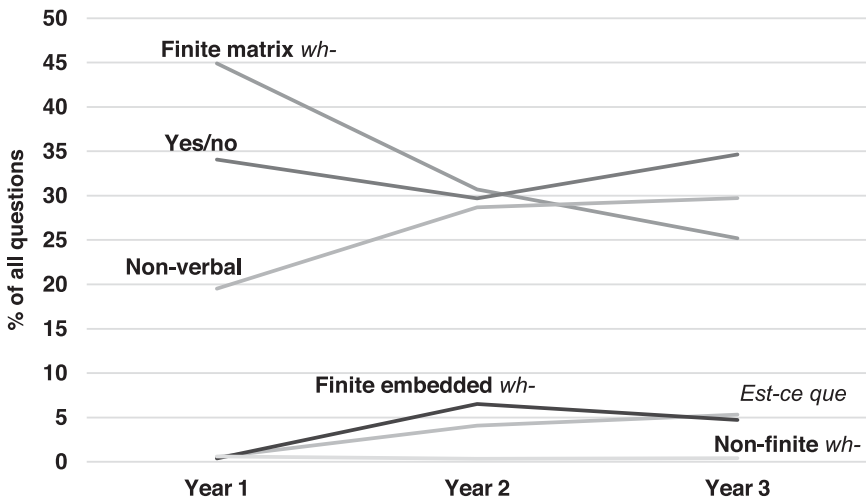


Figure 2. Distribution of question types.

(3) Six types of questions in the dataset:

a. Finite matrix *wh-* (34.5%):
Où il est Kelian ? (MS, 4;7)
 where he is Kelian
 ‘Where is Kelian?’

b. Yes/no (32.8%):
On peut lire des livres ? (MT, 4;11)
 one can read some books
 ‘Can we read books?’

- c. Non-verbal (25.5%):
Et le blanc ? (EL, 5;4)
and the white
'And (what about) the white one?'
- d. *Est-ce que* (3.7%):
Est-ce que t(u) as encore un tablier ? (MS, 5;0)
is it that you have still an apron
'Do you still have an apron?'
- e. Finite embedded *wh*- (3.1%):
Je sais pas comment i(ls) s'appellent. (LU, 4;9)
I know not how they REFL call
'I don't know what their names are'
- f. Non-finite *wh*- (0.4%):
À construire quoi ? (VI, 5;3)
at build what
'(they help) To build what?'

The longitudinal dataset in Figure 2 shows that the ratio of finite matrix *wh*-questions decreases during the period in favour of yes/no and non-verbal questions. To the contrary, the ratio of children's utterances compared to adults' utterances in the entire dataset increases during the period (Year 1: 50.8%, Year 2: 56.9%, Year 3: 68%). The progression shows that there are more child-child and fewer child-adult interactions in Year 3 than in Year 1. We hypothesize that the decrease in child *wh*-questions pertains to this change in addressee, children using more non-verbal and yes/no questions than matrix *wh*-questions when addressing another child than when addressing an adult.

Table 2 shows the initial dataset of 1,084 matrix *wh*-questions distributed according to *wh*-word (*quoi/que* 'what', *où* 'where', *qui* 'who', *pourquoi* 'why', *comment* 'how', *à qui* 'to whom', *à quoi* 'to what', *quand* 'when', *quel* 'which', *lequel* 'which one', *combien* 'how many') and according to syntactic structure. The syntactic structures stem from the Verb form hypothesis that suggested that verb form (i.e., Fixed *be* form *c'est* 'it is' vs. Free *be* and lexical forms) is a discriminating variable for *wh*-position in child data (Palasis *et al.*, 2019). Examples (4) illustrate the 11 different structures as a function of the position of the *wh*-item and the verb form in the dataset. Structures 1a, 1b, 4 and 6 display in-situ *wh*-questions and Structures 2a, 2b, 3a, 3b, 5, 7 and 8 display ex-situ *wh*-questions.

(4) Structures according to *wh*-position and verb form:

- a. Structure 1a (S Vlex *wh*-):
Je fais quoi ? (MT, 4;11)
I do what
'What do I do?'
- b. Structure 1b (S Vbe *wh*-):
Je suis où ? (LO, 5;9)
I am where
'Where am I?'

Table 2. Distribution of *wh*-questions according to *wh*-item and syntactic structure (initial 3-year dataset)

*	quoi/				comment	à		quand	quel	lequel	combien	Total
	que	où	qui	pourquoi		qui	quoi					
1a	80	14	12	0	8	0	2	0	6	2	3	127
1b	0	59	0	0	1	1	0	0	1	0	0	62
2a	0	8	45	94	62	1	0	0	2	2	1	215
2b	0	91	3	7	0	1	0	0	0	0	0	102
3a	1	0	1	0	1	0	0	0	0	0	0	3
3b	0	8	0	0	0	0	0	0	0	0	0	8
4	206	1	34	0	1	56	1	1	1	9	0	310
5	12	0	1	10	0	5	0	0	0	0	0	28
6	0	2	55	0	0	0	0	18	0	0	0	75
7	2	0	37	0	0	0	0	2	0	0	0	41
8	102	0	10	0	0	0	0	1	0	0	0	113
	403	183	198	111	73	64	3	22	10	13	4	1084

*Structures: 1a: S Vlex *wh*-, 1b: S Vbe *wh*-, 2a: *wh*- S Vlex, 2b: *wh*- S Vbe, 3a: *wh*- Vlex S, 3b: *wh*- Vbe S, 4: *c'est wh*-, 5: *wh-c'est*, 6: *c'est wh- que S V*, 7: *wh- c'est que S V*, 8: *wh-est-ce que S V*.

- c. Structure 2a (*wh*- S Vlex):
Combien ça coûte ? (LA, 5;2)
 how much that costs
 'How much does that cost?'
- d. Structure 2b (*wh*- S Vbe):
Où il est Kelian ? (MS, 4;7)
 where he is Kelian
 'Where is Kelian?'
- e. Structure 3a (*wh*- Vlex S):
Que se passe-t-il ? (VI, 3;5)
 what REFL happens it
 'What is happening?'
- f. Structure 3b (*wh*- Vbe S):
Elmer où es-tu ? (LS, 4;3)
 Elmer where are you
 'Elmer where are you?'
- g. Structure 4 (*c'est wh*-):
C'est qui Taz ? (LS, 5;1)
 it is who Taz
 'Who is Taz?'
- h. Structure 5 (*wh- c'est*):
À qui c'est ? (LI, 5;1)
 to whom it is
 'To whom does it belong?'

- i. Structure 6 (*c'est wh- que S V*):
Mais c'est quand que tu montres la maison ? (MT, 3;11)
 but it is when that you show the house
 'But when is it that you show the house?'
- j. Structure 7 (*wh- c'est que S V*):
Mais quand c'est qu'on va à l'école ? (MS, 3;9)
 but when it is that one goes to the school
 'But when is it that we go to school?'
- k. Structure 8 (*wh- est-ce que S V*):
Qu'est-c(e) qu'il a dit ? (VI, 5;4)
 what is it that he has said
 'What did he say?'

Previous research has put forward that some *wh*-words, contexts and syntactic structures favour either *ex-situ* or *in-situ wh*-questions. In order to investigate the factors that possibly interact with the position of *wh*-items, we needed to examine only the *wh*-items, contexts and syntactic structures that allow the alternative *in-situ/ex-situ wh*-positions. The following items ($n = 366$; mainly in shaded cells in Table 3) were hence discarded from the initial 1,084 set:

- The *wh*-item *pourquoi* 'why' ($n = 111$) appears only *ex-situ*. This *wh*-word is assumed to be generated in the left periphery, higher than the other *wh*-words (Rizzi, 2001; Hamann, 2006; Myers and Pellet, 2014).
- The *wh*-item *qui* 'who' always appears to the left of the finite verb when it is a subject ($n = 58$).
- Cleft structures (Structures 6 and 7; $n = 116$), because we do not address the specific matter of *wh*-fronting in clefts in this article (but see Oiry, 2011).
- The D-linked *wh*-items (mainly *quel* 'which', *lequel* 'which one', and variants thereof; $n = 59$, 23 *quel/lequel* + 36 with other *wh*-words), because of the possible interaction between D-linking and the *in-situ* position in French (Coveney, 1989; Obenauer, 1994; Chang, 1997; Boeckx, 2000; Cheng and Rooryck, 2000).
- Questions with subject-verb inversion (Structures 3a and 3b; $n = 11$), because they allow *ex-situ wh*-items only, and these questions belong to Standard French, which we do not examine in this article.
- Questions with any other characteristic of Standard French as opposed to Colloquial French (i.e., non-elided nominative clitics before consonants and discontinuous negation; Palasis, 2013): only one *wh*-question was discarded due to full *il* in front of a consonant (i.e., *et il [/] il voit quoi ?* 'and what does he see').
- Immediate (self-)repetitions ($n = 10$).

Table 3 shows the final dataset with 718 *wh*-questions produced by 16 children during their three kindergarten years. Year 3 ($n = 158$) is examined for the first time and is added to Year 1 ($n = 368$) and Year 2 ($n = 192$) that have been investigated in previous work.

Table 3. Distribution of *wh*-questions according to *wh*-item and syntactic structure (final 3-year dataset)

	<i>quoi/</i> * <i>que</i>	<i>où</i>	<i>qui</i>	<i>pourquoi</i>	<i>comment</i>	<i>à</i> <i>qui</i>	<i>à</i> <i>quoi</i>	<i>quand</i>	<i>quel</i>	<i>lequel</i>	<i>combien</i>	Total
1a	69	14	12	0	7	0	1	0	0	0	2	105
1b	0	57	0	0	1	1	0	0	0	0	0	59
2a	0	7	0	0	62	0	0	0	0	0	1	70
2b	0	86	0	0	0	0	0	0	0	0	0	86
3a	0	0	0	0	0	0	0	0	0	0	0	0
3b	0	0	0	0	0	0	0	0	0	0	0	0
4	200	1	34	0	1	48	1	1	0	0	0	286
5	12	0	1	0	0	5	0	0	0	0	0	18
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	93	0	0	0	0	0	0	1	0	0	0	94
	374	165	47	0	71	54	2	2	0	0	3	718

*Structures: 1a: S Vlex *wh*-, 1b: S Vbe *wh*-, 2a: *wh*- S Vlex, 2b: *wh*- S Vbe, 3a: *wh*- Vlex S, 3b: *wh*- Vbe S, 4: *c'est wh*-, 5: *wh-c'est*, 6: *c'est wh- que S V*, 7: *wh- c'est que S V*, 8: *wh-est-ce que S V*.

2.4 Predictions

Building on previous hypotheses detailed hereunder, we make the following predictions:

- i. Because there are no major changes in the grammar of the children between Years 1, 2 and 3 (as established in Section 2.3 with the near absence of markers of Standard French in Year 3), we predict that there should be no significant difference in the distribution of the in-situ versus ex-situ rates between Year 3 and the previously studied Years 1 and 2. We will address this matter in Section 3.1.
- ii. The Verb form hypothesis (Palasis et al., 2019) makes the prediction that the distribution of in-situ and ex-situ *wh*-questions in Year 3 of the dataset should be related to Verb form (Fixed vs. Free forms). We will test this prediction in Section 3.2.
- iii. Child datasets usually report an asymmetry in the position of the *wh*-items according to age (children vs. adults) and to grammatical category of the *wh*-item (pronouns vs. adverbs; Zuckerman and Hulk, 2001; Strik, 2007; Jakubowicz, 2011; Becker and Gotowski, 2015). Two hypotheses exist on this asymmetry in child questions: a relationship between the position of objects in the verbal phrase and the in-situ position of the corresponding *wh*-pronouns, and a relationship between the verb form in the sentence and the position of the *wh*-word whatever its grammatical category (Strik, 2007 and Palasis et al., 2019, respectively). The first hypothesis makes the prediction of a clear-cut asymmetry according to grammatical category

(i.e., in-situ pronouns vs. ex-situ adverbs); the second hypothesis predicts that *wh*-pronouns and *wh*-adverbs can appear in-situ or ex-situ depending on the verb form. We will test both predictions in Section 3.3 and ponder their respective developmental scopes.

3. RESULTS AND PRELIMINARY DISCUSSION

The final dataset in Table 3 ($n = 718$ *wh*-questions) was submitted to the Cochran-Mantel-Haenszel Test (henceforth CMH; Cochran, 1954; Mantel and Haenszel, 1959) for repeated 2x2 tests of independence. We investigated the relationship between the verb form (Fixed *be* form *c'est* 'it is' vs. Free *be* forms vs. Lexical forms) and the position of the *wh*-item (In-situ vs. Ex-situ) in each utterance. This was done by specifically programming an Excel spreadsheet. The data were submitted in as many 2x2 tables as there were participants with available data for each test in order to account for the quantitative and qualitative differences between participants (McDonald, 2011; see Figure 1 for the distribution per child). A significant result at the X^2 CMH test means that the variables are dependent (one has an effect on the other). In the following subsections, we will test the predictions stated in Section 2.4 on the evolution of *wh*-in-situ compared to *wh*-ex-situ (3.1), on the impact of the different verb forms (3.2), and on the asymmetry between *wh*-pronouns and *wh*-adverbs (3.3).

3.1 Distribution of the *wh*-questions according to the position of the *wh*-item (in-situ vs. ex-situ)

We examined the distribution of the *wh*-questions in Year 3 and longitudinally from Year 1 to Year 3 in order to describe the overall progression of the in-situ position in the dataset. The filtering of the data (described in Section 2.3) pointed towards an absence of radical change between Years 1 and 2 taken together and Year 3 in terms of grammatical characteristics (i.e., nominative clitics and negation). We therefore predicted no significant difference of distribution between both *wh*-positions in Year 3 compared to the first two years.

Figure 3 shows that *wh*-in-situ questions remain prevalent in Year 3 (55.7%), as they were in Year 1 and Year 2 (67.4% and 59.4%). A repeated-measures ANOVA including Year (1, 2, 3) and Position (ex-situ, in-situ) as factors confirmed that there are overall significantly more in-situ questions than ex-situ ones ($F(1,15) = 26.649$; $p < .001$; eta squared = .640) and no significant differences in the distribution between the three years ($F(2,30) = 1$; n.s.). Prediction (i) is thus borne out. Nevertheless, Figure 3 also shows a steady progression in the *wh*-ex-situ position during the period. Specific post-hoc comparisons using Bonferroni corrections showed that there are more in-situ questions than ex-situ ones in Year 1 ($t(15) = -4.418$, $p < .01$), while their proportions become equivalent in Years 2 and 3 (respectively: $t(15) = -1.575$, n.s.; $t(15) = -1.826$, n.s.).

We also considered the *wh*-words individually in order to uncover possible specific developmental patterns.⁶ There were no significant differences in the in-

⁶We thank an anonymous reviewer for this very interesting suggestion.

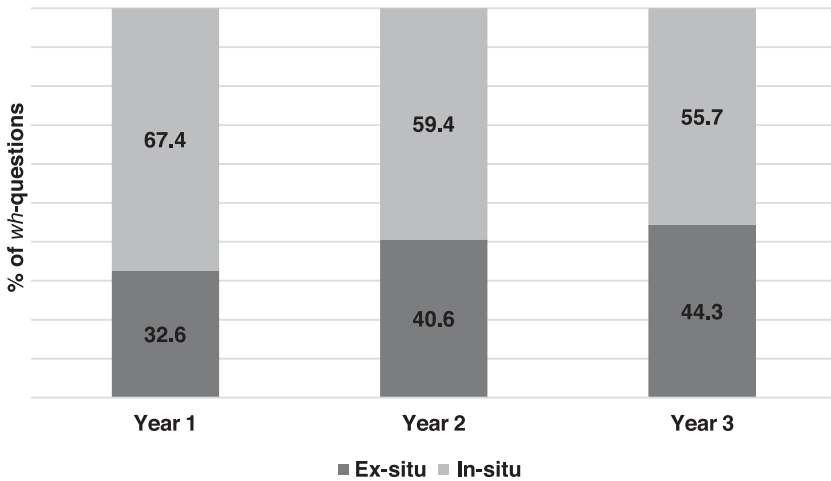


Figure 3. Longitudinal distribution of matrix *wh*-questions according to the position of the *wh*-item.

situ/ex-situ distributions between the three years, except for the *wh*-word *quoi/que* ‘what’ that showed a significant decrease in the in-situ position between Year 2 and Year 3 (73.5%–59.0%; Fisher’s Exact Test for Count Data: $p < .05$). In Section 3.2, we will investigate the detail of the in-situ/ex-situ distribution according to the verb form, and we will elaborate on the significant decrease in the in-situ *quoi* ‘what’ at the end of Section 3.3.

3.2 Distribution of the *wh*-questions according to the verb form (Fixed *be* form vs. Free *be* forms vs. Lexical verbs)

We examined the distribution of the *wh*-questions according to the verb form (Fixed *be* form vs. Free *be* forms vs. Lexical verbs) in Year 3 (Table 6) and longitudinally with Years 1 and 2 (Tables 4 and 5). We repeated the series of analyses carried out on the first two years on the third year: All *be* forms vs. Lexical verbs, Free *be* forms vs. Fixed *be* form, Free *be* forms vs. Lexical verbs, Fixed *be* form vs. Lexical verbs, and Fixed *be* form vs. All Free forms (Free *be* and Lexical).

The initial analyses of the first two years opposing All *be* forms to Lexical verbs (following the binary Lexical hypothesis on adult data; Coveney, 1995) had shown that the position of the *wh*-word in child data was also related to the type of verb, linking the in-situ position with *be* forms (without distinguishing Free *be* forms and Fixed *be* form) and the ex-situ position with Lexical verbs (Palasis et al., 2019: 225). In order to test if this is still the case in Year 3, we performed, as it had been done in Years 1 and 2, a CMH Test for repeated 2x2 tests of independence. The analysis of the independence between *wh*-position and verb type –that is Lexical Verbs vs. All *be* forms (free and fixed) – gave a X^2 CMH = 31.28, 1 *df*, $p < .0001$. Again, in the third year, the in-situ *wh*-position is significantly more frequent with *be* forms and the ex-situ *wh*-position is significantly more frequent with Lexical Verbs. Cramér’s

Table 4. Distribution of *wh*-questions according to verb form and *wh*-position (Year 1)

	Ex-situ		In-situ		Total	
	n	%	n	%	n	%
Free <i>be</i>	58	60.4	38	39.6	96	26.1
Fixed <i>be</i>	14	8.0	162	92.0	176	47.8
Lexical verbs	48	50.0	48	50.0	96	26.1
Total	120	32.6	248	67.4	368	100.0

Table 5. Distribution of *wh*-questions according to verb form and *wh*-position (Year 2)

	Ex-situ		In-situ		Total	
	n	%	n	%	n	%
Free <i>be</i>	20	62.5	12	37.5	32	16.7
Fixed <i>be</i>	2	2.7	71	97.3	73	38.0
Lexical verbs	56	64.4	31	35.6	87	45.3
Total	78	40.6	114	59.4	192	100.0

V, which determines the degree of association between variables, indicates a large effect size ($V = 0.44$; Cramér, 1946).

To refine this result in Year 3, we further tested the impact of each *be* form on the position of the *wh*-word separately. Since one child (NI) had no data for this test, data were submitted in only 15 2x2 tables. Analyzing the relationship between the *wh*-position and the Free *be* forms vs. the Fixed *be* form using CMH, the result was X^2 CMH = 12.97, 1 *df*, $p < .001$. This showed that, as it was the case in the first two years, the *wh*-position is significantly more likely to occur in-situ with the Fixed *be* form and ex-situ with the Free *be* forms. Cramér's V indicates a large effect size ($V = 0.533$). This result allows us to separate the *be* forms into two significantly different categories, hence distinguishing Free *be* forms and the Fixed *be* form *c'est* 'it is', as previously observed in Years 1 and 2.

Bearing in mind that Lexical verbs are also free verb forms, we then examined Lexical Verbs in relation to Free *be* forms. The result was not significant (X^2 CMH = 0.135, 1 *df*, $p = .7$). This result shows that, as in the first two years, the *wh*-position is not significantly different when comparing Free *be* forms and Lexical Verbs, which are also free forms. Therefore, we confirm that the type of free verb (Lexical vs. *be*) is not a discriminating variable for the position of the *wh*-words in neither year of the dataset.

To test if the form of the verb correlates with the position of the *wh*-word, we looked at the impact of the Fixed *be* form on the position of the *wh*-word and compared the Fixed *be* form with Lexical Verbs. The result was significant (X^2 CMH = 42.74, 1 *df*, $p < .0001$). As in the first two years, the *wh*-position is significantly more in-situ with the Fixed *be* form and more ex-situ with Lexical verbs. Cramér's V indicates a large effect size ($V = 0.649$).

Table 6. Distribution of *wh*-questions according to verb form and *wh*-position (Year 3)

	Ex-situ		In-situ		Total	
	n	%	n	%	n	%
Free <i>be</i>	8	47.1	9	52.9	17	10.8
Fixed <i>be</i>	2	3.6	53	96.4	55	34.8
Lexical verbs	60	69.8	26	30.2	86	54.4
Total	70	44.3	88	55.7	158	100.0

Finally, to confirm the importance of the Verb form (Fixed vs. Free) as a discriminating variable with respect to *wh*-position, we looked at the impact of the Fixed *be* form on the position of the *wh*-words by testing the Fixed *be* form against All Free forms, regardless of verb type. The result was significant (X^2 CMH = 46.516, 1 *df*, $p < .0001$) as it was for the first two years. The *wh*-position is significantly more in-situ with the Fixed *be* form and more ex-situ with All Free forms (Free *be* forms and Lexical verbs). Cramér's *V* indicates a large effect size ($V = 0.598$).

Altogether, this pattern of results is in line with what was observed for the first two years. In the third year too, the Verb form (Fixed vs. Free) is a discriminating variable for the *wh*-position, regardless of Verb type (All *be* vs. Lexical verbs), and the effect size is larger when investigating Verb form (Fixed vs. Free: $V = 0.598$) compared to Verb type (All *be* vs. Lexical verbs: $V = 0.44$). Finally, the different effect sizes between the three categories (Fixed *be* vs. Free *be*: $V = 0.533$ and Fixed *be* vs. Lexical: $V = 0.649$) illustrate the greater syntactic and semantic differences between Lexical verbs and the Fixed *be* form, which differ with regard to two features (i.e., lexicality and fixity), than between Fixed and Free *be* forms, which differ in one feature only (i.e., fixity).

The tests undertaken in Section 3.2 on Year 3 of the dataset allow us to confirm the Verb form hypothesis: the in-situ position of the *wh*-phrases is related to the Fixed *be* form *c'est* 'it is' against all Free forms. Prediction (ii) is thus borne out. Though significant, this link does not hold for the entire dataset. We will elaborate on the interface between syntax (the position of the *wh*-phrase) and semantics (the question type) in Section 4, when we discuss the outliers to the in-situ/Fixed *be* correlation (i.e., the *wh*-in-situ questions with Free verb forms).

3.3 Distribution of the *wh*-questions according to the grammatical category of the *wh*-word (pronouns vs. adverbs)

In Section 2.4 (Prediction iii), we reported that children usually utter more in-situ *wh*-pronouns than adults, whereas both groups favour the ex-situ position for *wh*-adverbs (Zuckerman and Hulk, 2001; Strik, 2007; Strik and Pérez-Leroux, 2011; Becker and Gotowski, 2015). We mentioned two hypotheses on the prevalence of in-situ *wh*-pronouns in children compared to adults. Strik (2007) suggested an effect of the embedded position of arguments in the verbal phrase compared

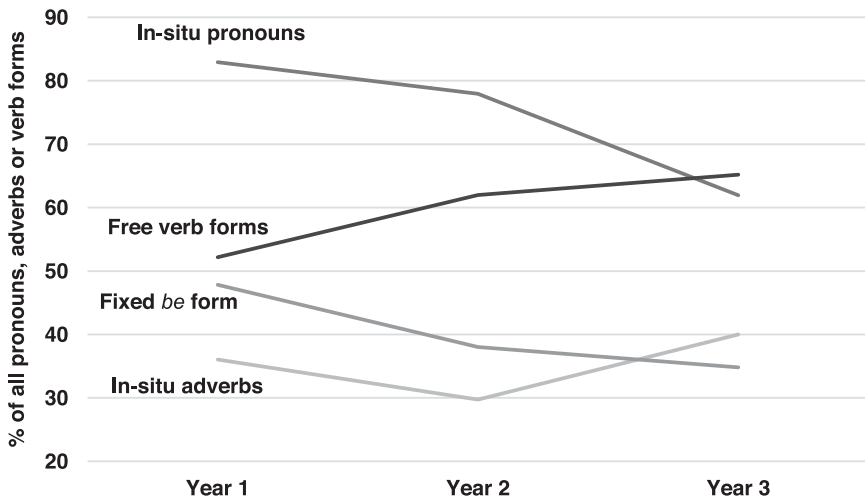


Figure 4. Evolution of ratios for *wh*-in-situ pronouns/adverbs and Fixed *be*/free verb forms.

to adjuncts that are merged higher in the structure. The VP internal position then renders pronouns more costly to move than adverbs. Palasis *et al.* (2019) reported a possible effect of the verb form on the position of the *wh*-phrase whatever the grammatical category of the *wh*-word. The ratio of the *wh*-in-situ phrases (pronominal and adverbial) is then expected to follow the ratio of the Fixed *be* form *c'est* 'it is', because the latter favours *wh*-in-situ questions. The first hypothesis makes the prediction that the ratio of *wh*-in-situ pronouns should decrease in favour of more *wh*-ex-situ pronouns as a function of the children's cognitive development. The hypothesis does not make any particular predictions on the evolution of the position of *wh*-adverbs. The second hypothesis makes the prediction that the ratio of *wh*-in-situ altogether (pronouns and adverbs) should decrease if the Fixed *be* form *c'est* 'it is' decreases, because verb form and *wh*-position are related. Conversely, *wh*-words co-occurring with Free verb forms are expected to be ex-situ. The developmental aspect of this hypothesis hinges upon the diversification of the verbal system with more *c'est* forms in children than in adults (Guillaume, 1927; Clark, 1978; Bassano *et al.*, 2005).

We examined the evolution of the in-situ/ex-situ ratios for *wh*-pronouns and *wh*-adverbs separately, and the evolution of the ratio of fixed and free verb forms during the three-year period. A repeated-measures ANOVA including Year (1, 2, 3) and Grammatical category (Pronouns, Adverbs) as factors indicates an effect of the grammatical category of the *wh*-item ($F(1,15) = 50.143$; $p < .001$; eta squared = .770), and that the distribution between the three years is not significantly different ($F(2,30) = 1.513$; n.s.).

Figure 4 shows the difference between *wh*-in-situ pronouns and adverbs. The ratio of *wh*-in-situ pronouns decreases constantly (82.9%–78.0%–61.9%) in favour of more ex-situ pronouns, as expected considering that adult Colloquial French is usually described as displaying more ex-situ *wh*-questions than child

Table 7. Distribution of *wh*-adverbs according to syntactic structure (Year 3)

		n	%		
1a	S Vlex <i>wh</i> -	7	15.6	35.6	95.6
1b	S Vbe <i>wh</i> -	9	20.0		
2a	<i>wh</i> - S Vlex	19	42.2	60.0	
2b	<i>wh</i> - S Vbe	8	17.8		
8	<i>wh</i> - ESK S V	0	0.0		
4	<i>c'est wh</i> -	2	4.4	4.4	4.4
5	<i>wh- c'est</i>	0	0.0	0.0	
	Total	45	100.0	100.0	100.0

French. The decrease could then well stem from the cognitive development suggested in Strik (2007). An intriguing fact then is the evolution of the ratio of in-situ *wh*-adverbs. They first decrease and then increase during the period (36.1%–29.7%–40.0%). Since the last year is not significantly different from the first year, there does not seem to be a developmental/cognitive constraint underpinning the ratio of in-situ/ex-situ *wh*-adverbs. We also note that despite the decrease in *wh*-in-situ pronouns in Figure 4, the observed decline is not significant either.

We then compared the evolution of the *wh*-positions – pronouns and adverbs – with the evolution of the verb forms – Fixed *be* and Free forms. It seems that the decrease in *wh*-in-situ pronouns parallels the decrease in Fixed *be* form, and that the decrease-increase in in-situ *wh*-adverbs mirrors the Free verb forms pattern, which means that the ex-situ *wh*-adverb pattern parallels it. The evolution of the in-situ/ex-situ ratios in pronoun and adverb *wh*-positions could then all be accounted for within the Verb form hypothesis that opposes the Fixed *be* form to All Free forms (Section 3.2).

In order to test the parallels observed in Figure 4, we examined the position of the adverb and pronoun *wh*-words as a function of the Verb form (Fixed *be* form vs. Free *be* forms vs. Lexical verbs). Table 7 shows that adverb *wh*-words ($n = 45$; *où* ‘where’, *comment* ‘how’, *quand* ‘when’, *combien* ‘how many’) massively co-occur with Free verb forms in the third year (95.6%; Structures 1a, 1b, 2a and 2b), similarly to the first two years (99.6%; same structures in Palasis et al., 2019). In the third year, two adverbs only appear with the Fixed *be* form (*c'est comment* ? ‘how is it’ VI, 5;2 and *c'est quand* ? ‘when is it’ CA, 5;3). The Verb form hypothesis predicts that we should not find a significant difference in the position of the *wh*-word when comparing Structures 1b and 2b (Free *be* forms) with Structures 1a and 2a (Lexical verbs), since the latter are also free verb forms.

We repeated the series of analyses carried out on the first two years on the third year. We examined the independence between the *wh*-position and the verb form with *wh*-adverbs. Only 11 children had sufficient data for this test ($n = 41$). The CMH test was carried on Lexical verbs vs. Free *be* forms, which represent 95.6%

Table 8. Distribution of *wh*-pronouns according to syntactic structure (Year 3)

		n	%		
1a	S Vlex <i>wh</i> -	19	16.8	16.8	53.1
1b	S Vbe <i>wh</i> -	0	0.0		
2a	<i>wh</i> - S Vlex	0	0.0	36.3	
2b	<i>wh</i> - S Vbe	0	0.0		
8	<i>wh</i> - ESK S V	41	36.3		
4	<i>c'est wh</i> -	51	45.1	45.1	46.9
5	<i>wh- c'est</i>	2	1.8	1.8	
Total		113	100.0	100.0	100.0

of the *wh*-adverb questions in Year 3. As expected, the result is not significant (X^2 CMH = 0.52, 1 *df*, $p = .471$). This result confirms that, as in the first two years and as in Section 2.3 for the third year, the *wh*-position is not significantly different when comparing Free *be* forms and Lexical verbs (also free forms) in this subgroup of adverbial *wh*-words. We conclude that adverbs are mainly ex-situ throughout the period (Year 1: 63.9% and Year 2: 70.3% in Palasis *et al.* 2019; Year 3: 60.0% in Table 7), because they massively and steadily co-occur with Free verb forms (Free *be* forms and Lexical verbs) in all three years.

Table 8 shows that *wh*-pronouns co-occur with Free and Fixed verb forms (53.1% and 46.9%, respectively), but that nearly half the pronouns (45.1%) appear in one configuration: in-situ with the Fixed *be* form (Structure 4). The rest of the pronouns mainly co-occur with Lexical verbs, either ex-situ with the *est-ce que* interrogative marker (Structure 8: 36.3%) or in-situ (Structure 1a: 16.8%). The *wh*-pronoun *quoi/que* 'what' is overwhelming ($n = 105$), and there are a few *qui* 'who' ($n = 5$), *à quoi* 'to what' ($n = 2$), and *à qui* 'to whom' ($n = 1$). Since *wh*-pronouns never appear with Free *be* forms, we examined the relationship between the position of the *wh*-pronouns with the Fixed *be* form as opposed to Lexical verbs only. The CMH test was performed on the data in 15 children ($n = 112$) and showed a significant result (X^2 CMH = 27.598, 1 *df*, $p < .00001$). This result confirms that the *wh*-items are more in-situ with the Fixed *be* form than with Lexical verbs in this subgroup of pronominal *wh*-words.

Taken together, these results seem to show that the grammatical category (adverb vs. pronoun) is not a discriminating variable for the position of the *wh*-phrase. We suggest that the in-situ/ex-situ ratios of *wh*-pronouns and *wh*-adverbs and the well-documented discrepancy with adult data pertain to the ongoing development of the verbal system of the children, who gradually flesh their system out by replacing *c'est* forms with lexical verbs.

In addition to this gradual verb diversification, the new data examined in this contribution draw our attention to a second developmental phenomenon. Indeed, we saw in Section 3.1 that there were no significant differences in the in-situ/ex-situ distributions between the three years, except for the *wh*-word *quoi/que* 'what' that showed a significant decrease in the in-situ position

Table 9. Outliers in the dataset according to *wh*-word and syntactic structure

		<i>quoi/que</i>	<i>où</i>	<i>qui</i>	<i>comment</i>	<i>à qui</i>	<i>à quoi</i>	<i>combien</i>	Total
1a	S Vlex <i>wh</i> -	69	14	12	7	0	1	2	105
1b	S Vbe <i>wh</i> -	0	57	0	1	1	0	0	59
5	<i>wh</i> - <i>c'est</i>	12	0	1	0	5	0	0	18
	Total	81	71	13	8	6	1	2	182

between Year 2 and Year 3 (from 73.5% to 59.0%; Fisher's Exact Test for Count Data: $p < .05$). Interestingly, the 'what' pronoun in French is the only *wh*-word with different morphological forms according to the position (i.e., in-situ *quoi* vs. ex-situ *que*). Moreover, the form of the pronoun varies between Colloquial French, which displays *qu'est-ce que* 'what is it that' without subject inversion in the question (e.g., *qu'est-ce qu'il a dit ?* 'what did he say'), and Standard French, which displays *que* 'what' with subject inversion (e.g., *qu'a-t-il dit ?*). The data thus illustrate that the children are fleshing out their *wh*-word inventory by gradually integrating the alternation between in-situ *quoi* and ex-situ *qu'est-ce que* 'what is it that', which results in the latter being more frequent over time. This development together with the verb diversification seem to account nicely for the decrease in the overall *wh*-in-situ ratio (Figure 3), the decrease in the pronoun *wh*-in-situ ratio (Figure 4), and the absence of decrease in the adverb *wh*-in-situ category (Figure 4).

4. OUTLIERS TO THE CORRELATION BETWEEN FIXED *BE* FORM AND *WH*-IN-SITU AND DEICTIC HYPOTHESIS

Finally, we listed the outliers to the correlation between the Fixed *be* form *c'est* 'it is' and the *wh*-in-situ position, that is the questions that display Free verb forms and *wh*-in-situ (Structures 1a and 1b) and the Fixed *be* form *c'est* 'it is' with the *wh*-ex-situ position (Structure 5). The total number of outliers amounted to 182 (25.3% of the total dataset, see Table 9).

In this section, we will examine the *wh*-in-situ questions only (Structures 1a and 1b, $n = 164$, 22.8% of the total dataset) in order to seek variables that could account for the unexpected *wh*-in-situ position in these structures. We explore two hypotheses:

- Outliers share with the Fixed *be* structure (Structure 4) a feature that triggers the usage of a *wh*-in-situ instead of a *wh*-ex-situ.
- Outliers exhibit types of *wh*-in-situ that are different from those found in the Fixed *be* questions.

Taking a closer look at the corpus, we observed that it is replete with in-situ questions such as (5) and (6).

Table 10. Distribution of *là* in *wh*-questions

	In-situ	Ex-situ	Total
Fixed <i>be</i>	13	0	13
Free <i>be</i>	11	3	14
Lexical verbs	18	13	31
Total	42	16	58

- (5) *Et là i(l) dit quoi?* (LA, 3;11)
and there he says what
'And there what does he say?'
- (6) *Là y a quoi?* (KE, 5;7)
there there is what
'What is there there?'

Note the presence of the adverb *là* 'there' in both cases. A survey of our corpus showed that *là* is surprisingly frequent among the outliers. We examined the presence of *là* in detail. Items like *celui-là* 'that one' or *ce chat-là* 'that cat' were excluded, since *là* is part of the discontinuous demonstrative morpheme *ce...là* and is hence not the adverb *là* (see 7a). Similarly, *là-dedans* 'there inside' and embedded versions like (7b) were not included.⁷

- (7) Presence of *là*:
- a. *Celui-là* 'that one':
Et celui-là ça va où lui? (NI, 3;0)
and that one that goes where him
'And where does that one go?'
- b. Embedded *là*:
Il est où le truc qui est là à sa vache euh? (MS, 3;0)
it is where the thing that is there to his cow uh
'Where is the thing that is there (and that belongs) to his cow uh?'

Based on the occurrences in Table 10, we ran a Fisher's Exact Test of independence. The result is significant ($p < .01$), which shows that the distribution of *wh*-in-situ and *wh*-ex-situ with respect to the presence of *là* is above chance and that there is a correlation between the presence of *là* and the presence of *wh*-in-situ.

Là is a complex marker in French. Although it is originally a locative adverb meaning 'here' or 'there', it developed a range of other deictic uses, including temporal, narrative, discursive and interactional ones. In its temporal and narrative uses, it serves as a linker between two parts of a story. This use is not

⁷Likewise, Kleiber (1995: 10) hypothesizes that these items have a different reference mode from lone *là* and treats them separately.

featured in our questions. As a discursive and interactional marker, *là* is used by the speaker to draw the addressee's attention to her/his discourse and her/his utterance. According to Smith (1995), *là* is the least specified of the French locative adverbs (in contrast with *ici* 'here' and *là-bas* 'over there'). It is characterized by the features [-far], [-speaker], which account for its local as well as its discourse usages (by default, it serves to involve the addressee). In contrast, Kleiber (1995: 23) suggests that *là* refers to something that is given, manifest or accessible. When it refers to a location, spatial referentiality is already active in the context. Note however that the two approaches are not incompatible. To draw the attention of the addressee to an object, this object must be salient or at least already accessible.

Be that as it may, the main questions are to discover what in *là* favours *wh*-in-situ, and if *là* and Fixed *be* structures have anything in common. To discuss the matter, it is useful to go back to what was introduced in Section 1. Recall that the target (adult) language is complex with respect to the *wh*-in-situ phenomenon, since it features three types of *wh*-questions (Baunaz, 2011, 2016). Type 1 is a neutral information-seeking question, without presupposition. Type 2 is presuppositional-partitive (with rise-fall intonation on the *wh*-phrase). Type 3 is presuppositional-specific (the *wh*-phrase forms its own prosodic phrase; Vergnaud and Zubizarreta, 2005).⁸ In adult speech, Faure and Palasis (2021) showed that *wh*-ex-situ is favoured by the presence of a contrastive feature whose specific semantics is exclusive, meaning that the question conveys the presupposition that at least one of the contextually available answers is a priori excluded.

Let us now compare our findings on child speech with this target situation. In Sections 1 and 3, we mentioned and confirmed one factor favouring *wh*-in-situ in child French: the presence of the Fixed *be* form *c'est* 'it is'. Palasis et al. (2019: 230) additionally suggested that the use of *wh*-ex-situ was favoured when the children's questions displayed a verb form with more semantic content (i.e., Free *be* forms and Lexical verbs). In the target adult language, more semantic content creates more context, in which a presupposed set can be identified. Such a situation is unlikely in a sentence with *c'est qui/quoi*, because it asks for identification or definition, and most of the time any answer is possible. Once again comparing with adult speech, this category of *wh*-in-situ sentences corresponds to Type 1 in Baunaz's (2011, 2016) classification.

What *c'est* and *là* share is that they display a deictic element (*c'* and *là*) linked to the *hic et nunc*. They pick out of the immediate context an element or a situation that was not yet in the scope of the attention of the addressee and thus initiate the identificational or definitional process that underpins the interrogative speech act. So, no relation is established with the previous context or a presupposed set, a necessary condition to have Types 2 or 3 *wh*-in-situ or *est-ce que*-less *wh*-ex-situ

⁸Here are Baunaz's (2016: 134 and 137) definitions (The reader is referred to the original article for details): 'A partitive *wh*-phrase is an object, which belongs to a presupposed set containing more objects. Each of the objects of the set can potentially be referents to the answer of the *wh*-phrase, i.e., all are alternatives.'

'Specificity narrows down the context to familiar individuals, excluding alternatives. A constituent question involving specificity entails an answer referring to a familiar individual that the interlocutor has in mind.'

(at this stage, children do not produce *wh-est-ce que*, apart from *qu'est-ce que* 'what is it that', as mentioned in Section 3.3).

Other data go in the same direction. In (8), the child begins the utterance with *là* and goes on with a *wh-ex-situ* question, but hesitates, stops and corrects the sentence to finally produce a *wh-in-situ* question, as expected under our analysis.

- (8) *Et là [/] et là où il est [/] il est où?* (MT, 4;4)
and there and there where he is he is where
'And there where is he?'

Consequently, this new dataset prominently features Type 1 *wh-in-situ*. Semantically, it does not carry a presupposition; syntactically, it does not feature fronting/movement of the *wh*-phrase. Although it is mostly materialised with *c'est* and *là* questions, unmarked Type 1 questions also occur in other guises, such as (9).

- (9) Absence of *là*:
Lui i(l) mange quoi? (MS, 3;9)
him he eats what
'What does he eat?'

Type 1 is most likely to appear first in child speech, since it is the least marked question type in the target language. Nevertheless, the prevalence of Type 1 is not sufficient to explain all the data, and the examples in (10) and (11) do not abide by the description of Type 1 but belong to Type 2 (so far, we have found no examples of Type 3). In each case, a presupposed set is given from which the answer must be picked out.

- (10) The children must pick an item out of a set of books and games:
Maintenant je prends quoi alors? (WI, 2;10)
now I take what so?
'So what do I take now?'
- (11) The children are discussing the different activities in a playground they are all familiar with:
T(u) as fait quoi? (meaning 'which activity'; LU, 5;4)
you have done what
'What have you done?'

Another interesting fact is that, while the proportion of *wh-in-situ* decreases from Year 1 to Year 3 (Figure 3), the proportion of outliers increases, particularly with lexical verbs (Table 11).

Summarizing our findings in this section, we found that *wh-in-situ* structures come in two types in child speech, corresponding to Type 1 and Type 2 questions in Baunaz's (2011, 2016) description of French adult speech. Type 1 is prevalent. It is a type of mere information-seeking questions, which take the form of Fixed *be* and *là* structures in child speech. In fact, these deictic structures are the most suitable to encode the operation of identification that

Table 11. Distribution of outliers among *wh*-in-situ (%)

	Lexical verbs	Free <i>be</i>	Total
Year 1	19.4	15.3	34.7
Year 2	26.3	10.5	36.8
Year 3	29.6	10.2	39.8
Total	23.1	13.1	36.2

underlies the interrogative process. Nevertheless, Type 2 (partitive questions) is also present. It may be tied to the other observation that we made, namely that the number of outliers with lexical verbs increases through time. A possibility is that children gradually develop more complex types of *wh*-in-situ, such as Type 2, which leads to a growing usage of this type and to the increase of the number of alleged outliers. This hypothesis is consistent with the fact that children gradually use structures with more material, a factor favouring the linking to the context and the construction of presupposed sets. We leave the exploration of this hypothesis for future research.

5. CONCLUSIONS

This article examined the production of matrix information-seeking *wh*-questions in 16 typically developing French children during their three kindergarten years ($n = 718$, Table 3). We aimed to explain the prevalence of the in-situ *wh*-position in child data compared to adult data and to disentangle the various factors that could explain the in-situ/ex-situ ratio and its evolution during the three-year period.

In Section 3.1, we observed that *wh*-in-situ questions remain prevalent during the period despite a steady increase in *wh*-ex-situ (Figure 3). There are no significant differences in the in-situ/ex-situ distributions between the three years. We further noticed that there are no significant differences in the in-situ/ex-situ distributions of each *wh*-item considered separately between the three years, except for the *wh*-word *quoi/que* ‘what’ that shows a significant decrease in the in-situ position between Year 2 and Year 3.

In Section 3.2, we statistically examined the position of the *wh*-phrases as a function of the Verb form in Year 3 (Fixed *be* forms vs. Free *be* forms vs. Lexical verbs, which are also free forms, following Palasis et al., 2019 in Year 1 and 2). The Verb form (Fixed vs. Free) is also a discriminating variable for the *wh*-position in Year 3: The Fixed *be* form favours the in-situ position (e.g., *c’est qui Taz ?* ‘who is Taz’), whereas the Free forms favour the ex-situ position (e.g., *combien ça coûte ?* ‘how much does it cost’). Taken together, these results allow us to make the prediction that the ex-situ rate will steadily increase as a function of the diversification of the child verbal system. We leave this matter for future research.

In Section 3.3, we investigated the well-documented asymmetry between in-situ *wh*-pronouns and ex-situ *wh*-adverbs. We confirmed that the asymmetry also

pertains to the Verb form. *Wh*-pronouns are more in-situ with the Fixed *be* form *c'est* 'it is' than with the Lexical verbs. *Wh*-adverbs are mainly ex-situ, because they massively co-occur with Free verb forms. In line with the prediction on the diversification of the child verbal system, we expect that *wh*-pronouns will be more ex-situ, when children produce fewer Fixed *be* forms and more Lexical verbs. Furthermore, the evolution of the data between Year 2 and Year 3 illustrates that the children are fleshing out their *wh*-word inventory by gradually integrating the alternation between in-situ *quoi* and ex-situ *qu'est-ce que* 'what is it that'. This development together with the verb diversification seem to account for the decrease in the overall *wh*-in-situ ratio (Figure 3), the decrease in the pronoun *wh*-in-situ ratio and the absence of decrease in the adverb *wh*-in-situ category (Figure 4).

Finally, we investigated the outliers to the Verb form hypothesis (i.e., the *wh*-in-situ questions with Free verb forms; 22.8% of the dataset) in Section 4. We observed the presence of the adverb *là* 'there' in the outliers and elaborated on the fact that *là* and the Fixed form *c'est* share deictic properties. Generalising, Fixed *be*, *là* and other (unmarked) questions all belong to Baunaz's (2011, 2016) Type 1 questions in the target adult language (i.e., mere information-seeking questions). We also observed that, while *wh*-in-situ questions tend to decrease throughout the three years, *wh*-in-situ outliers increase, which we tentatively attributed to the emergence of Type 2 (and maybe Type 3) *wh*-questions, which rest on a set of presupposed possible answers. This hypothesis is left for future research.

Another important field that requires future attention is the behaviour of ex-situ *wh*-questions in child speech. The verb diversification and the increase in the *wh*-pronoun *qu'est-ce que* 'what' pertain to this domain of investigation. Some Fixed *be* and *là* structures that unexpectedly exhibited ex-situ *wh*-items, though rarely ($n = 20$, 2% of the corpus), also require attention. More generally, the behaviour of D-linking and *wh*-ex-situ in child speech largely remains to be explored in future research.

Competing interests. The authors declare none.

Acknowledgements. We thank the editors and three anonymous reviewers for their very constructive comments on previous versions of this contribution. We also thank the audience at the Workshop 'L'interrogative *in situ* : aspects formels, pragmatiques et variationnels' organised by Caterina Bonan, Laurie Dekhissi and Alexander Guryev in June 2022.

References

- Adli, A. (2006). French *wh*-in-situ Questions and Syntactic Optionality: Evidence from Three Data Types. *Zeitschrift für Sprachwissenschaft*, 25, 163–203.
- Bassano, D., Eme, P.-E. and Champaud, C. (2005). A naturalistic study of early lexical development: General processes and inter-individual variations in French children. *First Language*, 25(1), 67–101.
- Baunaz, L. (2011). *The Grammar of French Quantification*. Dordrecht: Springer.
- Baunaz, L. (2016). French "quantifiers" in questions: Interface strategies. *Acta Linguistica Hungarica*, 63(2), 125–168.
- Bayer, J. and Cheng, L. L.-S. (2017). *Wh*-in-situ. In M. Everaert and H. Van Riemsdijk (eds), *The Wiley Blackwell Companion to Syntax* (Vol. 8: U–Z, pp. 4979–5022). Hoboken, NJ: Wiley Blackwell.

- Becker, M. and Gotowski, M.** (2015). Explaining children's *wh*-in situ questions: Against economy. In E. Grillo and K. Jepsen (eds), *Proceedings of the 39th annual Boston University Conference on Language Development [BUCLD 39]* (pp. 88–100). Somerville, MA: Cascadilla Press.
- Biezma, M.** (2020). Non-informative assertions: The case of non-optional *wh*-in-situ. *Semantics and Pragmatics*, 13(18), <https://doi.org/10.3765/sp.3713.3718>.
- Boeckx, C.** (2000). Decomposing French Questions. In J. Alexander, N.-R. Han and M. Minnick (eds), *Proceedings of the 23rd Annual Penn Linguistics Colloquium* (Vol. 6, pp. 69–80). Philadelphia, PA: University of Pennsylvania Working Papers in Linguistics.
- Chang, L.** (1997). *Wh-in-situ Phenomena in French*. (MA thesis), University of British Columbia.
- Cheng, L. L.-S. and Rooryck, J.** (2000). Licensing *wh*-in-situ. *Syntax*, 3(1), 1–19.
- Cheung, C. C.-H.** (2008). *Wh-fronting in Chinese*. (Ph.D dissertation), University of South California, Los Angeles.
- Clark, E. V.** (1978). Strategies for Communicating. *Child Development*, 49(4), 953–959.
- Cochran, W. G.** (1954). Some Methods for Strengthening the Common χ^2 Tests. *Biometrics*, 10(4), 417–451.
- Coveney, A.** (1989). Pragmatic constraints on interrogatives in spoken French. *York Papers in Linguistics*, 13, 89–99.
- Coveney, A.** (1995). The use of the QU-final interrogative structure in spoken French. *Journal of French Language Studies*, 5(2), 143–171.
- Cramér, H.** (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- Crisma, P.** (1992). On the Acquisition of *Wh*-Questions in French. *Geneva Generative Papers*, 0(1–2), 115–122.
- Déprez, V. and Pierce, A.** (1993). Negation and Functional Projections in Early Grammar. *Linguistic Inquiry*, 24(1), 25–67.
- Dryer, M. S.** (2013). Position of Interrogative Phrases in Content Questions. In M. S. Dryer and M. Haspelmath (eds), *The World Atlas of Language Structures Online* (<http://wals.info/chapter/93>). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Faure, R. and Palasis, K.** (2021). Exclusivity! *Wh*-fronting is not optional *wh*-movement in Colloquial French. *Natural Language & Linguistic Theory*, 39(1), 57–95. <https://doi.org/10.1007/s11049-020-09476-w>
- Gotowski, M. and Becker, M.** (2016). An information-structural account of children's *wh*-in situ questions in French. In L. Perkins, R. Dudley, J. Gerard, and K. Hitzenko (eds), *Proceedings of the 6th Conference on Generative Approaches to Language Acquisition North America [GALANA 2015]* (pp. 24–35). Somerville, MA: Cascadilla Press.
- Guasti, M. T.** (2000). An excursion into interrogatives in Early English and Italian. In M.-A. Friedemann and L. Rizzi (eds), *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics* (pp. 105–128). Harlow: Pearson Education.
- Guillaume, P.** (1927). Le développement des éléments formels dans le langage de l'enfant. *Journal de Psychologie*, 24, 203–229.
- Hammann, C.** (2006). Speculations about Early Syntax: The Production of *Wh*-Questions by Normally Developing French Children and French Children with SLI. *Catalan Journal of Linguistics*, 5, 143–189.
- Hamlouï, F.** (2011). On the role of phonology and discourse in Francilian French *wh*-questions. *Journal of Linguistics*, 47(1), 129–162.
- Higgins, F. R.** (1973/1979). *The pseudo-cleft construction in English*. (Ph.D dissertation; published by Garland), Massachusetts Institute of Technology, Cambridge, MA.
- Huang, C.-T. J.** (1982). Move *WH* in a language without movement. *The Linguistic Review*, 1, 369–416.
- Jakubowicz, C.** (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua*, 121(3), 339–351.
- Kleiber, G.** (1995). D'ici à là et vice versa ; pour les aborder autrement. *Le Gré des langues*, 8, 8–27.
- MacWhinney, B.** (2000). *The CHILDES Project. Tools for Analyzing Talk: Transcription Format and Programs* (3rd edn, vol. 1). Mahwah, NJ: Lawrence Erlbaum Associates, <http://childes.talkbank.org/>.
- Mantel, N. and Haenszel, W.** (1959). Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *Journal of the National Cancer Institute*, 22(4), 719–748.
- McDonald, J. H.** (2011). *Myths of Human Genetics*. Baltimore, MD: Sparky House Publishing.

- Myers, L. L. and Pellet, S.** (2014). *Pourquoi* in spoken French. In S. Katz Bourns and L. L. Myers (eds), *Perspectives on Linguistic Structure and Context* (pp. 157–182). Amsterdam: John Benjamins.
- Nguyen, A. D. and Legendre, G.** (2022). The acquisition of *wh*-questions: Beyond structural economy and input frequency. *Language Acquisition*, 29(1), 79–104. <https://doi.org/10.1080/10489223.2021.1968867>
- Obenauer, H.-G.** (1994). *Aspects de la syntaxe A-barre. Effets d'intervention et mouvements des quantifieurs*. (Ph.D dissertation), Université de Paris VIII.
- Oiry, M.** (2011). *L'acquisition des Questions à Longue-Distance par des enfants français. Stratégie à dépendance directe versus indirecte et questions alternatives*. Editions Universitaires Européennes.
- Palasis, K.** (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies*, 23(1), Special Issue 'L'hypothèse d'une diglossie en France', Massot, Benjamin and Paul Rowlett (eds), 17–35. <https://doi.org/10.1017/S0959269512000348>
- Palasis, K., Faure, R. and Lavigne, F.** (2019). Explaining variation in *wh*-position in child French: A statistical analysis of new seminataturalistic data. *Language Acquisition*, 26(2), 210–234. <https://doi.org/10.1080/10489223.2018.1513004>
- Pinheiro, J. C. and Bates, D. M.** (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Prévost, P.** (2009). *The Acquisition of French: the Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. Amsterdam: John Benjamins.
- Rasetti, L.** (2003). *Optional Categories in Early French Syntax: A Developmental Study of Root Infinitives and Null Arguments*. (Ph.D dissertation), Université de Genève.
- Rizzi, L.** (2001). On the position “Int(errogative)” in the left periphery of the clause. In G. Cinque and G. Salvi (eds), *Current Studies in Italian Syntax: Essays offered to Lorenzo Renzi* (pp. 287–296). Oxford: Elsevier.
- Roeper, T. and De Villiers, J. G.** (2011). The Acquisition Path for *Wh*-Questions. In J. G. De Villiers and T. Roeper (eds), *Handbook of Generative Approaches to Language Acquisition* (pp. 189–246). Dordrecht: Springer.
- Roy, I.** (2013). *Nonverbal Predication: Copular Sentences at the Syntax-Semantics Interface*. Oxford: Oxford University Press.
- Smith, J. C.** (1995). L'évolution sémantique et pragmatique des adverbes déictiques “ici”, “là” et “là-bas”. *Langue française*, 107, 43–57.
- Strik, N.** (2007). L'acquisition des phrases interrogatives chez les enfants francophones. *Psychologie française*, 52(1), 27–39.
- Strik, N.** (2012). *Wh*-questions in child bilingual acquisition of French: Derivational complexity and cross-linguistic influence. *Canadian Journal of Linguistics*, 57(1), 133–151.
- Strik, N. and Pérez-Leroux, A. T.** (2011). Jij doe wat girafe?: *Wh*-movement and inversion in Dutch-French bilingual children. *Linguistic Approaches to Bilingualism*, 1(2), 175–205.
- Stromswold, K.** (1995). The Acquisition of Subject and Object *Wh*-Questions. *Language Acquisition*, 4(1–2), 5–48.
- Thiberge, G.** (2020). *Acquisition et maîtrise des interrogatives partielles en français : La variation sociolinguistique comme outil interactionnel*. (Ph.D dissertation), Université de Paris.
- Van Kampen, J.** (1997). *First steps in wh-movement*. Delft: Eburon.
- Vergnaud, J.-R. and Zubizarreta, M. L.** (2005). The representation of focus and its implications: towards an alternative account of some ‘intervention effects’. In J. Koster, H. Van der Hulst and H. Van Riemsdijk (eds), *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk* (pp. 641–660). Berlin & New York: Mouton de Gruyter.
- Zuckerman, S. and Hulk, A.** (2001). Acquiring optionality in French *wh*-questions: An experimental study. *Revue québécoise de linguistique*, 30(2), 71–97.

Cite this article: Palasis K, Faure R, and Meunier F (2023). *Wh*-in-situ in child French: Deictic triggers at the syntax-semantics interface. *Journal of French Language Studies* 33, 273–298. <https://doi.org/10.1017/S0959269523000030>