



HAL
open science

Le développement de l'expression du genre grammatical en français L2

Amanda Edmonds

► **To cite this version:**

Amanda Edmonds. Le développement de l'expression du genre grammatical en français L2 : les cas de chose, famille, maison, nourriture et vie. *Langage, Interaction et Acquisition / Language, Interaction and Acquisition*, 2019, 10 (2), pp.229-254. 10.1075/lia.18010.edm . hal-03063654

HAL Id: hal-03063654

<https://hal.science/hal-03063654>

Submitted on 9 Feb 2021

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Le développement de l'expression du genre grammatical en français L2 : les cas de *chose*, *famille*, *maison*, *nourriture* et *vie*

Résumé

Les recherches portant sur l'acquisition de l'expression du genre grammatical en français langue seconde (L2) nous montrent, entre autres, que les apprenants réussissent plus souvent à marquer le genre avec un nom masculin (vs. un nom féminin) et sur les déterminants plutôt que sur les adjectifs. En revanche, il n'est pas clair si ces tendances générales s'appliquent de la même manière à l'ensemble des noms, adjectifs et déterminants concernés. En d'autres termes, nous pouvons nous demander à quel point l'expression du genre grammatical est déterminée par des spécificités lexicales. Est analysée dans cette étude l'évolution du marquage en genre produit par un groupe de 20 apprenants britanniques du français L2 sur une période de 21 mois. Des productions orales et écrites produites à trois périodes différentes ont été analysées, et le marquage en genre avec cinq noms féminins fréquents – *chose*, *famille*, *maison*, *nourriture*, *vie* – a fait l'objet d'une étude approfondie. Les résultats montrent que l'expression du genre grammatical se développe différemment avec ces cinq noms, résultat qui est interprété à l'aide des approches basées sur l'usage.

Mots clés : genre grammatical, spécificités lexicales, données longitudinales, approches basées sur l'usage

1. Introduction

Les recherches sur l'acquisition des langues secondes (L2) ont été pendant longtemps dominées par une focalisation sur des aspects morphologiques et syntaxiques de la nouvelle langue (Mitchell, Myles, & Marsden, 2013 : 6). Parmi ces recherches, des études sur l'acquisition du genre grammatical relevant de cadres théoriques variés (ex. générativisme, énonciativisme, connexionisme) sont nombreuses. Les chercheurs responsables de ces travaux divers s'accordent généralement sur le fait que l'acquisition du genre grammatical présente souvent des difficultés, et qu'elle est caractérisée par certaines régularités. Par exemple, en français L2, nous savons que les apprenants produisent un marquage du genre conforme à la norme plus souvent avec des noms masculins qu'avec des noms féminins, et que le marquage du genre est plus souvent correct sur les déterminants que sur les adjectifs. Si ces observations nous apprennent des choses importantes sur l'acquisition du genre grammatical en tant que système, les éventuelles spécificités lexicales y restent masquées. Et pourtant, des chercheurs comme Diessel (2016) et Sinclair (1991) soulignent la place importante des spécificités lexicales dans la grammaire. Pour le genre grammatical en français L2 plus particulièrement, Granfeldt (2005 : 181) affirme que « le développement de l'expression correcte du genre sur un type nominal particulier suit un parcours individuel. » Le nom pourrait donc constituer une unité d'analyse intéressante pour observer le développement du genre grammatical. Dans le présent article, je propose une analyse de l'évolution du marquage du genre grammatical tel qu'il est exprimé avec cinq noms féminins – *chose*, *famille*, *maison*, *nourriture*, *vie* – fréquemment utilisés dans le corpus étudié. Les données analysées dans cette étude proviennent du corpus longitudinal LANGSNAP, qui contient des productions orales et écrites recueillies auprès d'apprenants britanniques du français L2 avant, pendant

et après un séjour de 9 mois en France (Mitchell, Tracy-Ventura, & McManus, 2017).¹ L'analyse proposée prend appui sur les approches basées sur l'usage (*usage-based approaches*) pour examiner l'utilisation de ces cinq noms dans un corpus de référence et par les apprenants eux-mêmes. À travers une comparaison avec l'expression du genre pour l'ensemble des noms du corpus LANGSNAP, ainsi qu'une mise en lumière de l'évolution de l'expression du genre dans le temps, je tâcherai de démontrer la pertinence d'une analyse qui prend l'item lexical comme entrée pour une meilleure compréhension de l'acquisition L2 de ce phénomène morphosyntaxique.

2. Le cadre théorique et les recherches précédentes

Dans cette partie, je commencerai par une brève description du cadre théorique adopté dans cette étude, à savoir les approches basées sur l'usage. Je présenterai ensuite le genre grammatical en français. Cette présentation sera suivie d'un résumé des travaux portant sur l'expression du genre grammatical en français L2, avec une attention particulière portée aux résultats en lien avec des spécificités lexicales. Pour finir, je présenterai la question de recherche qui motive l'étude actuelle.

2.1 Les approches basées sur l'usage

Depuis la publication, en 2002, du numéro spécial de *Studies in Second Language Acquisition* sur *frequency effects* dans l'acquisition d'une L2, coordonné par Nick Ellis, les approches basées sur l'usage se sont installées parmi les théories dominantes dans le domaine de l'acquisition L2. Selon Tummers, Heylen, et Geeraerts (2005), le terme *approches basées sur l'usage* couvre en réalité de nombreuses théories et approches, allant de la grammaire de construction à l'émergentisme, en passant par certaines approches de l'analyse de l'acquisition d'une L1 (notamment les recherches de Tomasello et ses collègues). Dans leur discussion de cette famille d'approches, les auteurs soulignent plusieurs points de convergence, notamment (a) l'accent qu'elles mettent sur l'utilisation du langage (et, donc, sur l'importance d'étudier comment les locuteurs parlent, écrivent et comprennent), (b) la non-distinction entre *performance* et *compétence* et (c) le principe que le lexique et la grammaire existent sur un continuum, et non en tant que deux modules entièrement séparés.

Les deux premières caractéristiques des approches basées sur l'usage évoquées par Tummers et al. (2005: 228) révèlent un positionnement épistémologique selon lequel "l'attention se déplace automatiquement de la connaissance du système linguistique vers les processus qui sous-tendent l'emploi du langage et les processus de communication." En d'autres termes, pour les chercheurs travaillant dans ces approches, il n'y a pas de dichotomie entre une compétence abstraite et une utilisation concrète de la langue : ce sont les usages qui sont au centre de ces analyses, et c'est à travers l'étude des usages langagiers que les systèmes langagiers se révèlent. En discutant de l'application de ces approches aux recherches sur l'acquisition des L2, Tyler et Ortega (2018 : 6) écrivent que « il n'y a pas de règles innées, universelles et profondes à découvrir lors de l'apprentissage

¹ <http://langsnap.soton.ac.uk/>

d'une langue. Au contraire, la langue que construit l'apprenant est étroitement fondée sur la langue à laquelle il est exposé ; il s'agit essentiellement d'un phénomène de surface, mû par le traitement du sens. » Dans ces approches, l'apprentissage dépend crucialement de l'input, et les caractéristiques de celui-ci peuvent éclairer certains itinéraires acquisitionnels (Ellis, Römer, & O'Donnell, 2016 ; Eskildsen & Cadierno, 2007, 2015 ; Jach, 2018 ; Wulff, Ellis, Römer, Bardovi-Harlig, & Leblanc, 2009). La fréquence (d'un son, d'un mot, d'une structure) est une caractéristique de l'input très souvent étudiée. S'il est généralement accepté qu'une fréquence élevée facilite l'acquisition, elle ne détermine pas, à elle seule, celle-ci (Ellis & Wulff, 2015 : 84), car d'autres facteurs (ex. la saillance, la prototypicalité) rentrent en jeu. En plus, il n'existe, pour l'instant, aucun consensus sur comment définir et mesurer la fréquence (voir Linford, Long, Solon, & Geeslin, 2016). Pour ne prendre qu'un seul exemple, la fréquence examinée dans la plupart des études dans les recherches sur l'acquisition des L2 porte sur la fréquence de l'input ambiant auquel l'apprenant pourrait avoir accès. La fréquence, dans ce cas-là, est le plus souvent calculée à l'aide d'un corpus de référence (ex. Ellis et al., 2016). Il s'agit donc d'une approximation du « le langage auquel est exposé l'apprenant » mentionné par Tyler et Ortega. Or, comme souligné par Tummers et al. (2005), les productions du locuteur lui-même font également partie de son input. Ces auteurs parlent de *usage events* (des emplois individuels) qui « définissent et redéfinissent de façon continue le système linguistique sur un mode dynamique, et de ce fait, chaque emploi individuel contribue à redéfinir subtilement le système linguistique propre à chaque personne » (p. 228). Dans la présente étude, je ferai appel à la fréquence pour interpréter mes données et, en ce faisant, j'adopterai deux points de vue différents sur cette caractéristique de l'input : l'un extérieur, à l'aide d'un corpus de référence, et l'autre intérieur, pour lequel je m'intéresserai aux *usages events* des apprenants eux-mêmes.

2.2 Le genre grammatical en français

Dans les recherches sur le genre grammatical en français, une distinction est généralement faite entre l'attribution du genre (caractéristique du nom lui-même) et l'accord en genre (réalisation morphosyntaxique du genre grammatical). En ce qui concerne l'attribution du genre, chaque nom en français est soit masculin, soit féminin, avec certains noms (des épécènes, tel qu'*artiste*) pouvant avoir les deux genres. A l'exception de certains noms renvoyant à une entité avec un genre biologique (ex. *copine, copain*), le genre d'un nom n'est pas prévisible à partir de sa signification (cf. Nelson, 2005). Si l'attribution du genre grammatical pour la vaste majorité des noms français ne dépend pas d'une distinction sémantique, il a été suggéré que la terminaison orthographique ou phonétique d'un nom constitue un indice au genre du nom, et qu'autour de 80% des noms ont une terminaison qui prédit correctement son genre (Lyster, 2006 ; Tucker, Lambert, & Rigault, 1977). L'importance de la terminaison nominale pour l'attribution du genre ne fait pourtant pas l'unanimité (voir Carroll, 1989), et une étude récente pour laquelle un groupe de 168 locuteurs natifs devaient attribuer le genre à 295 noms fréquents et rares révèle de la variabilité dans l'attribution du genre grammatical, même chez des natifs (Ayoun, 2018).

Le genre grammatical de chaque nom peut être manifeste, à travers l'accord en genre, sur les déterminants et sur les adjectifs (les deux cas qui m'intéresseront dans la présente étude), ainsi que sur les participes passés (Gaucher, 2015) et sur les pronoms (Prodeau & Carlo, 2002). Or, dans de nombreux cas, l'accord en genre n'est pas marqué de manière ouverte, car les formes masculine et féminine du modificateur sont identiques. Tel est le cas, par exemple, pour les articles définis (*les*) et indéfinis (*des*) au pluriel (voir Ayoun, 2007 : 146-147 pour des détails sur l'ensemble des déterminants), ainsi que pour les adjectifs qui se terminent en <e> : *agréable*, *pauvre*. En plus, l'accord en genre dans certains cas est purement orthographique et, donc, inaudible, comme dans *quel_{masc}/quelle_{fém}* et *ravi_{masc}/ravie_{fém}*. Avec les noms à initial vocalique, des phénomènes d'élision (*petit ami/petite amie*), de liaison (*l'ami/l'amie*) et des formes supplétives (*son ami/son amie*) font que le genre de ces noms sont encore moins souvent audible et/ou lisible sur les modificateurs. L'input en français ne fournit donc pas toujours d'indication ouverte – audible dans le flux phonétique ou visible à l'écrit² – sur le genre grammatical. De cette manière, "les apprenants de français L2 sont confrontés à un input complexe et ambigu en ce qui concerne le genre grammatical" (Ayoun, 2007: 147).

2.3 L'expression du genre grammatical en français L2

Les recherches sur l'acquisition du genre grammatical en français L2 sont caractérisées par : (a) des études psycholinguistiques examinant comment les locuteurs non-natifs réagissent à des réalisations du genre qui sont erronées ou conformes à la norme (ex. Foucart & Frenck-Mestre, 2012 ; Vatz, 2009), (b) des recherches sur les connaissances sur le genre, pour lesquelles les apprenants attribuent le genre à certains noms (ex. Hardison, 1992 ; Kupisch, Akpınar, & Stöhr, 2013 ; SurrIDGE & Lessard, 1984) et (c) des études qui s'intéressent au marquage du genre dans des productions des non-natifs. L'étude actuelle se situe dans ce troisième groupe, et ce sont ces études-là qui feront l'objet de la présentation qui suit.

Les résultats des études précédentes nous permettent de dégager plusieurs tendances dans l'expression du genre, dont les deux plus importantes sont résumées en (1) :

- (1) Les apprenants produisent plus d'instances d'accord non-conformes à la norme
 - a. avec les noms féminins qu'avec les noms masculins (Bartning, 2000 ; Dewaele & Véronique, 2001 ; Edmonds & Gudmestad, 2018 ; Edmonds, Gudmestad, & Metzger, sous évaluation ; Harley, 1979)
 - b. sur les adjectifs (vs. les déterminants) (Ayoun, 2007 ; Bartning, 2000 ; Dewaele & Véronique, 2001 ; Edmonds & Gudmestad, 2018 ; Edmonds et al., sous évaluation ; Granfeldt, 2005 ; Harley, 1979 ; Holmes & Dejean de la Bâtie, 1999)

Ces tendances générales doivent toutefois être nuancées. En effet, certaines études fournissent des observations qui suggèrent l'existence de spécificités lexicales en lien avec l'expression du genre

² Voir Ayoun (2010) pour une analyse détaillée du marquage du genre dans un corpus de français écrit.

grammatical en français L2 (au niveau du nom et/ou au niveau des modificateurs), et ce sont ces études qui seront présentées plus en détails.

Dans une étude sur la production orale en français L2, Bartning (2000) s'intéresse aux différences dans l'expression du genre par des apprenants suédophones selon que le marquage se fait sur l'article défini, indéfini ou possessif. C'est avec l'article possessif que les apprenants pré-avancés et avancés montrent le taux de marquage conforme à la norme le plus élevé. Quant au maniement du marquage avec les articles définis et indéfinis, Bartning ne trouve aucune différence chez les apprenants pré-avancés, alors que chez les avancés, le taux de réussite est plus élevé pour les articles définis. Dans son livre de 2001, Hawkins présente brièvement des données sur l'acquisition du genre grammatical en français L2 qui laissent apparaître des spécificités lexicales en fonction du nom. Des apprenants anglophones, qui avaient un niveau avancé en français, devaient regarder un dessin animé et ensuite raconter son histoire. Hawkins constate que les apprenants ont tendance à sur-généraliser un seul article défini et indéfini, qui peuvent être du même genre (ex. *le / un*) ou pas (ex. *le / une*). Dans une analyse générativiste des données, Hawkins suggère que l'interlangue de ces apprenants ne dispose pas d'un *gender feature*, et que les tendances observées dans les données peuvent être expliquées grâce à une interprétation lexicale. Plus précisément, les apprenants anglophones auraient une seule forme de chaque article qui serait utilisée par défaut ; avec le temps, ils pourraient apprendre des exceptions à cette sur-généralisation, un apprentissage qui se ferait nom par nom. Granfeldt (2005), pour sa part, a suivi l'évolution dans l'expression du genre grammatical chez des apprenants suédophones ; l'analyse qu'il propose des productions d'un des apprenants adultes – Karl – est particulièrement pertinente pour l'étude actuelle. Granfeldt analyse cinq entretiens avec Karl, étalés sur 8 mois. Cet apprenant a été débutant à son arrivée en France, et les entretiens ont commencé 8 mois après son installation. L'analyse révèle la tendance de Karl à utiliser un seul article défini (*le*) et indéfini (*une*) quand il commence à employer un nom. Granfeldt suggère que, pour Karl, ces articles n'expriment pas (pour l'instant) le genre grammatical, mais l'opposition défini/indéfini, et il démontre que cette sur-généralisation peut évoluer dans le temps vers un marquage du genre. Même si le nombre d'occurrences pour chaque nom est parfois limité, Granfeldt constate, à partir de ces données, « un schème différent d'acquisition qui se réalise d'abord au niveau des types [nominaux] individuels » (p. 185).

Dans l'ensemble, ces trois études identifient des particularités dans l'expression du genre grammatical selon le modificateur (Bartning, 2000) ou le nom impliqués (Granfeldt, 2005 ; Hawkins, 2001). Si Hawkins et Granfeldt affirment tous les deux que l'expression du genre varie en fonction du nom, seul Granfeldt illustre ses propos avec une analyse empirique (des 73 noms utilisés par Karl dans au moins deux entretiens). A l'exception de l'étude de cas de Karl, nous savons encore peu sur l'évolution dans le marquage du genre en fonction du nom et à travers le temps. Par ailleurs, si Hawkins et Granfeldt ont identifié des différences en fonction du nom, ces auteurs ne se sont pas intéressés au rôle joué par les caractéristiques de l'input dans l'acquisition de ce phénomène

morphosyntaxique. Dans la présente étude, je tâcherai de contribuer aux discussions sur l'acquisition du genre grammatical avec un projet qui porte sur la question de recherche suivante : L'évolution dans l'expression du genre grammatical diffère-t-elle en fonction du nom chez un groupe de 20 apprenants britanniques du français L2 et, si oui, quel rôle la fréquence joue-t-elle dans ces différences ?

3. La présente étude

3.1 Le corpus LANGSNAP

Le corpus LANGSNAP est né du projet de recherche *Social networks, target language interaction, and second language acquisition during the year abroad: A longitudinal study*, mis en place et organisé à l'Université de Southampton au Royaume-Uni. Les chercheurs ont suivi deux groupes d'apprenants britanniques, l'un spécialisé en français ($n = 29$), l'autre en espagnol ($n = 27$), sur une période de 21 mois, qui a inclus une année universitaire dans un contexte d'immersion. Six recueils de données ont été effectués : avant le départ, trois fois pendant leur séjour à l'étranger et deux fois après leur retour au Royaume-Uni. Pour chaque recueil, trois tâches de production ont été mises en place : (a) un entretien à l'oral qui a duré à peu près 20 minutes et qui portait sur des sujets tels que les projets de l'apprenant, son rapport à la langue française et aux Français, ses attentes pour son année à l'étranger, etc., (b) une tâche narrative pour laquelle le participant devait regarder un livre d'images et ensuite raconter l'histoire à l'oral et (c) un essai argumentatif (d'environ 200 mots). Ces données ont été transcrites par l'équipe de LANGSNAP et sont librement disponibles sur leur site.

3.2 Participants

Pour la présente étude, les productions orales et écrites de 20 apprenants britanniques du français L2 suivis par l'équipe du projet LANGSNAP ont été analysées. Le groupe est constitué de 18 femmes et 2 hommes, qui étaient tous inscrits en licence de français dans une université anglaise. Au début du projet, les participants avaient entre 19 et 21 ans ($M = 20$) et avaient étudié le français entre 6 et 20 ans ($M = 10,45$).³ Une mesure indépendante de leur niveau de compétence en français a été obtenue par le biais d'une tâche d'imitation provoquée (*elicited imitation task*), sur laquelle le score maximal possible était de 120 points (Tracy-Ventura, McManus, Norris, & Ortega, 2014). Les 20 participants ont obtenu des scores allant de 36 à 97 points ($M = 59,25$, $SD = 14,58$). Pendant l'année universitaire passée en France, ces individus ont choisi trois activités principales différentes : trois ont trouvé un stage, cinq sont partis en tant qu'étudiant Erasmus et les 12 restants travaillaient en tant qu'assistant de langue anglaise dans des écoles primaires et secondaires.

3.3 L'analyse des données

Pour la présente étude, je me suis limitée aux données collectées lors de trois des six recueils : les données pré-séjour (mai/juin 2011), les données recueillies lors de la troisième visite pendant le séjour en France (mai/juin 2012) et la dernière collecte des données, qui a eu lieu à peu près 8 mois après le

³ Le participant 100, qui avait 20 ans au début du projet, a déclaré avoir étudié le français depuis 20 ans. Selon le site web du projet LANGSNAP, ce participant « a fait état de quelques contacts avec le français à domicile ».

retour des apprenants au Royaume-Uni (janvier/février 2013). Dans cette analyse, j'utiliserai les termes *pré-séjour*, *pendant-séjour* et *post-séjour* pour faire référence à ces trois recueils. Pour les recueils pré-séjour et pendant-séjour, les mêmes supports ont été utilisés pour la tâche narrative (Cat Story) et pour l'essai argumentatif (*Penses-tu que les couples homosexuels ont le droit de se marier et d'adopter des enfants ?*). Pour les données post-séjour, la tâche narrative a été basée sur Brothers Story, et la question posée pour l'essai argumentatif a été *Penses-tu que, de manière à inciter les gens à manger sainement, on devrait taxer les boissons sucrées et les aliments gras ?* L'ensemble des données des trois recueils a été analysé, et chaque instance d'un nom modifié par un déterminant et/ou par un ou des adjectifs a été repérée. Au total, 11 218 déterminants et adjectifs ont été relevés dans les données orales, et 3 059 dans les données écrites. A peu près la moitié des exemples implique un marquage ouvert du genre grammatical, c'est-à-dire, l'utilisation d'un déterminant ou d'un adjectif dont la forme (et la prononciation, pour les données orales) change en fonction du genre : 5 992 à l'oral et 1 601 à l'écrit (voir Edmonds & Gudmestad, 2018 et Edmonds et al., sous évaluation pour des analyses de l'ensemble des données).

J'ai ensuite examiné le profil lexical du corpus, afin d'identifier les noms qui seraient au centre de la présente étude. Vu le grand nombre de noms différents ($n = 827$), j'ai décidé de me focaliser sur les noms féminins les plus fréquemment utilisés avec un déterminant ou un adjectif sur lequel le genre grammatical est ouvertement marqué. Le choix des noms féminins a été dicté par le taux de réussite bien moindre avec les noms féminins comparé aux noms masculins dans ce corpus : Alors que 90,3% (3 898/4 317) des instances d'accord avec un nom masculin sont conformes à la norme, seul 76,5% des 3 275 occurrences avec un nom féminin le sont. Les noms féminins apparaissant au moins 100 fois dans le corpus ont été retenus pour la présente analyse, et sont présentés dans le tableau 1.⁴ Ces cinq noms sont employés par les 20 apprenants un total de 806 fois avec un modificateur sur lequel le genre grammatical est ouvertement marqué (c'est-à-dire, audible et/ou visible). Le sous-échantillon qui en résulte couvre presque un quart de toutes les occurrences de marquage ouvert avec un nom féminin dans le corpus ($806/3\,275 = 24,6\%$).

Tableau 1. *Les noms féminins les plus fréquents du corpus.*

Nom	Pré-séjour	Pendant-séjour	Post-séjour	Total
<i>chose</i>	83	67	55	205
<i>maison</i>	99	85	18	202
<i>famille</i>	81	55	14	150
<i>vie</i>	63	54	17	134
<i>nourriture</i>	33	30	52	115

⁴ Ces emplois viennent des trois tâches (entretien, narration, essai). Alors qu'il est possible que la tâche influe sur l'expression du genre grammatical, la taille du corpus actuel ne me permet pas d'examiner ce facteur.

L'analyse se fera en trois étapes. Dans un premier temps, j'exposerai le taux de réussite dans l'expression du genre grammatical avec chacun des cinq noms, en comparant les résultats avec les tendances globales pour l'ensemble des noms masculins et féminins du corpus. Cette présentation révélera des différences en fonction du nom, et les deux dernières parties de l'analyse seront consacrées à une exploration de ces différences à l'aide de deux approches complémentaires de la fréquence dans l'input. J'aurai d'abord recours à une caractérisation de la fréquence d'utilisation des cinq noms ciblés dans un corpus de référence, caractérisation qui fournira une indication de l'input ambiant auquel les apprenants ont accès. Pour finir, je m'intéresserai aux *usage events*, c'est-à-dire à comment les 20 apprenants utilisent eux-mêmes les cinq noms.

Avant de présenter les résultats, il est important de préciser il ne sera question ni d'attribution, ni d'accord en genre dans cette analyse. Au lieu de cela, les termes *marquage* ou *expression* seront employés, décision motivée à la fois par le type de données examinées dans cette étude (des données de production) et par le cadre théorique adopté. En effet, dans des données de production, un marquage non-conforme à la norme peut résulter de deux sources : l'attribution erronée du genre au nom ou une difficulté dans la réalisation morphosyntaxique de l'accord en genre. Or, il n'est pas chose aisée, à partir des productions, d'identifier de manière univoque la source d'un marquage non-conforme à la norme cible. Dans une étude sur l'accord en genre grammatical chez des apprenants de l'espagnol L2, Montrul, Foote et Perpiñán (2008 : 510) notent que le genre encodé sur le déterminant est souvent interprété comme reflétant l'attribution du genre, alors qu'un manque de concordance entre le genre marqué sur un déterminant et un adjectif révélerait une difficulté avec l'accord. En suivant cette logique, un apprenant qui prononce *le robe* aurait attribué le mauvais genre au nom *robe*, alors que si un apprenant prononce la séquence *la robe vert*, on considérera qu'il a correctement attribué le genre féminin à ce nom, mais qu'il rencontre des problèmes dans l'accord morphosyntaxique (voir Ayoun, 2007 : 159 ; Dewaele & Véronique, 2001 : 283 ; Kupisch et al., 2013 : 166 pour des études en français L2 ayant appliqué cette logique). Cependant, Montrul et al. reconnaissent que l'emploi d'un déterminant erroné (*le robe*) peut également résulter d'un problème d'accord, et que la distinction entre un problème d'attribution et d'accord dans les données de production « est très difficile à distinguer ». Cette difficulté est illustrée en (2a-d), où sont reproduits les huit emplois du nom *vie* par le participant 116 avant son départ pour la France.

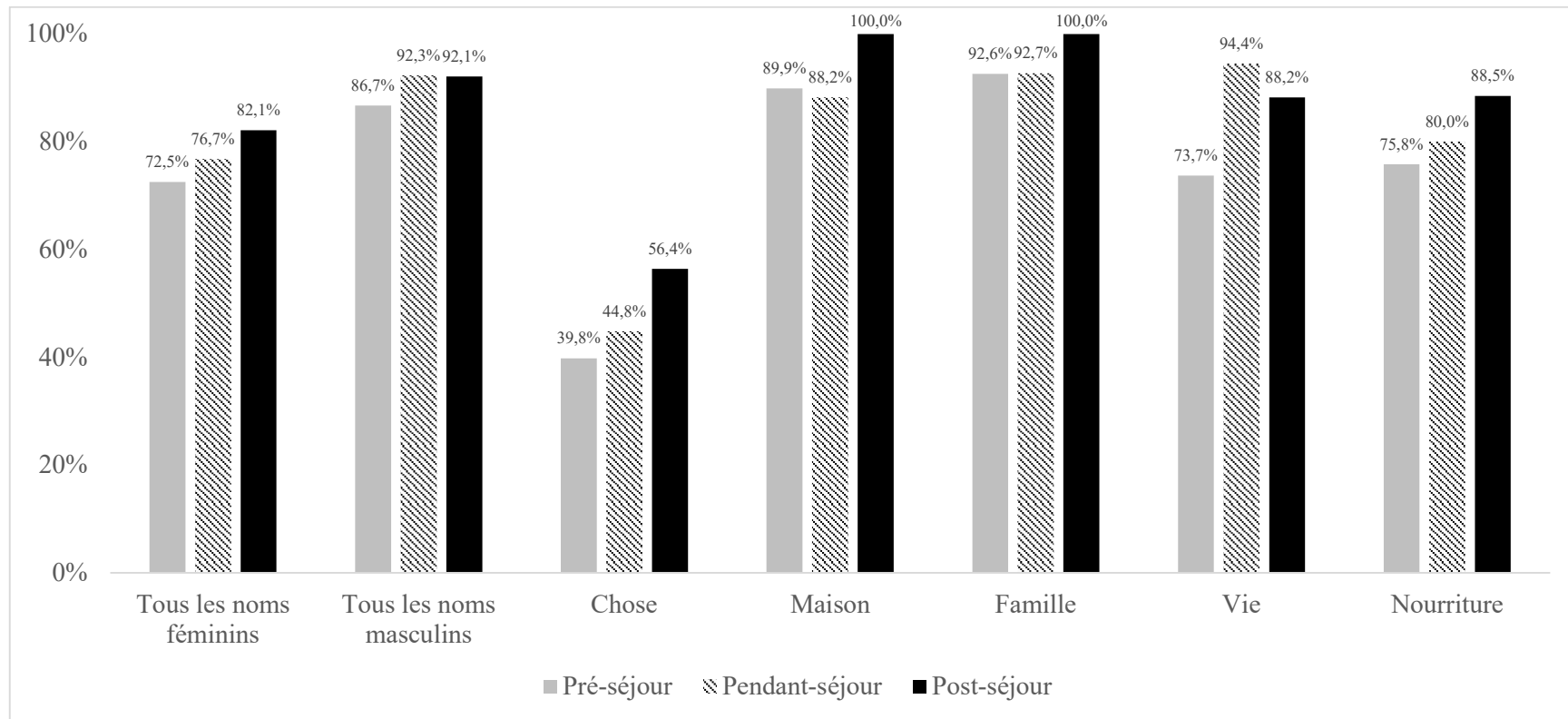
- (2) a. *la_{fém} vie_{fém} de chaque jour* (entretien)
sur le_{masc} la_{fém} vie_{fém} (entretien)
- b. *ehm la_{fém} vie_{fém} française_{fém} et ehm sort ehm* (entretien)
avec la_{fém} vie_{fém} française_{fém} en Angleterre ehm (entretien)
- c. *un peu plus sur la_{fém} vie_{fém} français_{masc}* (entretien)
une_{fém} nouveau_{masc} vie_{fém} (narration)
la_{fém} vie_{fém} ehm irait très différent_{masc} irait très différent_{masc} (entretien)
- d. *les droits pour vivre un_{masc} bonne_{fém} vie_{fém}* (essai)

Ces huit exemples illustrent une variabilité dans la forme du déterminant utilisée avec *vie* (féminin dans les exemples 2a-c, masculin en 2d), ainsi que dans la concordance entre le genre marqué sur le déterminant et l'adjectif (2b vs. 2c-d). Il semble donc périlleux, à partir de ces emplois, de tenter d'affirmer si ce participant a correctement attribué le genre féminin au nom *vie* et s'il est capable d'en faire l'accord. En revanche, dans une approche basée sur l'usage, il est pertinent d'examiner comment le marquage en genre se fait avec ce nom, sans pour autant se prononcer sur la source des éventuels usages qui ne sont pas conformes à la norme cible.

3.4 L'analyse : résultats et discussion

3.4.1. L'expression du genre : taux de réussite

Pour commencer l'analyse, je me suis intéressée à l'éventuelle évolution du taux de réussite dans le marquage du genre pour chacun des cinq noms, taux que l'on peut comparer aux résultats pour l'ensemble des noms masculins ($n = 4\ 318$) et féminins dans le corpus ($n = 3\ 275$). Le graphique 1 présente ces résultats pour chacun des trois recueils des données.



Graphique 1. Pourcentage du marquage du genre conforme à la norme.

Avant de partir pour la France, ce groupe d'apprenants a un taux de correction avec les noms masculins qui est de 86,7%. Un an plus tard, le marquage du genre pour les noms masculins a progressé, et atteint un niveau de correction de 92,3%. Ce niveau est maintenu huit mois plus tard, après le retour des apprenants au Royaume-Uni (92,1%). En ce qui concerne le marquage du genre grammatical avec les noms féminins, ce groupe d'apprenants s'améliore avec le temps, avec un taux de réussite de 72,5% au pré-séjour, 76,7% au pendant-séjour et 82,1% au post-séjour. Ces résultats sont à comparer à ceux obtenus pour *chose*, *maison*, *famille*, *vie* et *nourriture*. En ce qui concerne *chose*, avant de partir pour la France, ces apprenants produisent seulement 39,8% de marquage conforme à la norme, pourcentage qui augmente avec le temps pour finir à 56,4% en post-séjour. Viennent ensuite les noms *maison* et *famille*, avec lesquels nous constatons un marquage du genre relativement stable à travers le temps et conforme à la norme cible. En effet, le taux de réussite constaté au pré-séjour est déjà de 89,9% pour *maison* (avec seulement 10 exemples d'expression non-conforme à la norme produits par sept apprenants différents) et 92,6% pour *famille* (cinq cas d'expression non-conforme à la norme, produits par quatre apprenants différents). Au recueil post-séjour, le taux de réussite pour les deux noms atteint même les 100%. Il faut toutefois pondérer ce résultat par le fait que les échantillons pour ces deux noms au post-séjour sont petits : 18 emplois (de la part de huit participants) pour *maison* et 14 occurrences (produites par six individus) pour *famille*.

L'expression du genre avec les deux derniers noms – *vie* et *nourriture* – se range en quelque sorte entre les résultats pour *chose* et *maison/famille*. Au départ du projet, les apprenants produisent un marquage conforme à la norme dans 73,7% des cas pour *vie* et 75,8% pour *nourriture*, ce qui ressemble au taux global pour les noms féminins (72,5%). Un an plus tard, un net progrès est observé avec le marquage en genre en lien avec le nom *vie* : dans le recueil pendant-séjour, 94,4% des 54 occurrences nécessitant un marquage ouvert sont conformes aux attentes. Avec *nourriture*, un progrès moins important est visible : 80% des 30 emplois au pendant-séjour sont conformes à la norme. Au post-séjour, les taux de réussite avec *vie* et *nourriture* se ressemblent de nouveau : dans le cas de *vie*, le taux de marquage conforme à la norme recule par rapport au marquage au pendant-séjour, pour finir à 88,2%, alors que le marquage du genre avec *nourriture* continue à progresser, pour atteindre 88,5% au post-séjour.

Alors que les cinq noms sélectionnés pour cette analyse sont tous féminins et fréquents dans le corpus LANGSNAP, les résultats montrent clairement des différences importantes dans le taux de réussite dans l'expression du genre grammatical. Notamment, le nom le plus fréquent – *chose* – récolte le taux de réussite de loin le plus faible, alors que les taux de correction pour *maison* et *famille* dépassent la moyenne pour l'ensemble des noms féminins, mais également pour les noms masculins. La question se pose alors de savoir pourquoi les taux de réussite se différencient de manière si radicale. Dans les deux prochaines parties de l'analyse, je tenterai d'élucider ces différences lexicales en m'intéressant à une caractéristique de l'input – la fréquence –, avec d'abord une présentation des

usages faits de ces cinq noms dans un corpus de référence et, ensuite, une analyse des emplois par les 20 apprenants.

3.4.2 Une caractérisation des cinq noms dans un corpus de référence

Il est évidemment impossible de rendre compte de l'input exact auquel ces apprenants ont eu accès lors de leurs études et de leur séjour en France. Cela dit, en examinant l'utilisation des cinq noms dans un corpus de référence, il est possible d'identifier des tendances générales. Je considère qu'il est probable que ces tendances caractérisent également l'input français de manière générale et, donc, l'input ambiant auquel ces apprenants ont été exposés. Pour cette partie de l'analyse, le corpus de référence *Lexique 3.8* (corpus film) a été utilisé (New, Brysbaert, Veronis, & Pallier, 2007). Un corpus oral a été sélectionné pour permettre une estimation conservatrice des indices du genre grammatical dans l'input. En effet, un petit nombre de déterminants et d'adjectifs ont un marquage seulement orthographique du genre, ce qui signifie que les informations relatives au genre grammatical sont moins importantes à l'oral qu'à l'écrit.

Dans un premier temps, j'ai identifié le nombre d'occurrences par 1 million de mots pour chacun des cinq noms. Les résultats, présentés dans le tableau 2, révèlent qu'ils sont tous fréquents en français, le nom *chose* ayant une fréquence particulièrement élevée. Je me suis ensuite penchée sur les utilisations de ces noms et, en particulier, sur les modificateurs qui les accompagnent. Le but de cette analyse a été de déterminer avec quelle fréquence les modificateurs dans le corpus marquent ouvertement le genre de chacun des cinq noms. De cette manière, nous avons une idée de la présence d'indices sur le genre grammatical disponibles à l'apprenant dans l'input. Mille exemples de chaque nom (à l'exception de *nourriture*, pour lequel seules 779 occurrences étaient disponibles) ont été extraits du corpus. Ensuite, les occurrences pour lesquelles le nom était accompagné d'au moins un modificateur ont été codées selon l'expression du genre (les exemples sans modificateur – ex. *vivre en famille* – ont été exclus de l'analyse). Deux catégories ont été établies pour l'expression du genre grammatical : (a) un marquage ouvert et (b) un marquage non-ouvert. Pour les cas de marquage ouvert, l'expression du genre est audible sur le modificateur. Un marquage non-ouvert, en revanche, implique un modificateur invariable (ex. *c'est autre chose ; vous allez vraiment finir votre vie ici*) qui ne fournit aucune information sur le genre grammatical. Les occurrences où un nom a été accompagné d'un modificateur des deux catégories (ex. *toute votre vie*) ont été comptabilisées parmi les cas de marquage ouvert. Les résultats de cette analyse sont présentés dans le tableau 2.

Tableau 2. *Distribution de chose, maison, famille, vie, nourriture dans le corpus Lexique 3.8.*

Nom	Fréquence par 1 million de mots	Ouvert		Non-ouvert	
		<i>n</i>	%	<i>N</i>	%
<i>chose</i>	1 773,62	210	21,7	759	78,3
<i>maison</i>	384,92	899	94,9	48	5,1
<i>famille</i>	605,75	772	88,3	102	11,7

<i>vie</i>	835,47	779	92,3	65	7,7
<i>nourriture</i>	25,74	549	94,5	32	5,5

A partir de ce tableau, il est clair que, pour les noms *maison*, *famille*, *vie* et *nourriture*, la grande majorité (entre 88,3% et 94,9%) de tous les emplois avec un modificateur implique un marquage ouvert du genre. En revanche, le nom *chose*, lui, est rarement employé avec un modificateur qui montre son genre : seul 21,7% des emplois de *chose* avec un modificateur concernait un marquage ouvert.

Que signifient les informations dans le tableau 2 quant à l'input extérieur disponible à l'apprenant ? Tout d'abord, il est clair que les cinq noms devront faire partie de l'input, car tous les cinq sont fréquents. Or, tous les emplois ne seront pas instructifs quant au genre du nom. En effet, l'opposition entre marquage ouvert et non-ouvert quantifie les indices de genre auxquels l'apprenant est susceptible d'avoir accès. De cette manière, et malgré sa fréquence particulièrement élevée, les utilisations de *chose* sont pauvres en indications sur le genre. Pour les quatre noms restants, les emplois examinés révèlent que la grande majorité des utilisations marque le genre, fournissant ainsi des preuves explicites dans l'input qu'il s'agit de noms féminins. Nous avons ici une première piste pour éclairer les différences observées dans le graphique 1, où un écart important a été remarqué entre le taux de correction avec *chose* et les quatre autres noms dans les productions des apprenants. Un éclairage supplémentaire viendra sous la forme d'une seconde analyse de l'input. Dans la partie suivante, j'adopterai une perspective « intérieure » sur l'input, avec une analyse de l'utilisation de ces noms par les apprenants suivis dans le projet LANGSNAP.

3.4.3 Une caractérisation des usages events

Grâce à la caractérisation des emplois de *chose*, *maison*, *famille*, *vie* et *nourriture* dans un corpus de référence, nous avons une idée de certaines caractéristiques de l'input extérieur auquel les 20 apprenants du projet LANGSNAP ont été exposés. Or, l'input langagier ne se limite pas à l'input qui vient de l'extérieur ; il inclut également les *usage events* du locuteur lui-même. En effet, « les propres productions des apprenants de L2 dans le domaine du genre [...] finissent par constituer de l'input » (Hopp, 2016: 300). La caractérisation des *usage events* sera proposée en deux étapes. Pour commencer, et comme pour le corpus de référence, je partirai de l'ensemble des utilisations des cinq noms dans le corpus analysé ($n = 1\ 260$) pour présenter la distribution des emplois en fonction du type de marquage sur le modificateur (ouvert vs. non-ouvert). Dans un second temps, je proposerai une analyse des emplois de chacun des cinq noms en fonction du modificateur.

Dans le tableau 3 sont présentés le nombre de cas de marquage ouvert (où au moins un modificateur révèle le genre) et de marquage non-ouvert (impliquant des modificateurs invariables) dans les recueils pré-séjour, pendant-séjour et post-séjour du corpus LANGSNAP.

Tableau 3. *Distribution de chose, maison, famille, vie, nourriture dans le corpus LANGSNAP.*

	Oouvert	Non-ouvert
--	---------	------------

Nom	Total				
		<i>n</i>	%	<i>n</i>	%
<i>chose</i>	587	205	35,1	382	64,9
<i>maison</i>	219	202	92,2	17	7,8
<i>famille</i>	166	150	90,4	16	9,6
<i>vie</i>	164	134	81,7	30	18,3
<i>nourriture</i>	124	115	92,7	9	7,3

En comparant le nombre d'exemples « ouverts » et « non-ouverts », deux profils d'utilisation sont visibles, et ces profils correspondent à ceux identifiés dans les données tirées du corpus de référence (tableau 2). En effet, les noms *maison*, *famille*, *vie* et *nourriture* sont les plus souvent employés avec un déterminant ou un adjectif sur lequel le genre est audible ou visible : 92,2% (202/219) de tous les emplois pour *maison* ont un marquage ouvert du genre grammatical, 90,4% (150/166) pour *famille*, 81,7% (134/164) pour *vie* et 92,7% (115/124) pour *nourriture*. Le profil d'utilisation de *chose* est bien différent. S'il est de loin le nom le plus fréquent des cinq – on en trouve trace de 587 emplois avec un déterminant ou avec un adjectif dans le corpus – le genre grammatical n'est ouvertement marqué que dans 205 sur les 587 occurrences, soit 34,9%. Cela signifie que quand ces apprenants ont utilisé *maison*, *famille*, *vie* ou *nourriture* avec un déterminant ou un adjectif, les modificateurs choisis comportaient, dans une vaste majorité des cas, un marquage ouvert du genre. En revanche, pour le nom *chose*, qui a été utilisé trois fois plus fréquemment que les autres noms, la plupart des emplois ne concernaient pas un modificateur marqué ouvertement pour le genre. Ainsi les tendances distributionnelles dans les tableaux 2 (corpus de référence) et 3 (*usage events*) indiquent-elles que l'input pour le nom *chose* est sensiblement différent de l'input pour les quatre autres noms.

Pour terminer cette partie, je recentre l'analyse sur les 806 emplois de *chose*, *maison*, *famille*, *vie* et *nourriture* avec un modificateur qui exprime ouvertement le genre. Cette analyse des emplois individuels complétera les analyses précédentes et apportera de nouveaux éléments quant aux différences lexicales dans l'expression du genre avec les cinq noms sous étude. Plus particulièrement, nous verrons comment l'expression du genre grammatical varie en fonction du modificateur et à travers le temps. Ces tendances sont présentées dans le tableau 4.

Tableau 4. *L'expression du genre en fonction du modificateur et du temps.*

Modificateur	<i>chose</i>		<i>maison</i>		<i>famille</i>		<i>vie</i>		<i>nourriture</i>	
	<i>n</i>	%	<i>n</i>	%	<i>N</i>	%	<i>n</i>	%	<i>n</i>	%
L'article défini										
Pré-séjour	9/24	37,5	62/70	88,6	15/15	100	25/26	96,2	21/24	87,5
Pendant-séjour	10/20	50	64/67	95,5	10/10	100	24/24	100	13/14	92,9
Post-séjour	18/23	78,3	14/14	100	9/9	100	8/8	100	24/24	100
L'article indéfini										
Pré-séjour	11/16	68,8	17/17	100	22/23	95,7	8/9	88,9	3/3	100

Pendant-séjour	9/11	81,8	4/6	66,7	17/18	94,4	10/11	90,9	7/7	100
Post-séjour	3/3	100	3/3	100	4/4	100	—	—	1/1	100
Les déterminants possessifs <i>ma, ta, sa</i>										
Pré-séjour	—	—	7/7	100	23/23	100	5/6	83,3	1/3	33,3
Pendant-séjour	—	—	6/8	75	18/18	100	5/5	100	4/8	50
Post-séjour	1/1	100	—	—	2/2	100	6/6	100	—	—
Le démonstratif <i>cette</i>										
Pré-séjour	—	—	—	—	—	—	—	—	—	—
Pendant-séjour	1/1	100	—	—	—	—	—	—	—	—
Post-séjour	—	—	—	—	—	—	—	—	3/3	100
L'ensemble des déterminants										
Pré-séjour	20/40	50	86/94	91,5	60/61	98,4	38/41	92,7	25/30	83,3
Pendant-séjour	20/32	62,5	74/81	91,4	45/46	97,8	39/40	97,5	24/29	82,8
Post-séjour	22/27	81,5	17/17	100	15/15	100	14/14	100	28/28	100
L'ensemble des adjectifs										
Pré-séjour	13/43	30,2	3/5	60	13/16	81,3	10/22	45,5	0/3	0
Pendant-séjour	10/35	28,6	1/4	25	5/9	55,6	12/14	85,7	0/1	0
Post-séjour	9/28	32,1	1/1	100	1/2	50	1/3	33,3	18/24	75

Je souhaite faire deux observations concernant ce tableau. La première concerne l'évolution dans le marquage du genre à travers le temps. A commencer par les adjectifs, leur faible nombre avec certains noms à certains recueils (ex. *nourriture* au pré-séjour et au pendant-séjour) ne permet que des remarques préliminaires. Cela dit, et dans l'ensemble, aucune amélioration nette au cours des 21 mois représentés dans le corpus LANGSNAP n'est observable. Pour les déterminants, peu d'évolution est visible pour les noms *maison, famille, vie* et *nourriture*, car les taux de correction sont très élevés dès le recueil de pré-séjour. Avec l'article défini, une légère évolution est toutefois constatée entre pré-séjour et post-séjour pour les noms *maison* et *nourriture*. Une évolution plus tranchée est visible avec le nom *chose*, pour lequel la correction du marquage en genre sur les articles défini et indéfini progresse nettement dans le temps : alors que seul 37,5% des articles définis sont féminins au pré-séjour, 78,3% le sont au post-séjour. L'évolution est tout aussi spectaculaire avec l'article indéfini, qui passe d'un taux de correction de 68,8% (pré-séjour) au taux de 100% (post-séjour) ; toutefois le faible nombre d'occurrences oblige à interpréter ces tendances avec prudence. Comparé au taux de réussite global avec *chose* au pré-séjour (39,8%), au pendant-séjour (44,8%) et au post-séjour (56,4%), l'amélioration dans le marquage en genre est particulièrement remarquable sur les articles défini et indéfini. Pourquoi ces apprenants se seraient-ils améliorés précisément avec les séquences *la chose* et *une chose* ? Je suggère que les analyses de l'input fournissent des éléments de réponse à cette question. Dans l'analyse du corpus Lexique 3.8, seules 210 instances de *chose* (sur un échantillon

initial de 1 000) comportaient un modificateur qui marquait ouvertement le genre féminin. Or, presque tous ces exemples contenaient soit l'article défini ($n = 101$), soit l'article indéfini ($n = 91$). En plus, les articles défini et indéfini sont fréquemment utilisés par les apprenants eux-mêmes (voir le tableau 4). Considérés dans leur ensemble, ces éléments indiquent que les apprenants progressent dans leur expression du genre justement avec les modificateurs qui sont les plus fréquents, à la fois dans l'input extérieur et dans leurs propres usages de ces noms.

Pour la seconde observation, la différence entre déterminant et adjectif sera mise en exergue. La partie inférieure du tableau 4 permet une comparaison de l'expression du genre sur l'ensemble des déterminants et l'ensemble des adjectifs. Nous voyons ici une confirmation claire, et ce pour chacun des cinq noms, de l'un des résultats des recherches précédentes concernant l'acquisition du genre grammatical en français L2 : le niveau de correction est moins important sur les adjectifs que sur les déterminants (cf. Ayoun, 2007 ; Bartning, 2000 ; Dewaele & Véronique, 2001 ; Granfeldt, 2005 ; Edmonds & Gudmestad, 2018 ; Edmonds et al., sous évaluation ; Harley, 1979 ; Holmes & Dejean de la Bâtie, 1999). Deux explications de ce résultat, toutes les deux en lien avec l'input, seront proposées. Tout d'abord, l'adjectif – à la différence du déterminant – peut se situer à l'intérieur ou à l'extérieur du syntagme nominal. En d'autres termes, l'adjectif peut se trouver à une distance considérable du nom qu'il modifie, comme dans les exemples fournis en (3).

- (3) a. *la_{fem} vie_{fem} ici est plus euh plutôt différent_{masc} que qu'en France* (participant 110, entretien, pré-séjour)
- b. *il y a deux choses_{fem} à faire qui sont plus important_{masc}* (participant 106, essai, post-stay)
- c. *contre l'idée d'une_{fem} famille_{fem} pas tout à fait normale_{fem}* (participant 112, essai, pendant-séjour)

Or, comme cela a été suggéré par Finnemann (1992) dans une étude de l'expression du genre en espagnol L2, plus la distance entre nom est modificateur est importante, plus le travail mémoriel est coûteux pour le locuteur, ce qui peut expliquer un taux de réussite moindre dans l'expression du genre sur les adjectifs dans leur ensemble. Deuxièmement, la catégorie *adjectif* recouvre de nombreux types, il y a ainsi beaucoup plus de variété dans cette catégorie que dans celle des déterminants. Ceci est illustré dans le tableau 5, où sont présentés l'ensemble des adjectifs utilisés par les apprenants dans le corpus analysé pour cette étude.⁵

Tableau 5. *Les adjectifs nécessitant un marquage ouvert du genre : corpus LANGSNAP.*

Recueil	<i>chose</i>		<i>maison</i>	<i>famille</i>	<i>vie</i>	<i>nourriture</i>
Pré-séjour	<i>amusant</i>	<i>naturel</i>	<i>bon</i>	<i>aimant</i>	<i>bon</i>	<i>frais</i>
	<i>anglais</i>	<i>net</i>	<i>mignon</i>	<i>bon</i>	<i>différent</i>	<i>important</i>

⁵ Comme une partie des données provient des essais écrits, le marquage en genre pour certains adjectifs n'est pas audible.

	<i>bon</i>	<i>petit</i>	<i>petit</i>	<i>congolais</i>	<i>familial</i>	<i>tout</i>
	<i>créatif</i>	<i>principal</i>		<i>content</i>	<i>français</i>	
	<i>différent</i>	<i>pris</i>		<i>français</i>	<i>heureux</i>	
	<i>goûtant</i>	<i>stressant</i>		<i>gentil</i>	<i>meilleur</i>	
	<i>grand</i>	<i>tout</i>		<i>heureux</i>	<i>nocturne</i>	
	<i>important</i>			<i>soigneux</i>	<i>nouveau</i>	
				<i>tout</i>	<i>quotidien</i>	
				<i>traditionnel</i>	<i>tout</i>	
Pendant- séjour	<i>actif</i>	<i>offert</i>	<i>nouveau</i>	<i>accueillant</i>	<i>content</i>	<i>tout</i>
	<i>administratif</i>	<i>petit</i>	<i>petit</i>	<i>défavorisé</i>	<i>différent</i>	
	<i>bon</i>	<i>positif</i>		<i>équilibré</i>	<i>familial</i>	
	<i>contrôlé</i>	<i>premier</i>		<i>homosexuel</i>	<i>français</i>	
	<i>enrichissant</i>	<i>réglementé</i>		<i>meilleur</i>	<i>heureux</i>	
	<i>important</i>	<i>religieux</i>		<i>normal</i>	<i>important</i>	
	<i>intéressant</i>	<i>seul</i>		<i>tout</i>	<i>meilleur</i>	
	<i>naturel</i>	<i>tout</i>			<i>normal</i>	
	<i>normal</i>				<i>parisien</i>	
					<i>social</i>	
Post-séjour	<i>amusant</i>	<i>méchant</i>	<i>grand</i>	<i>français</i>	<i>quotidien</i>	<i>bon</i>
	<i>attractif</i>	<i>petit</i>		<i>heureux</i>		<i>cher</i>
	<i>bon</i>	<i>précieux</i>				<i>dégoûtant</i>
	<i>déterminant</i>	<i>premier</i>				<i>frais</i>
	<i>différent</i>	<i>quotidien</i>				<i>gras</i>
	<i>français</i>	<i>social</i>				<i>sain</i>
	<i>important</i>	<i>tout</i>				<i>sucré</i>
	<i>intéressant</i>					

En plus de devoir apprendre à marquer le genre sur de nombreux adjectifs, l'apprenant aura également moins d'occasions d'utiliser chaque combinaison. En d'autres termes, l'apprenant aura beaucoup plus d'occasions d'entendre et d'utiliser la suite *la chose* que *chose amusante*, *bonne chose* ou *chose créative*. Des cinq noms examinés, c'est *chose* qui est le plus souvent utilisé par les apprenants avec des adjectifs et ce sont donc les emplois de *chose* qui sont caractérisés par le plus grand nombre de séquences différentes. A l'opposé de *chose*, le nom *maison* se caractérise par des emplois répétitifs avec peu de variabilité. En effet, 74,7% (151/202) de tous les emplois représentés dans le tableau 4 concerne l'utilisation de *maison* avec l'article défini. Cette répétition de *la maison* crée un input régulier, prévisible, dans lequel le genre est presque toujours marqué ouvertement. Ces deux profils d'utilisation sont associés, dans cette étude, à des taux de réussite différents : alors que les apprenants

montrent un taux de réussite très bas avec le nom *chose*, le nom *maison*, lui, jouit d'un taux de correction très élevé.

3.4.4 Bilan

Dans l'étude actuelle, j'ai souhaité contribuer aux recherches sur l'expression du genre grammatical en français L2 avec une analyse par nom dans le cadre des approches basées sur l'usage. Pour ce faire, cinq noms féminins ont été sélectionnés, et la fréquence dans l'input – en français moderne et par les apprenants eux-mêmes – a été analysée. En effet, dans des approches basées sur l'usage, une attention particulière est accordée à l'input auquel l'apprenant a accès, car la caractérisation de celui-ci peut proposer des clés explicatives aux itinéraires acquisitionnels observés. Je suis maintenant en mesure de fournir une réponse à la question de recherche qui a guidé ce projet : l'évolution dans l'expression du genre grammatical diffère-t-elle en fonction du nom chez un groupe de 20 apprenants britanniques du français L2 et, si oui, quel rôle la fréquence joue-t-elle dans ces différences ? La réponse à la première partie de cette question est clairement oui. Aux trois recueils, le taux de correction avec le nom *chose* est clairement inférieur aux taux obtenus pour les quatre autres noms. Qui plus est, ce taux est largement en deçà de la moyenne pour les noms féminins dans leur ensemble. Pour les noms *vie* et *nourriture*, une progression dans le temps est observable, avec des taux de réussite qui ressemblent à ceux présentés pour l'ensemble des noms féminins. Le marquage du genre avec *maison* et *famille*, en revanche, est presque toujours conforme à la norme, que ce soit avant le départ pour la France ou 21 mois plus tard. En comparaison à l'ensemble des noms masculins et féminins du corpus, l'expression du genre avec ces deux noms est plus souvent conforme à la norme. En d'autres termes, s'il est vrai que le taux de correction est généralement moins élevé avec les noms féminins qu'avec les noms masculins en français L2, cette généralité recouvre des spécificités lexicales diverses, avec certains noms féminins (*maison* et *famille*) ayant un taux de correction qui dépasse la moyenne pour les noms masculins.

En quoi la fréquence de l'input peut-elle éclairer ces observations ? Pour cette analyse, j'ai adopté deux perspectives sur l'input : la perspective extérieure cherchait à caractériser l'input ambiant auquel l'apprenant pourrait avoir accès, alors que la perspective intérieure s'intéressait aux emplois individuels des cinq noms sous étude. Ces analyses de l'input se complètent et fournissent trois éléments principaux qui aident à comprendre les taux de correction variables. Premièrement, nous avons vu que *chose* est rarement utilisé avec un modificateur sur lequel le genre est ouvertement marqué, que ce soit dans un corpus de référence ou par les apprenants eux-mêmes. Cela signifie que, proportionnellement, les apprenants entendent rarement ce nom utilisé avec un modificateur qui indique son genre, et quand ils emploient eux-mêmes ce nom, la plupart du temps, ils utilisent un modificateur qui ne les oblige pas à marquer le genre (*les choses, des choses, autre chose*). Cela dit, et cela m'amène au deuxième point, une progression sensible dans le temps est visible pour ce nom, mais uniquement avec les articles défini et indéfini. Encore une fois, la fréquence dans l'input semble jouer un rôle : l'amélioration dans l'expression du genre concerne les deux modificateurs utilisés les

plus fréquemment avec *chose* dans un corpus de référence et dans les emplois individuels. Le dernier élément de réponse à la question de recherche porte sur les emplois individuels des cinq noms et, surtout, sur la variété de combinaisons utilisées par les apprenants. En effet, nous avons vu que les deux noms les plus fréquemment utilisés avec un modificateur ouvertement marqué pour le genre – *chose* ($n = 205$) et *maison* ($n = 202$) – ont des profils d'utilisation qui s'opposent. Alors que les emplois de *chose* sont caractérisés par la variété, surtout dans les adjectifs (voir le tableau 5), les usages de *maison* sont dominés par un nombre limité de combinaisons (*la maison*), ainsi réduisant la variabilité dans l'input. Ce dernier type de profil d'utilisation est caractérisé par la répétition, ce qui pourrait faciliter l'apprentissage du genre grammatical pour *maison*, car cela augmenterait la fiabilité entre « repère et interprétation » (Ellis, 2015 : 56).

Ces résultats permettent de nuancer les recherches précédentes sur l'expression du genre grammatical en français L2. Alors que plusieurs études (Bartning, 2000 ; Dewaele & Véronique, 2001 ; Edmonds & Gudmestad, 2018 ; Edmonds et al., sous évaluation ; Harley, 1979) ont trouvé que le marquage du genre est moins souvent conforme à la norme avec les noms féminins qu'avec les noms masculins, l'étude actuelle souligne la variabilité qui se cache derrière cette observation globale. En effet, les apprenants suivis pour cette étude ont atteint des niveaux de correction plus élevés pour deux des noms examinés – *maison* et *famille* – que pour l'ensemble des noms masculins, et ce aux trois recueils de données. En outre, plusieurs études précédentes ont présenté des preuves d'une différence dans l'expression du genre en fonction du modificateur (adjectif vs. déterminant). Les résultats pour *chose*, *maison*, *famille*, *vie* et *nourriture* tendent à confirmer cette observation. En plus de ces deux résultats globaux, quelques auteurs ont discuté de la variabilité au niveau lexical. Alors que Bartning (2000) a discerné des différences dans l'expression du genre en fonction du déterminant, aucune tendance semblable ne se dégage des résultats de la présente étude. Ce qui est observable dans le tableau 4, ce sont des emplois et des parcours particuliers en fonction du nom *et* du modificateur, ce qui suggère que les deux entrées apportent des informations complémentaires. Deux autres études – Hawkins (2001) et Granfeldt (2005) – ont suggéré l'existence d'itinéraires acquisitionnels différents par nom, et les deux se sont surtout intéressées à la sur-généralisation d'un seul article défini et indéfini. La méthodologie de l'étude actuelle, pour laquelle je n'ai étudié que cinq noms, ne me permet pas de me prononcer sur l'existence d'un éventuel marquage systématique par défaut chez ces apprenants. En revanche, cette étude démontre clairement que l'évolution de l'expression du genre diffère en fonction du nom, et qu'une analyse qui part du nom et qui s'intéresse à l'input peut contribuer à une meilleure compréhension du développement de ce phénomène morphosyntaxique.

4. Conclusion

Dans la présente étude, je me suis intéressée à la variabilité lexicale dans l'expression du genre grammatical dans un corpus longitudinal constitué de productions écrites et orales semi-authentiques. Les résultats de la présente étude complètent les observations tirées des recherches précédentes, et

soulignent à la fois le rôle des caractéristiques de l'input et la pertinence du nom comme unité d'analyse dans l'étude de ce phénomène morphosyntaxique. Cependant, ce type d'analyse ne peut être exhaustif – rappelons que ces apprenants ont employé 827 noms différents – mais il peut suggérer des pistes pour de futures recherches. D'abord, il est probable que d'autres facteurs qui n'ont pas été explorés ici pourront s'avérer utiles afin de mieux comprendre le développement dans l'expression du genre. En effet, dans l'étude actuelle, l'exploration de l'input dans le corpus Lexique 3.8 et dans les emplois individuels a surtout contribué à une meilleure compréhension de l'expression du genre avec *chose*. Ces analyses sont restées muettes sur, par exemple, les différences entre *famille*, d'un côté, et *vie* et *nourriture*, de l'autre. En effet, les profils d'utilisation pour les trois noms se ressemblent, mais le taux de correction avec *famille* dépasse largement ceux observés pour *vie* et *nourriture*. De futures analyses pourront chercher à élucider ces différences. A l'avenir, il serait également pertinent d'analyser l'expression du genre grammatical avec d'autres noms, afin de vérifier si les observations qui ont découlé de la présente analyse sont généralisables. Notamment, il sera important d'étudier comment les apprenants expriment le genre avec des noms masculins. Etant donné que les apprenants ont un taux de réussite plus important avec les noms masculins, la question se pose de savoir s'il y aurait autant de variabilité avec ceux-ci. Il sera également intéressant à l'avenir d'examiner le rôle de la tâche dans l'expression du genre. Alors que les données pour la présente analyse provenaient de trois tâches différentes (entretien oral, narration orale, essai écrit), il n'a été possible de prendre en compte ce facteur à cause de la taille du corpus. Et pourtant, il est possible que cette variable influe sur l'expression du genre, surtout dans une langue comme le français, qui est caractérisée par une morphologie silencieuse (Ågren, 2016). Une dernière piste que je mentionnerai concerne les différences individuelles. En effet, les conclusions présentées dans cet article se basent sur les productions d'un groupe de 20 apprenants. A l'exception de l'étude de cas présentée par Granfeldt (2005), pour l'instant nous savons peu sur les itinéraires acquisitionnels individuels dans le développement de l'expression du genre. A partir des données présentées dans cet article, il serait possible (et intéressant) d'examiner comment le marquage du genre se développe à travers le temps chez les différents apprenants. Pour ne prendre que l'exemple du nom *chose*, les données au niveau du groupe suggèrent que les séquences *la chose* et *une chose* ouvrent la marche, mais cela reste à confirmer. De futures recherches pourront profiter de la disponibilité du corpus LANGSNAP pour mener de telles analyses.

REFERENCES

- Ågren, M. (2016). Apprendre l'orthographe profonde du français comme langue seconde/étrangère. *Revue Française de Linguistique Appliquée*, 21(2). 95-108.
- Ayoun, D. (2007). The second language acquisition of grammatical gender and agreement. In D. Ayoun (Ed.), *French applied linguistics* (pp. 130-170). Amsterdam: Benjamins.

- Ayoun, D. (2010). Corpus data. Shedding the light on grammatical gender... or not. *EUROSLA Yearbook*, 10, 119-141.
- Ayoun, D. (2018). Grammatical gender assignment in French: Dispelling the native speaker myth. *Journal of French Language Studies*, 28(1), 113-148.
- Bartning, I. (2000). Gender agreement in L2 French: Pre-advanced vs. advanced learners. *Studia Linguistica*, 54, 225-237.
- Carroll, S. (1989). Second-language acquisition and the computational paradigm. *Language Learning*, 39(4), 535-594.
- Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, 4, 275-297.
- Diessel, H. (2016). Frequency and lexical specificity in grammar: A critical review. In H. Behrens & S. Pfänder (Eds.), *Experience counts: Frequency effects in language* (pp. 209-238). Berlin: De Gruyter.
- Edmonds, A., & Gudmestad, A. (2018). Gender marking in written L2 French: Before, during, and after residence abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 3(1), 59-82.
- Edmonds, A., Gudmestad, A., & Metzger, T. (sous évaluation). A longitudinal study of grammatical gender marking in French as an additional language.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, N. C. (2015). Cognitive and social aspects of learning from usage. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 49-73). Berlin: de Gruyter Mouton.
- Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Malden, MA: Wiley.
- Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 75-93). Berlin: de Gruyter Mouton.
- Eskildsen, S. W., & Cadierno, T. (2007). Are recurring multi-word expressions really syntactic freezes? Second language acquisition from the perspective of usage-based linguistics. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the first Nordic conference on syntactic freezes, 2006* (pp. 86-99). Joensuu, Finland: Joensuu University Press.

- Eskildsen, S., & Cadierno, T. (2015). Advancing usage-based approaches to L2 studies. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 1-16). Berlin: de Gruyter Mouton.
- Finnemann, M. D. (1992). Learning agreement in the noun phrase: The strategies of three first-year Spanish students. *International review of Applied Linguistics in Teaching*, 30(2), 121-136.
- Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, 66, 226-248.
- Gaucher, D. (2015). Sémantique temporelle et accord du participe passé en français parlé : une analyse variationniste. *Journal of French Language Studies*, 25, 65-83.
- Granfeldt, J. (2005). The development of gender attribution and gender concord in French: A comparison of bilingual first and second language learners. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language: Multidisciplinary approaches* (pp. 164-190). Clevedon, UK: Multilingual Matters.
- Hardison, D. M. (1992). Acquisition of grammatical gender in French: L2 learner accuracy and strategies. *The Canadian Modern Language Review*, 48, 292-306.
- Harley, B. (1979). French gender 'rules' in the speech of English-dominant, French-dominant, and monolingual French speaking children. *Working Papers in Bilingualism*, 19, 129-156.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. New York: Wiley/Blackwell.
- Holmes, V. M., & Dejean de la Bâtie, B. (1999). Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics*, 20, 479-506.
- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2), 277-307.
- Jach, D. (2018). A usage-based approach to preposition placement in English as a second language. *Language Learning*, 68(1), 271-304.
- Kupisch, T., Akpınar, D., & Stöhr, A. (2013). Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism*, 3, 150-179.
- Linford, B., Long, A., Solon, M., & Geeslin, K. L. (2016). Measuring lexical frequency: Comparison groups and subject expression in L2 Spanish. In L. Ortega, A. E. Tyler, H. I. Park, & M. Uno (Eds.), *The usage-based study of language learning and multilingualism* (pp. 137-154). Washington, DC: Georgetown University Press.
- Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*, 16, 69-92.
- Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. London: Routledge.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. London: Routledge.

- Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58(3), 503-553.
- Nelson, D. (2005). French gender assignment revisited. *Word*, 56, 19-38.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28, 661-677.
- Prodeau, M., & Carlo, C. (2002). Le genre et le nombre dans des tâches verbales complexes en français L2 : grammaire et discours. *Marges linguistiques*, 4, 165-174.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Surridge, M. E., & Lessard, G. (1984). Pour une prise de conscience du genre grammatical. *La Revue canadienne des langues vivantes*, 41, 44-52.
- Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). 'Repeat as much as you can': Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 143-166). Clevedon, UK: Multilingual Matters.
- Tucker, R., Lambert, W. E., & Rigault, A. (1977). *The French speaker's skill with grammatical gender: an example of rule-governed behavior*. Paris: Mouton
- Tummers, J., Heylen, K., & Geeraerts, D. (2005). Usage-based approaches in cognitive linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory*, 1-2, 225-261.
- Tyler, A. E., & Ortega, L. (2018). Usage-inspired L2 instruction: An emergent, researched pedagogy. In A. E. Tyler, L. Ortega, M. Uno, & H. I. Park (Eds.), *Usage-inspired L2 instruction: Researched pedagogy* (pp. 3-26). Amsterdam: Benjamins.
- Vatz, K. L. (2009). *Grammatical gender representation and processing in advanced second language learners of French*. Unpublished doctoral dissertation, University of Maryland.
- Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & Leblanc, C. J. (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal*, 93(3), 354-369.

Abstract

Research on the acquisition of gender marking in second language French has shown, among other things, that learners succeed more often in marking gender in a targetlike manner on masculine (vs. feminine) nouns and on determiners as opposed to on adjectives. However, it remains unclear as to whether these general tendencies apply in similar fashion to all nouns, adjectives, and determiners. In other words, it is fair to question to what extent gender marking may be determined by lexical specificities. In the current study, the evolution of gender marking as produced by a group of 20 British learners of French is analyzed, using data collected over 21 months. Oral and written productions collected at three points in time were analyzed, and gender marking with five frequent feminine nouns – chose, famille, maison, nourriture, vie – was examined. Results show that the expression of gender marking develops differently with these five nouns, and these findings are interpreted with the help of insights from usage-based approaches to language learning.

Key words: gender marking, lexical specificities, longitudinal data, usage-based approaches

