



HAL
open science

Etudier l'acquisition d'une L2. Quelles démarches méthodologiques ?

Sandra Benazzo, Pascale Leclercq

► **To cite this version:**

Sandra Benazzo, Pascale Leclercq. Etudier l'acquisition d'une L2. Quelles démarches méthodologiques ?. P. Leclercq, A. Edmonds & E. Sneed-German (éds.). Introduction à l'acquisition des langues étrangères., De Boeck., 2021. hal-03636438

HAL Id: hal-03636438

<https://hal.science/hal-03636438>

Submitted on 10 Apr 2022

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.

Benazzo, S. & Leclercq, P. (version préliminaire à paraître/2021) Etudier l'acquisition d'une L2. Quelles démarches méthodologiques ? In P. Leclercq, A. Edmonds & E. Sneed-German (éds.) *Introduction à l'acquisition des langues étrangères*. Bruxelles : De Boeck.

Chapitre 3 Etudier l'acquisition d'une L2 : quelles démarches méthodologiques ?

Sandra Benazzo & Pascale Leclercq

1. Introduction

L'acquisition d'une langue seconde est un processus complexe, qui implique des développements à différents niveaux : linguistique, paralinguistique, cognitif, interactionnel, culturel, *etc.*, et dont le déroulement peut être infléchi par un ensemble de facteurs, tels que les langues de l'apprenant (langues source (LS) et cible (LC), autres langues connues), l'âge du début de l'acquisition, le contexte d'acquisition (institutionnel, naturel, mixte) ainsi que des traits individuels dont la motivation. Dans ce chapitre nous allons nous focaliser surtout sur l'acquisition linguistique. Son étude vise à obtenir des données permettant d'interpréter : a) le système linguistique de l'apprenant en L2 à un moment donné de son apprentissage ; b) l'évolution de ce système linguistique ; c) les facteurs qui peuvent contribuer ou faire obstacle au rapprochement du lecte de l'apprenant à la LC (*cf.* Norris & Ortega, 2003). Comme pour toute recherche empirique, il est crucial d'obtenir des données qui rendent possible l'observation (et la quantification) des phénomènes acquisitionnels de manière fiable. Comment procéder pour observer le processus d'appropriation d'une nouvelle langue ? Quels types de données sont appropriées pour ce faire ? Ce sont les questions générales qui seront abordées dans ce chapitre.

Un premier point, crucial pour déterminer la méthodologie, concerne la *question de recherche* à laquelle on souhaite apporter des éléments de réponse. Au cours du développement en L2 il y a une interaction entre les progressions de l'apprenant dans différents domaines (phonologie, lexique, grammaire...). Il est pourtant difficile d'analyser l'évolution du système de l'apprenant dans tous ses aspects en même temps ; il en va de même pour l'analyse des nombreux facteurs pouvant l'infléchir. Pour des questions de *faisabilité*, la plupart des recherches se focalise sur le développement dans certains domaines et / ou l'influence de certains facteurs au détriment d'autres, même si ce découpage est, d'une certaine manière, artificiel. Par ailleurs, les questions de recherche, et la manière de les aborder, varient en fonction de l'orientation théorique choisie (fonctionnelle, formelle, interactionniste, émergentiste, *etc.*). Il n'est pas question ici d'entrer dans les détails de chaque approche ; il est cependant utile de distinguer entre les recherches qui visent à décrire ou explorer les données pour établir des hypothèses sur l'acquisition (recherche *descriptive* ou *exploratoire*), et des recherches qui sont guidées par une hypothèse précise, issue d'une théorie ou des résultats d'études précédentes, afin d'en confirmer la validité (recherche *confirmative*). C'est seulement une fois la question de recherche et les facteur(s) à investiguer choisis que se pose la question du type de *données pertinentes* à recueillir sur lesquelles baser les analyses : comment conçoit-on un corpus d'observables, en compréhension ou en production, à l'écrit ou à l'oral ? Quel type d'apprenants sélectionner ? Faut-il faire une étude de cas ou de groupes d'apprenants ?

Dans ce chapitre, nous proposons un survol de démarches méthodologiques générales, puis nous proposerons un exemple de différents protocoles expérimentaux appliqués à un même objet d'étude, à savoir les particules de portée.

2. Concevoir un dispositif de recueil de données

Notre but ici est de présenter quelques approches courantes dans la conception d'un dispositif expérimental. Nous aborderons la question de la nature et de l'architecture du corpus en fonction du but de la recherche, puis celle du codage et de l'interprétation des données.

2.1 Données « écologiques », « contrôlée », « expérimentales » ou « introspectives »

Si la constitution d'un corpus poursuit toujours le but d'obtenir des informations sur le processus acquisitionnel, les types de données utilisés pour ce faire peuvent être de nature différente.

Certaines études visent à obtenir des données « spontanées », « authentiques » ou encore « écologiques » dans la mesure où elles seraient représentatives de l'usage de la L2 dans un contexte naturel. En d'autres termes, on voudrait observer le comportement des apprenants lorsqu'ils ne sont pas observés, même si ce but est difficile à atteindre du fait de la présence d'un chercheur ou d'un dispositif d'enregistrement des données (*cf. paradoxe de l'observateur*, Labov, 1972 / 1978).

D'autres études s'appuient, en revanche, sur des données plus contrôlées, à savoir obtenues en utilisant une tâche et / ou un support spécifique, qui permettent l'observation de structures linguistiques ciblées. En fonction des contraintes présentes dans la tâche utilisée, certaines de ces données sont qualifiées d'« expérimentales ». La nature « contrôlée » / « expérimentale » des données n'est pas toujours tranchée : elle correspond plutôt aux extrêmes d'un *continuum*¹. En effet, certaines données « contrôlées » sont plutôt semi-guidées et motivées par un but communicatif, ce qui les apparente aux données spontanées, alors que d'autres demandent un emploi du langage complètement décontextualisé.

À côté de ces procédures, centrées sur l'emploi de la L2 en production et / ou en compréhension, il convient de mentionner les données introspectives (introspection, rétrospection, verbalisations de la pensée lors de l'accomplissement des tâches, *etc.*), qui donnent un aperçu sur différents types de représentations de l'apprenant concernant l'apprentissage et l'emploi des langues.

Chaque procédure présente des points forts et des faiblesses, que nous allons présenter ci-dessous, d'où l'intérêt de les combiner pour étudier un même phénomène acquisitionnel.

2.1.1 Données « écologiques »

Parmi les procédures de recueil de données « écologiques » figure l'enregistrement d'interactions libres entre le sujet enquêté et son entourage. Cette procédure est surtout appliquée lorsque l'observation porte sur des sujets très jeunes, que ce soit dans le cas de l'acquisition monolingue (L1) ou de l'acquisition bilingue (2L1). Il en est ainsi par exemple dans les études de (Leopold, 1939-1949), qui a pris des notes sur l'acquisition simultanée de l'anglais et de l'allemand par sa fille Hildegard de sa naissance jusqu'à seize ans.

La prise de notes utilisée à cette époque a cédé le pas à l'enregistrement audio ou vidéo : l'enfant est ainsi enregistré (ou filmé) pendant qu'il joue ou interagit spontanément avec ses parents, sa fratrie ou d'autres personnes de son environnement (voir par exemple les corpus de la banque de données CHILDES ou, pour le français, ceux du projet *Colaje* : <http://colaje.scicog.fr/>).

Il est, en revanche, plus compliqué d'obtenir des données « écologiques » avec des sujets adultes : d'une part, l'adulte est plus conscient d'être observé et, d'autre part, ses interactions quotidiennes sont bien plus complexes et diversifiées. Pour ce cas de figure, les données « naturelles » peuvent prendre la forme de conversations libres (plus souvent semi-guidées) entre apprenants ou entre apprenants et locuteurs natifs, ou encore d'enregistrement de sorties accompagnées dans différents milieux. Une partie des données ESF (voir chapitre Véronique) est ainsi composée d'enregistrements de différentes situations de communication (dont des conversations libres avec un chercheur, des sorties accompagnées au bureau de poste...), que l'on suppose représenter les différents contextes et usages de la L2 en dehors de la situation de recherche.

¹ Le classement des types de données varie selon les auteurs et les critères adoptés. (Chaudron, 2003 : 765) classe les procédures de collectes de données en trois catégories (« *naturalistic data collection procedures, elicited production procedures, experimental procedures and tasks* »), tout en soulignant la difficulté de distinguer entre les deux dernières catégories, alors que (Jarvis & Pavlenko, 2008 : 33-34) distinguent 5 catégories : (1) *natural use data* (*i.e.*, observation de l'emploi de la L2 non sollicité) ; (2) *clinical elicitation data* (*i.e.*, emploi de L2 sollicité mais non guidé, comme dans les récits de films et dans d'autres tâches permettant de documenter des performances verbales et non-verbales) ; (3) *experimental elicitation data* (*i.e.*, observation de performances linguistiques guidées, comme dans les tests à trous, et d'aspects non linguistiques, comme les temps de réaction) ; (4) *metalingual judgments* (*i.e.*, tâches de jugement de grammaticalité ou d'acceptabilité) ; (5) *self-report data* (*i.e.*, introspection, rétrospection ou verbalisation de la pensée).

Des données « écologiques » sont également attestées en classe de langue, du moment où le chercheur essaie de « préserver les conditions naturelles de leur production » (Pallotti, 2002 : 165), c'est-à-dire de ne pas influencer sur le déroulement de la classe et les interactions qui en découlent entre l'enseignant et les apprenants et / ou les apprenants entre eux.

Les recueils de données spontanées sont particulièrement intéressants pour étudier les phénomènes liés aux interactions en situation exolingue (*cf.* Py & Porquier, 2004, et chapitre Pekarek-Doelher), notamment les stratégies pour atteindre l'intercompréhension (adaptation des participants aux niveaux de langue respectifs, résolution de malentendus, alternances codiques, *etc.*) ainsi que les séquences où l'apprenant a l'opportunité de saisir des nouveaux éléments de la langue cible (« séquences potentiellement acquisitionnelles », *cf.* De Pietro, Matthey & Py, 1989).

En revanche, il s'agit de données peu comparables, toute production libre ou spontanée étant très liée au contexte spécifique dans lequel elle a lieu et pouvant évoluer sur des sujets différents (*cf.* document 1 en annexe portant sur la recette d'une boisson). En particulier, si l'objet d'étude concerne l'emploi particulier d'une structure spécifique, il se peut que la conversation n'ait pas donné l'opportunité de l'utiliser et donc d'en observer l'emploi par l'apprenant, ce qui pose le problème de comment interpréter l'absence d'une structure (stratégie d'évitement ? pas de contexte favorable à sa production ?). À titre d'exemple, si le chercheur s'intéresse au subjonctif, la conversation reportée dans l'annexe 1 n'est pas pertinente puisqu'elle ne présente pas de contextes demandant son emploi.

2.1.2 Données plus contrôlées, ou expérimentales

Il existe de nombreuses procédures de recueil des données permettant d'orienter les sujets vers la production de structures linguistiques ciblées. En fonction de la nature de la tâche et de la liberté de manœuvre qu'elle laisse au locuteur, il convient de distinguer entre les données « semi-guidées », qui s'apparentent aux données écologiques, et les données « fortement guidées », plus expérimentales, qui seront traitées plus bas. Le premier cas inclut une variété de tâches de production orale ou écrite, telles que raconter une histoire (personnelle ou fictive), décrire des images ou un itinéraire, donner des instructions, ou encore effectuer un jeu de rôles, qui donnent lieu à des discours structurés.

Ce type de données présente deux caractéristiques importantes. D'une part, l'activité est motivée par un but communicatif : le sujet s'adresse à un interlocuteur qui est le destinataire des informations verbalisées. D'autre part, il s'agit de tâches verbales complexes (*cf.* Levelt, 1989) dont l'accomplissement demande la construction d'un discours cohérent et cohésif : pour ce faire, le locuteur doit sélectionner et segmenter les informations à transmettre, décider de l'ordre dans lequel elles seront verbalisées, construire chaque énoncé et les relations inter-énoncés. À chaque étape, il jouit d'une certaine liberté de choix dans la réalisation de la tâche.

En même temps, ses choix sont contraints, dans la mesure où chacune de ces tâches demande de mobiliser des structures discursives ciblées. Si on reprend la distinction classique de types de discours (*instructionnel, narratif, descriptif, argumentatif*), chacun est caractérisé par l'emploi prioritaire de certaines structures au détriment d'autres. À titre d'exemple, la tâche de décrire un itinéraire (« comment aller de X à Y ? ») demande la mobilisation de structures encodant le déplacement, alors que les tâches descriptives sont caractérisées par la mention d'entités et leur localisation. Les tâches narratives – récit fictif ou personnel – impliquent, en revanche, un discours organisé autour d'événements se succédant sur l'axe temporel (« qui a fait quoi à quel moment ? »). Ces dernières (élicitées par un support d'images, une vidéo ou autre) sont très répandues, sans doute parce que le discours (écrit ou oral) qui en découle se prête à l'observation de plusieurs domaines (référence au temps, à l'aspect et aux personnages de l'histoire, structure de l'énoncé, liens inter-énoncés, structure narrative, *etc.*) (Voir chapitre Lenart & Leclercq).

Certains supports ont par ailleurs été repris dans de nombreuses études. Ainsi *Frog Story* (Mayer, 1969), une série de 24 images sans texte, a été utilisée d'abord pour étudier de manière comparative l'acquisition de la L1 dans différentes langues (Berman & Slobin, 1994), puis pour observer l'acquisition d'une L2 (en particulier en ce qui concerne l'expression des déplacements) ; la vidéo *Les Temps Modernes* (un montage du film homonyme muet avec Charlie Chaplin) a été utilisée dans le projet ESF (*cf.* Perdue, 1993), puis dans de nombreuses autres études, permettant ainsi la comparaison des résultats obtenus.

Dans les tâches décrites ci-dessus le participant produit un discours qui se rapproche des monologues (en fonction du protocole, l'interlocuteur peut ne pas intervenir ou se limiter à des questions de clarification), alors que d'autres sont conçues pour induire une plus forte interaction entre les

participants, chacun ne détenant qu'une partie de l'information nécessaire pour l'accomplir (tâches de déficit d'information). Il en est ainsi, par exemple, lorsque deux participants doivent décrire ensemble un itinéraire sur la base d'images partiellement différentes (*map tasks*), où chacun a accès à une image incomplète du même itinéraire, ou encore lorsqu'ils doivent identifier ensemble les différences entre deux images (*spot-the-difference task*), sans avoir accès à l'image de l'autre.

Indépendamment du degré d'interaction, ces types de données présentent l'avantage de permettre l'observation de structures linguistiques mises en œuvre dans le discours. Il est toutefois possible de ne pas obtenir assez d'occurrences de la structure linguistique qui intéresse le chercheur, malgré le caractère relativement contraint de la tâche. C'est le cas lorsqu'on s'intéresse à des structures peu fréquentes (par exemple, le participe présent) ou lorsqu'on souhaite obtenir des données qui donnent un aperçu des différents emplois d'une structure (par exemple, les différents sens véhiculés par l'imparfait). On peut alors recourir à des tâches plus fortement guidées, qui en « forcent » l'emploi, telles que la complétion de discours ou de phrases (plus ou moins contextualisées), les exercices à trous (*cloze tests*) ou le choix de réponses pré-déterminées (voir Mackey & Gass, 2005 pour une liste plus exhaustive).

L'augmentation de contraintes dans la tâche peut impliquer une diminution (voire une annulation) du but communicatif dans son accomplissement. Par ailleurs, on peut se demander si le sujet qui a utilisé une certaine structure dans ce type de tâche serait capable de la mobiliser également dans des situations plus écologiques d'emploi du langage. Les données ainsi obtenues donnent néanmoins des indications sur les connaissances des participants. Des tâches de jugement d'acceptabilité permettent aussi de tester les intuitions des locuteurs à propos des structures cibles.

Pour tester des hypothèses bien précises, il est également possible d'avoir recours à des méthodes psycholinguistiques, empruntées à la psychologie expérimentale, qui peuvent être appliquées aussi bien à des tâches verbales (par ex., décisions lexicales) que non verbales (par ex., appariement d'images). On peut ainsi mesurer, les temps de réaction du sujet dans des tests d'appariement de phrases (*sentence matching tests*) : en visionnant des phrases, présentées l'une après l'autre sur un écran, le sujet doit décider si les phrases sont identiques ou pas en appuyant le plus vite possible sur le bouton correspondant à sa décision. On peut aussi observer les mouvements oculaires, notamment les fixations du regard lors du visionnage d'une image ou d'une vidéo. Ces techniques, généralement centrées sur la compréhension / interprétation du stimulus, permettent de mesurer les différentes stratégies de traitement de l'information, qu'elle soit de nature verbale ou non verbale (voir chapitre Colonna).

2.1.3 Données introspectives ou rétrospectives

Une autre source d'information sur le processus acquisitionnel est fournie par des données qui ne concernent pas directement l'usage de la L2 (en production ou en compréhension), mais plutôt un retour réflexif de la part de l'apprenant sur son expérience d'apprentissage ou d'utilisation de la L2. Voici quelques procédures visant l'obtention de données de ce type :

- la tenue d'un journal de bord : l'apprenant est censé prendre des notes, au fil du temps, sur sa situation d'apprentissage, les difficultés rencontrées et les stratégies déployées pour les dépasser (par exemple la comparaison avec d'autres langues, la consultation de dictionnaires, *etc.*) ou sur d'autres éléments typiques d'un cours de langue ;
- la verbalisation à voix haute de la pensée pendant l'accomplissement d'une tâche : un sujet devant résoudre des mots croisés verbalise au fur et à mesure les stratégies mises en œuvre pour récupérer les mots en question ;
- les activités de verbalisation « à chaud », juste après l'accomplissement d'une tâche en L2 permettent d'éclairer des comportements verbaux ou non verbaux observés lors de la tâche, qui se prêtent à différentes interprétations. À titre d'exemple, si le sujet a effectué de longues pauses dans une tâche narrative en L2, il est intéressant de lui en demander la cause (recherche de mots, incertitudes sur la structure à utiliser, indécision sur les événements à relater...). Le document 2 en annexe illustre l'emploi de cette procédure.

On peut apparenter à ce type de données toutes sortes de questionnaires : en fonction des points abordés, ils permettent d'obtenir des informations sur les attitudes, opinions, motivation des apprenants, ou plus en général sur leurs représentations concernant l'apprentissage des langues ou d'une langue

spécifique (cf. Mackey & Gass, 2005 : 93-94, et chapitre Berdal-Masuy sur les profils d'apprenants). Il n'est pas clair jusqu'à quel point les données introspectives (*verbal report data*) sont valides et fiables. Toujours est-il qu'elles nous donnent un aperçu de processus mentaux qui ne seraient pas accessibles autrement.

2.2 Etudes de cas vs. observation de groupes d'apprenants

Lors de la conception d'un corpus, un des choix à faire concerne le nombre de sujets à observer et la durée de l'observation au fil du temps. Sur ce point on peut distinguer les études centrées sur le développement chez un sujet (ou un groupe restreint de sujets) – ce qu'on appelle des « études de cas » – par rapport à celles qui impliquent l'observation de groupes d'apprenants numériquement plus importants.

Les études de cas sont souvent basées sur des données *longitudinales* : les mêmes sujets sont observés plusieurs fois sur une période de temps relativement longue (par exemple six mois ou un an). La dimension temporelle de l'enquête donne un aperçu sur l'évolution du système de l'apprenant au fil du temps, à savoir sa complexification suite à l'intégration de nouvelles structures lexicales et grammaticales au fur et à mesure qu'il se rapproche de la langue cible, ou bien sa stabilisation à un certain stade (ce qu'on appelle aussi *fossilisation*). A titre d'exemple, nous pouvons citer les études de Leopold mentionnées plus haut (un sujet suivi sur seize ans !) ou les données longitudinales du projet ESF, qui couvrent trente mois d'observation pour trois ou quatre apprenants par combinaison de LS / LC.

La richesse des données obtenues auprès d'un même sujet permet de décrire (et interpréter) les caractéristiques de son lecte à un moment donné et son évolution au fil du temps. Ces recueils sont précieux pour des recherches « descriptives » ou « exploratoires », notamment pour établir des hypothèses sur un phénomène acquisitionnel qui n'a pas encore été étudié, hypothèses qui seront à vérifier ensuite sur des populations plus larges. En effet, les résultats ainsi obtenus sur quelques sujets sont souvent peu généralisables : on ne peut savoir dans quelle mesure l'itinéraire acquisitionnel observé est représentatif des parcours d'autres sujets ou bien spécifique au sujet / contexte en question.

La prise en compte de groupes de sujets plus importants donne typiquement lieu à des recueils de données *transversales*, qui consistent à observer plusieurs individus à un moment spécifique de leur apprentissage. Ce type de recueil implique la sélection d'un groupe relativement nombreux de participants, se situant au même stade et partageant les mêmes conditions d'apprentissage (par ex., une classe de 2^e année de français en Chine) : on n'essaie pas de voir le changement du système linguistique au niveau de l'individu, mais plutôt de faire ressortir les usages en L2 partagés par l'ensemble du groupe. Pour ajouter une dimension développementale, il est possible de considérer des groupes d'apprenants représentatifs de niveaux acquisitionnels successifs, tels que débutant, intermédiaire, avancé (données *quasi-longitudinales* ou *pseudo-longitudinales*). En reprenant l'exemple donné plus haut, il s'agirait de recueillir des données d'étudiants chinois en première année de français, en 2^e année de français et en 3^e année de français. Chaque cohorte illustre un moment de l'apprentissage, et les différents groupes de niveau sont censés donner un aperçu de l'évolution au fil du temps, en partant de l'hypothèse qu'ils passent par les mêmes étapes. Bien que la quantité de sujets observés soit variable – on peut considérer des groupes plus ou moins nombreux en fonction du phénomène analysé – cette procédure vise à fournir des résultats plus généralisables que les études de cas. Les données transversales sont par ailleurs mises en place surtout pour valider des hypothèses précises (recherche confirmative).

Cette modalité de recueil de données présente certes des avantages, tels que le gain de temps (recueils plus rapides), mais aussi une certaine difficulté dans la composition d'un groupe homogène de sujets étant donné le nombre de variables potentiellement impliquées (niveau, langues connues, motivation, milieu d'apprentissage, etc.). Il est ainsi usuel de recueillir des données socio-biographiques sur les apprenants, permettant de connaître plus en détail leur parcours d'apprentissage (âge de début, durée, contexte : milieu guidé ou naturel...).

Pour garantir la fiabilité des données transversales, il est aussi essentiel de contrôler le niveau de langue des apprenants. Le choix du test est crucial : il doit s'agir d'une mesure valide (c'est-à-dire qui teste réellement la compétence ciblée), fiable et pratique à mettre en œuvre, faute de quoi les résultats de l'étude risquent d'être ininterprétables (cf. Leclercq & Edmonds, 2014). Parmi les moyens fréquemment utilisés figurent le jugement intuitif (auto-évaluation ou jugement d'experts), le statut institutionnel (généralement le nombre d'années d'étude de la L2) ainsi que les tests d'évaluation « maison » ou standardisés (pour le français, DELF ou DALF, alignés sur les niveaux du CECRL). Les

jugements intuitifs sont néanmoins peu fiables, le nombre d'années d'études de la L2 n'est pas très révélateur, en raison de la grande hétérogénéité des apprenants ayant suivi le même programme ; quant aux tests, tout dépend des habiletés évaluées : un test portant sur les connaissances grammaticales ne pourra prédire que de manière limitée la compétence d'un apprenant en compréhension orale. Parmi les autres techniques figurent aussi des tests de vocabulaire (par exemple, Lex 30, qui évalue le vocabulaire productif : <http://www.lognostics.co.uk/tools/Lex30/>), la répétition d'un texte oral (le degré d'exactitude de la répétition reflète le niveau de l'apprenant, dans la mesure où il ne peut répéter correctement les énoncés entendus que s'il les a compris et segmentés correctement), des tests à trous pour évaluer les connaissances lexicales et grammaticales, ainsi que des questionnaires d'auto-évaluation (comme par exemple le *Bilingual Language Profile* : <https://sites.la.utexas.edu/bilingual/>).

Généralement, les chercheurs préconisent une forme de « triangulation » des évaluations, en utilisant différents outils pour pouvoir cerner au plus près le niveau de l'apprenant.

2.3 Approche comparative et contrôle des variables

Les études en acquisition impliquent souvent des comparaisons, dans la mesure où l'on considère la distance entre les usages des apprenants et ceux des locuteurs natifs de la même LC, à parité de tâche. Ce genre de comparaison est précieux pour déterminer, par exemple, des phénomènes de sous-emploi ou sur-emploi d'une structure, qui passeraient inaperçus sans la prise en compte d'un groupe de contrôle (justement celui des natifs). En fonction du phénomène analysé, il peut y avoir par ailleurs un certain écart entre les structures disponibles dans une langue (telles que décrites dans les grammaires) et l'usage réel qu'en font les locuteurs. A titre d'exemple, plusieurs grammaires prévoient l'emploi du subjonctif dans la subordonnée régie par *après que*, alors que dans les usages natifs le subjonctif alterne avec l'indicatif.

Une dimension comparative plus complexe est celle qui vise à identifier l'influence d'un facteur spécifique parmi ceux qui affectent le processus acquisitionnel. Comment décider si certains usages en L2 sont dus par exemple à l'âge du sujet, à sa motivation, à son niveau en L2, à l'influence de la L1 ou d'autres langues connues, ou encore à l'input auquel le sujet est exposé ? Afin de répondre à ces questions, il devient nécessaire d'adopter une démarche qui permette d'observer les effets du facteur en question en l'isolant par rapport aux autres variables.

Ainsi, pour identifier l'influence de la L1 (ou d'autres langues antérieurement acquises), un des facteurs cruciaux en L2, on ne peut se contenter de considérer les performances d'un groupe d'apprenants ayant la même L1 par rapport à des locuteurs natifs de la LC. Il est en effet possible que les écarts attestés en L2 soient partagés par des apprenants de L1 différentes : autrement dit, ils ne seraient pas motivés par l'influence de la L1 en question, mais plutôt par la complexité de la langue à apprendre et, partant, typiques d'un certain niveau des variétés d'apprentissages de cette LC. Par exemple, si un apprenant produit « *he walk* », l'omission du -s de 3^e personne est-elle due à l'influence de la langue maternelle ou est-elle typique des apprenants débutants en anglais L2, quelle que soit leur L1 ?

Jarvis & Pavlenko (2008) listent trois conditions de base pour attribuer un comportement en L2 à l'influence de la L1 : a) un groupe d'apprenants, partageant la même L1 et le niveau de connaissances en L2, manifestent un comportement linguistique homogène en L2 (*homogénéité à l'intérieur du groupe*) ; b) ce comportement est spécifique au groupe ayant la même L1 et n'est pas observé chez les apprenants ayant une autre L1 (*hétérogénéité entre les groupes*) ; c) le comportement linguistique en L2 rappelle l'emploi d'une structure de leur L1 (*crosslinguistic performance congruity*). Il s'ensuit que, pour isoler le facteur « influence de la L1 », il devient nécessaire de prendre en compte des apprenants d'au moins deux L1 différentes pour une même LC et de procéder par des comparaisons successives.

L'impact d'autres facteurs peut être investigué de manière similaire, à savoir en comparant les performances d'apprenants qui diffèrent (autant que possible) uniquement pour la variable correspondante au facteur en question : à titre d'exemple, les effets du contexte guidé vs. non-guidé (Schlyter, 2003 : suédophones apprenant le français L2 en immersion en France vs. dans des cours de langue en Suède) ou d'un séjour à l'étranger (Llanes & Muñoz, 2013 : apprenants hispanophones effectuant un séjour au Royaume Uni vs. suivant des cours d'anglais en Espagne), l'impact du degré de littéracie (Mocciaro, 2020 : migrants débutants avec différents degrés de scolarisation et de littéracie),

l'influence de la L2 sur l'acquisition d'une L3 (voir chapitre Trévisiol) ou encore l'effet de l'âge d'exposition à la LC (cf. Munoz, 2003 sur des enfants hispanophones exposés à l'anglais à partir de différents âges). Certaines recherches essaient d'établir des corrélations entre différents facteurs au lieu d'en isoler un : par exemple, la correspondance entre le degré d'authenticité dans la prononciation et la motivation envers la LC (cf. Birdsong, 2003 sur la prononciation en français L2 par des apprenants anglophones ayant tous été exposés tardivement à la L2).

La dimension comparative est également utilisée pour tester l'efficacité de différentes techniques d'enseignement (types d'activités, de feed-back, etc.) dans le contexte « classe de langue », en prenant soin d'avoir un groupe expérimental (sujets exposés aux techniques à tester) et un groupe de contrôle (sujets exposés aux techniques classiques), ainsi que des tests mesurant les connaissances des apprenants avant et après l'expérience (cf. Mackey & Gass, 2005 ; Ellis, 2012).

3. Du recueil des données à leur interprétation

Après avoir déterminé l'architecture générale des données, d'autres choix méthodologiques s'imposent afin de rendre possible leur analyse et interprétation : ces choix concernent la transcription des données ainsi que le codage et la quantification des structures ciblées.

3.1 La transcription

Pour que l'étude soit valide, le chercheur doit donc avoir posé une question de recherche claire et précise, qui aura donné lieu à des hypothèses de travail, lesquelles auront permis l'élaboration d'un dispositif expérimental débouchant sur un recueil de données pertinent.

Le mode de recueil et le type de données dépend de la modalité étudiée (orale, écrite, gestuelle) : les données écrites peuvent être manuscrites ou tapuscrites, et l'on peut même enregistrer le déroulement du processus d'écriture, comme par exemple avec le logiciel *Inputlog* (cf. chapitre de Barbier dans ce volume).

Quant aux données orales, une fois qu'elles sont enregistrées ou filmées (ce qui permet aussi d'étudier la communication non-verbale), il faut ensuite les transcrire : cette étape amène à faire des choix de catégorisation des productions en L2, qui conditionnent l'interprétation des données. Par exemple, comment transcrire des formes verbales ambiguës, comme celles produites par des apprenants du français de niveau débutant ou intermédiaire telles que /zue/, /tombe/, /sove/ (voir transcription en annexe 3) ? Faut-il les représenter à l'aide de formes orthographiques (*jouer, joué, jouait, jouaient...*), ce qui implique de décider dès la transcription si l'apprenant maîtrise la conjugaison du français ? Ou bien utiliser l'alphabet phonétique, ce qui permet d'interpréter de manière plus ouverte les formes produites ? Le degré de finesse de la transcription va aussi varier en fonction de la question de recherche : on peut décider de faire apparaître certains phénomènes comme les hésitations ou les répétitions, par exemple pour étudier la fluence. Une discussion approfondie des conséquences des choix de transcription se trouve dans Mondada (2000), Leclercq (2020) et Benazzo & Watorek (sous presse / 2020).

D'un point de vue technique, les transcriptions sont le plus souvent effectuées avec des logiciels spécialisés, téléchargeables gratuitement, tels que CLAN (<https://childes.talkbank.org/>), ELAN (<https://archive.mpi.nl/tla/elan>), ou ExMARALDA (<https://exmaralda.org/en/>), qui permettent d'aligner le support vidéo ou audio et la transcription, ainsi que d'annoter les données.

3.2 Codage, quantification et interprétation

Une fois les données transcrites, il faut identifier les formes cibles, qui pourront alors être codées en fonction de la question de recherche : par exemple, dans une étude sur les temps et aspects en discours, on peut relever tous les prédicats employés, et indiquer pour chacun d'entre eux l'aspect lexical (état, activité, achèvement ou accomplissement), l'aspect grammatical (accompli, inaccompli, progressif...), et le temps (présent, passé). Lorsque les données sont ainsi étiquetées, le chercheur pourra quantifier les

phénomènes, à l'aide des statistiques. Un premier niveau, descriptif, permet de caractériser les données : il s'agit des comptages d'occurrences, des pourcentages, moyennes, écart-types.

Un deuxième niveau, principalement utilisé lorsque les corpus sont suffisamment volumineux, consiste en l'utilisation de statistiques inférentielles pour tenter d'établir si le nombre d'occurrences d'une certaine structure ou phénomène est significatif (c'est-à-dire que les conclusions tirées des données observées pourront très probablement s'appliquer à d'autres). La description des outils statistiques dont dispose le chercheur dépasse le cadre de ce chapitre. Les lecteurs pourront consulter sur ce thème, entre autres, Grosjean & Dommergues (2011), et en anglais, Norris, Plonsky, Ross & Schoonen (2015).

Il est essentiel de documenter avec précision et transparence la méthodologie adoptée, pour le recueil des données comme pour leur transcription, leur annotation et leur description chiffrée, car cela garantit leur fiabilité, leur comparabilité avec d'autres études, ainsi que leur réplication, notamment la possibilité pour d'autres chercheurs de reproduire, exactement ou partiellement, un dispositif expérimental déjà mis en œuvre, afin de vérifier si les résultats obtenus sont identiques ou divergents. Un outil précieux, IRIS (*Instruments for Research into Second Languages* : <http://www.iris-database.org>), permet de partager des outils de recherche et des dispositifs expérimentaux en vue d'en faciliter la réplication (cf. Marsden, Mackey & Plonsky 2016).

Finalement, les résultats sont analysés à la lumière du cadre théorique pour répondre à la question de recherche initiale. Cela donne lieu à une évaluation de la généralisation des résultats : des éléments de réponse sont ainsi apportés, mais souvent des recherches complémentaires s'avèrent nécessaires pour apporter de nouvelles preuves (ou contredire les résultats). La section suivante illustre cette dynamique de la recherche autour de l'étude des particules de portée.

4 Evolution dans l'étude d'un même phénomène linguistique pour en saisir l'acquisition

Cette section donne un exemple concret de différents types de données utilisés pour saisir l'acquisition en L2 d'une même structure linguistique. L'accent est mis sur les aspects acquisitionnels que chaque méthodologie permet de faire ressortir, ainsi que sur les questions qui émergent au fur et à mesure et qui amènent à considérer d'autres types de données afin d'y répondre. Les particularités de la structure en question sont présentées avant de passer à la partie méthodologique.

4.1 Des particules à portée sémantique variable

L'objet d'étude initial est constitué par un ensemble d'items de nature adverbiale tels que *aussi*, *également*, *seulement*, *même*, qui partagent un ensemble de propriétés sémantico-syntaxiques (cf. König, 1991). Pour comprendre leur contribution sémantique, prenons l'exemple de la phrase « Marie parle chinois » et regardons comment la présence de ces particules la modifie :

- (1) contexte : quelles langues parle Marie ?
- a) Marie parle *aussi* chinois. (additif)
 - b) Marie parle *même* chinois. (additif et scalaire)
 - c) Marie parle *seulement* chinois. (restrictif)

La valeur sémantique de *aussi* est dite « additive » : sa présence dans l'énoncé (1a) implique que Marie parle d'autres langues en plus que le chinois. *Même* en (1b) a également une valeur additive, mais il ajoute une nuance de plus, à savoir que le chinois n'est pas la langue la plus probable à laquelle on s'attendait (valeur scalaire). *Seulement* (1c) a en revanche une valeur restrictive, dans la mesure où cette particule exclut que Marie parle d'autres langues. La présence de ces items dans l'énoncé implique par conséquent une présupposition sur une autre assertion.

Ces particules présentent une forte mobilité syntaxique et la possibilité d'affecter sémantiquement différents constituants de l'énoncé : un changement de leur position syntaxique peut s'accompagner d'une variation de portée sémantique mais ce n'est pas toujours le cas. Chaque particule présente en effet une configuration spécifique de positions admises et de portées possibles. Par exemple, « même Marie parle chinois » implique que d'autres personnes parlent chinois, la particule s'associant au sujet « Marie », alors qu'en (1b) elle s'associait au complément « chinois » (« Marie parle même

chinois ») : en simplifiant, on peut dire que *même* précède le constituant dans sa portée sémantique². En revanche, *aussi* peut le précéder (comme en (1a) ou le suivre. En effet, lorsque cette particule porte sur le sujet, elle lui est toujours postposée : pour indiquer que Marie parle chinois en plus d'autres personnes, *aussi* peut ainsi se placer directement après le sujet (« Marie aussi parle chinois »), mais également après le verbe (cf. (1a) « Marie parle aussi chinois ») ou en fin d'énoncé (« Marie parle chinois aussi »). Il en résulte une forte dépendance du contexte pour leur interprétation ou, en d'autres termes, une interaction importante avec la structure informationnelle de l'énoncé (d'où leur désignation de *focus particles*, König, 1991).

Etant donné leur mobilité syntaxique et complexité structurelle (interface syntaxe / sémantique), l'étude de ces particules est particulièrement intéressante pour analyser la structure de l'énoncé en L2.

4.2 Etudes exploratoires dans des données longitudinales

Un premier ensemble d'études s'est intéressé à ce sujet de recherche sur la base de données longitudinales recueillies auprès d'un public de migrants. Il s'agit d'une partie des données ESF pour le français, anglais et allemand L2 (Benazzo, 2005 ; Giacomi, Meyfren, Stoffel, Tissot & Véronique, 2000), des données P-Moll pour l'allemand L2 (Dimroth, 1998) et des données du projet de Pavia pour l'italien L2 (Andorno, 2005)³. Ces différents corpus avaient été initialement recueillis pour avoir un panorama général sur l'acquisition en milieu naturel par des migrants, dont la plupart étaient de vrais débutants au début de la période d'observation. Chaque recueil comporte des données écologiques (conversations libres) ainsi que des données plus contrôlées (tâches discursives plus ciblées : récits de films, descriptions, etc.).

Etant donné le niveau initial des apprenants et leur suivi longitudinal, ces corpus sont très appropriés pour observer l'émergence des structures linguistiques dans les lectures des apprenants et l'évolution dans leur emploi au fil du temps. L'étude des particules de portée a été ainsi guidée par les questions suivantes : À quel stade apparaissent ces items ? Avec quelles fonctions ? Comment sont-ils intégrés dans l'énoncé (positions des particules et du constituant dans leur portée sémantique) ?

Les analyses ont mis en évidence l'apparition d'au moins un item additif et d'un item restrictif (en français il s'agit respectivement de *aussi* et *seulement*) aux stades initiaux de l'acquisition, à savoir lorsque les énoncés en L2 sont constitués d'items nominaux juxtaposés (absence de formes verbales clairement identifiables) et la production des apprenants nécessite de l'étayage de l'interlocuteur. Ce résultat amène à s'interroger sur les raisons pouvant expliquer leur précocité d'émergence.

À ce propos, la disponibilité de données issues de différentes tâches a permis de constater que ces particules sont plus nombreuses dans les conversations libres que dans les autres types de données. C'est dans les séquences conversationnelles que l'on a pu remarquer leur utilité chez les débutants (Benazzo, 2008), ce qui peut en expliquer l'émergence précoce : leur valeur anaphorique permet à l'apprenant (AP) de s'appuyer sur les propos de l'interlocuteur expert (LN), sans devoir construire toute la proposition sous-jacente (cf. ex. 2-3). En même temps elles permettent de moduler une réponse qui serait sinon totalement affirmative ou négative.

(2) (à propos de ne pas apprécier les achats dans les supermarchés)

LN : je deviens énervée, agressive

AP : oui oui moi aussi

(3) LN : elle parle espagnol ?

AP : non rien d'espagnol, /solamã/ français

La richesse des données a permis de mettre à jour des parcours acquisitionnels détaillés concernant la grammaire de la portée (positions des particules par rapport à l'élément dans leur portée) et son évolution avec la complexification de l'énoncé qui accompagne la progression en L2.

Bien que l'ordre d'acquisition soit largement partagé, il a été constaté une certaine disparité du nombre d'occurrences en fonction de l'enregistrement, de la LC et / ou de l'apprenant : étant donné la nature semi-spontanée des données, on peut se demander si ces inégalités sont motivées par les aléas

² Pour le mettre en évidence, l'élément dans la portée de la particule est reporté en italiques.

³ Ensemble d'études qui se sont déroulées en parallèle, ce sujet faisant partie des thèmes de recherche des projets européens sur l'acquisition d'une langue seconde, coordonnés par l'Institut Max-Planck de Nimègue.

des sujets abordés dans les conversations, par l'influence de la L1 ou bien à une part de variation individuelle. Pour approfondir l'étude de leur acquisition, il a paru nécessaire d'utiliser des supports permettant d'éliciter certains types de portée et de disposer en même temps d'un groupe témoin de locuteurs natifs.

4.3 Une histoire par images pour éliciter différentes particules additives

Un premier support a été ainsi conçu pour éliciter différents types de particules additives (équivalant à *aussi*, *encore*, *toujours*). Il s'agit d'une histoire par images (*histoire additive*, Dimroth, 2002), initialement mise au point pour un public d'adultes en L2, puis adaptée à un public enfantin⁴, avec une histoire dont le but était de deviner qui a libéré une princesse emprisonnée dans un château.

La difficulté principale dans la conception du support concernait la création de contextes encourageant l'emploi des particules additives, étant donné que leur présence n'est jamais obligatoire. Pour ce faire, les images illustrent les actions similaires de deux personnages, qui sont toutefois effectuées en succession temporelle (4), avec la consigne de raconter l'histoire image par image.

- (4) (extrait du récit d'un locuteur natif adulte)
Mr. Bleu entre dans un magasin
Il achète une échelle
Il se dirige vers le château (...)
Mr. Rouge achète aussi une échelle
Il se dirige lui aussi vers le château (...)

Les comparaisons effectuées entre les productions d'enfants de différents âges et de différentes L1 ont montré des spécificités développementales en fonction de la LC, mais aussi confirmé la moindre présence des particules additives (dont *aussi*) dans les données en français L1 par rapport à leurs équivalents dans les données en allemand L1, aussi bien chez les enfants que chez les adultes (Benazzo, Dimroth, Perdue & Watorek, 2004). Cette différence quantitative se retrouve également dans les productions en L2 : les germanophones utilisent plus de particules additives en français L2 que, par exemple, les polonophones (transfert de la L1). Une fois déterminée cette inégalité, qui avait déjà été soupçonnée dans les données longitudinales, il se pose la question de son interprétation : pour quelle raison certaines particules devraient être plus fréquemment utilisées dans une langue que dans une autre ?

4.4 Une vidéo pour observer l'expression des relations additives et contrastives

Cette question a été abordée dans le cadre du projet franco-allemand *Langacross* (http://ircom.humanum.fr/site/description_projet.php?projet=LANGACROSS), qui était centré sur l'influence du facteur typologique dans la construction du discours natif et non natif en différentes langues. L'étude a porté d'abord sur les différences chez les natifs de langues germaniques et romanes, avant d'aborder la question de leur acquisition en L2.

Contrairement aux langues romanes, les langues germaniques disposent d'un ensemble de particules permettant de moduler l'assertion : pas seulement additives / restrictives / scalaires, mais aussi contrastives (telles que *doch* en allemand, équivalent du *si* en français), qui servent justement à contraster une assertion précédente négative.

Un deuxième support a été ainsi mis au point pour observer simultanément l'expression des relations additives et contrastives. Il s'agit d'une vidéo composée de trente-et-un courtes séquences sans paroles (*The Finite Story*), mettant en scène les mésaventures de trois personnages : monsieur bleu,

⁴ Ce support a été utilisé dans le cadre du projet *Construction du discours chez des apprenants de langue, enfants et adultes*, centré sur les productions d'enfants de différents âges (4, 7 et 10 ans) en L1 et d'adultes de différents niveaux en L2 (cf. Chapitre Watorek).

monsieur vert et monsieur rouge⁵. Le sujet est invité à construire son récit en racontant ce qui se passe dans la vidéo, juste après avoir visionné chaque séquence : c'est donc une narration « en ligne », dont la planification est fortement contrainte par la succession de séquences pré-segmentées. Les points saillants de l'histoire concernent les réactions des trois personnages lors d'un incendie dans leur immeuble et leur sauvetage par les pompiers qui les font sauter par la fenêtre. À la différence de l'histoire par images, ce support présente une série de contextes où les protagonistes effectuent systématiquement des actions similaires et / ou opposées, comme en (5).

- (5) (l'immeuble est en flammes, les pompiers invitent les trois messieurs à sauter)
- | | |
|---|------------------------------------|
| a. M. Bleu saute par la fenêtre. | |
| b. M. Vert <i>fait de même</i> / saute <i>aussi</i> . | Changement d'entité |
| c. <i>Par contre</i> M. Rouge ne veut pas sauter. | Changement d'entité et de polarité |
| d. Mais finalement il saute <i>aussi</i> . | Changement de polarité |

L'influence du facteur typologique dans les productions natives a été étudiée à travers un premier recueil de données chez les locuteurs de deux langues germaniques (allemand et néerlandais) et deux langues romanes (français et italien) (cf. Dimroth, Andorno, Benazzo & Verhagen, 2010). Ces analyses ont été guidées par la question de recherche suivante : comment sont exprimés les contextes additifs et contrastifs ? L'analyse a pris en compte tous les moyens anaphoriques mobilisés pour marquer ces relations par rapport à des énoncés « neutres » tels que : « M. Bleu saute » / « M. Vert saute » / « M. Rouge ne veut pas sauter » / « après il saute ».

La comparaison des productions natives révèle un taux similaire de marquages des deux types de relations, alors que les moyens pour ce faire varient nettement entre les langues considérées. À titre d'exemple, dans les contextes contrastifs (changement d'entité et de polarité de l'assertion: *M. Vert ne saute pas* / *M. Bleu saute*), les germanophones et les néerlandophones marquent le contraste avec l'énoncé précédent à travers des particules telles que *doch* en allemand (6a) ou une intonation contrastive sur le verbe fini (6b), réalisant ainsi un lien anaphorique au niveau de l'assertion, alors que les francophones / italophones mettent l'accent sur le changement d'entité, en utilisant des adverbes tels que *par contre* (6c) ou la reprise pronominale de l'entité (6d).

- (6) contexte précédent : M. Vert ne saute pas
- | | |
|----|---|
| a. | <i>Herr Blau ist DOCH gesprungen</i>
(M. Bleu a BIEN sauté) |
| b. | <i>Herr Blau IST gesprungen</i>
(M. Bleu A sauté) |
| c. | Par contre M. Bleu saute |
| d. | M. Bleu LUI il saute |

Dans les contextes additifs, les locuteurs natifs des deux langues germaniques utilisent massivement les particules (*auch* en allemand et *ook* en néerlandais). En revanche, les francophones et les italophones expriment souvent la similarité (« X fait de même » / « X fait la même chose ») au lieu de l'addition (« X aussi fait Y »), ce qui donne une première explication au nombre plus faible d'occurrences de particules additives en français par rapport à l'allemand.

L'analyse de l'ensemble des données met ainsi en évidence un clivage entre la perspective discursive germanique vs. romane dans l'unité informationnelle choisie pour assurer la cohésion discursive. À la lumière de ces résultats, le recueil en L2 a été fait en opposant systématiquement, pour une même langue cible, des apprenants d'une LS typologiquement proche et d'une LS typologiquement éloignée (par exemple, pour le français L2 : italien L1 et allemand L1), de manière à vérifier l'impact de la proximité / distance typologique dans l'influence translinguistique. L'analyse des productions en français et en italien L2 révèle que la proximité génétique ne favorise pas l'intégration des structures propres à la LC, du moins au niveau avancé : en effet, les apprenants germanophones arrivent généralement à des productions plus proches de la cible que les apprenants de L1 romane (italien L1 en français L2 et *vice versa*) (cf. entre autres Benazzo, Andorno, Interlandi & Patin, 2012).

⁵ La vidéo est consultable sur IRIS : <http://www.iris-database.org>

4.5 Lecture d'énoncés pour étudier la prosodie des énoncés additifs

Pour terminer ce survol rapide, il est à noter que l'une des différences importantes entre les particules additives fr. *aussi* / all. *auch* concerne la manière d'indiquer leur portée sémantique, lorsque la position est ambiguë. En français, la portée de *aussi* (*en italique* dans l'exemple) est claire uniquement lorsqu'il se trouve après le sujet (« *Pierre aussi* boit une bière ») ; les autres positions sont structurellement ambiguës, mais il est possible de signaler l'élément auquel s'associe *aussi* par des moyens syntaxiques, notamment une reprise pronominale lorsque celui-ci est le sujet (« *Pierre* boit *lui* aussi une bière » / « *Pierre* boit une bière *lui* aussi »). En allemand, il s'agit d'indices prosodiques : si *auch* précède le constituant dans sa portée (ein Bier), ce dernier est accentué (Peter trinkt auch *ein BIER*...) ; lorsque la particule le suit (*Peter*), c'est *auch* qui est réalisé avec une prééminence prosodique (*Peter* trinkt AUCH ein Bier).

Pour étudier la question de l'influence translinguistique chez les apprenants allemands du français, il devient intéressant de vérifier s'ils auront tendance à « transférer » la variation intonative propres aux particules additives de l'allemand. Cette question mène à son tour à s'interroger, plus en général, sur le rôle de la prosodie en français : est-il possible que les locuteurs natifs utilisent également un marquage prosodique pour signaler la portée de *aussi* ?

Les données de *The Finite Story* ne se prêtent pas à l'étude systématique de la prosodie, les énoncés additifs en français présentant à la fois trop de variation dans leur structure et pas assez d'occurrences pour chaque position des particules. Une analyse fine au niveau prosodique demande, en effet, de pouvoir comparer un énoncé additif avec une phrase de référence non additive, les deux produites par le même locuteur.

Pour ce faire, une tâche beaucoup plus expérimentale a été mise au point, à savoir la *lecture d'énoncés contextualisés* (cf. Benazzo & Patin, 2017). Les énoncés à lire sont construits de manière à alterner, pour une même phrase, la position de la particule (cf. 7-8), sa portée sémantique à parité de position (cf. 8a-b), ainsi que la présence / absence de particule dans un contexte additif (9). Chaque énoncé est contextualisé par l'énoncé précédent, qui permet au participant d'interpréter la portée sémantique de l'addition (indiquée en italique), et les contextes cible sont alternés avec des distracteurs.

- (7) Pierre aime le vin. *Marie* aussi aime le vin.
- (8) a. Pierre parle roumain. *Marie* parle aussi roumain
b. Marie parle anglais. Marie parle aussi *roumain*
- (9) Marie aime la vie. *Pierre* aime la vie

A priori, il semblerait logique d'avoir un marquage prosodique dans les cas où la position de la particule est structurellement ambiguë, comme en (8a) vs. (8b). L'analyse des productions natives révèlent en revanche la présence d'un contour prosodique marqué principalement dans les occurrences où la portée de la particule est structurellement claire (position après le sujet, ex.7), mais aussi beaucoup de variation entre les locuteurs : en d'autres termes, seuls certains locuteurs l'utilisent et pas d'autres, ce qui le rend de fait optionnel (indice peu fiable). Si le marquage prosodique ne semble pas crucial en français natif, il reste à voir ce qu'il en est en français L2.

On voit ainsi un parcours sur le même objet d'observation qui, partant des résultats de recherches exploratoires sur les particules de portée en L2 (données longitudinales), a engendré des questionnements conduisant à la conception de différents dispositifs expérimentaux, conçus pour en éliciter leur production dans des contextes de plus en plus contrôlés (données transversales) en L1 et en L2. Parallèlement, leur étude a été abordée en relation avec d'autres phénomènes linguistiques. Les recueils de données successifs ont ainsi permis de tester différentes hypothèses aussi bien sur leur acquisition en L2 que sur leur emploi natif.

5. En résumé

La conception d'un dispositif de recueil de données demande la définition préalable d'une question de recherche, qui permet d'émettre des hypothèses explicatives à tester expérimentalement. Les données sont collectées, transcrites et analysées à la lumière du cadre théorique/interprétatif choisi.

1. Comme on l'a vu, les données varient en fonction de la tâche à accomplir par l'apprenant et du type de connaissances qu'elle demande de mobiliser. Chacune des techniques illustrées (écologiques, contrôlées, introspectives) présente ainsi des points forts et des faiblesses ; aucune ne peut fournir le « meilleur » type de preuve du phénomène acquisitionnel considéré car on voit des aspects différents de l'acquisition en fonction de la tâche choisie. C'est pour cette raison que, pour étudier l'acquisition d'un L2, les chercheurs combinent souvent des types de données différentes ;
2. Les choix méthodologiques d'une recherche et leur transparence sont cruciaux pour garantir la fiabilité des résultats ainsi que pour permettre la comparabilité des résultats et la réplification ;
3. Il existe de nombreux outils de qualité à disposition des chercheurs sur Internet, que ce soit en termes de supports, de tests d'évaluation des apprenants, ou encore de corpus existants qui peuvent être téléchargés.

6. Pour aller plus loin

CHAUDRON, Craig (2003), « Data collection in SLA research », dans DOUGHTY, Catherine J. & LONG, Michael H. (éds.), *The Handbook of second language acquisition*, Oxford, Blackwell, pp. 762-828.

GROSJEAN, François & DOMMERGUES, Jean-Yves (2011), *La statistique en clair*, Paris, Ellipses.

ELLIS, Rod (2012), *Language Teaching Research and Pedagogy*, West Sussex, Wiley-Blackwell.

MACKEY, Alison & GASS, Susan (éds.) (2012), *Research methods in SLA. A practical guide*, Oxford, Blackwell.

L'encadré ci-dessous recense des corpus disponibles en français L2 : les sites correspondants permettent de voir leur architecture, les questions de recherche qui en ont guidé la conception, ainsi que de consulter la totalité ou une partie des données recueillies.

❖	CHILDES : Child Language Data Exchange System https://childes.talkbank.org/ Banque de données de productions d'enfants dans différentes langues
❖	Banque de données du projet ESF : « Adult Language Acquisition » https://www.mpi.nl/world/tg/lapp/esf/esf.html ou https://slabank.talkbank.org/access/Multiple/ESF/ Apprenants de 5 LC (anglais, français, allemand, néerlandais, suédois), 2 LS pour chaque LC, pour un total de 10 combinaisons de langues. Public de migrants en milieu naturel.
❖	Projet InterFra : « Interlangue française – développement, interaction et variation » https://www.su.se/romklass/interfra Apprenants suédophones (étudiants et lycéens) du français L2
❖	Projet FLLOC : « French Learner Language Oral Corpora » http://www.flloc.soton.ac.uk/ Apprenants anglophones du français L2, âgés entre 13 et 15 ans
❖	Projet LANGSNAP : « Social networks, TL interaction and SLA during the year abroad » http://langsnap.soton.ac.uk/ Etudiants anglophones du français et de l'espagnol L2, effectuant un séjour à l'étranger

Références

ANDORNO, Cecilia (2005), « Additive and restrictive particles in Italian as a second language. Embedding in the verbal utterance structure », dans HENDRIKS, Henriëtte (éd.), *The Structure of Learner Varieties*, Berlin, Mouton De Gruyter, pp. 405-460.

- BENAZZO, Sandra (2005), « Le développement des lectures d'apprenants et l'acquisition de la portée à distance en L2 », *Acquisition et interaction en langue étrangère*, 23, pp. 65-93.
- BENAZZO, Sandra (2008), « Marqueurs de focus, structure informationnelle et lectures d'apprenants adultes d'une L2 », dans DURAND, Jacques, HABERT, Benoît & LAKS, Bernard (éds.), *Actes du 1^{er} Colloque mondial de linguistique française (CMLF)*, Paris, EDP Sciences, pp. 665-678.
- BENAZZO, Sandra, DIMROTH, Christine, PERDUE, Clive & WATOREK, Marzena (2004), « Le rôle des particules additives dans la construction de la cohésion discursive en langue maternelle et en langue étrangère », *Langages*, 155, pp. 76-105.
- BENAZZO, Sandra, ANDORNO, Cecilia, INTERLANDI, Grazia & PATIN, Cédric (2012), « Perspective discursive et influence translinguistique : exprimer le contraste d'entité en français et en italien L2 », *Language, Interaction & Acquisition*, 3(2), pp. 173-201.
- BENAZZO Sandra & PATIN Cédric (2017). « French additive *aussi*: does prosody matter? », dans DE CESARE, Anna Maria & ANDORNO, Cecilia (éds.), *Focus on Additivity. Adverbial Modifiers in Romance, Germanic and Slavic languages*, Amsterdam, John Benjamins, pp. 107-136.
- BENAZZO, Sandra & WATOREK, Marzena (sous presse/2020). « Transcriptions de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques », dans SPREAFICO, Lorenzo, BERNINI, Giuliano, VALENTINI, Ada & SATURNO, Jacopo (éds.), *L'evanescenza del parlato. Un vademecum per il trattamento digitale di dati linguisticie*, Bergamo, Sestante.
- BERMAN, Ruth A. & SLOBIN, Dan I. (éds.) (1994), *Relating Events in Narrative. A Crosslinguistic Developmental Study*, Hillsdale, NJ, Laurence Erlbaum.
- BIRDSONG, David (2003), « Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones : analyses segmentales et globales », *Acquisition et interaction en langue étrangère*, 18, pp. 17-36.
- CHAUDRON, Craig (2003), « Data collection in SLA research », dans DOUGHTY, Catherine J. & LONG, Michael H. (éds.), *The Handbook of Second Language Acquisition*, Oxford, Blackwell, pp. 762-828.
- DE PIETRO, Jean-François, MATTHEY, Marinette & PY, Bernard (1989), « Acquisition et contrat didactique : les séquences potentiellement acquisitionnelles dans la conversation exolingue », dans WEIL, Dominique & FUGIER, Huguette (éds.), *Actes du 3^e colloque régional de linguistique*, Strasbourg, Université Louis Pasteur, pp. 99-124.
- DIMROTH, Christine (1998), « Indiquer la portée en allemand L2 : une étude longitudinale de l'acquisition des particules de portée », *Acquisition et interaction en langue étrangère*, 11, pp. 11-34.
- DIMROTH, Christine (2002), « Topics, assertions, and additive words. How L2 learners get from information structure to target language syntax », *Linguistics*, 40(4), pp. 891-923.
- DIMROTH, Christine, ANDORNO, Cecilia, BENAZZO, Sandra & VERHAGEN, Josje (2010), « Given claims about new topics. How Romance and Germanic speakers link changed and maintained information in narrative discourse », *Journal of Pragmatics*, 42(12), pp. 3328-3344.
- ELLIS, Rod (2012), *Language Teaching Research and Pedagogy*, West Sussex, Wiley-Blackwell.
- GIACOMI, Alain, MEYFREN, Nathalie, STOFFEL, Henriette, TISSOT, Hélène & VERONIQUE, Daniel (2000), « Grammaire et discours en L2 : l'appropriation des phénomènes de portée en français par des arabophones », dans GIACOMI, Alain, STOFFEL, Henriette & VERONIQUE, Daniel (éds.) *Appropriation du français par des Marocains arabophones à Marseille*, Aix-en-Provence, Université de Provence, pp. 245-269.
- GROSJEAN, François & DOMMERGUES, Jean-Yves (2011), *La statistique en clair*, Paris, Ellipses.
- JARVIS, Scott & PAVLENKO, Aneta (2008), *Crosslinguistic Influence in Language and Cognition*, New York, Routledge.
- KÖNIG, Ekkehard (1991), *The Meaning of Focus Particles: A Comparative Perspective*, London, Routledge.
- LABOV William (1978), *Le parler ordinaire* (KHIM, Alain, traducteur). Paris, Éditions de Minuit. (Œuvre original publié 1972)
- LLANES, Àngels & MUÑOZ, Carmen (2013), « Age effects in a study abroad context: children and adults studying abroad and at home », *Language Learning*, 63(1), pp. 63-90.

- LECLERCQ, Pascale (2020), « Transcribing interlanguage: The case of verb-final /e/ in L2 French », dans EDMONDS, Amanda, LECLERCQ, Pascale & GUDMESTAD, Aarnes (éds.), *Interpreting language-learning data*, EuroSLA Studies Series.
- LECLERCQ, Pascale & EDMONDS, Amanda (2014), « How to assess L2 proficiency? An overview of proficiency assessment research », dans LECLERCQ, Pascale, EDMONDS, Amanda & HILTON, Heather (éds.), *Measuring L2 proficiency: perspectives from SLA*, Clevedon, Multilingual Matters, pp. 3-23.
- LEOPOLD, Werner F. (1939-1949), *Speech Development of a Bilingual Child*, Evanston, IL, Northwestern University Press.
- LEVELT, Willem J.M. (1989), *Speaking: from Intention to Articulation*, Cambridge Mass., MIT Press.
- MACKEY, Alison & GASS, Susan (2005), *Second Language Research: Methodology and Design*, Mahwah, NY, Lawrence Erlbaum Associates.
- MARSDEN, Emma, MACKEY, Alison & PLONSKY, Luke (2016), « The IRIS Repository: Advancing research practice and methodology », dans MACKEY, Alison & MARSDEN, Emma (Eds.), *Advancing Methodology and Practice: The IRIS Repository of Instruments for Research into Second Languages*, New York, Routledge, pp. 1-21.
- MAYER, Mercer (1969). *Frog, where are you?*, New York, The DIAL press.
- MOCCIARO, Egle (2020), *The Acquisition of L2 Italian Morphosyntax by Low - or Not - literate Adult Learners*. Thèse de doctorat, Université de Palerme.
- MONDADA, Lorenza (2000), « Les effets théoriques des pratiques de transcription », *Linx*, 42, <http://journals.openedition.org/linx/902>, (consulté le 19 avril 2019).
- MUÑOZ, Carmen 2003. « Le rythme d'acquisition des savoirs communicationnels chez des apprenants guidés. L'influence de l'âge », *Acquisition et interaction en langue étrangère*, 18, pp. 53-77.
- NORRIS, John & ORTEGA, Lourdes (2003), « Defining and measuring SLA », dans DOUGHTY, Catherine J. & LONG, Michael H. (éds.), *The Handbook of Second Language Acquisition*, Oxford, Blackwell, pp. 717-761.
- NORRIS, John, PLONSKY, Luke, ROSS, Steven J. & SCHOONEN, Rob (2015), « Guidelines for Reporting Quantitative Methods and Results in Primary Research », *Language Learning*, 65(2), pp. 470–476.
- PALLOTTI, Gabriele (2002), « La classe dans une perspective écologique de l'acquisition », *Acquisition et interaction en langue étrangère*, 16, pp. 165-197.
- PERDUE, Clive (éd.) (1993), *Adult Language Acquisition: Crosslinguistic Perspectives*, Cambridge, Cambridge University Press.
- PY, Bernard & PORQUIER, Rémy (2004), *Apprentissage d'une langue étrangère : contextes et discours*, Paris, Didier.
- SCHLYTER, Suzanne (2003), « Stades de développement en français. Exemples d'apprenants suédophones, guidés et non-guidés, du "Corpus Lund" ». Lunds universitet : <http://www.rom.lu.se/durs> (consulté le 01/07/2020)

Annexe 1 Exemple communication exolingue (annexe numérique possible pour le fichier son)
Extrait d'une conversation entre deux amies, dont l'une allemande (BAR) et l'autre française (MAR), portant sur une boisson, appelée « *Goldene Milch* » en allemand. Données recueillies par L. Gilloire dans le cadre d'un travail universitaire portant sur les stratégies d'intercompréhension, les séquences potentiellement acquisitionnelles et les alternances codiques.

Conventions de transcription : + indique une pause ; le soulignement signale des superpositions dans les tours de parole ; /.../ segment transcrit phonétiquement ; en **gras** les mots en anglais et en ***gras*** ceux en allemand.

00-00'48

*BAR: alors Margot ++ il / il y a quelque chose ++ e: ++ c'est quelque + ah + **I need to take it**
c'est quelque chose + à boire + et + je vais faire ++ pour toi.

*MAR: d'accord.

*BAR: c'est quelque chose + e: + e: + pour la santé.

*MAR: ouais.

*BAR: **it's very healthy** + et et + ähm + on a besoin + de ***Kurkuma***.

*MAR: m: m:.

*BAR: poivre + **pepper** + ähm + alo:rs , + ***kokosöl***.

*MAR: de l'huile de coco.

*BAR: /ly:./ de coco , /ly:::l / (rire)

*MAR: l'huile de coco.

*BAR: (rire) /ly:l / de coco.

7'27-9'37

*MAR: et ça vient d'où ça ? +++ cette recette ?

*BAR: +++ qu'est que tu veux ? (rire)

*MAR: (rire)

*BAR: (en riant) et quoi ?

*MAR: pardon c'était trop vite !

*BAR: (rire)

*MAR: d'où ça vient ?

*BAR: /du:savjẽ/ ?

*MAR: **where does it come from ?**

*BAR : a:h + d'où ça vient ! (rire) ä:hm le **drink ?**

*MAR: **yes**.

*BAR: ähm + je sais pas

*MAR: la boisson

*BAR: la boisson + ähm ++ ma coloc !

*MAR: je crois ++ que c'est le ***goldene Milch***.

*BAR: **ja**.

*MAR: c'est mon acupuncteur qui me l'a dit donc j'ai pas vérifié

*BAR: /ma/ **that makes a lot of sense** + ähm + parce que /se/ + **golden / *golden* ?**

*MAR: c'est **golden / *golden* ?** et c'est <xxx>

*BAR: et c'est pas non ! c'est pas du lait.

*MAR: c'est du **fake *Milch***.

*BAR: **fake *Milch***.

et oui + et + je suis très ++ **excited ?**

*MAR: exc ++ euh ++

*BAR: très:s

*MAR: (tousse, mimique d'embarras, et rire) e++++uh

*BAR: très excitée euh + non ! (rire)

je suis très:s

*MAR: la traduction littérale ? c'est très ex/ contente ?

*BAR: m::

*MAR: impatiente ?
 *BAR: très impatiente ? non +
 *MAR: qu'est-ce que tu::
 *BAR: je suis très **excited** / euh non + ***gespannt *** en allemand c'est ***gespannt ***
 *BAR: et c'est pas la même que **excited**
 *MAR: non
 *BAR: mais je veux dire je suis très **excited** + ähm + **if you like it.**
 *MAR: ah, je suis <xxx?> + j'ai envie d(e) savoir si t / tu aimes
 *BAR: /ʒɑ̃/ /ʒɑ̃vi/ de savoir ++
 *MAR : j'aimerais bien savoir si tu aimes ? + quelque chose comme ça ?
 *BAR: si tu aimes , oui
 ähm + en allemand *** ich bin gespannt ob du / ob es dir schmeckt ***
 *MAR: a:h.
 *BAR: *** ich bin gespannt ob es dir schmeckt ***
ja, ähm + parce que moi + j'aime bien !
 *MAR: ok.
 *BAR: et + maintenant + on va + mettre + le **fake** lait ?
 *MAR: (rire) le faux lait.
 *BAR : le faux lait + faux lait l'avoine
 *MAR: ah oui c'est vraiment <golden / ***golden***>
 *BAR: oui !

Annexe 2 Réflexion métalinguistique sur l'accomplissement d'une tâche narrative

Le texte reporté ci-dessous est la traduction en français d'un entretien qui s'est déroulé en japonais après que le sujet avait effectué un récit de film (tiré de Kerrou Nadine 2019. « Expression de la temporalité dans le discours d'apprenants japonais du français : une approche acquisitionnelle et didactique », thèse de doctorat, Université de Lille. Vol. Annexes : p. 103-104)

INV: Lorsque tu as fait ton récit en français, qu'est-ce qui a été le plus difficile ?

SBJ: *Je n'arrivais pas à parler car les mots que je connais en japonais ne me venaient pas du tout en français.*

INV: Et ensuite ?

SBJ: *La structure des phrases ne me venait pas facilement.*

INV: D'accord, ensuite, lorsque tu as fait ton récit en français, as-tu pensé à la conjugaison des verbes ?

SBJ: *Euh, pas vraiment, en fait je n'y ai pas du tout pensé pour la simple raison que les verbes eux-mêmes ne me sont pas du tout venus à l'esprit, donc c'était impossible.*

INV: Lorsque tu as fait ton récit en français, as-tu pensé un peu en anglais ?

SBJ: *Oui, j'ai pensé que cependant les mots me venaient beaucoup en anglais m: drop into the water par exemple, playing, boy and a dog playing the..., playing on the ice entre autres me sont venus à l'esprit.(...)*

INV: Lorsque tu as fait ton récit en français, y a-t-il des expressions ou des phrases que tu voulais dire en français, mais tu n'as pas réussi à les dire ?

SBJ: *Euh, de toute façon, je connais tellement peu de verbes que je n'arrive pas du tout à dire ce que je veux dire.*

INV: Par exemple, qu'est-ce que tu voulais dire ?

SBJ: *Par exemple, « s'amuser », « patiner », « tomber », je voulais dire ça mais je ne connaissais pas les mots. (...)*

Récit qui venait d'être effectué par le même sujet

*INV: alors est-ce que tu peux me raconter ce qui se passe dans la vidéo ?

*SBJ: + e: un garçon et **inu tte nani** +++ **inu tte nan tte iu** ? (Jap= chien c'est quoi ? comment on dit chien ?)

*INV: un chien.

*SBJ: chien + un chien ++ un chien ++ m: ++ m: ++.

*SBJ: **asobu tte nan desu ka ? play tte eigo** (Jap= s'amuser, qu'est-ce que c'est ? play c'est en anglais)

*INV: jouer.

*SBJ: /ʒuE/ un glace.

*SBJ: le garçon ++ e: ++ **ochiru tte nan desu ka** ? (Jap= tomber, qu'est-ce que c'est ?)

*INV: tomber.

*SBJ: /tōbE/.

*SBJ: /tōbE/ dans + dans glace.

*SBJ: et le chien ++ **tasukeru tte nan desu ka ? sumimasen** (jap= sauver, qu'est-ce que c'est ? désolée)

*INV: sauver.

*SBJ: /sovE/ le garçon.

*SBJ: **ii desu ka?** (Jap= c'est bon ?)

*INV: ok merci.

Annexe 3 Transcription récit Reksio

Tâche de narration orale après visionnage du dessin animé Reksio. La participante est une apprenante canadienne anglophone du français de 31 ans, vivant en France depuis 15 mois au moment de

l'enregistrement, ayant débuté l'apprentissage du français à l'école à l'âge de 3 ans. (annexe numérique possible pour le fichier son). L'enregistrement a eu lieu par visioconférence. Le symbole & précède les verbes pour lesquels la terminaison est ambiguë (-é ? a-is ? -er ? d'où sa représentation par E).

*ERI: voilà ça enregistre .

*ERI: pouvez-vous me raconter .

*ERI: ce qui s' est passé .

*VIR: d'accord euh il y a un petit chien dans le vidéo .

*VIR: il a il a s' &enlevE .

*VIR: il a partir son petit maison de chien .

*VIR: mais il est tombé sur la glace .

*VIR: donc il a] il a appelé un petit garçon avec les cheveux blonde .

*VIR: mais les deux &tombE après ça .

*VIR: mm le garçon a jeté les petits les petits pierres .

*VIR: je sais pas le mot pour *ashes* petit cendre .

*VIR: il a jeté les petits pierres .

*VIR: donc les deux &marchE à la glace .

*VIR: eum le petit garçon donne a donné un manteau une chemise comme ça à le petit chien .

*VIR: mm ils se marchent à la lac .

*VIR: et le garçon il a prise les euh les patine .

*VIR: je pense .

*VIR: que ça c'est le mot eum .

*VIR: c'est pas parfait .

*VIR: il patine il a ils ont patiné eum sur la lac .

*VIR: mais le chien a tombé mm quelques fois .

*VIR: mm le garçon est plus bien .

*VIR: il était plus bien que le chien .

*VIR: mais il est tombé sur ou en bas de la glace .

*VIR: mais euh le chien a sauvé le garçon .

*VIR: euh et il a retourné le garçon chez il .

*VIR: et le garçon a .

*VIR: c'est pas réchauffer .

*VIR: parce que ça c' est le mot pour euh pour manger les nourritures .

*VIR: mais comme ça le garçon a réchauffé chez il .