



HAL
open science

Transcription de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques

Sandra Benazzo, Marzena Watorek

► To cite this version:

Sandra Benazzo, Marzena Watorek. Transcription de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques. In Lorenzo Spreafico, Giuliano Bernini, Ada Valentini & Jacopo Saturno (éds.) *Superare l'evanescenza del parlato. Un vademecum per il trattamento digitale di dati linguistici* (pp. 127-165). Bergamo: Sestante, pp.127-165, 2021. hal-03312984

HAL Id: hal-03312984

<https://hal.science/hal-03312984>

Submitted on 3 Aug 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.

Transcription de corpus oraux d'apprenants débutants en français L2 : quelques enjeux théoriques

1. Introduction

L'étude de la production langagière orale nous pose face aux problèmes de sa transcription à l'écrit, transcription qui devient nécessaire dès lors qu'on veut observer et revenir sur un objet qui est par définition 'évanescant'. La reproduction à l'écrit de tous les phénomènes typiques de l'oral (volume, débit, intonation, etc.), si jamais elle était possible, rendrait la transcription illisible. La transcription de l'oral implique par conséquent une série de choix, dont une partie est déterminée par l'objet d'étude, à savoir le phénomène de langue à analyser pour lequel le corpus a été conçu, et par la nécessité de trouver un compromis entre lisibilité et fiabilité de la parole transcrite. Ainsi, le degré de finesse de la transcription sera différent si l'on s'intéresse à des phénomènes de nature phonologique (par ex. la liaison) ou bien à des phénomènes lexicaux ou encore morphosyntaxiques (Baude 2006).

Si ces questions méthodologiques concernent la transcription du parler de tout locuteur, ils se posent de manière plus aiguë lorsque l'objet d'étude concerne (a) la production de sujets qui sont en cours d'apprendre la langue, que ce soit la langue maternelle (L1) ou une langue seconde (L2), et (b) lorsque la langue en question présente un grand écart entre oral et écrit, comme c'est le cas en français (cf. Blanche Benveniste 2000). Pour illustrer le type de difficultés rencontrées en français L2, nous pouvons citer l'exemple suivant mentionné par cette auteure :

(1) [mwa vule pa partir] *moi (voulais ? vouler ? voulez ?) pas partir* (Blanche Benveniste 2000: 29)

La reproduction à l'écrit de certains items de cet énoncé est relativement univoque, pour d'autres elle l'est moins. En l'occurrence, la transcription orthographique d'une forme verbale avec un [e] final - comme [vule] en (1) - oblige à choisir entre différentes désinences qui sont homophones alors qu'elles portent des informations grammaticales (accord, temps, aspect, mode) différentes : imparfait 1^{ère} p.s. *-ais*, infinitif idiosyncrasique *-er*, présent 2^e p.pl. *-ez*, etc. L'acte de transcrire implique un premier travail d'interprétation du discours en L2 qui s'avère être particulièrement délicat puisqu'il peut influencer l'analyse ultérieure du chercheur. En effet, le choix d'une forme au détriment des autres comporte le risque de sur-estimer (ou sous-estimer) les connaissances de l'apprenant à ce moment de son apprentissage.

Notre contribution est centrée justement sur les enjeux théoriques liés aux choix de transcription de la morphologie verbale dans les corpus oraux en français L2. En particulier, nous allons reprendre les critères méthodologiques appliqués dans un vaste corpus de productions d'apprenants débutant en L2 en différentes langues (projet ESF), afin d'illustrer pourquoi le choix de transcrire selon l'orthographe standard de la langue cible (LC) peut s'avérer dangereux, du moins en ce qui concerne l'étude de la morphologie verbale en français L2, que ce soit dans le but d'en étudier le développement ou de déterminer à quel stade se trouve un apprenant de cette LC.

Il est utile de souligner que les choix de transcription varient également en fonction de l'orientation théorique adoptée, comme le faisait déjà remarquer Ochs (1979). A ce propos, nous nous situons ici dans

l'approche fonctionnaliste : notre contribution sera ainsi limitée à la discussion des enjeux pertinents pour les analyses/études qui adoptent cette orientation théorique.

L'article est structuré en deux parties. La première partie sera consacrée, d'une part, à la présentation du projet ESF (structure, objectifs, approche théorique) et, d'autre part, à la description des spécificités du français, qui expliquent les choix méthodologiques sous-jacents les transcriptions en français L2 appliqués dans ce projet. Dans la seconde partie, nous allons montrer les enjeux théoriques de ces choix à travers la présentation de deux cas de figure, notamment l'acquisition de la morphologie verbale et l'acquisition de la négation. Toutes les études mentionnées ici se situent dans l'approche fonctionnaliste, dite approche des lectes d'apprenants (Klein & Perdue 1997).

2. Lectes d'apprenants en L2 et corpus oraux

L'étude de l'acquisition d'une langue seconde est un domaine de recherche relativement récent, qui a été profondément marqué par l'émergence de la notion d'*interlangue* : en effet, depuis les travaux pionniers de Corder (1967) et Selinker (1972), qui postulaient l'existence d'un système 'idiosyncrasique' propre à l'apprenant, de nombreuses études empiriques ont démontré que l'apprenant d'une L2 développe un système linguistique dont les règles de fonctionnement sont, du moins en partie, indépendantes de celles des langues en contact (la L1, ainsi que d'autres langues connues, et la langue cible).

Le projet ESF (« Adult Second Language Acquisition » cf. Perdue 1993), caractérisé par l'étude comparative de parcours acquisitionnels dans plusieurs L2 (cf. description plus détaillée en 2.1), a joué un rôle crucial dans la description de principes développementaux communs, partagés par des apprenants de langues secondes (LS)/LC différentes, ce qui a conduit à la théorisation de la variété linguistique produite par des apprenants L2 sous la forme de l'approche des lectes d'apprenants (*Learner Variety Approach*, Klein & Perdue 1997, Perdue 1993, etc.).

Les recherches menées dans le cadre du projet ESF, qui ont servi à identifier les stades acquisitionnels présentés en 2.1., sont basées sur des productions orales d'apprenants débutants. Ainsi, la réflexion sur les critères de transcription dans différentes langues cibles a été fondamentale et étroitement liée aux principes de l'approche *Learner Variety*. Selon cette approche, les apprenants adultes d'une L2 possèdent la compétence de communication en tant que locuteurs natifs d'au moins une langue et ont des besoins de communication relativement complexes qui constituent le « moteur de l'acquisition » (*factors 'pushing' L2 acquisition*). L'apprenant adulte exposé à une nouvelle langue (langue cible) élabore des hypothèses sur son fonctionnement à partir des données linguistiques de la L2 auxquelles il a accès. Les facteurs structurels provenant de l'input de la langue cible « façonnent » ainsi l'acquisition (*factors 'shaping' L2 acquisition*) (Watorek & Perdue 2005). Le processus acquisitionnel est donc envisagé comme une interaction entre des facteurs communicatifs (relativement similaires) et des facteurs structurels (dépendant en partie des spécificités de la LC). Le lecte d'apprenant reflète le travail cognitif fourni par l'apprenant dans la construction du nouveau système linguistique. Ce dernier est grammatical dans le sens où, à côté d'éléments variables, il présente des traits systématiques qui permettent de décrire les règles sous-jacentes à son fonctionnement à un moment donné de l'apprentissage. Les éléments sujets à variation reflètent l'évolution des hypothèses de l'apprenant au fur et à mesure qu'il prend conscience d'autres aspects de la langue à apprendre. Les lectes d'apprenants sont ainsi dotés d'une dynamique que le chercheur doit décrire pour pouvoir comprendre le processus d'acquisition.

Les choix de transcription appliqués dans le projet ESF tiennent compte de cette approche théorique envisageant les lectes des apprenants comme des systèmes à part entière, par définition dépourvus d'erreurs, qui seraient à décrire comme une langue inconnue :

« Learner varieties are not imperfect imitations of a ‘real language’ – the target language- but systems in their own right, error-free by definition » (Klein & Perdue 1997: 308)

Pour ce faire, deux dangers sont à éviter. Le premier consiste à juger le lecte de l’apprenant en termes de déviations ou d’erreurs par rapport à la LC (cf. la notion de ‘*comparative fallacy*’, Bley-Vroman 1983), à savoir de considérer sa façon de s’exprimer comme une imitation plus ou moins réussie de la langue à apprendre ; le deuxième concerne la tentation d’interpréter la production en L2 en termes de catégories propres à la LC (cf. la notion de ‘*proximity fallacy*’, Perdue 1993) et attribuer ainsi le statut et fonctions des noms, verbes, auxiliaires, etc. de la LC aux unités linguistiques produites par l’apprenant sur la base de leur ressemblance formelle.

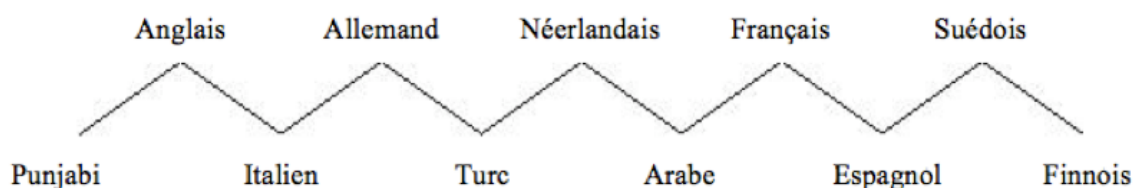
On est ici face au problème bien connu des chercheurs en acquisition de L2 : comment transcrire les productions des apprenants de manière lisible, sans en donner pour autant une image faussée par les catégories de la LC ?

Même si toute transcription implique inévitablement un travail d’interprétation, l’idée est de fournir une reproduction des segments ambigus en L2 qui serait relativement neutre, afin de ne pas sur-estimer ou sous-estimer le système linguistique que l’apprenant est en train de construire. C’est en effet l’analyse successive de ses productions, établissant des liens entre les formes produites en L2 (niveau phrastique) et leurs fonctions en contexte d’énonciation (niveau discursif), qui permet de mieux interpréter la dynamique du développement de son lecte.

2.1 Le projet ESF et les stades initiaux de l’acquisition

Le projet européen « Adult Language Acquisition » (Perdue 1993), connu aussi comme projet ESF puisque financé par la Fondation Européenne de la Science (*European Science Foundation*), s’est donné comme objectif de décrire le processus d’acquisition d’une L2 à travers les productions d’apprenants adultes de L2 en milieu naturel. Il s’agit d’une population de migrants, sans guidage systématique lié à l’enseignement, qui ont été observés de manière longitudinale. Le croisement des langues sources et cibles (cf. schéma ci-dessous) a été conçu de manière à pouvoir comparer l’acquisition de deux langues cibles différentes par des apprenants qui ont la même langue maternelle et, inversement, l’acquisition de la même langue cible par des apprenants qui ont des langues maternelles différentes. Ainsi, ce projet se caractérise par une dimension translinguistique forte.

Schéma 1 - Combinaisons de LS (en bas) et de LC (en haut) dans le projet ESF



Dans ce projet, les apprenants ont été suivis et enregistrés pendant une période de 30 mois dans différents pays européens. Les résultats des travaux translinguistiques sont comparables car la même méthodologie de recueil des données et d’analyse a été adoptée. Les apprenants ont réalisé des tâches communicatives diverses comme une description d’affiche, un récit de fiction et d’expériences personnelles, des indications d’itinéraires, des indications scéniques, etc¹.

¹ L’ensemble du corpus est disponible à l’adresse suivante : <http://www.mpi.nl/world/tg/lapp/esf/esf.html>

Un des résultats les plus robustes de ce projet a été l'explication du processus d'acquisition de L2 en termes de développement graduel du lecte de l'apprenant indépendamment des spécificités des langues en présence, sources et cibles, ce qui n'exclut pas que ces spécificités interviennent en dehors du développement partagé. En ce qui concerne le développement partagé, le premier stade (variété prébasique) comprend un répertoire restreint, composé surtout d'éléments qui ressemblent aux noms, adverbes, adjectifs (et parfois prépositions) de la langue cible (cf. ex. 2). En réalité, ces items véhiculent essentiellement des valeurs lexicales, alors qu'ils sont dépourvus d'informations grammaticales (genre, nombre, etc.). Par ailleurs, il est parfois difficile de catégoriser précisément leur statut. A titre d'exemple, un énoncé tel que « de man + ausgang » (litt. l'homme + sortie) en allemand L2, bien que compréhensible en contexte, peut se reconstruire de différentes manières (l'homme se dirige vers la sortie ? sort ? est sorti ? veut sortir ?). De manière similaire, l'item « cooking », présent en (2), est employé par le même apprenant dans différents contextes ayant comme référent le lieu 'cuisine', l'action de cuisiner ou encore la nourriture (cf. Perdue 1995). Etant donné l'absence de formes verbales clairement identifiables, ce stade se caractérise par une organisation nominale de l'énoncé. La relation entre les items qui le composent ne peut être comprise que dans des contextes discursifs précis.

Variété prébasique²

- (2) BE : *y*après [a al + a] la femme *al* camion de la police
(et après la femme (monte) dans le camion de la police)
- AN : every people happy in the cooking + tea + biscuit
(tous les gens (sont) contents dans la cuisine/de cuisiner ? + thé + biscuit)

Le passage graduel vers le stade suivant se caractérise par l'acquisition de lexèmes verbaux qui explicitent les relations entre les constituants nominaux. Les schémas des énoncés attestés sont très simples et ne contiennent que des verbes sans morphologie fonctionnelle et leurs actants : l'énoncé est ainsi structuré autour d'un verbe non conjugué, comme en (3). Les constituants suivent des ordres déterminés par un petit nombre de principes : un principe pragmatique concernant la structure informationnelle de l'énoncé (focus en dernier) et un principe sémantique, selon lequel l'actant le plus agentif se trouve en position pré-verbale (agent en premier). Ce stade d'acquisition, appelé « variété de base » (cf. Klein & Perdue 1997), marque un moment charnière dans l'acquisition de L2 dans la mesure où, à ce stade, l'apprenant devient autonome dans sa production.

Variété de base

- (3) (Berta à propos de ses deux filles)
BE: [la du fil] [se] à lycée
(les deux filles sont/vont au lycée)
- (récit de fiction *Calling 999*)
AN: another woman help for ring (...)
(une autre femme aide pour téléphone)
AN: after + come back the brigade fire
(après + arrivent les pompiers)

² Dans les extraits des apprenants, les conventions suivantes sont adoptées : les deux premières lettres font référence à l'apprenant (en l'occurrence : BE = Berta ; AN = Andrea) ; + indique une pause non remplie ; les astérisques *...* entourent des segments produits dans la langue maternelle de l'apprenant et les crochets [...] la transcription phonétique de segments ambigus. Une glose entre parenthèses signale l'interprétation contextuelle de l'énoncé produit en L2.

Le stade suivant (Variété postbasique) se caractérise par la complexification syntaxique du discours en L2 (émergence de la subordination) et, surtout, par le développement progressif de la morphologie verbale : l'énoncé est ainsi structuré autour d'un verbe qui commence à être fonctionnellement fléchi.

Variété postbasique initiale

- (4) BE: jamais je [swi perdu] dans le métro
 (jamais je ne me suis perdue dans le métro)
- AN: he has finished the work
 (il a fini le travail)
- LA: when I was young + I had a job in a shop
 (quand j'étais jeune + j'avais un travail dans un magasin)

Cependant, ce stade n'est pas atteint par tous les apprenants observés tandis que celui de la structuration à verbe non conjugué l'est : un tiers des apprenants analysés dans le projet ESF fossilise même à ce niveau. Au-delà de cette étape le développement du lecte d'apprenant se diversifie tant au niveau individuel qu'au niveau de l'influence des spécificités des langues sources et cibles.

En résumé, on peut voir que l'évolution des stades initiaux implique de manière cruciale la catégorie du verbe (émergence des formes verbales dans le passage de la variété prébasique à la variété de base) et le développement d'une morphologie verbale fonctionnelle (transition entre la variété de base et la variété postbasique).

Afin de limiter le danger de la surinterprétation, les chercheurs du projet ESF ont opté pour un certain nombre de principes communs de transcription, qui ont été toutefois adaptés par chacune des équipes des cinq pays participant au projet.

« During the early phase of the project, the need was for all-purpose, theory-neutral transcriptions, serving as general basis for different types of more specific analysis in the different research areas of the project (...) transcription of the verbal material was a compromise between an orthographic and a phonetic transcription reflecting systematic deviations from the standard TL realization of tokens. How exactly this framework was filled differed from team to team » (Feldweg 1993: 109)

La variation à laquelle Feldweg fait référence en termes d'*équipes* s'avère être plutôt une différence en termes de *langue cible*. Pour illustrer l'impact de cette dernière sur la manière de transcrire, nous proposons de comparer deux extraits de transcriptions, dont l'un en anglais L2 (apprenant italoophone Santo) et l'autre en français L2 (apprenante hispanophone Berta), qui correspondent à des conversations libres entre un enquêteur du projet ESF (LN dans les extraits) et l'apprenant.

- (5) LN: *so you didn't have any holiday ...*
 me for holiday er no september
 because er ++ er *se* I go in september for holiday no possible christmas
 you understand?
 allora I no like london christmas...
 er last christmas in london
 and next + in my country...
 when you when holiday you?

LN *I've just had a holiday*
 and when you going another one?

- (6) LN *tu étais partie en vacances + tu peux me raconter un peu ça?*
qu'est ce que tu as fait là bas?
 moi + [ale] + [a fe] de ski
 [se] très très très dur * por por primera * fois
 [nEpa] possible [ZE môte] *sobre* les [eski]

LN et euh tu avais déjà fait du ski ?
 jamais jamais + [sE] la première fois...
LN et alors ? ça s'est bien passé ces vacances ?
 oui oui + [solamâke] ++ [nEpa] de chance * por * gladys
 eh + la jambe + la + [se kase] *en* [do] fois
LN comment elle a fait ça ? + comment + c'est arrivé ?
 le ski elle + elle [mõt] + * en * la montagne
 y après + *se desliza* seule *y + y* [el tōb / tōbe]

Les deux extraits présentent les traits typiques de la variété de base, dont l'organisation des énoncés autour d'une forme verbale qui est encore dépourvue de valeur fonctionnelles (soit de marques de temps, mode, aspect et/ou accord avec le sujet). Chez Santo (et tous les apprenants italophones du projet, à ce stade) la plupart des verbes apparaissent sous la forme de la racine verbale (*go, like*) dans tout contexte temporo-aspectuel ; ils alternent parfois avec la forme Ving (comme *go – going*), mais de manière aléatoire, à savoir la variation formelle ne marque pas l'aspect imperfectif. Chez Berta (et les apprenants du français en général) les verbes lexicaux sont produits souvent avec un [e] final, qui alterne parfois avec la racine verbale ([mõtē] - [mõt]).

Dans l'extrait on note aussi la forme invariable [se], correspondant à la structure *c'est*, mais ne présentant aucune variation formelle de personne, temps, etc. (cf. section sur la négation 3.2). Un autre trait typique de ce stade est l'absence des modaux : à ce propos on peut remarquer que, dans les deux cas, la valeur modale du verbe *pouvoir* est exprimée de manière lexicale grâce à l'adjectif 'possible' et à sa négation : *not possible* en (5) et [nEpa] *possible* en (6). Par ailleurs, la négation est clairement préverbale en anglais (*I no like London Christmas*), alors qu'on peut avoir plus de doutes sur son positionnement en français, ou mieux sur la présence d'une forme verbale dans des séquences telles que [nEpa] *de chance* (ce point sera développé dans la section 3.2 suivante).

En revanche, les deux extraits diffèrent de manière remarquable en ce qui concerne le recours à la transcription orthographique vs. phonétique : toutes les formes verbales de la production en français L2 sont transcrites en symboles phonétiques (bien que simplifiés), alors qu'elles ne le sont pas dans l'extrait homologue en anglais L2. Cette différence est dû à l'ambiguïté phonologique inhérente aux formes verbales produites en français : la transcription orthographique obligerait le transcripteur du français à choisir entre des désinences bien précises, porteuses d'informations grammaticales distinctes. En effet, le français oral présente beaucoup de formes homophones et hétérographes, comme nous le verrons dans la section suivante.

2.2 Français écrit / français parlé

Le français présente un écart considérable entre l'écrit et l'oral, qui concerne non seulement la faible correspondance graphèmes / phonèmes³, mais aussi le fonctionnement des marques grammaticales ainsi que la structuration de la phrase et du texte (cf. Blanche Benveniste 2000 ; Riegel *et al.* 2009). Aborder l'oral à partir de l'écrit donnerait une image faussée de son fonctionnement, puisqu'une bonne partie de la morphologie écrite n'est pas réalisée à l'oral (cf. la notion de morphologie silencieuse : Agren 2008).

³ La non correspondance phonème / graphèmes est reconduite à trois cas de figure (cf. Riegel *et al.* 2009: 31-33) : (a) nombre de phonèmes différant du nombre des lettres (par exemple, graphème complexe équivalant à un phonème unique, tel que *eau* = [o], ou graphème unique correspondant à divers phonèmes, tel que x = [ks] dans *taxi* mais [gz] dans *exact*) ; (b) certaines lettres (dites 'muettes') ne correspondent à aucun phonème (par ex. -t final dans *doucement*, de -nt du pluriel dans *aiment*, etc.) ; (c) des ressemblances dans un système correspondent à des différences dans l'autre : ainsi, le phonème [s] correspond à des graphèmes aussi variés que 's' (*son*), 'ss' (*poisson*), 'c' (*cette*), 'x' (*soixante*), 'ç' (*ça*). En français, il existe de très nombreux homophones se différenciant uniquement par la graphie, tel que [vεʁ], qui correspond à l'objet *verre* ('un verre à vin') ou la matière dont est composé un objet ('un vase en verre'), l'animal *ver* ('un ver de terre'), l'adjectif *vert* ('un T-shirt vert'), la préposition *vers* ('vers la gare'), etc.

Dans ce qui suit nous nous limitons à attirer l'attention sur la variation dans le nombre de marques grammaticales et leur distribution qui caractérise les deux modalités expressives. Le français est, en effet, une langue qui présente une morphologie 'riche' à l'écrit mais relativement plus 'pauvre' à l'oral.

La phrase reportée en (7) illustre cette différence pour ce qui est des marques du pluriel (cf. Riegel *et al.* 2009: 33). Dans la version écrite, le pluriel est marqué sur 3 items (le déterminant *les*, le nom *enfant* et le verbe *jouer*), alors que la version orale est beaucoup plus économique : il n'y a qu'une seule marque perceptible, notamment au niveau de l'article (qui, dans ce cas, donne lieu à la liaison [lez]).

(7) les enfants jouent dans la cour [lezãfãzudãlakuʁ]

L'écart est plus étendu en ce qui concerne l'ensemble de la morphologie verbale. Prenons à titre d'exemple la conjugaison au présent de l'indicatif de *parler*, verbe régulier du premier groupe (Tableau 1) : l'écrit présente 5 marques distinctes, alors que l'oral se contente de 3, le radical /paʁl/ pouvant correspondre à la 1^{ère}, 2^e et 3^e personne du singulier ainsi qu'à la 3^e personne du pluriel.

Tableau 1- *Présent du verbe parler*

		ECRIT	ORAL
1 ^{ère} p. s.	je	parl-e	[paʁl]
2 ^e p. s.	tu	parl-es	[paʁl]
3 ^e p. s.	il / elle	parl-e	[paʁl]
1 ^{ère} p. pl.	nous	parl-ons	[paʁlɔ̃]
2 ^e p. pl.	vous	parl-ez	[paʁle]
3 ^e p. pl.	ils/elles	parl-en	[paʁl]

Par ailleurs, étant donné qu'à l'oral 'nous' (1^{ère} p. pl.) est souvent remplacé par 'on' (*on part*), le nombre de désinences à l'oral peut se réduire à 2, [paʁl]-[paʁle] (cf. Riegel *et al.* 2009: 34).

Plus en général, le système verbal est caractérisé par une opposition entre formes simples – présent, imparfait (*je parlais* [zəpaʁle]), futur (*je parlerai* [zəpaʁl(ə)re]), subjonctif (*que je parle* [zəpaʁl]), conditionnel (*je parlerais* [zəpaʁləʁe]) – et formes composées grâce à l'auxiliaire *être* ou *avoir*, telles que le passé composé (*j'ai parlé* [zəpaʁle] / *je suis venu* [zəsɥivəny]) et le plus-que-parfait (*j'avais parlé* [zavəpaʁle] / *j'étais venu* [zɛtɛvəny]), etc. (cf. Riegel *et al.* 2009 pour un tableau plus exhaustif).

La prise en compte de l'ensemble des temps verbaux (simples et composés) conduit Noyau *et al.* (1995) à souligner l'opacité globale qui caractérise le système verbal temporo-aspectuel en français du fait que, dans la plupart des cas, il n'y a pas de relation directe entre un morphème et une valeur de mode / temps / personne / nombre.

On peut constater cette opacité dans l'aire des suffixes verbaux. En reprenant l'exemple du verbe *parler*, la désinence [e] peut correspondre à l'infinitif, au participe passé, ainsi qu'à la 2^e personne au pluriel du présent.

(8) [paʁle] → Infinitif: 'parl-er'
 → Participe passé: 'parl-é'
 → Présent 2^e p.s. / impératif: 'parl-ez'

Le suffixe [ɛ] marque à son tour plusieurs personnes de l'imparfait (1^e, 2^e, 3^e p.s. ainsi que 3^e p.pl.).

(9) [paʁlə] → Imparfait: *je parlais, tu parlais, il/elle parlait, ils/elles parlaient*

Cependant, la distinction [e] vs. [ɛ] est neutralisée chez la plupart des locuteurs natifs (Blanche Benveniste 2000: 31) si bien que plusieurs linguistes préfèrent utiliser l'archiphonème /E/ pour les deux cas, ce qui augmente le nombre de formes associées à ces deux phonèmes.

Quant aux verbes du deuxième groupe, l'homophonie est plus réduite, mais elle implique toutefois la coïncidence entre les trois personnes au singulier du présent ainsi que le participe passé.

- (10) [fini] → présent : *je finis, tu finis, il finit*
 → participe passé : *fini*

La zone préverbiale est également opaque, puisqu'il s'agit de la position d'un ensemble de marqueurs non accentués dont le pronom sujet (souvent réduit dans le parlé informel : [il] → [i], [ty] → [t]), le morphème de la négation 'ne', le pronom objet ou oblique et la forme fléchie de l'auxiliaire. Les auxiliaires *être* et *avoir* présentent à leur tour une faible opposition autour des sons [e] / [ɛ] pour certaines personnes du singulier ('j'ai', 'tu es' / 'il est').

Si on raisonne en termes d'input (essentiellement oral) auquel les apprenants sont exposés, il en résulte de nombreuses séquences dont les informations grammaticales sont assez peu perceptibles (cf. 11, exemples tirés de Dietrich *et al.* 1995 : 148) :

- (11) [zəpɑːle] vs. [zɛpɑːle] → 'je parlais' vs. 'j'ai parlé'
 [zəse] vs. [zɛseje] → 'j'essaie' vs. 'j'ai essayé'
 [ilɛse] vs. [ilɛse] → 'il essaie' vs. 'il l'essaie' ou 'il les sait'
 [ilapɔrt] vs. [illapɔrt] vs. [illapɔrte] → 'il apporte' ou 'il la porte' vs. 'il l'a porté'
 [ilavole] vs. [illavole] → 'il a volé' vs. 'il l'a volé'

Les problèmes de segmentation du parler natif sont également signalés dans le corpus oral du Groupe Aixois de Recherche en Syntaxe (GARS). Blanche Benveniste (2000: 31) souligne ainsi les possibilités multiples de transcription de segments discursifs (indiquées entre barres obliques) lorsque deux transcriptions sont également plausibles en contexte, comme c'est le cas en (12) (cf. faible opposition entre [e] vs. [ɛ], ainsi que présence du *ne* de négation qui se confond avec liaison en *n-*).

- (12) un monsieur qui /se fait, s'est fait/ élire
 /j'étais, j'ai été / frappé de voir que ...
 ils disaient qu'en France on (n') osait pas trop

Compte tenu de l'écart entre français écrit et français parlé, l'orthographe française donnerait une image déviante du fonctionnement de (la morphologie à) l'oral chez les natifs⁴. Cette image serait d'autant plus faussée pour la production d'apprenants débutants qui sont en train d'apprendre la LC, tels que ceux enregistrés dans le projet ESF. Il est bien connu, par ailleurs, que les oppositions phonologiques basées sur le degré d'aperture entre [e] vs. [ɛ] ou sur l'arrondissement des lèvres entre [e, ɛ] vs. [ø, œ], sont développées très tardivement en français L2 et constituent un véritable défi aussi pour les apprenants plus avancés et/ou exposés à un apprentissage en milieu institutionnel (cf. Véronique *et al.* 2009).

Le choix méthodologique appliqué dans le cadre de ce projet pour le français L2 est ainsi motivé, d'une part, par l'ambiguïté inhérente à certains segments verbaux et, d'autre part, par la volonté de retarder leur interprétation afin de l'appuyer sur des critères plus solides.

⁴ Par ailleurs, la maîtrise des différentes graphies de /E/ constitue un défi aussi pour les locuteurs natifs, comme le montrent Brissaud *et al.* (2006) dans une étude sur les productions d'enfants entre 8 et 15 ans.

Il est utile de souligner à ce propos que, dans les transcriptions originales ESF, l'emploi du symbole E a été parfois adopté, même si pas de manière systématique, pour éviter les différenciations subtiles entre la voyelle [e] et [ɛ] en français L2. En effet, il est difficile de présupposer d'emblée qu'un apprenant débutant puisse discriminer ces deux sons en langue cible. Nous optons pour ce codage dans les exemples cités dans cet article.

3. Illustration des enjeux théoriques par deux études de cas

Les choix de transcription appliqués dans le projet ESF reflètent, comme dit plus haut, l'approche *learner variety* qui envisage les lectures des apprenants comme des systèmes à part entière, dont on essaie de décrire les principes de fonctionnement indépendamment de la LC, à savoir sans attribuer aux formes produites en L2 les catégories des formes correspondantes en LC.


La transcription en phonétique des formes verbales en français L2 s'impose, surtout avec la production d'apprenants débutants, afin de ne pas sur-interpréter d'emblée leurs connaissances de la LC. Ceci dit, ce choix présente également des enjeux théoriques sur l'interprétation du parcours acquisitionnel, que nous allons illustrer à travers deux exemples sur l'acquisition du français L2 : les deux portent sur l'émergence de la morphologie verbale qui sera mise en relation, d'abord, avec l'expression des relations temporelles et, ensuite, avec l'acquisition de la négation.

3.1 Morphologie verbale et relations temporo-aspectuelles

L'acquisition de la temporalité a fait l'objet de nombreuses études. Dans ce qui suit nous allons focaliser l'attention sur l'émergence de la morphologie verbale dans l'expression des relations temporo-aspectuelles, soit sur le développement qui caractérise le passage entre la Variété de Base – permettant d'exprimer un ensemble riche de relations temporo-aspectuelles grâce à des moyens discursifs ou lexicaux, mais limité par l'absence de flexion verbale - et le stade suivant, la Variété Postbasique, où la temporalité commence à être encodée par des moyens grammaticaux (émergence de la finitude) (cf. tableau 2).

Cette transition a donné lieu à un débat concernant la valeur des premières oppositions verbales : s'agit-il d'oppositions encodant une valeur temporelle (*marquage de l'opposition entre présent vs. passé / futur*) ou bien une valeur aspectuelle (*marquage de l'aspect perfectif vs. imperfectif*) ?

Tableau 2 - Expression des relations temporelles

<p>VARIETE DE BASE</p> <p>Tps/Asp.: Moyens lexicaux et discursifs</p> <p><i>(adverbes temporels, principes d'organisation discursive, marqueurs bornes.... mais absence de flexion verbale fonctionnelle)</i></p>		<p>VARIETE POSTBASIQUE</p> <p>Tps/Asp. : Moyens grammaticaux</p> <p><i>Flexion verbale :</i></p> <p><i>- Temps (prés. / passé / futur) ?</i></p> <p><i>- Aspect (perf. / imperfectif)?</i></p>
---	---	--

Sur ce point on constate des résultats divergents. Dans la synthèse des résultats pour les différentes combinaisons de LS/LC du projet ESF, Dietrich *et al.* (1995) arrivent à la conclusion générale que le marquage grammatical du temps (*présent/passé*) précède le marquage grammatical de l'aspect :

« **Tense marking precedes aspect marking.** All target languages of this study have grammatical tense marking, only some of them have grammatical aspect marking, but all can mark aspect by various types of periphrastic constructions. In all cases, tense comes first » (Dietrich *et al.* 1995 : 270).

Cette généralisation ne va pas contre le fait que pour certaines langues cible puissent apparaître des formes encodant l'aspect grammatical (par exemple, l'apparition précoce de *V-ing* en anglais), puisque leur l'emploi

n'est pas fonctionnel, ou que certains apprenants puissent développer une opposition indéterminée entre le système aspectuel et temporel, comme c'est le cas des apprenants arabophones du français.

Le projet de Pavia, qui dispose d'un large corpus sur l'apprentissage de l'italien L2 par un public similaire à celui du projet ESF (migrants adultes en immersion)⁵, arrive en revanche à la conclusion opposée, à savoir que le marquage grammatical de l'aspect (opposition perfectif/imperfectif) précède l'expression grammaticale du temps. En prenant en compte les trois catégories sémantiques grammaticalisées dans les verbes de la LC, le développement en italien L2 est synthétisé de la manière suivante : aspect > temps > mode (Banfi & Bernini (2003: 95).

Sans remettre en question les conclusions des études mentionnées ci-dessous, dans les sections suivantes nous allons en reprendre les résultats concernant l'italien et le français L2 pour comprendre comment la manière de transcrire les données peut jouer un rôle dans les analyses et leur interprétation.

3.1.1 Morphologie verbale et relations temporelles : italien L2

Les études sur l'acquisition de l'italien L2 (cf. Giacalone Ramat 1992, Banfi & Bernini 2003, Bernini 2010) mettent en évidence la séquence acquisitionnelle suivante concernant le développement de la morphologie verbale : presente / infinito > (ausiliare +) participio passato > imperfetto > futuro > condizionale ... (*présent /infinitif* > (*auxiliaire* +) *participe passé* > *imparfait* > *futur* > *conditionnel*...).

Avant de revenir sur les débuts de ce parcours acquisitionnel, il est utile de souligner que, tout en étant deux langues romanes, le français et l'italien présentent des différences notables : contrairement au français, l'italien est une langue à sujet nul (*pro drop*) et, surtout, elle ne présente pas de morphologie silencieuse. A titre d'exemple, dans le cas du présent d'un verbe régulier, à l'oral comme à l'écrit, une désinence spécifique encode chaque personne ainsi que le marquage de l'opposition singulier/pluriel.

Tableau 3 - *Présent du verbe parlare* (fr. parler)

Italien 'parlare'	ECRIT	ORAL
1 ^{ère} p. s.	parl-o	['par.lo]
2 ^e p. s.	parl-i	['par.li]
3 ^e p. s.	parl-a	['par.la]
1 ^{ère} p. pl.	parl-iamo	[par. 'lja.mo]
2 ^e p. pl.	parl-ate	[par. 'la.te]
3 ^e p. pl.	parl-ano	['par.la.no]

Les premières formes verbales apparaissant en italien L2 ressemblent à des verbes fléchis : elles correspondent le plus souvent à la 3^e personne (*parla*) ou la 2^e personne du singulier (*parli*) du présent, plus rarement à l'infinitif (*parlare*). La désinence de la forme verbale en question est aisément identifiable du point de vue de la langue cible, mais elle est initialement dépourvue de valeur fonctionnelle en italien L2 : l'apprenant l'utilise comme une forme figée qui peut apparaître dans tout contexte temporo-aspectuel. En effet, comme on peut le voir dans les exemples suivants, il n'y a pas de correspondance entre la désinence verbale utilisée par l'apprenant et ses fonctions (personne, temps) en LC.

(13) io *parla* + mh – tigrigna e am/ amarigna + dell'Itiopia
je parle-3^ep.s. + mh – *tigrinya et am/ amarinya + de l'Étiopie* (Banfi & Bernini 2003: 92)

(14) appena in Italia + non + non *capisco* idalia tutto

⁵ Le corpus du projet de Pavia comporte des données d'apprenants de l'italien L2 ayant différentes L1, dont entre autres le chinois mandarin, le tigrinya et l'allemand (cf. Andorno & Bernini 2003).

à peine en Italie + NEG + NEG comprends-1^e p.s. italien tout

++ poi + *vai* + *va* in libreria *complale* un libro

++ *après* + *vas-2^e p.s.* + *va-3^e p.s.* dans librairie acheter un livre

chiama lingua per ai/ + la lingua italiana per i stranieri

(s') appelle langue pour aux/ la langue italienne pour les étrangers

in casa + stasera+ in casa *studiare* ++ questo libro due anni

en maison + ce soir + en maison *étudier-INF* + ce livre deux ans

(quand je suis arrivé en Italie je ne comprenais pas l'italien, ensuite je suis allé

dans une librairie acheter un livre qui s'appelait la langue italienne pour étrangers.

A la maison le soir, j'ai étudié ce livre pendant deux ans)

(Giacalone Ramat 1992 : 155)

Il s'agit donc de 'formes de base' n'ayant qu'une valeur lexicale. Pour mettre en évidence cet écart par rapport à la LC, dans les travaux sur l'italien L2 elles sont parfois représentées avec le symbole \emptyset au lieu de la flexion verbale : *parla- \emptyset* ; *studiare- \emptyset* (cf. par exemple Banfi & Bernini 2003).

L'étape successive correspond à l'émergence de formes équivalentes au participe passé (forme de base + suffixe *-to*), qui entrent en opposition avec les formes du présent, établissant ainsi une opposition entre perfectif (et/ou passé) vs. imperfectif. L'extrait (15) illustre bien ces valeurs. L'apprenant décrit des vignettes (dessins) et commente le fait qu'une femme lave des casseroles (*lava pentola*) ; une fois la casserole lavée (*lavato pentola*), elle se regarde dedans comme dans un miroir. Le participe passé encode la fin de l'opération de laver et l'état qui en résulte de propreté (valeur résultative de l'action de laver)⁶.

(15) *lava* eh pentola +++
lave (3^e p.s.) casserole

la/ eh + lavato eh pentola +++

la/ eh + lavé-part.passé casserole +++

guarda eh +++ come eh +++ specchio

regarde eh +++ comme eh +++ miroir

(Banfi & Bernini 2003 : 94)

En résumant l'évolution des plusieurs apprenants du corpus en question, on constate que tous (à part un) arrivent à disposer de formes du présent (et infinitif) en opposition avec des formes du participe passé (« a parte Hagos (...) tutti gli apprendenti dispongono di forme di presente (e infinito) in opposizione a forme di participio passato », Banfi & Bernini 2003: 90).

Le développement au-delà de ce stade est marqué par l'émergence des auxiliaires *essere* ('être') et *avere* ('avoir') qui contribuent à la composition du *passato prossimo* (Aux+ participe passé), construction équivalente au passé composé en français. Dans cette phase, il n'est pas rare de constater la production de formes verbales auxiliées qui ne sont pas encore conformes à la langue cible (cf. en 16a choix de l'auxiliaire 'avoir' au lieu de 'être' et accord au singulier au lieu du pluriel), ou encore l'émergence de formes analytiques encodant séparément les significations grammaticales (accord, temps de la validité de l'assertion) et lexicales (16b) au lieu d'être fusionnées dans une seule forme (dans ce cas l'imparfait du verbe 'aller'= *andavano*).

(16a) tanti persone *ha-morto* qui (= tante persone sono morte)
beaucoup de gens *a-3^e p.s. mort* (= sont mortes) ici

⁶ Il est à noter que la valeur du participe passé n'est pas toujours aussi claire : dans d'autres occurrences, son emploi présente une certaine ambiguïté entre la lecture perfective et le passé, étant donné leur solidarité sémantique (cf. Giacalone Ramat 1992 ; Banfi & Bernini 2003).

(beaucoup de personnes sont mortes ici)

(Giacalone Ramat 1992 : 158)

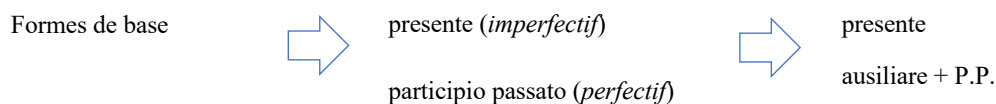
(16b) erano andare (= andavano)
Etaient-3^e p.pl. IMPF. aller-INF (= ils allaient)

(Banfi & Bernini 2003 : 107)

Toujours est-il que l'apparition de l'auxiliaire marque un moment charnière dans le développement des lectures d'apprenants en direction de l'organisation temporelle de l'italien : c'est l'auxiliaire (et/ou la copule) qui encodent l'intervalle temporel de validité de l'assertion.

Le tableau suivant visualise l'évolution de ces premières étapes : l'emploi d'une forme verbale initialement indifférenciée se réduit avec l'encodage d'une première opposition aspectuelle entre imparfaitif (présent)/perfectif (participe passé). Le marquage de l'opposition temporelle (présent/passé) est liée à l'émergence de l'auxiliaire qui participe à la construction du *passato prossimo*.

Tableau 4 - *Emergence de la morphologie verbale pour les relations temporelles en italien L2*



3.1.2 Morphologie verbale et relations temporelles : français L2

Les premières formes verbales apparaissant en français L2 peuvent ressembler également à des verbes fléchis, mais il est plus compliqué de les faire correspondre à une forme spécifique de la LC, comme le demanderait une transcription orthographique. Il s'agit notamment de formes qui, du point de vue de la LC, semblent soit comporter la présence d'un suffixe (en particulier [e] ou [i]), soit correspondre au radical du verbe, ce qui équivaldrait à différentes personnes du présent pour les verbes réguliers. Ce sont en réalité des formes figées, invariables dans le système de l'apprenant, puisqu'elles apparaissent dans tout contexte temporel ou personnel. A titre d'exemple, l'extrait (17) montre un contexte, induit par la question de l'intervieweur sur la matinée de l'apprenante, qui demanderait en LC le marquage du passé dans la réponse. Etant donné l'opacité du système verbal français à l'oral (et surtout les nombreux cas d'homophonie), une transcription orthographique des segments verbaux poserait de sérieux problèmes d'interprétation, puisque le même segment pourrait être transcrit de manières différentes. Ainsi, en (17a) pour « [preparer] » il s'agirait de choisir entre l'infinitif (*préparer*) ou le participe passé (*préparé*) - sachant qu'on ne peut exclure non plus des formes d'imparfait (*préparais / préparait / ...*) dont la prononciation est phonétiquement similaire -, en (17b) pour « je [le fe] » entre présent (*je fais* 1^{ère} p.s. présent) et passé composé (*j'ai fait*) ou participe passé (*fait*), et en (17c) pour « [sorti] » entre le participe passé (*sorti*) du verbe *sortir* et une forme potentiellement possible de présent idiosyncrasique (*je *sortis*), construit sur le modèle du verbe *finir* (*je finis*).

- (17) LN: *qu'est ce que tu as fait alors depuis ce matin ?*
- | | | |
|-----|--|---|
| (a) | je [prepare] le *por* [manze] | → Vinf <i>préparer</i> ? P.P. <i>préparé</i> ? |
| (b) | après je [le fe] un petit peu de ménage... | → Présent <i>fais</i> ? p.p. <i>fait</i> ? |
| (c) | *y* je [sorti] à [un or trent] | → P.P. <i>sorti</i> ? Présent * <i>sortis</i> ? |

A l'inverse, l'exemple (18) montre une forme verbale que l'on serait tenté de transcrire orthographiquement comme *je travaille*, puisque sa forme phonétique pourrait correspondre à la 1^{ère} p.s. présent du verbe *travailler*

en LC⁷. Cependant, il est évident que l'apprenante l'utilise pour faire référence au passé, comme l'indiquent clairement les adverbiaux utilisés (*avant, jamais au Chili = jamais quand j'étais au Chili*)

- (18) LN: *qu'est-ce que tu fais ici ? tu travailles ?*
 BE: avant je [travaj] maintenant non → présent *travaille?*
 (avant je travaillais/j'ai travaillé, maintenant non)
 LN: *tu sais bien faire la cuisine?*
 BE: oui [solamã] ici
 (oui, seulement ici = en France elle travaille dans la cuisine d'un foyer)
 BE: je jamais jamais je [travaj] au chili
 (je n'ai jamais travaillé au Chili)

Afin d'éviter toute identification abusive par rapport aux multiples désinences possibles de la LC, les formes verbales en français L2 ont été toutes transcrites en symboles phonétiques. Dans le même but, elles sont également représentées en symboles phonétiques dans les analyses successives ayant conclu qu'à ce stade, celui de la variété de base, les lexèmes verbaux en français L2 apparaissent typiquement sous les trois formes de base suivantes (cf. Noyau *et al.* 1995)⁸:

V-e	comme [vole]
V-i	comme [sorti]
V-∅	comme [travaj]

Dans les trois cas, il s'agit de formes dépourvues d'informations grammaticales, à savoir qui ne présentent pas d'opposition fonctionnelle entre elles (comme c'est le cas en anglais pour la variation initiale entre V et Ving), ni de correspondance spécifique au temps ou à la personne demandés par le contexte discursif.

Pour les apprenants dépassant ce stade, le développement est marqué, d'une part, par la spécialisation de V-∅ dans le marquage du présent et, d'autre part, par l'émergence des auxiliaires *être/avoir* + V-e / V-i, qui permet de discriminer la référence aux contextes présents vs. passés.

Les formes auxiliées peuvent, cependant, coexister longuement avec les formes de base en V-e. Ainsi, dans l'extrait (19), qui implique un contexte passé, il est à noter que l'apprenante BE utilise dans le même passage aussi bien de formes correspondantes au passé composé (je [swi arive] = je suis arrivée ; je [swi perdu] = je me suis perdue), appropriés au contexte temporel (récit de son arrivée en France), que des formes en V-e ([komense], [reste]). Pour ces dernières, il est difficile de décider s'il s'agit de formes de base persistantes du stade précédent, d'infinitifs, ou bien de premières formes d'imparfait.

- (19) LN *vous saviez où aller quand vous êtes arrivée ici?*
 je [swi arive] dans un foyer de transit
 * y * [ʒe me di] « [se] pas possible [reste] toute la journée ici »
 [ilja] de sortir de ici...
 après [ʒe komense] à sortir toute seule
 y jamais jamais je [swi perdu] dans le métro (Noyau *et al.* 1995 :195-6)

⁷ NB : il n'est pas à exclure à ce stade que le segment en question puisse être le substantif 'travail', homophone des formes fléchies (1^{ère} p.s., 2^e p.s., 3^e p.s. et pl.) du verbe *travailler* au présent.

⁸ L'emploi initial de verbes en Ve/Vi/V∅ semble bien être une caractéristique du français L2, puisqu'elle a été attestée chez différents types d'apprenants, aussi bien en milieu naturel qu'institutionnel (cf. Bartning & Schlyter 2004 ; Véronique *et al.* 2009).

C'est l'analyse longitudinale de ses productions qui nous permet de voir l'évolution plus complète de cette apprenante. Le tableau 5, tiré de Benazzo & Starren (2007), reporte toutes les formes verbales produites lors d'une tâche de narration (le récit d'un extrait du film *Les Temps modernes*) qui a été répétée trois fois dans le recueil des données ESF.

Tableau 5 - *Formes verbales attestées chez BE dans le récit des Temps Modernes*

APPRENANT	BERTA (esp.> français)		
CYCLE	I	II	III
Formes de base V- <i>e</i> – V- <i>i</i>	3 [tSerSe], [sorti], [vole]	7 [aSEte], [marSe], [parti], [prepare], [sorti], [entruve], [tombe]	6 [sorti], [prãde], [vãdr], [truve], [uvre], [vole]
V-∅	3 a, va, [di (ke)]	7 [demãd], [atrap], [pas], [vjen], [rest], [vi]/[vy], [fE], [return]	21 [apel], [searete], [ariv], [demãd], [di], [dorm], [komãs], [krwa], [mont], [pas], [prãd], [profit], [rest], [revej], [sør], [suy], [se(n)truve], [turn], [(se)tõb], [vwa], [(sã)va]/[ve]
<i>Avoir</i> + P.P.	-	-	5 [avole], [avy], [adone], [asorti], [atõbe]
<i>Être</i> + P.P.	-	-	5 [ilesorti], [eparti] / [sonparti], [sontõbE], [sonperdy], [sefrape]
Proto-modaux + Vinf	-	-	[vudra] / [vudre]
<i>Aller</i> + Vinf	-	-	Va + Vinf
<i>Etre en train de</i> + Vinf	-	-	[il entre] de / [sonentre] de V
Imparfait	-	-	-

Le récit du premier cycle montre l'emploi de 6 verbes différents, qui représentent l'actualisation typique des formes de base V-*e*/V-*i*/V-∅. La production du cycle II permet d'apprécier l'accroissement du répertoire verbal produit pour la même tâche et, en parallèle, l'augmentation de formes en V-∅. Toutefois, ce n'est qu'au cycle III qu'on peut constater un changement qualitatif dans la morphologie verbale avec des signes fiables de progression, notamment à travers la production de formes auxiliées (*être* ou *avoir* + V-*e* ou V-*i*), qui s'opposent aux formes du présent (par ex. [sør] vs. [asorti]), ainsi que la présence de proto-modaux (*vouloir*)⁹ et des périphrases aspectuelles (futur proche *aller* + Vinf; progressif *être en train de*).

Malgré cette progression évidente, la proportion des formes de base en V-*e* se réduit mais ne disparaît pas jusqu'à la fin du recueil des données. Par ailleurs, une analyse plus fine de leurs contextes d'apparition dans différentes tâches conduit à la conclusion que la valeur temporo-aspectuelle de l'imparfait n'est pas maîtrisée par cette apprenante (en d'autres termes il n'y a pas d'opposition au passé entre PC et IMP) :

« if there is an emergent pattern V-*e* as opposed to V-∅ in the very last data, it does not support an aspectual contrast perfective/imperfective, as in the French Passé Composé/imparfait (...), but rather the marking of hypothetical (non-referential) situation vs. actual » (Noyau *et al.* 1995: 196)

⁹ [vudra] / [vudrE] sont en réalité utilisés dans un contexte demandant le présent *veut*.

Etant donné l'opacité du français aussi bien dans la zone des suffixes que des préfixes (cf. Section 2.2), on ne considère comme signes fiables de la présence d'un auxiliaire que les occurrences où une forme auxiliée clairement perceptible s'oppose à une forme simple du même verbe :

- (20) [vole] → [avole]
[tɔ̃be] → [sɔ̃tɔ̃be]

En effet, on peut toujours hésiter entre forme de base ou forme auxiliée dans l'interprétation de cas qui impliquent comme préfixe les sons [e] / [ɛ] / [ə] - tels que [sefrapɛ]: *s'est frappé* ou *se frapper* ? [ʒeparle] : *j'ai parlé* ou *je parlais* ? - oppositions que les hispanophones en particulier (mais aussi les apprenants avec d'autres L1) maîtrisent tardivement en français.

Les premiers stades de développement d'une morphologie verbale fonctionnelle en français L2, du moins en ce qui concerne les relations temporo-aspectuelles, peuvent donc être schématisés de la manière suivante (Tableau 6) :

Tableau 6 - *Emergence de la morphologie verbale pour les relations temporelles en français L2*

Formes de base			Présent	V-∅ [ilvol]
V-e	[vole]	➔	Passé composé	Aux-Vlex [ilavole]
V-i	[sorti]		Futur proche	va + Vlex [ilvavole]
V-∅	[travaj]			

La transition entre la variété de base et le stade suivant est ainsi caractérisée par l'apparition d'un ensemble de marqueurs morphologiques préverbaux. La séquence acquisitionnelle comporte ensuite le développement de formes telles que l'imparfait (et plus tard le plus-que-parfait) pour le passé ainsi que du futur simple pour le futur (stades avancés, cf. Véronique *et al.* 2009 ; Bartning & Schlyter 2004).

La comparaison des itinéraires développementaux en italien vs. français L2 met en évidence des tendances communes, telles que l'absence d'informations grammaticales dans les premières formes verbales, mais aussi quelques différences liées à la LC, dont la difficulté à catégoriser la désinence des premières formes verbales en français et l'absence d'une étape caractérisée par l'opposition entre le présent (inaccompli) vs. participe passé (accompli), qui précéderait l'apparition de l'auxiliaire. A ce propos il est à souligner qu'en fonction de la manière de transcrire, on pourrait arriver à la conclusion inverse, notamment si les formes en V-e / V-i étaient transcrites comme des participes passés. L'emploi de la transcription phonétique montre en revanche que, même si on voulait supposer un développement en français L2 similaire à celui constaté en italien L2, cette étape ne serait pas détectable à l'oral en français, puisque la plupart des formes de base coïncident phonétiquement avec celles du participe passé ([parle] – [sorti]).

En revanche, ce qu'on remarque pour exprimer la valeur perfective ou passée c'est l'association de la forme verbale non-finie avec le marqueur de bornes [fini] et, en particulier chez les apprenants hispanophones, avec l'adverbe 'déjà' (cf. Benazzo & Starren 2007). Le recours à cet adverbe est illustré dans l'exemple suivant, tiré d'une conversation libre pendant laquelle BE raconte la fête d'anniversaire d'un de ses enfants.

- (21) (contexte : récit de la fête d'anniversaire d'un enfant de BE)
BE: [ʒarive] à la maison à 8 h *y* la fête déjà [fini]
(je suis arrivée chez moi à 8h et la fête était déjà finie)

De même, après un long passage dans lequel un autre apprenant hispanophone (AL) explique ses problèmes dans les démarches administratives à la préfecture et que le LN n'arrive pas à comprendre si la situation décrite

s'est déroulée dans le passé ou doit avoir lieu dans le futur (*tu y est allé ou tu iras ?*), c'est la présence de *déjà* (avec une forme verbale non-finie) qui permet de situer la situation dans le passé.

- (22) LN: *tu retourneras à la préfecture pour le logement ?...tu y es allé + ou tu iras?*
AL: .. je déjà je [ale] là bas oui

La fonction de *déjà* se prête à plusieurs interprétations qui rappellent la double lecture du participe passé en italien L2 : d'une part, on pourrait le considérer comme un moyen parmi d'autres qui contribue à localiser dans le passé la situation décrite par le verbe non-fini (voir un passé dans le passé en 21) ; d'autre part, il n'est pas à exclure une valeur purement aspectuelle d'accompli. Il n'en reste pas moins que le passage suivant confirme de manière assez convaincante la valeur discriminante de *déjà* avec des formes verbales de base pour véhiculer la valeur perfective du passé composé.

- (23) AL: je [te done] mon numéro téléphone...
LN: *y si dices por ejemplo has anotado mi teléfono en un papel ?*
AL: tu déjà [ekrive] mon numéro de téléphone

Face au segment verbal ambiguë que AL produit en L2 - [te done] pouvant correspondre aussi bien à la forme fléchie avec PC (*je t'ai donné*) qu'à l'énième emploi de la forme de base en V-e ([done]) -, l'enquêteur demande à l'apprenant de dire en français 'has anotado', soit l'équivalent du passé composé ('tu as pris note du numéro de téléphone'). Dans sa traduction en français, AL semble établir une équivalence entre le passé composé espagnol 'has anotado' et *déjà* + V non fini [ekrive].

Il est à noter que si l'on suivait une transcription orthographique standard, on ne saurait décider comment transcrire la forme [ekrive], qui pourrait être interprétée comme une tentative d'imparfait ('*tu écrivais*'), cependant inappropriée au contexte, ou encore comme un participe passé idiosyncrasique (**écrivé*, au lieu de *écrit*, sur le modèle de *parlé*). L'évolution générale du lecte de l'apprenant nous indique en revanche qu'elle fait partie des formes de base en V-e et que l'association avec l'adverbe *déjà* représente l'un des moyens pour la désambiguïser.

Ce survol rapide sur l'émergence de la morphologie verbale en L2 donne lieu à deux remarques : au niveau acquisitionnel, il est tout à fait plausible que les caractéristiques de la LC influent sur les étapes développementales en L2 (la comparaison des trajectoires en l'italien et en français L2 montrent des tendances similaires, modulées par l'exploitation de moyens différents), mais au niveau méthodologique il est évident que, en fonction des choix de transcription des formes verbales, on pourrait arriver à des conclusions assez différentes concernant le degré de morphologie présente dans les productions en français L2 et ses fonctions.

3.2 Acquisition de la négation en français L2

L'acquisition de la négation a fait l'objet de plusieurs travaux se basant sur le corpus du projet ESF (p.ex. Becker 2005, 2012 ; Giuliano & Véronique 2005 ; Stoffel & Véronique 2003 ; Silberstein 2001). Nous l'illustrons ici par l'étude de Giuliano (2005) en reprenant son analyse de l'acquisition de la négation en français L2 par des apprenants hispanophones du projet ESF. Giuliano (2005) éclaire certains phénomènes liés à l'émergence de la morphologie verbale en français L2 en relation avec le développement de la négation discontinue en français.

La négation en français présente des difficultés d'analyse à la fois du point de vue de l'apprenant que de celui du chercheur qui doit découvrir le système émergent en L2. Ces difficultés sont en grande partie liées à l'ambiguïté phonologique dont il a été question dans la section précédente.

Ainsi, la négation en français est discontinue et construite grâce à deux éléments qui englobent la forme finie du groupe verbal (*ne + V_{fin} + pas*), alors qu'en espagnol, langue maternelle des apprenants analysés par Giuliano, la négation est typiquement pré-verbale (*no + V_{fin}*). Pour un apprenant débutant, l'acquisition de la négation discontinue représente un défi important dans la mesure où la maîtrise de celle-ci va de pair avec le processus de grammaticalisation du lecte de l'apprenant et l'émergence de la morphologie flexionnelle productive. Autrement dit, tant que l'apprenant ne parvient pas à décomposer l'énoncé en constituants majeurs (Sujet – Verbe – Complément) et à identifier, grâce à la morphologie, la copule, les auxiliaires et les modaux, les règles de la négation en LC ne pourront pas être opérationnalisées. Si on prend en considération le système phonologique du français particulièrement ambigu (cf. section 2.2. ci-dessus), l'identification des formes verbales à valeur d'auxiliaire et de copule, combinées avec les négateurs, devient très difficile. En effet, certaines distinctions phonologiques sont très subtiles ou nulles (cf. exemple 24 ci-dessous) :

- (24) [ilnapa] 'il n'a pas'
 [ilnjapa] 'il n'y a pas'
 [ilnɛpa] 'il n'est pas'
 [ilnɛpa] 'il n'ait pas'

L'apprenant doit faire face à une difficulté de segmentation liée à la faible perception de la copule et à la proximité phonologique avec d'autres formes.

Par rapport aux étapes acquisitionnelles initiales présentées dans la section 2.1., l'acquisition de la négation qui se dégage de l'analyse de Giuliano (2005) montre une intégration du négateur français dans la structure du lecte d'apprenant dans les 3 étapes reportées ci-dessous.

Tableau 7 - *Evolution de la négation chez Berta (apprenante hispanophone, corpus ESF)*

(1) Variété prébasique	(2) Variété de base	(3) Variété postbasique
Neg + X	Neg + Vnonfin	Aux/Mod + Neg + Vnonfini
X + Neg		

Nous illustrons les difficultés que pose la transcription des séquences contenant la négation en français L2, en nous basant sur des exemples provenant des productions de Berta, une des apprenants hispanophones du projet ESF.

Variété prébasique

Les énoncés que l'apprenante construit au niveau du premier palier acquisitionnel, la variété prébasique, ne possèdent pas encore d'élément verbal. Deux négateurs idiosyncrasiques sont produits par l'apprenant, [no/non] et [nɛpadɛ], qui portent sur un des constituants nominaux de l'énoncé. Cependant, on n'atteste que très peu d'occurrences négatives (7 au total) dans les productions de Berta.

En suivant la logique de l'approche des lectes d'apprenants, le négateur [nɛpadɛ], est transcrit phonétiquement et est considéré comme une forme non analysée. Si on optait pour une transcription filtrée par l'orthographe de la LC, une surinterprétation conduirait à présupposer l'existence d'une structure auxiliée complexe (*n'est pas de*). Une telle interprétation va à l'encontre de l'analyse globale du lecte de cette apprenante à ce stade de l'acquisition où elle ne produit pas encore d'énoncés structurés autour d'un lexème verbal. L'emploi de ce négateur ainsi que de [no/non] s'explique par la structure informationnelle de l'énoncé où il apparaît. Ainsi, lorsque Berta produit des énoncés négatifs avec [no/non], l'élément nié est celui qui constitue le topique de l'énoncé, comme dans l'exemple ci-dessous :

- (25) LN: *il y a des taxis ?*
BE: non + taxis non

En revanche, lorsque l'énoncé négatif contient le négateur [nEpadE], c'est l'expression de focus qui est dans sa portée (cf. exemple 26, ci-dessous).

- (26) LN: *c'est un peu une imprimerie alors ?*
BE: non non [nEpadE] imprimerie + *otra forma*

L'analyse des négateurs au niveau de la variété prébasique pose un certain nombre de problèmes liés aux structures souvent ambiguës à cette étape. De plus, comme nous l'avons précisé, les énoncés négatifs sont peu nombreux, ce qui ne permet pas de déduire facilement une systématisme dans leur emploi. Cependant, le domaine d'application du négateur est transparent puisqu'il se trouve en position adjacente par rapport à l'élément nié sans qu'il corresponde à un verbe

Voici donc la distribution des négateurs au niveau de la variété prébasique:

- (27) Neg [nEpadE] + X
X + Neg [no/non]

Variété de base

La complexification de l'énoncé au niveau de la variété de base, qui résulte de l'apparition d'un constituant clairement verbal, même s'il correspond à une forme non finie, va avoir des conséquences sur l'acquisition de la négation. Au niveau quantitatif, on observe, à ce stade, un accroissement des énoncés négatifs à la fois avec le négateur *no/non*, à valeur surtout anaphorique, et le négateur à structure [nEpa(dE)].

Nous poursuivons ici notre illustration de la relation entre la production orale des apprenants débutants et les choix de transcription en nous intéressant au négateur [nEpa(dE)].

Dans la période où le lecte de l'apprenante Berta atteint les caractéristiques de la variété de base, le nombre des énoncés négatifs augmente nettement (99 énoncés négatifs). Le passage de la variété prébasique à la variété de base est marqué par l'évolution de « Neg + X » vers « Neg+Vnon fini ». En effet, dans les énoncés construits autour d'un constituant à valeur verbale, les structures comme [nEpa/nEpadE] conduisent à des ambiguïtés phonétiques relevant du système du français, ce qui rend l'analyse des productions en L2 compliquée. Une transcription phonétique du groupe verbal avec ce négateur s'impose afin d'éviter toute interprétation *a priori*. Seule l'analyse de ces éléments dans différents contextes de discours de l'apprenant permettra de trouver les arguments pour préférer une interprétation à l'autre.

Ainsi, une première question qui se pose concerne la décodification de la forme [nE] dans la structure du négateur [nEpa/nEpadE]. Deux interprétations possibles sont proposées par Giuliano (2005). D'une part, cette forme correspondrait à la particule négative [nə] où on atteste un remplacement de [ə] par un son se rapprochant de [ɛ]. Cette interprétation serait fondée sur le fait qu'on atteste ce même phénomène dans la réalisation phonétique par l'apprenante des pronoms personnels, tels que *je* [ʒə] plutôt comme [ʒɛ] et *le* [lə] se rapprochant de [lɛ].

D'autre part, on peut envisager [nE] comme résultat de la contraction de la particule négative et de la forme du verbe *être* : « [nE] → ne + est (*être*) ».

Les deux interprétations sont importantes pour l'analyse de la nature des deux formes [nEpa/nEpadE], qui peuvent être considérées comme deux formes interchangeable [nEpa(dE)], dans la mesure où elles

apparaissent dans les mêmes contextes d'énonciation. On admet donc qu'il s'agit d'une même structure avec deux variantes libres.

Les emplois les plus fréquents de ce négateur correspondent aux fonctions principales de *être* et *avoir*, à savoir celles d'attribution des propriétés, d'existence et de possession. Dans ce cas-là, le négateur précède un constituant à valeur nominale ou adjectivale.

Deux interprétations sont proposées pour rendre compte de la structure interne du négateur. Ainsi, [nEpa(dE)] pourrait correspondre à une négation discontinuée où [E] entre [n] et [pa(dE)] aurait la valeur de la forme verbale « est » surgénéralisée à toutes les personnes et tous les temps d'*être*, *avoir*, *y avoir*. Il s'agirait donc d'une segmentation [n+E+pa] où [E] constitue un élément à valeur verbale (*est*). Ou bien, le sémantème négatif réalisé par la consonne [n] n'étant pas perçu de façon indépendante de [pas], le [E] serait amalgamée en constituant une négation continue [nEpa(dE)]. Celle-ci serait une forme figée et non analysée.

Afin de trancher entre l'une ou l'autre possibilité d'interprétation, Giuliano (2005) propose d'examiner les emplois de l'apprenante lorsqu'elle construit les énoncés négatifs avec le verbe lexical, en excluant de cette analyse des expressions formulaïques de type « je (ne) sais pas », « je (ne) comprends pas ». La fréquence d'emploi de ces structures dans l'input natif en français conduit à leur apparition précoce en L2 en tant que formes non analysées et, en tant que telles, elles ne sont pas représentatives du développement de la syntaxe négative.

En revanche, l'augmentation des énoncés avec le verbe lexical non fini, typique de la variété de base, permet de comprendre la nature et le fonctionnement du négateur [nEpa(dE)] dans le lecte basique de Berta. Deux structurations des énoncés négatifs avec le verbe lexical sont attestées : le négateur est ainsi placé soit en position pré-verbale initiale ([nEpa(dE)] + Vlex nonfin), comme en (28), soit entre le SN et le verbe (SN + [nEpa(dE)] + Vlex nonfin + SN), comme en (29).

- (28) LN: *est-ce qu'il y a un travail que vraiment vous n'aimeriez pas du tout faire ?*
BE: ah oui + [nEpadE komprende] por français à travail [de kusin]
(ah oui + je ne comprends pas à cause de mon français le travail de cuisine)

- (29) BE: mon mari eh [eskri] *y* [kompri] bien le français
(mon mari écrit bien et comprend le français)
BE: mais moi [nEpadEkribir]
(mais moi je n'écris pas)

L'interprétation du négateur dans ces exemples permet d'appuyer l'hypothèse selon laquelle [nEpa(dE)] est une forme non analysée. Étant donné que le contexte d'énonciation exige les verbes au présent et non pas au passé composé, il serait difficile de postuler la décomposition de la part de l'apprenante de type « ne+AUX+pas+Vfini ».

Parallèlement, on atteste également des énoncés négatifs avec un verbe implicite ([nEpa(dE)] + SN/Sprép) comme dans l'exemple ci-dessous.

- (30) BE: non + un mois après *y* moi [nEpa] l'école
(non + un mois après et moi je ne (suis) pas (allée) à l'école)

Le sens de l'énoncé produit ici par Berta ne peut être reconstruit que si on admet la présence du verbe lexical n'apparaissant pas en surface. Le négateur marque donc la négation d'un verbe lexical (*aller/faire*) qui n'est pas produit dans l'énoncé.

En résumant le fonctionnement de la négation au stade de la variété de base chez Berta, nous pouvons affirmer que [nEpa(dE)] est une structure figée. Son emploi très systématique dans la position pré-verbale montre qu'il remplit trois fonctions : celle de la négation d'existence, de possession, d'attribution des propriétés, celle de la négation des verbes lexicaux et celle du marquage de l'assertion négative dans les énoncés sans le verbe lexical. La présence du négateur dans des énoncés négatifs uniquement dans un contexte au présent rend impossible de supposer un éventuel découpage en « Neg+Aux+V ».

Variété postbasique

L'évolution du lecte de Berta de la variété de base vers la variété postbasique est lente et progressive : elle est marquée par une émergence des distinctions morphologiques de plus en plus productives, même si la flexion verbale reste toujours instable et souvent idiosyncrasique.

Pour ce qui est de la négation, on observe un passage de la structure des énoncés négatifs

« SN + [nEpa(dE)] + Vlex nonfin + SN », attestée au stade de la variété de base, vers une intégration progressive du négateur dans le groupe verbal analysé de type « SN + V + neg + SN ». Autrement dit, la négation continue à laisser la place à la mise en œuvre de la négation discontinue. Cependant, pendant un certain temps les anciennes fonctions coexistent avec les nouvelles distinctions qui apparaissent graduellement. A la même période, nous trouvons des productions de Berta, qui montrent qu'elle emploie les deux types de négation. Le premier exemple (31) illustre l'emploi de la négation continue précédant la copule « être ».

- (31) BE: [ʒe nEpa swi] content parce que [medi] mes enfants [kElja] des problèmes
(je ne suis pas contente parce que mes enfants me disent qu'il y a des problèmes)

Dans l'exemple (32) ci-dessous, en revanche, nous attestons une tentative de construire la négation discontinue, même si l'homophonie entre la structure non analysée [nEpa] et les formes *n'ai pas/n'est pas* rend difficile, voire impossible, de décider clairement de la stratégie adoptée par l'apprenante : « [nEpa]+V » ou « ne+AUX+pas + participe passé ».

- (32) BE: [ʒe nEpa veny] parce que [ʒEnEpa sy] l'adresse
(je ne suis pas venu parce que je ne savais pas l'adresse)

Et finalement, l'exemple suivant atteste d'un travail cognitif intéressant de la part de Berta qui se manifeste par les auto-reformulations portant sur les structures négatives.

- (33) BE: [ilija] de personnes [kE nEpa naze + nE save pas naze]
(il y a des personnes qui ne savent pas nager)

Dans cet exemple, on voit que la reformulation lui permet de produire la négation discontinue, dans la mesure où le verbe « savoir » se trouve bien entre « ne » et « pas ».

Le choix de la transcription phonétique des structures verbales négatives permet aboutir à une analyse de la cohérence interne du lecte de l'apprenant sans un filtrage préalable par des règles de la langue cible.

4. Conclusions

Ce chapitre propose une réflexion concernant la transcription des données orales d'apprenants adultes d'une L2 et, plus particulièrement, les problèmes spécifiques que pose la reproduction à l'écrit du français parlé en L2.

Nous avons vu que l'objectif de reproduire de manière relativement neutre la parole en L2, propre à l'approche des lectures d'apprenants (*learner variety approach*) implique des précautions à prendre afin de ne pas interpréter le lecture de l'apprenant selon les catégories de la LC. Ces précautions sont d'autant plus cruciales pour les langues qui sont marquées à la fois par une faible correspondances graphème-phonème et par une forte présence de formes grammaticales homophones et hétérographes, comme c'est le cas en français (cf. Jaffré & Brissaud 2006). Ainsi, la transcription phonétique peut s'imposer dans certains cas, notamment en fonction des caractéristiques de la langue cible (cf. différence français vs italien).

En exposant les études de cas sur le début de la morphologie verbale en français L2, nous avons essayé de montrer que la manière de transcrire peut influencer l'interprétation des formes employées par des apprenants en L2 à un moment donné de l'acquisition : c'est l'évolution du lecture au fil du temps qui permet de mieux comprendre quels éléments morphologiques deviennent fonctionnels. Le risque de sur-interprétation est sensiblement plus fort dans le cas de données transversales, à savoir lorsque le chercheur ne dispose que de peu de données d'un même apprenant.

Par ailleurs, notre réflexion se limite à pointer les problèmes qui se posent dans le cadre d'études issues d'une approche fonctionnaliste, mais les choix de transcription, ainsi que l'interprétation des mêmes données, varient également en fonction des présupposés théoriques qui guident l'analyse. Sur ce point nous renvoyons à Leclercq (à paraître/2020), dont l'analyse comparative de différentes études portant sur les formes verbales avec [e] final (V-[e]) en français L2 montre l'influence du type de théorie sur les choix de transcription de cette structure hautement ambiguë à l'oral : elle oppose ainsi les études issues des théories dites « data driven », qui choisissent des solutions permettant d'éviter toute présupposition *a priori* sur les productions des apprenants, à des « theories constrained approaches », où ce choix est davantage guidé par des *a priori* théoriques. Réfléchir sur la manière dont les données sont transcrites et analysées permet de contribuer au débat plus large sur la description et l'interprétation de l'interlangue des apprenants L2, ce que propose par exemple Ortega (2014), en prenant comme exemple l'acquisition de la négation.

Notre contribution se base sur les données d'un des premiers grands corpus oraux en L2, qui a été recueilli dans les années '80, époque à laquelle la transcription était initialement faite à la main¹⁰. Feldweg (1993) résume les phases et les écueils qui ont jalonné le processus de mise à disposition du corpus correspondant en version électronique (banque de données de l'Institut Max Planck de Nimègue) ainsi que les divergences attestées dans la transcription de chaque LC - compromis entre transcription orthographique / phonétique, contraintes de temps pour homogénéiser les transcriptions initiales -, malgré le fait que les chercheurs impliqués aient suivi le principe général d'une transcription relativement 'neutre'.

Depuis, le nombre de corpus oraux en L2 (et des recherches correspondantes) s'est accru de manière importante. En ce qui concerne le français L2, nous pouvons citer au moins deux grands projets ayant donné lieu à des corpus de taille importante : le projet Interfra¹¹ (« *Interlangue française – développement, interaction and variation* », cf. Bartning 1997), centré sur les apprenants suédophones, et le projet FLLOC¹² (« *French Learner Language Oral Corpora* », cf. Marsden *et al.* 2002), fondé sur l'observation d'apprenants anglophones.

Les deux ont été conçus et transcrits de manière à être utilisés pour des analyses ultérieures variées. Face aux problèmes de transcription de l'oral en français L2, les auteurs ont élaboré des solutions partiellement différentes, mais les productions en L2 sont en grande partie transcrites suivant l'orthographe standard, avec

¹⁰ Comme le fait remarquer Feldweg : « When data transcription started in 1983, the research teams did not even have a personal computer available to them, and transcriptions were handwritten » (1993:108).

¹¹ <https://www.su.se/romklass/interfra>

¹² <http://www.flloc.soton.ac.uk/conventions.html>

une sensibilité plus ou moins forte pour l'ambiguïté potentielle des segments en L2. Ceci s'explique par des raisons pratiques telles que la lisibilité, la taille importante du corpus, la nécessité de respecter la période précise couverte par des financements reçus, ou l'objectif de les analyser de façon automatique grâce à des logiciels de traitement des données linguistiques.

Ainsi, par exemple, le corpus Interfra est transcrit pour l'essentiel de manière orthographique, mais s'en écarte lorsque le même segment discursif se prête à deux interprétations également possibles en contexte : on recourt notamment à des parenthèses (par ex. *il(s) prenai(en)t*) et, surtout, au symbole « E » pour indiquer la variation possible dans la réalisation en L2 des sons [ɛ/e] dans les formes verbales ambiguës, telles que « *passé / passait* » (*passE*)¹³.

En revanche, le corpus FLLOC opte pour la transcription en format CHAT¹⁴, avec des adaptations nécessaires pour des données en L2¹⁵. La ligne principale, qui reporte les énoncés des apprenants, suit la transcription orthographique 'normalisée' (en d'autres termes, les erreurs sont corrigées) pour que le programme puisse reconnaître les mots produits. La transcription n'est pas donc fidèle à la forme effectivement produite par l'apprenant, mais la consultation des fichiers sons correspondants, disponibles et reliés à la transcription, permet d'accéder aux productions originales de l'apprenant, et d'adapter, si nécessaire, la transcription aux besoins de l'analyse.

La base de données *Corpus Inter Langue*, composée des productions d'apprenants du français L2 ayant un niveau de A2 à C1 (Arbach 2015), constitue un autre exemple, plus récent, où un compromis entre la taille du corpus, d'une part, et la lisibilité et les exigences techniques liées à l'analyse grâce aux outils informatiques, d'autre part, a conduit au choix de la transcription orthographique. En effet, les auteurs de ce corpus ont pris cette décision en considérant que la transcription phonétique pose des problèmes de lisibilité et rend très difficile, voire impossible les analyses automatiques.

Le choix du type de transcription au sein de différents grands corpus en L2 conduit donc à des compromis entre fidélité et lisibilité, mais est également conditionné par des contraintes relatives aux moyens matériels (logiciel de traitement des données linguistiques) et au temps à disposition.

La revue de ces différents corpus et leurs critères de transcription des données orales en L2 n'a pas comme but de critiquer les choix adoptés, mais plutôt de réfléchir aux problèmes de la transcription, en distinguant soigneusement entre la tâche de transcrire et celle de l'analyse successive, sachant que l'une peut influencer l'autre. De plus, la transcription et l'analyse ne sont pas toujours accomplies par les mêmes personnes, ce qui amplifie la question de la « neutralité » relative de la transcription.

La discussion proposée dans ce chapitre nous a permis d'aborder la question suivante, selon nous primordiale pour toute recherche en acquisition de L2 : « Les choix de transcription ont-ils un impact sur la description du fonctionnement et de l'évolution des lectures d'apprenants, sachant que cette description est essentielle pour comprendre le processus complexe d'acquisition de L2 ? ».

Références bibliographiques

Agren, Malin. 2008. *A la recherche de la morphologie silencieuse : sur le développement du pluriel en français L2*. Lund: Université de Lund. (Thèse de doctorat.)

¹³ Cf. manuel de transcription disponible sur le site suivant : <https://www.su.se/romklass/interfra>

¹⁴ Le format CHAT est un code de transcription proposé par la banque de données en acquisition CHILDES, Child Language Data Exchange System (MacWhinney 2000), ce qui permet d'appliquer à des données transcrites dans ce format les programmes du logiciel CLAN pour le traitement automatique.

¹⁵ A titre d'exemple, le format CHAT conçu pour des analyses automatiques par le logiciel CLAN, demandent des ajustements pour la transcription de l'oral en L2 (ajout de lignes de glose, etc.) qui alourdissent de manière considérable la tâche du transcripteur, surtout lorsque le corpus a été recueilli pour l'analyse des phénomènes linguistiques de différente nature.

- Andorno, Cecilia & Bernini, Giuliano. 2003. Premesse teoriche e metodologiche. In Giacalone Ramat, Anna (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 27–36. Milano: Carocci.
- Arbach, Najib. 2015. *Constitution d'un corpus oral de FLE : enjeux théoriques et méthodologiques*. Thèse de doctorat, Université Rennes 2.
- Banfi, Emanuele & Bernini, Giuliano. 2003. Il verbo. In Giacalone Ramat, Anna (a cura di), *Verso l'italiano. Percorsi e strategie di acquisizione*, 70–115. Milano: Carocci.
- Bartning, Inge. 1997. L'apprenant dit avancé et son acquisition d'une langue étrangère : Tour d'horizon et esquisse d'une caractérisation de la variété avancée. *Acquisition et Interaction en Langue Etrangère* 9. 9–50.
- Bartning, Inge & Schlyter, Suzanne. 2004. Itinéraires acquisitionnels et stades de développement en français L2. *French Language Studies* 14. 281–299.
- Baude, Olivier. 2006. *Corpus oraux. Guide de bonnes pratiques*. Paris: CNRS éditions.
- Becker, Angelika. 2005. The semantic knowledge base for the acquisition of negation and the acquisition of finiteness. In Hendriks, Henriëtte (ed.), *The Structure of Learner Varieties*, 263–314. Berlin: Mouton de Gruyter.
- Becker, Angelika. 2012. Finiteness and the Acquisition of Negation. In Watorek, Marzena & Benazzo, Sandra & Hickmann, Maya (eds.), *Comparative Perspectives on Language Acquisition : A tribute to Clive Perdue*, 54–72. Clevedon: Multilingual Matters.
- Benazzo, Sandra & Starren, Marianne. 2007. L'émergence de moyens grammaticaux pour exprimer les relations temporelles en L2. *Acquisition et Interaction en Langue Etrangère* 25. 129–158.
- Bernini, Giuliano. 2010. Acquisizione dell'italiano come L2. In Simone, Raffaele (a cura di), *Enciclopedia dell'italiano*, vol. 1, 139–140. Roma: Istituto dell'Enciclopedia Italiana G. Treccani.
- Blanche-Benveniste, Claire. 2000. *Approches de la langue parlée en français*. Paris: Ophrys.
- Bley-Vroman, Robert W. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(4). 1–17.
- Brissaud, Catherine & Chevrot, Jean-Pierre & Lefrançois, Pascale. 2006. Les formes verbales homophones en [E] entre 8 et 15 ans : contraintes et conflits dans la construction des savoirs sur une difficulté orthographique. *Langue française* 151. 74–93.
- Corder, S. Pit. 1967. The significance of learner's errors. *IRAL* 5(4). 161–170.
- Dietrich, Rainer & Klein, Wolfgang & Noyau, Colette (eds.). 1995. *The acquisition of temporality in a second language*. Amsterdam: John Benjamins.
- Feldweg, Helmut. 1993. Transcription, storage and retrieval of data. In Perdue, Clive (ed.), *Adult Language Acquisition: crosslinguistic perspectives*, vol. I, 108–130. Cambridge: Cambridge University Press.
- Giacalone Ramat, Anna. 1992. Sur quelques manifestations de la grammaticalisation dans l'acquisition de l'italien comme deuxième langue. *Acquisition et Interaction en Langue Etrangère* 1. 143–170.
- Giuliano, Patrizia. 2005. *La négation linguistique dans l'acquisition d'une langue étrangère*. Berne: Peter Lang.
- Giuliano, Patrizia & Véronique, Daniel. 2005. The acquisition of negation in French L2. An analysis of Moroccan Arabic and Spanish 'Learner Varieties'. In Hendriks, Henriëtte (ed.), *The Structure of Learner Varieties*, 355–404. Berlin: Mouton de Gruyter.
- Jaffré, Jean-Pierre & Brissaud, Catherine (éds.). 2006. *Morphographie et hétérographie*. Numéro thématique de *Langue française* 151.
- Klein, Wolfgang & Perdue, Clive. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13(4). 301–347.
- Leclercq, Pascale. 2020. Transcribing interlanguage. The case of verb final [e] in L2 French. In Edmonds, Amanda & Leclercq, Pascale & Gudmestad, Aarnes (eds.), *Interpreting language-learning data*, 169–196. Amsterdam: Eurosla Studies Series.
- MacWhinney, Bryan. 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3^a ed. Mahwah: Lawrence Erlbaum Associates.
- Marsden, Emma & Mitchell, Rosamond & Myles, Florence & Rule, Sarah. 2002. Oral French INterlanguage Corpora: Tools for Data management and analysis. *Occasional Papers* 158.
- Noyau, Colette & Houdaïfa Et-Tayeb & Vasseur, Marie-Thérèse & Véronique, Daniel. 1995. The acquisition of French. In Dietrich, Rainer & Klein, Wolfgang & Noyau, Colette (eds.), *The acquisition of temporality in a second language*, 145–209. Amsterdam: John Benjamins.

- Ochs, Elinor. 1979. Transcription as theory. In Ochs, Elinor & Schieffelin, Bambi B. (eds.), *Developmental pragmatics*, 43–72. New York: Academic Press.
- Ortega, Lourdes. 2014. Trying out theories on interlanguage : Description and explanation over 40 years of L2 negation research. In Han, ZhaoHong & Tarone, Elaine (eds.), *Interlanguage: Forty years later*, 173–201. Amsterdam: John Benjamins.
- Perdue, Clive (ed.). 1993. *Adult Language Acquisition: Crosslinguistic perspectives*, vol. I et vol. II. Cambridge: Cambridge University Press.
- Perdue, Clive. 1995. *L'acquisition du français et de l'anglais par des adultes*. Paris : CNRS Editions.
- Riegel, Martin & Pellat, Jean-Christophe & Rioul, René. 2009. *Grammaire méthodique du français*. Paris : PUF.
- Selinker, Larry. 1972. Interlanguage. *IRAL* 10(3). 209–232.
- Silberstein, Dagmar. 2001. Facteurs interlingues et spécifiques dans l'acquisition non guidée de l'anglais L2. *Acquisition et Interaction en Langue Etrangère* 14. 25–58.
- Stoffel, Henriette & Véronique, Daniel. 2003. L'acquisition de la négation en français par des adultes arabophones. *Marges linguistiques* 5. 242–252.
- Véronique, Daniel & Carlo, Catherine & Kim, Jin-Ok & Granget, Cyrille. 2009. *L'acquisition de la grammaire du français langue étrangère*. Paris : Didier.
- Watorek, Marzena & Perdue, Clive. 2005. Psycholinguistic Studies on the Acquisition of French as a Second Language: The 'Learner Variety' Approach. In Dewaele, Jean-Marc (ed.), *Focus on French as a Foreign Language*, 1–16. Clevedon: Multilingual Matters.