



LUND UNIVERSITY

Les débutants parlent-ils à l'infinifif? : Influence de l'input sur la production des verbes par des apprenants adultes du français

Thomas, Anita

2009

[Link to publication](#)

Citation for published version (APA):

Thomas, A. (2009). *Les débutants parlent-ils à l'infinifif? : Influence de l'input sur la production des verbes par des apprenants adultes du français*. [Centre for Languages and Literature].

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

ÉTUDES ROMANES DE LUND 87

Anita Thomas

LES DÉBUTANTS PARLENT-ILS À L'INFINITIF ?

**INFLUENCE DE L'INPUT
SUR LA PRODUCTION DES VERBES
PAR DES APPRENANTS ADULTES DU FRANÇAIS**



LUNDS
UNIVERSITET

Språk- och litteraturcentrum
Franska

THOMAS, ANITA, Les débutants parlent-ils à l'infinitif? Influence de l'input sur la production des verbes par des apprenants adultes du français (Do beginners speak in the infinitive? Influence of input on verb production by adult learners of French), Études Romanes de Lund, 87, Lund 2009. 255 pages. Written in French. Monograph

This dissertation deals with the characteristic variation found in adult French L2 oral production. In their production of regular verbs, beginners use a short form, like *parle*, and a long form, like *parlE*, in contexts of both the present (*je parl/*je parlE*) and the infinitive (*je veux parlE/*je veux parl*). This sometimes results in target-like uses and sometimes not.

The question addressed in the thesis is if there is any systematicity in learners' use of the verb forms. In attempting to account for the variation, some previous studies have pointed at specific linguistic factors while others have suggested that the observed variation can be more or less arbitrary. One kind of systematicity has been found in earlier studies focusing on the development of past tense morphology. These studies showed that production of the forms could be predicted from the prototypical lexical aspect of the verbs. Based on French L2 studies of this kind, I present a first hypothesis stating that, also when referring to non-past contexts, i.e. present and infinitive contexts, learners use the short form with static verbs and the long form with dynamic verbs.

I further propose a second, contrasting hypothesis, namely that learners are influenced by the frequency of surface forms in the input. This hypothesis emerges from a usage-based perspective (Bybee, 2008) where input frequency is a key factor (Ellis, 2002).

A first investigation of 16 regular verbs in different input sources (for example, corpora of spoken French and teacher talk) is conducted in order to obtain specific data for the second hypothesis. Firstly, this examination shows that the distribution of verb forms in the input is uneven: some verbs occur more often in the short form and others more often in the long form. A gradual scale can thus be established for how frequent each of the studied verbs is in the short and the long form respectively. Secondly, I found that lexical aspect and frequency are closely related. The verbs more frequent in the short form turned out to be static verbs while the verbs more frequent in the long form turned out to be dynamic verbs.

The analysis of the same verbs in learner data from a corpus of oral production shows tendencies in favour of the influence of input frequency. This investigation is completed by using experimental data for 12 regular verbs from a free production task, an elicited imitation task and a (serial) recall task. The results of the imitation task, where the verbs are presented in sentences, showed a possible influence of input frequency from the beginning stages. An influence of lexical aspect could only be detected in the production by learners having passed the beginning stages. The recall task, where the verbs are presented in series of five at a time, reveals a weaker correlation with input frequency and no significant differences between the verbs with respect to the lexical aspect.

In conclusion, the data show that lexical aspect and input frequency are closely related in input as well as in learner data and the results speak in favour of input frequency for the verbs where the two hypotheses can be distinguished.

Språk- och litteraturcentrum
Lunds universitet
Box 201
SE-221 00 LUND, Suède

© Anita Thomas

ISSN 0347-0822
ISBN 978-91-978017-2-0

Imprimé en Suède
Media-Tryck
Lund 2009

TABLE DES MATIÈRES

Index des figures et tableaux	7
Abréviations et conventions de transcription	8
Remerciements	9
1. INTRODUCTION	11
PARTIE THEORIQUE	15
2. LES VERBES REGULIERS EN –ER EN FRANÇAIS.....	17
3. VARIATION DES FORMES ET FINITUDE	21
3.1. Approche fonctionnelle : la variété de base.....	21
3.2. Stades de développement en français LE/L2.....	22
3.3. Perspective de la grammaire générative : formes et syntaxe.....	24
3.4. Synthèse et discussion	28
4. VARIATION DES FORMES ET ASPECT LEXICAL	33
4.1. Présentation de l’aspect lexical	33
4.2. Influence de l’aspect lexical en français LE/L2.....	38
4.2.1. Présentation de l’hypothèse de l’aspect	39
4.2.2. Études à partir de l’hypothèse de l’aspect et français LE/L2.....	41
4.2.3. Aspect lexical dans le corpus ESF – approche fonctionnelle	47
4.3. Résumé et discussion.....	50
5. CADRE THEORIQUE : PERSPECTIVE BASEE SUR L’USAGE DE LA LANGUE (USAGE-BASED).....	53
5.1. Principaux concepts de la perspective basée sur l’usage de la langue.....	53
5.1.1. Expérience de la langue.....	53
5.1.2. Constructions.....	54
5.1.3. Fréquence	54
5.1.4. Input et usage de la langue.....	57
5.2. Autres facteurs en jeu	58
5.2.1. Attention	58
5.2.2. Saillance.....	60
5.2.3. <i>Intake</i>	61
5.3. Résumé et discussion.....	61
6. HYPOTHESES	63
7. PRINCIPAUX CHOIX METHODOLOGIQUES	65
7.1. Provenance des données empiriques	65
7.1.1. Données d’input	65

7.1.2. Données d'apprenant/es.....	66
7.1.3. Données expérimentales.....	68
7.2. Choix des verbes.....	69
7.2.1. Présentation des verbes choisis.....	69
7.2.2. Classement des verbes selon leur aspect lexical.....	70
7.3. Traitement des données.....	74
7.3.1. Formes verbales et contextes.....	74
7.3.2. Codage des formes.....	75
7.3.3. Format des données (logiciels).....	75
7.3.4. Méthodes statistiques.....	76

PARTIE DONNEES DE CORPUS79

8. DONNEES D'INPUT81

8.1. Choix méthodologiques.....	81
8.1.1. Sources de données d'input.....	81
8.1.2. Choix des verbes.....	88
8.1.3. Compilation des formes.....	88
8.2. Analyse empirique des données d'input.....	89
8.2.1. Verbes réguliers en <i>-er</i> dans l'input.....	89
8.2.2. Verbes irréguliers dans l'input oral.....	97
8.3. Résumé et prédictions.....	101

9. DONNEES D'APPRENANT/ES DE CORPUS105

9.1. Choix méthodologiques.....	105
9.1.1. Présentation des apprenant/es.....	105
9.1.2. Choix des verbes.....	105
9.1.3. Calcul des occurrences.....	106
9.2. Résultats des apprenant/es de corpus.....	107
9.2.1. Verbes réguliers en <i>-er</i>	107
9.2.2. Verbes irréguliers.....	115
9.3. Résumé et discussion.....	120

PARTIE EXPERIMENTALE123

10. PRESENTATION DE LA PARTIE EXPERIMENTALE125

10.1. Introduction.....	125
10.2. Déroulement de l'expérience.....	125
10.3. Sélection des verbes pour la partie expérimentale.....	127
10.4. Présentation des règles éthiques suivies.....	128
10.5. Présentation des participant/es.....	129

10.5.1. Apprenant/es	129
10.5.2. Groupe contrôle de natifs du français	131
11. TACHE DE PRODUCTION LIBRE.....	133
11.1. Présentation de la tâche de production libre	133
11.2. Traitement des données.....	134
11.3. Évaluation du niveau des apprenant/es.....	136
11.3.1. Choix des méthodes.....	136
11.3.2. Procédure	137
11.3.3. Résultats de l'évaluation du niveau des apprenant/es.....	140
11.4. Résultats de la production libre.....	145
11.4.1. Calcul des occurrences	145
11.4.2. Distribution des formes et contextes chez les participant/es.....	146
11.4.3. Vérification des hypothèses	150
11.5. Résumé des données de la tâche de narration.....	153
12. TEST D'IMITATION	155
12.1. Présentation du test d'imitation.....	155
12.2. Déroulement et construction du test d'imitation.....	158
12.2.1. Construction du matériel sonore (phrases stimulus).....	159
12.2.2. Tâche de suppression articulatoire	164
12.2.3. Images indices	165
12.2.4. Consignes.....	166
12.3. Traitement des données.....	166
12.3.1. Conventions de transcription	166
12.3.2. Codage des phrases.....	168
12.4. Résultats du test d'imitation.....	168
12.4.1. Outputs sans verbes	169
12.4.2. Vérification de l'hypothèse 1 sur l'influence de l'aspect lexical	171
12.4.3. Vérification de l'hypothèse 2 sur la fréquence des formes	173
12.4.4. Analyse qualitative de quelques exemples	176
12.5. Résumé des résultats du test d'imitation	183
13. TEST DE RAPPEL.....	187
13.1. Présentation du test de rappel.....	187
13.2. Construction du test de rappel.....	189
13.2.1. Construction des séries de formes verbales.....	189
13.2.2. Déroulement et consignes.....	193
13.3. Conventions de transcriptions et codage	193
13.4. Résultats du test de rappel.....	196
13.4.1. Vérification de l'hypothèse 1 sur l'influence de l'aspect lexical	196
13.4.2. Vérification de l'hypothèse 2 sur la fréquence des formes	198
13.5. Résumé des résultats du test de rappel	200

SYNTHESE.....	203
14. SYNTHESE FINALE	205
14.1. Introduction	205
14.2. Résultats des données de l'input et prédictions	207
14.3. Résultats des données d'apprenant/es de corpus.....	209
14.4. Résultats de la partie expérimentale	210
REFERENCES BIBLIOGRAPHIQUES	223
ANNEXES.....	235

Index des figures et tableaux

Tableau 1-1 :	Répartition des formes verbales de l'exemple (1-1).....	12
Tableau 2-1 :	Formes courtes et longues en français	18
Tableau 2-2 :	Forme longue “parlE” dans trois langues romanes.....	18
Figure 3-1 :	Illustration simplifiée de la montée du verbe en français	25
Figure 3-2 :	Montée du verbe en L2 (à partir de Prévost & White, 2000b).....	27
Tableau 3-1 :	Morphologie des formes verbales en suédois à l'oral.....	30
Tableau 4-1 :	Catégories liées à l'aspect lexical.....	35
Tableau 7-1 :	Corpus Thomas : nombre de semestres par apprenant/e.....	67
Tableau 7-2 :	Listes des verbes étudiés dans ce travail	70
Tableau 7-3 :	Répartition des verbes selon aspect lexical prototypique	72
Tableau 7-4 :	Codage de la production des verbes réguliers en –er.....	75
Tableau 8-1 :	Données sur les enseignants du corpus Flyman Mattsson	84
Tableau 8-2 :	Fréquence des formes des verbes en –er dans l'input.....	91
Tableau 8-3 :	Comparaison des sources d'input	94
Tableau 8-4 :	Corrélation entre les pourcentage de formes courtes dans l'input.....	95
Tableau 8-5 :	Test statistique pour l'aspect lexical	96
Tableau 8-6 :	Fréquence des formes des verbes irréguliers dans l'input	99
Tableau 8-7 :	Formes attendues chez les apprenant/es – verbes en –er.....	102
Tableau 8-8 :	Formes attendues chez les apprenant/es – verbes irréguliers	103
Tableau 9-1 :	Production des verbes en –er par les apprenant/es de corpus.....	109
Tableau 9-2 :	Pourcentages de formes courtes : apprenant/es de corpus et input.....	113
Tableau 9-3 :	Corrélations entre input et apprenant/es de corpus.....	114
Tableau 9-4 :	Production des verbes irréguliers par les apprenant/es de corpus	116
Tableau 9-5 :	Pourcentages de formes du présent: apprenant/es et input.....	119
Tableau 10-1 :	Déroulement de l'expérience pour les apprenant/es	126
Tableau 10-2 :	Verbes de test de la partie expérimentale.....	128
Tableau 10-3 :	Présentation des apprenant/es.....	129
Tableau 10-4 :	Méthodes de français utilisées dans les classes de lycées.....	131
Figure 11-1 :	Série d'images “Le voyage en Italie”	134
Tableau 11-1 :	Classement des participant/es par stade et nombre de mots/minutes	140
Tableau 11-2 :	Tests statistiques pour le regroupement des participant/es	144
Tableau 11-3 :	Évaluation manuelle du stade de six récits par trois juges.....	145
Figure 11-2 :	Contextes du présent et de l'infinitif – tâche de narration	147
Figure 11-3 :	Formes dans les contextes pour le présent – tâche de narration.....	148
Figure 11-4 :	Formes dans les contextes pour l'infinitif – tâche de narration.....	149
Tableau 11-4 :	Production des verbes selon l'aspect lexical – tâche de narration.....	151
Tableau 11-5 :	Pourcentages de formes courtes – tâche de narration et input	152
Tableau 12-1 :	Phrases du test d'imitation par ordre alphabétique des verbes.....	159
Tableau 12-2 :	Codage des formes verbales dans le test d'imitation.....	168
Tableau 12-3 :	Outputs sans verbes et sans mots par stade	170
Tableau 12-4 :	Test statistique pour les outputs sans verbes par stade	170
Tableau 12-5 :	Production des verbes selon l'aspect lexical – test d'imitation.....	172
Tableau 12-6 :	Pourcentages de formes courtes – test d'imitation et input.....	173
Figure 12-1 :	Test d'imitation – corrélation entre le stade 1 et l'input.....	175
Tableau 13-1 :	Liste A du test de rappel libre.....	191

Tableau 13-2 : Séries d'exercices	192
Tableau 13-3 : Codage des formes verbales dans le test de rappel.....	194
Tableau 13-4 : Formes verbales discutables dans le rappel libre.....	195
Tableau 13-5 : Production des verbes selon l'aspect lexical – test de rappel.....	197
Tableau 13-6 : Pourcentages de formes courtes – test de rappel et input	199
Figure 13-1 : Corrélation input et données du test de rappel au total	200
Tableau 14-1 : Input et prédictions pour la production des apprenant/es.....	208
Tableau 14-2 : Pourcentages de formes courtes par tâche au total	213
Tableau 14-3 : Pourcentages de formes courtes par aspect lexical et par stades.....	216
Annexe 1 : Description des stades 1 à 3 selon Bartning & Schlyter (2004).....	235
Annexe 2 : Formes des verbes en –er dans l'input.....	237
Annexe 3 : Formes des verbes irréguliers dans l'input oral.....	241
Annexe 4 : Tâche de narration : verbes en –er par participant/es	243
Annexe 5 : Test d'imitation : données par participant/es	245
Annexe 6 : Test d'imitation : verbes par stade.....	246
Annexe 7 : Test de rappel : données par participant/es	247
Annexe 8 : Test de rappel : verbes par stade.....	249

Abréviations et conventions de transcription

-E : l'archiphonème –E représente les sons [e] et [ɛ]

LE : langue étrangère

*ABC : locuteur

*INT : interlocutrice (intervieweuse)

@x : mot dans une autre langue que le français

[//] : reformulation avec changement

[/] : reformulation sans changement

+/. : énoncé coupé

+, : poursuite de l'énoncé

+... : énoncé laissé en suspens

: pause silencieuse

: pause silencieuse prolongée

(...) : partie de texte non transcrite

www : partie de texte non transcrite. Dans un exemple plus long, le signe www entre deux énoncés indique qu'une partie de la conversation a été coupée, mais que les énoncés présentés font partie de la même séquence.

Remerciements

Ce livre est le fruit de plusieurs années de travail et je suis reconnaissante d'avoir été si bien entourée. Je remercie du fond du coeur toutes les personnes, collègues, amies, amis, qui m'ont encouragée et soutenue durant ces années. Merci d'avoir été là !

Je tiens à exprimer ma profonde gratitude à Suzanne Schlyter et Jonas Granfeldt, qui ont dirigé cette thèse. Merci de votre enthousiasme et pour tout ce que j'ai eu la chance d'apprendre lors de nos discussions. Merci Suzanne pour ta générosité, ta créativité, ton ouverture à ce qui est nouveau et pour la mémoire de ce qui est passé. Merci Jonas pour m'avoir encouragée et guidée en particulier dans l'aventure expérimentale de cette thèse. J'ai beaucoup appris grâce à ta manière d'expliquer simplement les choses.

Le contenu de ce travail a été enrichi par les suggestions et commentaires de nombreuses personnes. Merci pour l'empreinte que vous avez laissée sur le texte. Je remercie tout particulièrement Marie Labelle pour son rapport critique lors de la soutenance de mon travail de licenciat, mais aussi pour sa présence active lors de plusieurs séminaires centraux pour ce travail. Je suis reconnaissante pour tous les commentaires que j'ai reçus lors des séminaires du département de langues romanes et du séminaire sur l'acquisition des langues. Je tiens tout particulièrement à remercier les personnes qui ont participé à mon dernier séminaire, Petra Bernardini, Verner Egerland, Patrik Fridlund, Frédérique Granath, Disa Holmlander, et j'adresse ma profonde reconnaissance à Valéria Molnár et Eva Wiberg pour leur lecture critique et constructive. Merci aussi à Susan Sayehli et Malin Ågren pour leurs commentaires sur mon texte pré-final. Je suis également reconnaissante pour l'accueil chaleureux et l'inspiration que j'ai reçue lors de ma visite au Laboratoire de Psychologie Sociale et Cognitive à Clermont-Ferrand. Merci aussi à Lisa Christensen pour ses commentaires sur l'aspect lexical des verbes, à Paul Touati pour son aide lors de la construction des phrases du test d'imitation et ses conseils pour l'enregistrement du matériel sonore et à Malin Ågren pour m'avoir autorisée à utiliser la série d'image "Le voyage en Italie".

Écrire une thèse à partir de données d'apprenant/es implique l'engagement de nombreuses personnes. Je tiens ici à remercier toutes les personnes qui m'ont prêté leur corpus et tout particulièrement Anna Flyman Mattsson et Suzanne Schlyter. Je remercie aussi les apprenant/es qui ont accepté de se faire enregistrer pour la création de mon corpus, les personnes qui ont participé aux

tests pilotes et celles qui ont accepté de participer à la partie expérimentale en tant qu'apprenante ou native.

Je suis reconnaissante pour l'aide que j'ai eue pour me mettre en contact avec des enseignant/es de lycée et des étudiant/es d'échange, merci tout particulièrement à Jonas Granfeldt, Frédérique Granath, Malin Ågren et Christine Bozier. Je remercie les enseignants et enseignantes de français Annika Backemyr Nyberg, Heléne Bäckström, Katarina Friström, Helene Johansson, Torun Johansson, John Jönsson, Piera Karlsson, Charlotta Nord, Nils Wemmenhed et Madeleine Zachrisson pour leur engagement, non seulement pour me mettre en contact avec leurs élèves, mais aussi au niveau de l'organisation pratique des salles et des rendez-vous.

Je tiens aussi à remercier le personnel du laboratoire expérimental (*Humanistlaboratoriet*) du Centre de Langues et de Littérature de l'Université de Lund : Kenneth Holmqvist pour m'avoir expliqué les procédures éthiques à suivre lors d'une expérience psycholinguistique ; Stefan Lindgren pour son aide technique lors de l'enregistrement du matériel sonore ; Joost van der Weijer pour avoir patiemment répondu à mes questions de statistique.

Merci à Flore Buchaillet qui a accepté de prêter sa voix au matériel sonore des tests expérimentaux et pour sa patience au moment des deux enregistrements.

Je remercie Ralph Thomas pour sa relecture attentive du manuscrit final et les petits commentaires en marge.

Je remercie la fondation Knut et Alice Wallenberg et l'Université de Lund pour le financement de mon salaire et la fondation Crafoord pour le financement de mon ordinateur.

Merci à mes collègues doctorantes Disa Holmlander et Nina Bengtsson pour le bon temps passé à la 608. Merci à Malin Ågren pour l'amitié au quotidien et pour la présence dans les bons et les mauvais moments de la vie académique.

Je remercie mes proches pour leur présence malgré la distance, en particulier ma sœur Karin Thomas pour les discussions sur la recherche, et mon amie Berta Staedler. J'adresse ici un remerciement particulier à ma mère pour m'avoir encouragée puis donné la possibilité de faire des études ; merci pour cet engagement cohérent !

Merci Patrik, pour le couple que nous avons la chance de former et qui me permet chaque jour de me ressourcer.

Lund, le 25 août 2009

1. Introduction

Ce travail est consacré à l'étude de la variation des formes dans la production orale d'apprenant/es¹ en début d'acquisition du français langue étrangère ou langue seconde (dorénavant LE/L2)². La variation étudiée ici est celle de la production de formes courtes du type *parle* et de formes longues du type *parlE* dans les contextes du présent et de l'infinitif, parfois correctement, parfois non. Par convention l'archiphonème *–E* représente les sons [e] et [ɛ]. Pour illustrer ce phénomène de variation, voici l'exemple (1-1) de production d'une apprenante qui a commencé à apprendre le français au lycée à l'âge de 16 ans. Il s'agit d'un extrait de la tâche de production libre dans laquelle elle raconte *Le voyage en Italie* à partir d'une série d'images³.

- (1-1) (P08 – stade 3) C'est deux, c'est deux personnes que va **va voyagE** à en Italie. (...) Et elle **fait** du soleil en le, oui. En la nuit il elle elle **mange** en un restaurant. Et l'italienne, et deux deux hommes **aime** les filles. Et les hommes **va parle** avec les filles et **boire** le vin et la bière. Et ils il, il **dansE** ensemble. (...)

Cet extrait montre qu'on trouve quatre combinaisons différentes dans l'utilisation des formes verbales. Ces combinaisons sont présentées dans le tableau 1-1 pour les verbes réguliers en *–er*, avec d'une part les contextes syntaxiques du présent et de l'infinitif et d'autre part les formes utilisées, la forme courte et la forme longue, indépendamment de l'accord sujet-verbe. L'astérisque * devant certaines formes indique qu'elles ne sont pas correctes selon les normes du français.

¹ Etant donné que l'on trouve au moins 50% de femmes dans ces groupes, une trace du féminin sera exprimée sur ce terme, sur celui de *participant/e* ainsi que sur les adjectifs épithètes accompagnant ces deux termes.

² Dans ce travail, les termes de LE – langue étrangère et L2 – langue seconde seront utilisés en parallèle sans distinction. En effet, les apprenant/es en question apprennent le français dans un pays où la langue n'est pas parlée et en ce sens elle est étrangère, mais l'approche de ce travail se situe dans celle de l'acquisition d'une L2 par opposition à une première langue, L1.

³ Pour une description détaillée de la tâche voir chapitre 11.

Tableau 1-1 : Répartition des formes verbales de l'exemple (1-1)

		présent		infinitif
forme courte	1	<i>elle mange</i> <i>deux hommes aime</i>	3	<i>*les hommes va parle</i>
forme longue	2	<i>*il dansE</i>	4	<i>deux personnes va voyageE</i>

La production de formes correctes (cases 1 et 4 du tableau 1-1) et incorrectes (cases 2 et 3) est un phénomène discuté depuis une vingtaine d'années et il est considéré comme une des caractéristiques du français des apprenant/es LE/L2.

L'objectif de ce travail est de savoir si l'on peut observer une certaine systématisme dans la production des formes par les apprenant/es. La question sera ici posée en termes d'influence de l'aspect lexical d'une part et d'influence de l'input d'autre part.

La question de la variation des formes a souvent été étudiée à partir des catégories de la finitude, en analysant la forme courte comme une forme finie (du présent) et la forme longue comme une forme non-finie (de l'infinitif) notamment dans le cadre de la grammaire générative (Prévost & White, 2000 a et b ; Prévost 2004 ; 2007). Si ces travaux offrent une bonne description de base du phénomène de la variation des formes, ils ne permettent pas de voir de systématisme dans la production des apprenant/es. En effet, comme on l'a vu dans le tableau 1-1, les apprenant/es utilisent les deux formes dans les deux contextes. Ces études tendent ainsi à suggérer que la variation est aléatoire. Mon propre travail commence donc là où ces travaux s'arrêtent, en considérant la question d'un autre point de vue que de celui de la syntaxe et de la finitude.

Une certaine systématisme dans la production des formes par les apprenant/es a par contre été mise en évidence dans des travaux sur l'expression de la temporalité et en particulier du passé (Andersen, 1991, 1993, 2002 ; Dietrich *et al.*, 1995). Ces travaux ont pu montrer que les apprenant/es semblaient guidés par l'aspect lexical des verbes (*Aktionsart*) dans leur production des formes du passé. Bien que le présent travail ne traite pas du passé, on peut supposer que les apprenant/es soient influencés par l'aspect lexical des verbes quel que soit le contexte dans lequel ils produisent les formes.

Ces dernières années, on a vu le développement d'une théorie basée sur l'usage de la langue (*usage-based*), qui propose une influence directe de la fréquence des formes sur la production des apprenant/es (Bybee, 2008 ; Ellis, 2002). De ce point de vue, la variation des formes verbales produites par les apprenant/es s'explique par le fait que les formes fréquentes dans l'input sont plus aisément accessibles par les apprenant/es que les formes moins fréquentes.

La question de la systématique dans la production des formes verbales par les apprenant/es sera ainsi traitée à partir de deux hypothèses : l'une à partir de l'aspect lexical des verbes et l'autre à partir de la fréquence des formes de surface.

Pour étudier ces questions, j'ai choisi un nombre limité de verbes réguliers en *-er* (16) ainsi que quelques verbes irréguliers (12) à titre de contrôle, le nombre de formes distinctes à l'oral étant plus élevé que pour les verbes en *-er*. Une difficulté propre aux verbes en *-er* est en effet la multifonctionnalité de la forme longue comme on le verra au chapitre 2.

Dans le but de concrétiser les données sur la fréquence des formes dans l'input, les verbes choisis pour ce travail seront tout d'abord étudiés dans différentes sources de données d'input, couvrant les principales variétés de français auxquelles des apprenant/es débutant/es peuvent être exposés.

Les hypothèses seront ensuite vérifiées sur différentes données d'apprenant/es : d'une part de corpus de conversation et d'autre part de production en condition expérimentale. L'avantage de données de corpus est qu'elles reflètent une utilisation libre de la langue par les apprenant/es, mais il est souvent difficile de contrôler les contextes d'utilisation des formes. Pour cette raison, la question de la systématique de la production des formes par les apprenant/es sera vérifiée de manière ciblée avec des tests psycholinguistiques et un choix précis de verbes réguliers en *-er*. Dans le test d'imitation, les verbes sont présentés dans des phrases alors que dans le test de rappel, ils sont présentés dans des séries de verbes.

La question de la systématique dans la production des formes par les apprenant/es sera donc traitée à partir d'une variété de données empiriques. Une telle analyse permettra de discuter si et à quel moment les apprenant/es sont influencés par l'aspect lexical ou la fréquence des formes de surface.

Ce travail est divisé en trois grandes parties, une partie théorique, une partie de données de corpus et une partie expérimentale ; il se termine par une synthèse.

La partie théorique s'ouvre sur une présentation du paradigme des verbes réguliers en *-er* pour mettre en valeur les difficultés propres à l'acquisition de ces verbes en français (chapitre 2). Le chapitre 3 présente des études antérieures qui ont considéré la question de la variation des formes chez les apprenant/es à partir de la finitude. Le chapitre 4 sera consacré aux études antérieures qui proposent que la production des formes par les apprenant/es soit influencée par l'aspect lexical des verbes en particulier dans l'expression du passé. Ce chapitre présentera les différents concepts reliés à l'aspect lexical ainsi que des études

qui ont été faites sur le français LE/L2. Il se terminera sur la formulation d'une première hypothèse pour ce travail. Le cadre théorique sera présenté au chapitre 5. Il s'agit d'une approche basée sur l'usage de la langue (*usage-based approach*) qui met en avant l'influence de la langue telle qu'elle est utilisée dans l'input, ainsi que les processus cognitifs de traitement de la langue par les apprenant/es. Le rôle clé de la fréquence sera présenté et servira de base à la formulation de la deuxième hypothèse. Les deux hypothèses seront résumées dans le chapitre 6. Cette partie théorique se terminera sur une présentation générale des choix méthodologiques au chapitre 7.

La partie des données de corpus comprend les deux premiers chapitres d'analyse empirique. Le chapitre 8 présentera une étude de l'input à partir de trois sources de données orales : celles de deux interlocutrices en conversation avec des apprenant/es débutant/es, d'enseignants en classe de français LE/L2 au lycée ainsi que du corpus de français parlé en France du DVD C-ORAL-ROM (2005). Ces données orales seront complétées par celles d'une méthode de français pour apprenant/es débutant/es suédophones. Ce chapitre se terminera sur des prédictions concrètes au niveau de l'influence de la fréquence des formes de surface sur la production des apprenant/es. Le chapitre 9 sera consacré à une première vérification des hypothèses, cette fois sur des données d'apprenant/es de corpus. Les chapitres 8 et 9 comprendront également l'analyse de quelques verbes irréguliers à titre de comparaison aux verbes réguliers en –*er*.

La partie expérimentale comprend quatre chapitres : un chapitre de présentation générale, le chapitre 10, et trois chapitres d'analyse. Le premier chapitre d'analyse, le chapitre 11, est consacré aux données de la tâche de production libre. Cette tâche a été insérée dans la partie expérimentale dans le but de permettre la comparaison entre des données de production libre et expérimentales par les mêmes apprenant/es. Le chapitre 12 présente les résultats du test d'imitation et le chapitre 13 ceux du test de rappel.

Ce travail se terminera sur une synthèse finale au chapitre 14, qui présentera les principaux résultats de l'analyse et une discussion sur l'influence de l'aspect lexical et de la fréquence des formes dans l'input sur la production des formes verbales par les apprenant/es.

PARTIE THÉORIQUE

2. Les verbes réguliers en *-er* en français

L'objectif de ce chapitre est de présenter les principales caractéristiques des verbes réguliers en *-er* en français, avec une attention particulière à la réalisation de ces verbes à l'oral.

Les verbes réguliers dont l'infinitif se termine en *-er* sont traditionnellement présentés comme faisant partie du premier groupe. Selon Riegel *et al.* (1994 : 263) ce groupe contient environ 90% des verbes de la langue française soit environ 4000 verbes différents. Le groupe des verbes en *-er* est un groupe ouvert, son paradigme s'applique à la plupart des nouveaux verbes qui entrent dans la langue française (par exemple *smser*, *printer*, *downloader*)⁴. Il est tellement fréquent (*productif*, cf. Bybee, 2001 ; cf. section 5.1.3.) que les natifs l'appliquent parfois sur d'autres verbes comme dans *vous faisez* (au lieu de *vous faites*) ou *il conclue* (au lieu de *il conclut*). Les principales variations dans la construction des formes du paradigme des verbes en *-er* sont d'ordre phonologique (variation vocalique ou consonantique) ou orthographique (par exemple *préférer*, *préfère* ; *manger*, *mangeons* ; *appeler*, *appelle*) (voir par exemple Riegel, 1994 : 272ss).

Les différentes formes du paradigme des verbes en *-er* se construisent à partir de la forme racine, *parl-*. À l'oral, cette forme est aussi celle du présent singulier et de la troisième du pluriel (*je parle*, *tu parles*, *il parle*, *ils parlent*). Une des spécificités du français par rapport aux autres langues romanes est que les formes de l'infinitif *parler*, du participe passé *parlé*, de la 2^e du pluriel du présent *parlez* et de l'imparfait singulier et 3^e du pluriel *parlais* (*parlait*, *parlaient*) sont homophones. Bien qu'il y ait en principe une différence phonétique entre *parler* et *parlais*, elle n'est pas toujours audible en français (David *et al.*, 2006 : 109). Le tableau 2-1 résume les différentes utilisations de la forme courte et de la forme longue en français.

⁴ Selon Riegel *et al.* (1994 : 263) le deuxième groupe de verbes, en *-ir* comme *finir*, (300 verbes) est faiblement ouvert. Dans leur présentation, tous les autres verbes ont été classés dans un troisième groupe.

Tableau 2-1 : Formes courtes et longues en français

	forme courte <i>parle</i>	forme longue <i>parlE</i>
forme racine	parl-	
présent	sujet + parle/s/nt	vous parlez
infinitif		parler
participe passé		parlé
imparfait		sujet + parlais/t/nt

La forme courte n'est utilisée que pour le présent en français alors que la forme longue correspond à un grand nombre de fonctions différentes. Les utilisations de la forme courte en français forment un faisceau plus dense que celui de la forme longue ou en d'autres termes on trouve une relation univoque entre la forme courte et le présent alors que la forme longue est multifonctionnelle.

Les différentes fonctions de la forme longue sont par contre clairement distinctes dans les autres langues romanes. Le tableau 2-2 présente une comparaison de ces différentes formes entre le français, l'italien et l'espagnol. Toutes les lettres se prononcent en italien et en espagnol sauf le /h/ qui est muet.

Tableau 2-2 : Forme longue “*parlE*” dans trois langues romanes

terme en français	infinitif	participe passé	présent 2 ^e du pluriel	imparfait
français	parler	parlé	parlez	parlais
italien	parlare	parlato	parlate	parlavo
espagnol	hablar	hablado	habláis	hablaba

La multifonctionnalité de la forme longue *parlE* est donc une caractéristique des verbes réguliers en français. Les différentes utilisations des formes longues des verbes en *-er* ne posent d'ailleurs pas que des difficultés aux apprenant/es, mais aussi aux natifs du français, à l'écrit, comme en témoigne un numéro complet de la revue *Langue française* dédié à ce phénomène (Jaffré & Brissaud, 2006). Bien que les formes longues soient présentées de différentes manières à l'écrit, l'absence de correspondance dans le rapport phonème et graphème en français donne à penser que l'écrit ne sera pratiquement d'aucune aide aux apprenant/es LE/L2. Il est même probable que la distinction entre la forme courte *parle* se terminant par un *-e* et la forme du participe passé *parlé* se terminant par un *-é* ne soit pas notée, l'accent aigu passant inaperçu. Ces questions dépassent néanmoins le cadre de la présente étude. Pour une discussion sur l'impact de l'écrit voir Granget (2004 : 46ss).

Ensemble, la forme courte *parle* et la forme longue *parlE* couvrent les utilisations les plus fréquentes des verbes réguliers en *-er* à l'oral : le présent de

l'indicatif (de l'impératif et du subjonctif⁵), l'infinitif (et donc le futur périphrastique *va parler*), le participe passé, ainsi que l'imparfait.

Les formes qui manquent – et qui ne seront pas traitées dans ce travail – sont celles du conditionnel *je parlerais* etc., du futur simple *parlerai, parlera* etc. ainsi que la forme de la première du pluriel du présent *nous parlons*. La forme *nous parlons* n'est que rarement utilisée à l'oral, probablement en raison de sa longueur, la construction avec *on* + *forme courte* étant préférée (Weinrich, 1989 : 78-80).

Une présentation des formes courtes et longues comme dans le tableau 2-1 permet de visualiser le fait que les formes non-finies (infinitif et participe passé) sont toujours des formes longues de même que les formes du passé (participe passé et imparfait). La forme courte et la forme longue permettent donc de marquer l'opposition fini vs non-fini ainsi que présent vs passé ou plus généralement le présent vs le non-présent, sauf au niveau de la 2^e du pluriel. Mais les formes longues ne s'utilisent pas uniquement dans des contextes non-finis, puisqu'elles s'utilisent au présent (2^e du pluriel) et à l'imparfait. C'est une des raisons pour lesquelles je me tiens dans ce travail aux termes de formes *courtes* et *longues* et non de formes *finies* et *non-finies*.

Cette présentation du groupe des verbes réguliers en –er en français donne une idée de la difficulté d'acquisition inhérente au paradigme de ce verbe, en particulier celle de la multifonctionnalité de la forme longue à l'oral.

⁵ Les occurrences de l'impératif et du subjonctif ne seront pas traitées dans ce travail, même si elles pourraient avoir une influence au niveau de la fréquence des formes.

3. Variation des formes et finitude

Ce chapitre présentera quelques études antérieures sur la variation dans la production des formes courtes et longues en français LE/L2, notamment dans les contextes du présent et de l'infinitif, à partir des catégories de la finitude. La section 3.1. présentera la contribution de l'approche fonctionnelle (projet ESF) au niveau du statut général des premières formes verbales produites. La section 3.2. présentera les stades d'acquisition proposés par Bartning & Schlyter (2004), qui considèrent la production des formes courtes et longues dans une perspective du développement du français LE/L2. La section 3.3. sera consacrée aux études sur la variation des formes à partir de la perspective de la grammaire générative. La dernière section (3.4.) présentera un résumé des principaux apports de ces recherches tout en montrant de quelle manière le présent travail permet de passer à une nouvelle étape dans l'étude de la variation des formes.

3.1. Approche fonctionnelle : la variété de base

Les recherches effectuées dans le cadre du projet ESF des années 1980-1990 ont entre autres dégagé trois variétés ou stades d'interlangue qui servent encore aujourd'hui de référence pour situer le niveau d'apprenant/es. Ces trois variétés se distinguent par la structure des énoncés produits par les apprenant/es (Perdue, 1993 – Vol. II : 104ss ; Klein, 1989 ; Perdue *et al.* 2002).

- la première variété est la *variété pré-basique (pre-basic variety)* dans laquelle les énoncés des apprenant/es ont une structure essentiellement nominale, sans verbe énoncé (*Nominal Utterance Organisation*)
- la deuxième variété, la variété de base (*basic variety*), est caractérisée par l'apparition du verbe dans l'énoncé, mais le verbe ne semble pas être fléchi (*Infinite Utterance Organisation*)
- finalement, la variété la plus proche des normes de la langue cible est la *variété post-basique (post-basic variety)* dans laquelle les apprenant/es produisent des énoncés contenant un verbe fini (*Finite Utterance Organisation*) ; les apprenant/es produisent en outre différentes formes morphologiques, qui sont de plus en plus appropriées à la fonction visée.

La variété de base est une variété simple et versatile tout en étant hautement efficace pour la plupart des objectifs communicatifs (Klein et Perdue, 1997 : 303). Comme on vient de le voir, elle est caractérisée par l'introduction du verbe dans les énoncés. Mais les verbes produits ne sont pas conjugués, "[...] there is no trace of *finite* verbs, in whatever function" (Klein & Perdue, 1997 : 314 – leur soulignement ; Perdue *et al.* 2002, 859). Il s'agit plutôt d'une forme de *base*, par exemple une forme de l'infinitif ou une forme racine, qui est utilisée de manière généralisée, une forme invariable (Klein *et al.*, 1995 : 265 ; Noyau *et al.*, 1995 : 202 ; Klein & Perdue, 1997 : 311).

Mais l'absence de morphologie verbale n'empêche pas l'apprenant/e d'exprimer des relations qui sont marquées de manière morphologique dans la langue cible, comme par exemple la temporalité : "Learners' expression of temporality therefore relies on other, lexical and discourse means [...]" (Noyau *et al.*, 1995 : 205). Pour cette raison, les tenants de l'approche fonctionnelle soulignent la différence entre la *finitude morphologique* du stade au-delà de la variété de base et la *finitude sémantique* présente dès le début (Gretsch & Perdue, 2007 ; Perdue *et al.*, 2002), un point qui sera repris au chapitre suivant.

3.2. Stades de développement en français LE/L2

Bartning et Schlyter (2004) s'inspirent entre autres des travaux de l'approche fonctionnelle pour proposer des itinéraires d'acquisition pour le français L2/LE au niveau de différents phénomènes morphosyntaxiques, dont la morphologie verbale. Par itinéraire (ou parcours), elles entendent que les éléments d'une langue cible sont acquis dans un certain ordre "indépendamment de l'ordre dans lequel ces phénomènes sont enseignés ou quelle que soit la manière dont la langue est acquise, acquisition guidée ou non-guidée" (Bartning & Schlyter, 2004 : 281). Les itinéraires qu'elles proposent se fondent sur les résultats empiriques de travaux menés sur l'acquisition du français par des suédophones adultes à l'oral⁶.

Elles font l'hypothèse d'un itinéraire qui comporte six stades, du stade initial au stade avancé, basé sur un certain nombre de critères linguistiques, dont le développement de la finitude, le système temporel et modal (temps, mode, aspect), le développement de la négation, les constituants de la phrase nominale

⁶ Ågren (2008) et Granfeldt (2007) ont montré que les stades pouvaient aussi être utilisés pour décrire les itinéraires d'acquisition à l'écrit.

(pronoms, genre, accord) et la subordination (Bartning et Schlyter, 2004 : 283). Les principales caractéristiques des trois premiers stades ainsi que des stades avancés sont résumées ci-dessous (pour une citation complète de la description des stades 1 à 3 voir l'annexe 1).

- Le stade 1 correspond *grosso modo* à la variété pré-basique du corpus ESF “ainsi qu’à la transition vers la *variété basique*” (Bartning & Schlyter, 2004 : 295 – leur soulignement). Au niveau des verbes, les apprenant/es produisent des formes courtes et des formes longues, mais il n’y a pas d’opposition entre les contextes finis et non-finis, ni entre les différentes personnes des formes verbales (Bartning & Schlyter, 2004 : 294). En d’autres termes, différents contextes sont exprimés par les mêmes formes.
- Le stade 2 “correspond en partie à la variété basique”. On voit l’apparition de la production de quelques formes avec un auxiliaire suivi du verbe, futur périphrastique ou passé composé. “Les formes verbales non-finies dans un contexte fini sont encore fréquentes mais le nombre de ‘formes finies courtes’ augmente” (Bartning & Schlyter, 2004 : 295).
- Au stade 3, dit intermédiaire, la langue des apprenant/es est devenue “plus systématique et régulière”, mais elle contient toujours “des surextensions et des régularisations” en désaccord avec la langue cible. Le marquage de la référence au passé et au futur se fait de plus en plus et le nombre de formes “non-finies dans des contextes finis” diminue (Bartning & Schlyter, 2004 : 295).
- Les stades 4 à 6 sont les stades dits *avancés*, durant lesquels les apprenant/es continuent de développer leur langue de plus en plus en accord avec la langue cible. Au niveau des verbes, on voit notamment le développement du plus-que-parfait, du conditionnel et du subjonctif, ce qui implique aussi un développement au niveau de la complexité syntaxique et discursive.

Bartning et Schlyter décrivent également un itinéraire détaillé pour les verbes réguliers en –er, au niveau de ce qu’elles appellent les “‘formes finies courtes’ (étudier) et les formes non-finies (étudier/é)” dans un contexte fini du présent, où fini signifie “après un sujet ouvert pronominal ou nominal” (Bartning & Schlyter, 2004 : 285).

A. Les apprenants utilisent dans différents contextes des formes invariables, ‘finies courtes’ (type *parle*) ou ‘non-finies’ (type *parler/parlé*) ;

B.-C. Il y a un développement graduel vers une distinction de telle sorte que les ‘formes finies courtes’ sont utilisées dans les contextes finis et les formes non-finies dans les contextes non-finis.

D. Quand l’apprenant a atteint un niveau assez avancé, cette distinction est faite et les formes sont utilisées en accord avec la langue cible. (Bartning et Schlyter 2004 : 285)

Selon Bartning & Schlyter, la forme courte et la forme longue des verbes réguliers en *-er* sont donc utilisées de manière invariable en début d’acquisition et finissent par être utilisées correctement dans leurs contextes finis et non-finis respectifs. Les données de Schlyter (2003) montrent que le développement complet de la finitude se fait durant les quatre premiers stades, avec 50-75% d’occurrences de formes courtes correctes au stade 1, 70-80% au stade 2, 80-90% au stade 3 et finalement 90-98% au stade 4. Ainsi l’étude des formes courtes et longues semble être un phénomène pertinent pour différencier les apprenant/es des trois premiers stades (*cf.* aussi Granfeldt, 2005).

3.3. Perspective de la grammaire générative : formes et syntaxe

Dans une perspective générativiste, la syntaxe et la production des formes sont étroitement liées. La syntaxe comprend deux grandes catégories : d’un côté les catégories lexicales qui comprennent les mots de contenu tels que les noms, les verbes et les adjectifs et de l’autre les catégories fonctionnelles représentées par les déterminants, les temps et les flexions (Herschensohn, 2000 : xi – préface). L’acquisition d’une langue implique la maîtrise des traits morphologiques et la capacité de relier ces traits aux catégories fonctionnelles tout en élargissant la connaissance du lexique (Herschensohn, 2000 : 6).

Dans l’approche minimaliste (Chomsky, 1995), on considère le contrôle d’un paramètre en L2 comme étant fonction directe de la maîtrise des traits d’une catégorie fonctionnelle donnée :

More recent research in the Minimalist Program looks at syntax as a function of morphological features of lexical items, and thus it sees the control of an L2 parameter as a direct function of the mastery of the features of a given functional category. (Herschensohn, 2000 : 80)

Pour le français, on fait l'hypothèse d'un paramètre de la *montée du verbe* (*Verb Raising Parameter*) selon lequel le verbe conjugué doit monter au-delà de la position hiérarchique de la négation et des adverbes de phrase. Cette montée est liée au fait que le trait verbal *temps* (*Tense*) est dit *fort* de sorte qu'il force la montée du verbe (Herschensohn, 2000 : 74 ; Prévost, 2007 : 53-54). La figure 3-1 illustre la montée du verbe *manger* au-dessus de l'adverbe de phrase *souvent*.

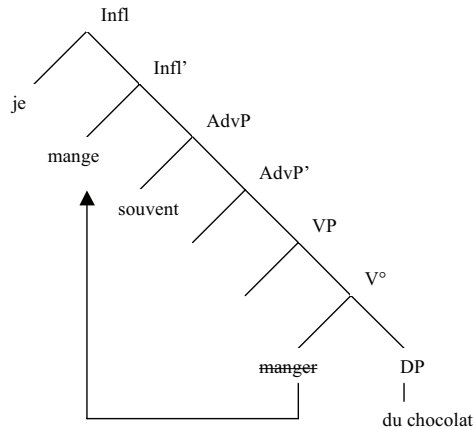


Figure 3-1 : Illustration simplifiée de la montée du verbe en français

Au niveau de la finitude, on s'attend aux deux principaux cas suivants en français standard :

- la production du verbe lexical (non-fini) à l'infinitif après une préposition ou un verbe modal comme dans *je veux manger du chocolat*. Dans ce cas, le verbe reste dans la phrase verbale VP, le point d'insertion lexicale du verbe.
- la production d'un verbe lexical (fini) conjugué, par exemple une forme courte du présent comme dans *je mange*. Dans ce cas, le verbe est monté dans *Inflection* comme illustré dans la figure 3-1 ci-dessus.

Chez les enfants, qui apprennent leur(s) L1, la maîtrise de la morphologie et des mouvements syntaxiques se développe parallèlement et la mise en place du paramètre de la montée du verbe en français se fait à peu près en même temps que l'enfant commence à conjuguer les verbes (Herschensohn, 2001 : 275 ;

Meisel, 1997⁷). Il semble que les enfants passent par une période dite d'infinitif optionnel (*Optional Infinitive*). Durant cette phase, on trouve des phrases contenant des infinitifs racine (*root infinitives*)⁸, mais où le verbe est toujours à droite de la négation comme on peut le voir dans l'exemple (3-1) présenté et glosé par Herschensohn (2001 : 276).

- (3-1) Pas manger la poupée.
'La poupée ne mange pas.'

En même temps l'enfant produit des phrases contenant des verbes fléchis (*Moi, je tousse encore, Il est pas là* – exemples de Herschensohn, 2001 : 276) dans lesquelles le verbe est monté et où il y a projection des catégories fonctionnelles supérieures (temps et accord). Dans cette période d'infinitif optionnel, les phrases contenant un verbe conjugué contiennent la morphologie verbale du temps visé alors que les phrases contenant des infinitifs racine sont en référence au présent, passé ou futur (Herschensohn, 2001 : 276-277). La *variation* que l'on trouve chez les enfants en L1 est donc différente de ce que l'on trouve chez les adultes en LE/L2. En effet, chez les enfants, on ne trouve pas de phrases du type *il dansE* ou *les hommes va parle* comme dans l'exemple (1-1) en introduction.

La variation dans l'utilisation des formes dans les contextes du présent et de l'infinitif comme présentée dans le tableau 1-1, propre aux apprenant/es adultes, pose la question de l'accès de ceux-ci à la grammaire universelle, ainsi que celle de l'état initial de leur grammaire (*cf.* White, 2003 pour une vue d'ensemble). La production de formes longues dans un contexte pour une forme courte du présent (*il dansE*) pose en effet la question de savoir si le verbe est monté dans *Inflection* ou s'il est resté dans la phrase verbale, puisque les traits morphologiques ne correspondent pas aux catégories fonctionnelles impliquées (temps, accord).

Prévost et White (2000 a et b) ont étudié la syntaxe des apprenant/es adultes à partir de "l'Hypothèse de la flexion superficielle manquante" (termes de Prévost, 2007 : 57 ; *Missing Surface Inflection Hypothesis – MSIH*) qui présuppose que la syntaxe et la morphologie sont séparées. Ainsi, l'acquisition

⁷ "[...] verb placement is acquired with considerable ease once the grammatical notion of [\pm finite] is available to the child [...]" (Meisel, 1997 : 228).

⁸ "In French, children use root infinitives with VP internal subjects and preverbal negation [...], indications that the verb does not raise out of the VP." (Herschensohn, 2001 : 275)

de la morphologie d'une L2/LE serait indépendante du développement de la syntaxe (Herschensohn, 2001 : 279 ; Lardiere, 1998).

L'étude de Prévost et White (2000 a et b) à partir des données d'Abdelmalek et de Zahra, deux apprenants arabophones du Maroc du français (corpus ESF ; cf. aussi section 4.2.3.) a montré que le verbe lexical était effectivement monté dans la catégorie *Inflection* même s'il était produit à une forme apparemment non-finie (forme longue en *-E*) conformément à leur hypothèse de *MSIH*. La montée du verbe est entre autres visible dans le placement de celui-ci au-dessus de la négation *pas*, comme on peut le voir dans l'illustration simplifiée ci-dessous (3-2).

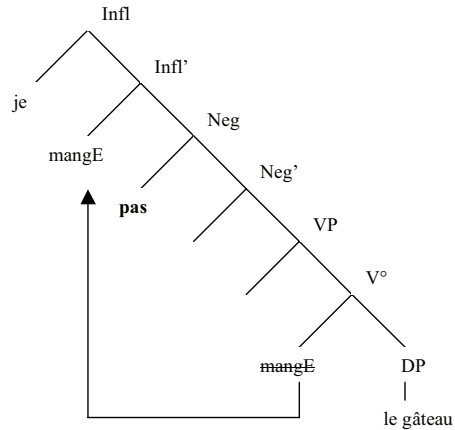


Figure 3-2 : Montée du verbe en L2 (à partir de Prévost & White, 2000b)

Du moment où le verbe est à la bonne place, l'hypothèse de la *MSIH* permet de dire que les apprenant/es LE/L2 semblent utiliser des formes non-finies pour les substituer aux formes finies. Ces formes non-finies ne sont pas des formes non-finies en tant que telles, mais servent de forme par défaut. “In such cases, then, main verbs bearing the infinitival marker are not in fact non-finite but are being used as default forms exhibiting properties of finite verbs” (Prévost & White 2000b : 109; cf. aussi Herschensohn, 2001 : 297 dans une étude longitudinale de deux apprenants anglophones du français L2).

La production de phrases du type *les hommes va parle* de l'exemple (1-1) dans lesquelles on trouve une forme courte à la place de la forme de l'infinitif, n'a pas été considérée dans ces premiers travaux dans le cadre de la *MSIH*, car leur fréquence est très faible dans le corpus ESF. La production de telles phrases

n'est pas prévue dans la perspective générativiste, la forme attendue dans la phrase verbale VP étant la forme non-finie (Prévost, 2007 : 57). En outre, de telles phrases ne sont pas produites par des enfants en L1.

Dans Prévost (2004), on trouve une analyse de données d'apprenant/es anglophones débutant/es du français qui produisent aussi bien des phrases du type *il dansE* que *les hommes va parle*. Encore une fois, l'analyse semble confirmer la *MSIH*, les apprenant/es ont connaissance des catégories fonctionnelles, mais ils ne réussissent pas à produire les formes attendues. Selon Prévost, l'utilisation de formes dites finies après une préposition ou un verbe modal est le fruit de l'influence de la L1 des apprenant/es :

The fact that some (superficially) finite forms occur in typically nonfinite positions, I contend, is due to L1 influence, whereby the lack of morphological marking in the L1 may yield the production of bare (nonfinite) forms in the L2. (Prévost, 2004 : 171)

Le fait que l'anglais ne marque pas l'opposition entre la forme finie et la forme non-finie (*I speak, to speak*) aurait donc une influence sur la production des formes dans les contextes de l'infinitif.

Une nouvelle étude, cette fois comparant les données d'apprenant/es anglophones et hispanophones a confirmé cette hypothèse de l'influence de la L1 (Prévost, 2007). Les deux groupes d'apprenant/es ont produit des formes non-finies en position finie (*il dansE*) alors que la production de formes finies en position non-finie (*les hommes va parle*) est le fait des apprenant/es anglophones. Prévost (2007 : 72) conclut que "la L1 semble jouer un rôle dans la production des formes non-tensées mais pas dans celle des formes tensées."

En résumé, ces études à partir d'une perspective générativiste de la production verbale d'apprenant/es L2 du français ont confirmé les prédictions de la *MSIH* selon laquelle le développement de la morphologie et celui de la syntaxe se font de manière indépendante. Les apprenant/es connaissent les catégories fonctionnelles, mais ne réussissent pas toujours à produire les formes morphologiques attendues.

3.4. Synthèse et discussion

Ce chapitre a présenté un certain nombre de recherches antérieures qui ont étudié le phénomène de la variation des formes à partir des catégories de la finitude.

- Les travaux de l'approche fonctionnelle (Klein & Perdue, 1997 ; Perdue *et al.*, 2002) ont montré que la variation des formes était une caractéristique des apprenant/es de la *variété de base* : les verbes lexicaux commencent à apparaître dans les énoncés, mais ils sont produits à une forme de base, une forme non conjuguée, sans marques morphologiques pour le temps et l'accord.
- Bartning & Schlyter (2004) ont présenté un itinéraire de développement de l'utilisation de la forme courte et de la forme longue des verbes aux premiers stades de l'acquisition du français montrant que les premières formes produites étaient utilisées de manière invariable. Ce n'est qu'à un niveau avancé que les formes finissent par être utilisées correctement au niveau de la finitude.
- Les travaux dans une perspective générative à partir des présupposés de la *Missing Surface Inflection Hypothesis* ont montré que la variation des formes chez les apprenant/es adultes n'était pas due à un problème de déficit au niveau de la syntaxe ; elle reflète plutôt le fait que le développement de la morphologie et de la syntaxe se fait de manière indépendante chez les apprenant/es adultes.

L'étude de la variation des formes à partir de la finitude et de la syntaxe ne permet donc pas d'observer de systématisme dans la production des apprenant/es. Comme on l'a vu dans l'introduction (tableau 1-1), les apprenant/es produisent des formes courtes et longues dans les contextes pour le présent et des formes longues et courtes dans les contextes pour une forme de l'infinitif. Il est donc nécessaire de poursuivre la recherche sur d'autres facteurs.

Comme on l'a vu, Prévost (2004) propose différentes explications selon l'erreur à laquelle on a à faire. Ces propositions seront discutées dans ce qui suit.

L'une des propositions de Prévost est l'influence de la L1. Il mentionne les travaux de Goad *et al.* (2003) qui proposent que les apprenant/es transfèrent les propriétés de phonologie prosodique de leur L1 en L2. Cette hypothèse s'appuie sur des données d'apprenant/es de l'anglais ayant le chinois mandarin comme L1. En anglais la flexion doit être ajoutée au mot prosodique, donc à la fin du mot, ce qui n'est pas permis en mandarin où la flexion doit être incorporée au radical. Certains apprenant/es ont simplement omis la flexion alors que d'autres l'ont mise lorsqu'il était possible de l'intégrer au radical, comme par exemple dans le cas de verbes irréguliers (Goad *et al.*, 2003 : 255).

Comme le souligne Prévost, cette hypothèse est peu probable dans le cas où les flexions sont ajoutées à la fin du mot prosodique dans la langue source et la langue cible. Le fait que des apprenant/es anglophones par exemple ont parfois de la peine à produire des formes de l’infinitif n’est pas le fait de la difficulté de produire ces formes en soi – elles sont notamment produites dans des contextes pour une forme finie – le problème réside dans la production de la bonne forme dans le bon contexte (Prévost, 2007 : 76).

Selon Prévost, l’influence de la L1 serait tout particulièrement visible dans la production d’énoncés de type *les hommes va parle*, dans lesquels la forme courte est produite dans un contexte pour une forme longue. Selon lui, ce genre d’erreur est une caractéristique des apprenant/es dont la L1 ne marque pas la différence entre l’infinitif et les formes finies, comme l’anglais. Mais cette différence est marquée en suédois (Stroh-Wollin, 1998 : 30 ; Florby & Fält, 1998 : 21) comme le montre le tableau 3-1 ci-dessous. Ce tableau présente les différentes formes verbales en suédois pour un verbe régulier, *titta* (*regarder*), et un verbe irrégulier *sova* (*dormir*) à l’oral. Les formes de ces exemples peuvent *grosso modo* se prononcer comme en français mais en prononçant toutes les lettres. Les parenthèses indiquent que la finale peut tomber à l’oral. La dernière ligne présente les formes correspondantes en français parlé et écrit.

Tableau 3-1 : Morphologie des formes verbales en suédois à l’oral

	présent	imparfait	infinitif	participe passé
suédois régulier	titta(r)	titta(de)	titta	titta(t)
suédois irrégulier	sove(r)	sov	sova	sovi(t)
français oral – écrit	parle / parlE – parlez	parlE – parlais	parlE – parler	parlE – parlé

Le suédois présente donc une distribution morphologique nette des formes entre le présent et l’infinitif, même si la prononciation de certaines finales peut tomber à l’oral. On peut en outre mentionner que le verbe fini monte au-dessus de la négation en suédois (Platzack, 1998 : 89ss ; 2004) comme en français et contrairement à l’anglais. La montée du verbe est donc déjà établie chez les apprenant/es suédophones. Pourtant, comme on l’a vu, Bartning & Schlyter (2004) ont constaté la production de formes finies et non-finies indépendamment du contexte \pm fini chez les apprenant/es suédophones. L’étude de Gunnarsson (2006) sur des apprenant/es suédophones, mais cette fois à l’écrit, a elle aussi montré que les apprenant/es avaient de la difficulté à produire les formes adéquates, dans les contextes finis comme dans les non-finis. Elle a même pu constater que le contexte non-fini posait plus de problèmes que le

contexte fini, i.e. les apprenant/es du corpus Gunnarsson ont produit plus d'énoncés du type *les hommes va parle* que *il dansE*.

On peut par contre noter en passant que le suédois n'a plus de morphologie pour différencier les personnes. L'utilisation d'une seule forme pour toutes les personnes pourrait être un effet de transfert. D'un autre côté, comme on l'a vu dans la présentation du paradigme des verbes réguliers en *-er*, il n'y a pas d'accord sujet – verbe à l'oral.

On peut ainsi conclure que l'influence de la L1 ne semble pas être une explication qui mérite d'être poursuivie pour traiter de la variation des formes par des apprenant/es suédophones, la langue suédoise tout comme les données d'apprenant/es ne présentant aucun argument en faveur d'une telle hypothèse.

On pourrait par contre penser à une influence de la première L2 des apprenant/es suédophones, l'anglais. Une hypothèse intéressante, mais qui ne sera pas retenue ici.

Pour ce qui est de la production d'énoncés du type *il dansE*, donc de la production erronée d'une forme longue dans un contexte pour une forme courte du présent, les tenants de la perspective générativiste ont suggéré que les apprenant/es aient recours à des formes *par défaut* (Prévost & White, 2000 a et b ; Herschensohn, 2001). Prévost (2007) propose que le recours à des formes par défaut soit lié à la charge cognitive de traitement des formes conjuguées.

Dans le domaine tense, les apprenants doivent prendre des décisions concernant la personne et le nombre de la terminaison, de manière à ce qu'il y ait correspondance avec les propriétés du sujet. Cette opération peut être très coûteuse sur le plan computationnel en production orale pour les apprenants, surtout dans les premiers stades d'acquisition. Cette difficulté se traduirait par l'emploi de formes défauts (sic). (Prévost, 2007 : 73)

Selon Prévost (2007 : 73) la forme longue (de l'infinitif) est “plus facilement accessible” car chaque verbe ne possède qu'une forme de l'infinitif et qu'il n'y a pas de décisions à prendre au niveau de la personne et du nombre. C'est aussi pour cette raison que la forme courte remplace plus rarement la forme longue dans un contexte pour l'infinitif, toujours selon Prévost.

Pourtant comme on l'a vu dans le chapitre précédent sur la présentation des verbes réguliers en *-er*, la forme courte et la forme longue couvrent la plupart des formes du paradigme de ces verbes. La production de la forme courte du présent ne demande pas plus de décisions que celle de la forme longue de l'infinitif, puisqu'il n'y a pratiquement pas d'accord sujet – verbe à l'oral. On

peut donc supposer que la facilité d'accès est la même pour les deux formes à l'oral.

La question de savoir pourquoi l'apprenant/e choisit plutôt une forme qu'une autre reste donc ouverte. Dans ces travaux antérieurs, l'étude de la variation a été faite à partir du *contexte syntaxique* suggérant que la variation se fasse de manière aléatoire, l'apprenant/e utilisant une fois une forme une fois une autre. Dans le présent travail la question sera posée à partir des *verbes*. Du moment où les verbes semblent être produits à une forme de base ou à des formes invariables comme le suggèrent les tenants de l'approche fonctionnelle et Bartning & Schlyter (2004), on peut partir de l'idée que les apprenant/es produisent essentiellement une forme par verbe et que cette forme est influencée par l'input, soit parce qu'elle correspond à l'aspect lexical du verbe soit parce qu'il s'agit de la forme la plus fréquente dans l'input. C'est ce qui sera étudié dans le présent travail.

4. Variation des formes et aspect lexical

L'objectif de ce chapitre est de présenter les études qui proposent que la production des formes verbales par les apprenant/es soit influencée par l'aspect lexical des verbes (*Aktionsart*). La section 4.1. introduira tout d'abord les principaux concepts théoriques autour de l'aspect lexical et l'expression de l'aspect lexical en L1. La section 4.2 sera consacrée à l'influence de l'aspect lexical en LE/L2. Elle commencera par une présentation de l'hypothèse de l'aspect proposée par Andersen (1991, 2002) selon laquelle l'aspect lexical permet de prédire le développement de la morphologie du passé. Cette section présentera également quelques études empiriques sur le développement de la morphologie verbale en français à partir de l'aspect lexical, dont celle du projet ESF. La dernière section, 4.3., sera consacrée à un résumé et une discussion de l'influence de l'aspect lexical sur les apprenant/es du français LE/L2 ainsi qu'à l'esquisse d'une hypothèse à partir de l'aspect lexical.

4.1. Présentation de l'aspect lexical

Dans le sens large du terme, le domaine de l'aspect comprend trois niveaux. Le niveau le plus proche du lexique est ce que l'on appelle l'aspect lexical (*Aktionsart*) qui concerne les caractéristiques inhérentes d'un verbe (*savoir, lire, partir*) ou d'une situation (*gagner une course, construire une maison*) et qui sera présenté plus en détail dans cette section. L'aspect lexical se distingue de l'aspect grammatical qui comprend l'opposition perfectif vs imperfectif (*viewpoint* chez Smith, 1991). Cette opposition est une opposition grammaticale, ce qui veut dire qu'elle est exprimée par des marques morphologiques dans les langues, en français par exemple par l'imparfait pour l'imperfectif et le passé composé (ou à l'écrit le passé simple) pour le perfectif (Bonnotte, 2008). Le troisième niveau concerne le temps : présent, passé ou futur. En français, le perfectif est uniquement relié au passé, exprimé par le passé composé à l'oral, alors que l'imperfectif est relié au passé comme au non passé exprimé par les formes du présent et de l'imparfait (Bybee & Dahl, 1989 ; Giacalone-Ramat, 2002 : 226).

L'aspect lexical, *Aktionsart*⁹, concerne le type de caractéristique temporelle inhérente (Sasse, 2002 : 203), tel que statique, dynamique, duratif, télique, ponctuel etc. Andersen parle d'aspect lexical inhérent qu'il considère comme étant les propriétés sémantiques des verbes (2002 : 80). En principe l'aspect lexical concerne la situation, mais c'est le verbe qui la représente :

More specifically, it is the situation or event that has inherent aspect and the verb or predicate merely represents that situation or event and the aspect inherent in it. (Andersen, 1993 : 313)

Contrairement à l'aspect grammatical, l'aspect lexical est une catégorie au-delà des langues spécifiques. Il est basé sur les aptitudes cognitives de l'être humain comme le dit Smith :

I will assume that aspectual categories are not language dependent, but are based in human cognitive abilities. People distinguish between the basic situation types on the basis of their perceptual and cognitive faculties, [...]. Human beings make aspectual distinctions automatically. States, activities, etc. impress themselves on our notice, organizing the way we see the world. (Smith, 1991 : xvii)

Selon Smith, l'aspect lexical est donc une disposition cognitive innée et automatique qui permet à l'être humain d'organiser sa vision du monde.

Parler de l'influence de l'aspect lexical dans la production des verbes par les apprenant/es présuppose que l'on puisse catégoriser les verbes selon les propriétés temporelles qui les caractérisent. Pour cela on utilise souvent les catégories de situations qui ont été systématisées par Vendler (1957) : les états, les activités, les accomplissements et les achèvements. Ces situations sont caractérisées par la dynamicit , la durativit  et la t licit  (Smith, 1991 : 6; cf. aussi Andersen, 1991, 310ss ; Andersen, 1993 : 314ss ; Bardovi-Harlig, 2000 : 213ss)¹⁰. Le tableau 4-1 ci-dessous pr sente ces cat gories ainsi que la distribution de l'aspect grammatical et des formes correspondantes en fran ais parl .

⁹ Les termes utilis s pour parler de l'aspect lexical sont nombreux : *caract re aspectuel*, *classe de verbes*, *sens inh rent du verbe*, *type aspectuel*, *sorte d' v nement*, etc. ( num r s par Sasse, 2002 : 203), *type de situation* (Smith, 1991), *mode d'action* ou *actionalit * (en fran ais).

¹⁰ Je pr sente ici les quatre situations traditionnellement utilis es dans le cadre de l'hypoth se de l'aspect. Pour une description d'autres classes de verbes voir Christensen (  para tre).

Tableau 4-1 : Catégories liées à l'aspect lexical

aspect lexical	statique	activités	télique	télique ponctuel
Vendler	états	activités	accomplissements	achèvements
dynamacité	-	+	+	+
durativité	+	+	+	-
télicité	-	-	+	+
aspect grammatical	imperfectif non marqué	imperfectif vs perfectif	perfectif non-marqué	perfectif non-marqué
formes en français	présent, imparfait	présent, imparfait, passé composé	passé composé	passé composé
exemples	avoir, vouloir, aimer	travailler, lire, marcher	construire une maison, écrire un livre, grandir	tomber, partir, trouver (qc), gagner la course

Les quatre catégories de Vendler correspondent aux trois oppositions majeures au niveau de l'aspect lexical : les verbes statiques vs les verbes dynamiques, les verbes téliques ponctuels vs les verbes duratifs et finalement les verbes atéliques (d'états et d'activités) vs les verbes téliques (d'accomplissement et d'achèvement).

Bien qu'ils soient distincts, l'aspect grammatical et l'aspect lexical sont étroitement liés. Selon Sasse (2002 : 204ss) dans son approche typologique, il y a un lien intime entre l'aspect lexical statique et l'imperfectif d'une part et entre l'aspect lexical télique/ponctuel et le perfectif dans le sens où ces liens sont non-marqués (*unmarked*). L'aspect perfectif implique une limitation ou un terme contrairement à l'aspect imperfectif qui n'en présente pas (Riegel, *et al.*, 1994 : 293ss ; Bybee & Dahl, 1989 : 55).

1. Les verbes statiques décrivent un état, une situation stable qui existe sans apport d'énergie et se termine au moment où quelque chose change l'état (Bonnotte *et al.*, 1993 ; Andersen, 1991, 1993, 2002 ; Vendler, 1957). Ainsi, si *j'aime la musique jazz* je l'aimerai jusqu'au jour où quelque chose arrive qui fasse que je ne l'aime plus. Le verbe *aimer* ne porte donc pas de limitation inhérente au niveau lexical. L'absence de limite est inhérente aux états comme à l'imperfectif.
2. Les verbes d'activités décrivent des activités avec une durée inhérente. Pour ces verbes, on peut utiliser les points de vue perfectif et imperfectif (Sasse, 2002 : 206). On peut décrire une activité en cours sans qu'une limitation soit impliquée (*je travaillais*) ou présenter une version dans laquelle l'activité est entourée de limites (*j'ai travaillé, après j'ai fait du*

sport...)¹¹. Les verbes d'activités n'ont donc pas de limitation inhérente au verbe en tant que lexème comme les verbes téliques, la présence ou non d'une limitation est fournie par l'aspect grammatical.

3. Les verbes téliques se divisent en deux groupes de verbes selon que la situation comporte une durée inhérente comme dans *construire une maison* ou non comme dans *gagner la course*. Ces verbes contiennent tous de manière inhérente la fin naturelle de l'événement qu'ils décrivent, comme par exemple dans *j'ai acheté une maison* ou *le vase est tombé*. La présence d'une limitation peut se tester en introduisant une pause virtuelle dans l'événement : si je suis en train d'acheter une maison et que je demande quelques jours de réflexion, je n'ai pas 'acheté de maison', car l'événement exprimé directement par le verbe n'est pas encore accompli.

Comme on le verra au chapitre 7, il existe un certain nombre de tests qui permettent de répartir les verbes entre les différentes catégories aspectuelles. Mais un verbe peut changer de catégorie selon l'utilisation que l'on en fait. Sasse parle de *recatégorisation* : un verbe d'activité peut par exemple être recatégorisé en verbe d'accomplissement en ajoutant du matériel tel qu'une phrase prépositionnelle ou un syntagme nominal. Il peut aussi y avoir recatégorisation par la morphologie aspectuelle, par exemple l'utilisation du progressif recatégorise un événement en état (Sasse, 2002 : 215 ; cf. aussi les opérations de *coercion* chez de DeSwart, 1998). De fait, la différence entre les verbes d'activités et les verbes d'accomplissement par exemple se fait essentiellement au niveau de l'objet, délimité ou non (*écrire des lettres* vs *écrire une lettre* ; cf. Kihlstedt, 1998 : 78-79 ; Platzack, 1979). La catégorie des verbes d'accomplissements ne comporte donc pas ou peu de verbes sans matériel supplémentaire.

Les travaux de Bonnotte et Fayol (e.a. Bonnotte *et al.*, 1991 ; 1993 ; Bonnotte & Fayol, 2000 ; 2001 ; Bonnotte, 2008) ont étudié les propriétés sémantiques des situations aspectuelles ("types de procès" dans leur vocabulaire). Dans différentes études sur le français L1, ils ont pu constater que l'effet de l'aspect lexical sur la sélection des formes se mettait en place au cours du développement de l'enfant. Ce n'est qu'à l'âge adulte que la sélection de temps (des formes) selon l'aspect lexical est en place (Bonnotte & Fayol, 2001 : 180).

¹¹ Ces exemples sont les miens. Sasse (2002 : 206) utilise l'exemple "*I was working*" pour l'absence de limitations et "*I worked (and then...)*" pour une activité délimitée.

Les travaux de Bonnotte et Fayol, à partir de différentes méthodes, ont aussi montré que les sous-catégories \pm duratif et \pm résultatif semblaient être des propriétés particulièrement importantes des situations aspectuelles. Les propriétés de résultativité (télicité) et de durativité “permettent de faire des prédictions relatives à l’emploi de temps grammaticaux du passé en français [...]” (Bonnotte & Fayol, 2001 : 195). Les sous-catégories de +duratif et –résultatifs sont celles qui apparaissent le plus souvent pour les verbes d’activités par exemple ; on peut ainsi considérer qu’elles définissent le prototype des activités (Bonnotte & Fayol, 2001 : 184). Comme le dit Bonnotte (2008 : 204), un concept verbal peut varier d’un contexte d’utilisation à l’autre, mais il est possible d’extraire un sens primaire (*primary sens*) et de le distinguer d’un sens secondaire, qui peut être activé dans des contextes spécifiques. L’étude de Bonnotte (2008) avec des adultes natifs du français à partir de verbes réguliers en *-er* par les méthodes de codage graphique et de *priming* a confirmé les travaux antérieurs de Bonnotte & Fayol :

Durative, non-resultative verbs [états, activités] attract more frequently imperfective tenses and less frequently perfective tenses than non-durative, resultative verbs [achievements]. (Bonnotte, 2008 : 211)

Ainsi, selon Bonnotte (2008 : 211), les propriétés de classification des verbes peuvent être interprétées comme des traits sémantiques.

Ces études indiquent donc que les adultes ont mis en place un système complet de sélection préférentielle des temps à partir de l’aspect lexical.

L’étude de Bronckart et Sinclair (1973) à partir de données de 74 enfants de 2,11 à 8,7 ans ayant le français comme L1 a observé l’utilisation du passé composé, de l’imparfait et du présent dans la description d’événements qu’ils appellent perfectifs et imperfectifs. Leurs données ont montré des utilisations préférentielles du passé composé avec les événements perfectifs (78%), selon leurs termes, et du présent avec les événements imperfectifs (76%) toujours selon leurs termes, alors que l’imparfait était beaucoup plus rare. Les enfants devaient décrire des événements présentés à l’aide de jouets, le présent utilisé était donc à référence au présent. Ils ont également pu voir que les actions menant à un résultat étaient décrites avec le passé composé alors que les actions sans résultat étaient décrites au présent.

The choice of tenses used in the descriptions shows that at all ages the subjects take into account the difference between perfective, aperfective [*shouting*] and imperfective events. Actions that obtain a clear result are mostly described in the *passé composé*,

actions without an intrinsic aim are described in the *présent* or the *passé composé*, and actions that do not lead to any result are described in the *présent*. (Bronckart & Sinclair, 1973 : 125 – leurs soulignements)

Les enfants apprenant le français comme L1 semblent donc de préférence utiliser le passé composé avec les verbes téléiques et le présent avec les verbes atéléiques. Les enfants semblent être influencés par l'aspect lexical dès leur plus jeune âge dans leur choix des temps.

Finalement, l'étude de Ferdinand sur l'acquisition du français L1 a montré une distribution des formes finies (notamment du présent) avec les verbes statiques (*non-eventive*) et les verbes non-finis avec les verbes dynamiques (*eventive*) chez les enfants dans leurs premières productions de verbes en production spontanée. Cette distribution s'appliquait même sur les différents sens du verbe *aller*, le sens statique comme auxiliaire étant toujours produit à une forme finie et le sens téléique comme verbe lexical à une forme non-finie (Ferdinand, 1996 : 85-129 ; cf. aussi Hoekstra & Hyams, 1998 : 90 qui ont montré que les verbes statiques n'apparaissent pas dans les infinitifs racines, *root infinitives*, chez les enfants).

4.2. Influence de l'aspect lexical en français LE/L2

Cette section est consacrée aux études sur l'influence de l'aspect lexical sur la production des formes par des apprenant/es adultes. Une des caractéristiques des apprenant/es adultes par rapport aux enfants est qu'ils ont déjà acquis toutes les catégories impliquées dans la production des verbes y compris les catégories de l'aspect lexical, qui, comme le rappelle Giacalone-Ramat ci-dessous, correspondent à des distinctions universelles.

We thus expect that L2 learners start their learning path endowed not only with the temporal categories of past, present and future, but also with verb types and actionality [Aktionsart, inherent lexical aspect] distinctions which should belong to the universal perceptual and cognitive faculties (Smith 1991). (Giacalone-Ramat, 2002 : 242, note 1)

Les principales études sur l'influence de l'aspect lexical proviennent de l'hypothèse de l'aspect qui sera présentée à la section suivante, mais, comme on le verra plus loin, les travaux dans le cadre du projet ESF, dans une approche fonctionnelle, ont également étudié cette influence. Comme le montre la citation ci-dessous, selon les tenants de l'approche fonctionnelle, l'expression de

l'aspect lexical se ferait dès l'introduction du verbe dans l'énoncé, même s'il est produit à une forme non-conjuguée, de base (cette approche a été présentée à la section 3.1.).

The adverbial and discourse-organizational means for expressing temporality are already present in NUO but are enriched at IUO by the fact that the verb-argument structure allows the consistent expression of inherent lexical aspect (*Aktionsart*). (Perdue *et al.*, 2002 : 860)

4.2.1. Présentation de l'hypothèse de l'aspect

L'hypothèse de l'aspect¹² (Andersen, 1991, 1993, 2002) propose que le développement de la morphologie du passé en LE/L2 se fasse à partir de l'aspect inhérent des verbes : “in the beginning stages of language acquisition *only inherent aspectual* distinctions are encoded by verbal morphology, not tense or grammatical aspect” (Andersen, 1991 : 307 – son soulignement). En d'autres termes, les apprenant/es débutant/es utilisent des formes pour exprimer les distinctions au niveau de l'aspect lexical. Cette hypothèse comporte trois parties, dont une sur le développement du progressif qui ne sera pas considérée ici. La formulation ci-après est faite à partir de celle d'Andersen (2002 : 79).

1. Les apprenant/es marquent d'abord le passé sur les verbes d'achèvements et d'accomplissements, puis sur les activités et ensuite sur les états.
2. Chez les apprenant/es de langues qui encodent la distinction perfectif *vs* imperfectif (comme le français), le passé imperfectif (en français l'imparfait) apparaît plus tard que le passé perfectif (en français le passé composé). L'hypothèse propose aussi que le marquage du passé imperfectif commence avec les verbes d'états et d'activités, puis s'étende aux verbes d'accomplissement et d'achèvement.

Selon cette hypothèse, on devrait donc d'abord voir des marques du passé avec les verbes téliques et seulement à la fin sur les verbes d'états. En outre, le passé imperfectif serait tardif, mais commencerait par les verbes d'état. Cela implique que les verbes pour lesquels on ne voit pas de marque du passé sont produits à une autre forme.

¹² Dans Andersen (1991), l'hypothèse est appelée “the defective tense hypothesis”.

Andersen (1990) propose un certain nombre de principes cognitifs qui permettent d'expliquer le processus de sélection des formes pour exprimer l'aspect lexical par les apprenant/es. Voici un résumé de la version présentée dans Andersen (2002 : 90-91).

1. Selon le principe de correspondance univoque (*One to one principle of interlanguage construction*) les apprenant/es présupposent que toute nouvelle forme (mot ou marque grammaticale – *gram*¹³) perçue puis utilisée n'a qu'un seul sens. Il faut du temps pour que les apprenant/es découvrent d'autres sens.
2. Les apprenant/es (comme les natifs probablement) ont plus naturellement et plus facilement accès au sens *prototypique* d'un verbe, quelle que soit la marque grammaticale qu'ils y associent. Les sens non-prototypiques sont moins accessibles, avec un décroissement graduel selon l'augmentation de la distance entre l'occurrence et le noyau prototypique.
3. Les premières marques grammaticales utilisées par les apprenant/es sont les marques ayant la plus haute pertinence (*relevance*) par rapport au sens du verbe¹⁴. "In choosing to mark a verb with a gram, learners choose a gram with the greatest relevance to the meaning of the verb" (Andersen, 2002 : 91).
4. Parmi les marques grammaticales disponibles dans l'input, les apprenant/es vont choisir la marque la plus congruente (*congruent*) entre le sens de la marque grammaticale et celle du verbe : en espagnol, le prétérit avec les verbes d'événements, le présent et l'imparfait avec les états.

Les apprenant/es soumettent la langue cible à un processus de simplification en réduisant les possibilités et en concentrant leur attention sur ce qui est prototypique, pertinent et congruent au niveau du sens.

L'hypothèse de l'aspect a été vérifiée sur différents corpus pour différentes langues cibles principalement l'anglais et l'espagnol LE/L2 (cf. Bardovi-Harlig,

¹³ Le terme de *gram* provient de Bybee & Dahl (1989). Il est ici traduit par *marque grammaticale*. Il fait référence à une flexion, un auxiliaire ou une construction périphrastique (Andersen, 2002 : 90).

¹⁴ Ce principe est inspiré de Bybee (1985) qui propose que l'aspect soit marqué en premier en raison de sa pertinence au niveau de l'item lexical auquel il est attaché suivi du temps et seulement à la fin l'accord sujet-verbe (Andersen, 1990 : 60 ; cf. aussi Bybee, 1991 : 80 avec une échelle hiérarchique complète).

2000 : 206ss pour une vue d'ensemble). La section suivante présentera quelques études qui ont été faites pour le français LE/L2.

4.2.2. Études à partir de l'hypothèse de l'aspect et français LE/L2

L'objectif de cette section est d'une part de voir comment l'hypothèse de l'aspect se vérifie pour le français, puis de noter le rôle de l'aspect lexical en français LE/L2. Les points communs entre les différentes études pourraient révéler les systématiquités liées à l'aspect lexical dans la production des formes verbales par les apprenant/es de français LE/L2.

a) Kaplan (1987)

Les données de Kaplan (1987) proviennent d'apprenant/es adultes en milieu scolaire, d'un niveau initial à intermédiaire, ayant l'anglais comme L1. L'objectif de son étude est de vérifier l'impact de l'enseignement sur la production des formes. Elle vérifie avant tout la première partie de la deuxième section de l'hypothèse de l'aspect, c'est-à-dire l'ordre d'apparition des temps du passé.

- Bien que l'imparfait soit introduit avant le passé composé dans les manuels utilisés par ces apprenant/es, les données provenant de brèves interviews semi-structurées indiquent que le passé composé apparaît plus tôt, plus souvent et plus correctement que l'imparfait.
- Kaplan (1987 : 54) constate en outre que l'imparfait est sous-représenté dans ses données d'apprenant/es. En effet, dans 82% des cas d'erreurs dans les contextes pour l'imparfait, c'est la forme du présent qui a été produite (les autres cas étant des formes du passé composé).

Kaplan propose trois principales explications pour ces résultats :

- L'utilisation d'un auxiliaire pour le passé est saillante de sorte que l'on trouve une sur-utilisation du passé composé dans les contextes pour l'imparfait.
- La fréquence dans l'input semble être la raison pour une utilisation plus correcte du passé composé et l'absence de marques de l'imparfait. Le passé composé semble plus fréquemment utilisé par les enseignants et élicité auprès des élèves que l'imparfait.
- La complexité sémantique de l'imparfait le rend plus difficile à utiliser.

- Finalement, elle souligne le fait que l'utilisation du présent vs passé composé correspond à une opposition temporelle en français standard, mais qu'il pourrait s'agir pour les apprenant/es d'une opposition au niveau de l'aspect grammatical, les apprenant/es ayant recours à un paradigme réduit pour exprimer cette opposition.

b) Bergström (1997)

L'étude de Bergström a pour objectif "de voir si le mode d'action joue un rôle dans l'acquisition de la temporalité" (Bergström, 1997 : 61). Elle présente les données de 75 étudiants anglophones du français qui ont rédigé un récit à partir d'un extrait du film *Modern Times*. La même tâche a été soumise à un groupe contrôle de natifs du français. Les apprenant/es ont été répartis en 3 groupes selon le nombre de semestres de français. Le nombre et le pourcentage de formes du passé composé, de l'imparfait, du présent et de "autres" a été calculé pour chaque groupe d'apprenant/es et chaque catégorie de verbes. Les occurrences des verbes *être* et *avoir* sont incluses dans les données.

- Les formes du présent pour le passé sont utilisées par tous les apprenant/es ainsi que les natifs¹⁵, mais surtout par les plus débutants et pour les verbes d'états. Selon Bergström, l'utilisation du présent correspond à un usage par défaut qui permet d'exprimer le trait aspectuel [-perfectif] (Bergström, 1997 : 75). Selon elle, l'utilisation du présent correspond bien aux verbes d'état, "l'absence de structure temporelle/aspectuelle des verbes d'état pourrait très bien influencer l'usage plus répandu du présent [...]" (Bergström, 1997 : 77).
- Le passé composé est utilisé dès le début avec tous les verbes dynamiques par les apprenant/es, mais aussi avec les verbes d'état, alors que les locuteurs natifs l'ont surtout utilisé avec les verbes dynamiques.
- L'imparfait a rarement été produit par les apprenant/es. Il a surtout été utilisé avec les états, mais aussi avec les autres catégories de verbes. Les locuteurs natifs quant à eux ont avant tout utilisé l'imparfait avec les verbes d'état et d'activité.
- Ainsi les apprenant/es de cette étude "font une distinction dynamique/non dynamique dès le début de l'acquisition des temps du passé [...]" (Bergström, 1997 : 76). Même si l'hypothèse de l'aspect ne

¹⁵ Six des douze natifs avaient écrit tout le texte au présent. Ils ont été exclus du calcul.

se vérifie pas à tous les niveaux, l'usage des temps du passé, passé composé et imparfait, "semble être en corrélation avec le mode d'action des verbes utilisés." (Bergström, 1997 : 76).

c) Kihlstedt (1998)

L'étude de Kihlstedt (1998) est basée sur les données de quatre apprenant/es adultes suédophones en production orale ayant un niveau avancé de français, ainsi que de quatre natifs du français. Les données proviennent de 4-5 interviews par apprenant/e sur différents sujets, pour la plupart argumentatifs.

- Dans cette étude, le passé composé est utilisé avec tous les verbes. Selon Kihlstedt (1998 : 108), "l'impact du mode d'action est négligeable dans l'emploi du PC [passé composé]. Ce temps serait alors perçu par les apprenant/es comme la forme passée par excellence, indépendamment du sémantisme du verbe".
- L'imparfait est utilisé dans une large mesure avec les verbes d'état. En fait les formes *était* et *avait* correspondent à 70% des formes de l'imparfait pour les apprenant/es et 65% pour les natifs (Kihlstedt 1998 : 93).
- Les données de Kihlstedt indiquent également une préférence pour l'utilisation du présent avec les verbes d'état, alors que les apprenant/es analysés se trouvent à un stade avancé de l'acquisition du français.

Toujours est-il qu'une FdB [une forme basique au présent à référence passée¹⁶] avec des verbes d'état semble persister même à des stades avancés d'acquisition. Il semble s'agir d'une forme interlangagière de portée générale, vu que différents types d'apprenants, à différentes étapes d'acquisition, en font usage. (Kihlstedt 1998 : 155)

- Chez les natifs, l'imparfait est avant tout utilisé avec les verbes d'état et le passé composé avec tous les verbes, mais plus souvent avec les verbes téliques (Kihlstedt, 1998 : 116).
- Selon Kihlstedt, le passé composé sert donc à marquer le passé alors que pour l'utilisation de l'imparfait avec *était* et *avait*, on ne peut savoir si c'est la fréquence dans l'input ou le mode d'action qui guide les apprenant/es (Kihlstedt, 1998 : 110).

¹⁶ Kihlstedt 1998 : 45.

d) Duperron (2003)

Les données de Duperron proviennent de 46 apprenant/es intermédiaires du français ayant l'anglais L1. Les tâches comprenaient d'une part un test de jugement aspectuel, mais surtout un test à trous dans lequel les participant/es devaient choisir la forme de l'imparfait ou du passé composé (pour éviter que les contextes pour l'imparfait soient exprimés par du présent). L'utilisation du présent a donc clairement été exclue dans cette étude. Les résultats ont été analysés en fonction du pourcentage de choix corrects pour le test à trous.

- Le passé composé a été utilisé correctement avec les verbes d'accomplissement à 84%, d'achèvement à 77% et d'activité à 74% pour descendre à 55% avec les verbes d'état (Duperron, 2003 : 164ss).
- L'imparfait a été utilisé correctement avec les verbes d'activité à 74% et les verbes d'état à seulement 62% ; à 51% avec les verbes d'achèvements.
- Selon Duperron, l'hypothèse de l'aspect se vérifie donc faiblement : les utilisations du passé composé correspondent à ce qui est prédit. Par contre on ne voit pas l'émergence de l'utilisation de l'imparfait avec les verbes d'état. L'imparfait semble en effet mieux utilisé avec les activités.
- L'hypothèse se vérifie par contre sur le point du retard de la production de l'imparfait par rapport au passé composé.
- Duperron n'a pas pu observer le développement linéaire allant des états aux achèvements et vice-versa. En effet, les verbes d'activités étaient les premiers à avoir été utilisés correctement au passé composé et à l'imparfait.

e) Howard (2004)

Dans une étude avec des apprenant/es avancé/es du français L2, Howard 2004 a testé simultanément l'hypothèse de l'aspect inhérent et l'hypothèse discursive à partir de données de 60 minutes de conversation par personne sur différents sujets, formels et informels. Selon l'hypothèse discursive, la production morphologique des apprenant/es dépend de la fonction narrative du verbe, avant-plan (*foreground*) vs arrière-plan (*background*).

- Les résultats (sans les occurrences des verbes *être* et *avoir*) suggèrent que les deux hypothèses interagissent plutôt que de s'exclure mutuellement.

- Ainsi le cas le plus prototypique pour l'utilisation du passé composé par les apprenant/es est celui des verbes d'achèvements en avant-plan et le moins prototypique celui des verbes d'états en arrière-plan.
- À l'inverse, le présent et l'imparfait apparaissent le plus prototypiquement avec les verbes d'état en arrière-plan, alors que les contextes de verbes d'achèvement et en avant-plan y seraient le plus résistants (Howard 2004 : 330).

Ces résultats correspondent à ce qui a été montré par Bonnotte et Fayol (2001 ; Bonnotte, 2008). Ces apprenants avancés du français reprennent donc les associations que l'on trouve chez les adultes natifs.

f) Labeau (2005)

Labeau a vérifié l'hypothèse de l'aspect à partir de données d'étudiants en français LE/L2 à un niveau intermédiaire – avancé. Pour cela, elle s'est servie de différents tests spécifiques à l'hypothèse de l'aspect (Labeau, 2005 : 81 ; cf. Bardovi-Harlig 2000 : 197-205 pour une présentation et une discussion de ceux-ci).

- Les résultats obtenus ont montré que l'hypothèse de l'aspect semblait se vérifier pour le passé composé dans la mesure où il était utilisé de préférence avec les verbes téléliques.
- L'imparfait a par contre été utilisé avec toutes les catégories de verbes, y compris les verbes téléliques, mais principalement avec des expressions comprenant les verbes *être* et *avoir* comme dans l'étude de Kihlstedt (1998) (Labeau 2005 : 85).
- Labeau souligne le rôle de la fréquence des formes dans l'input ainsi que leur distribution pour expliquer les résultats obtenus, en particulier pour le passé composé.

Generally speaking, the data seem to show the tendencies predicted by A[spect] H[ypothesis] but, because of the high frequency of the combination of PC with telic verbs in native usage, one may suggest that other factors than lexical aspect, such as frequency of the input or formulaic use, are at work in the interlanguage development. (Labeau, 2005 : 82)

- Finalement, les apprenant/es ont également eu recours au présent dans les contextes du passé, à la place de l'imparfait comme du passé composé. Les données semblent indiquer que le recours à une forme du

présent à référence passée soit surtout un phénomène de l'oral (84% des utilisations du présent à référence passée ; Labeau 2005 : 94).

g) *Synthèse des études à partir de l'hypothèse de l'aspect*

Les études à partir de l'hypothèse de l'aspect sur des données d'apprenant/es du français LE/L2 ont montré les tendances suivantes :

- a) le marquage du passé se fait plus rapidement avec le passé composé qu'avec l'imparfait conformément à l'hypothèse de l'aspect.
- b) l'imparfait est avant tout utilisé avec les verbes d'état, mais son utilisation n'est pas limitée aux verbes atéliques d'état et d'activités contrairement à ce qui est prévu par l'hypothèse de l'aspect. Plusieurs études ont en fait révélé que l'imparfait était surtout utilisé avec les formes *était* et *avait*. On a aussi pu voir que l'imparfait était rarement utilisé avec les verbes d'achèvement comme attendu par l'hypothèse.
- c) que le présent était souvent utilisé à la place de l'imparfait, même à des niveaux avancés, et souvent avec les états. Ce point correspond à l'hypothèse de l'aspect dans la mesure où le passé imperfectif, l'imparfait, est attendu plus tardivement que le passé perfectif, le passé composé.
- d) que le passé composé était utilisé avec *tous* les verbes dynamiques dans la plupart des études et plus rarement avec les verbes d'état. Contrairement à ce qui est attendu par l'hypothèse de l'aspect, le passé composé semble très rapidement être utilisé avec tous les verbes dynamiques, ce qui implique que la progression du marquage du passé d'une catégorie de verbes à l'autre ne se fait pas de manière visible en français. L'opposition atélique vs télique est rarement attestée dans les données d'apprenant/es, mais régulièrement dans celles des natifs.

En somme, l'hypothèse de l'aspect ne se vérifie que faiblement pour le français LE/L2. L'influence de l'aspect lexical en français LE/L2 semble plutôt suivre une opposition statique vs dynamique, avec d'un côté les formes du présent (imperfectif non passé) et de l'autre les formes du passé composé, l'imparfait étant presque uniquement utilisé avec deux formes verbales.

Le fait que l'hypothèse de l'aspect ne se vérifie pas au niveau de l'opposition aspectuelle prototypique chez les apprenant/es du français peut être lié au fait que l'hypothèse a avant tout été développée à partir d'apprenant/es de l'espagnol (Andersen, 1991 ; 1993). Contrairement au français (et à l'italien), l'espagnol

possède trois formes du passé : le pretérito imperfecto (*hablaba*) qui correspond à l'imparfait français, et deux formes pour le perfectif, le pretérito perfecto compuesto (*ha hablado*) qui correspond au passé composé français et le pretérito perfecto simple (*habló*) qui correspond au passé simple et, sémantiquement, au passé composé en français. Toutes ces formes sont couramment utilisées à l'oral. Ainsi l'opposition télique – atélique est peut être plus saillante en espagnol LE/L2 qu'en français LE/L2 en raison de la variété des formes disponibles et d'une opposition morphologique entre le temps de l'imparfait et du prétérît à l'oral.

On peut d'ailleurs mentionner l'étude de Wiberg (1997) sur des jeunes bilingues suédois et italien ainsi que des monolingues italiens, qui présente des résultats semblables à ceux pour le français. Le système du passé en italien est similaire à celui du français, avec un passé composé (*passato prossimo*) et un imparfait (*imperfetto*). Les données d'utilisation du passé montrent que le passé composé est utilisé comme le temps de base du passé avec tous les types de verbes dans le dialogue, alors que l'imparfait est moins fréquent (Wiberg, 1997 : 255-256). Néanmoins, la différence entre le français et l'italien est que l'on trouve des formes de l'imparfait avec un plus grand nombre de verbes, notamment dans la narration, et qu'il est utilisé avec les verbes d'états et d'activités (Rocca, 2002 ; Giacalone-Ramat, 2002).

Les différences entre le français et ces langues romanes pourraient être liées au fait que les différentes formes du passé sont distinctes en italien et en espagnol alors que la forme de l'imparfait et du participe passé sont homophones pour 90% des verbes de la langue française comme on l'a vu au chapitre 2 (tableaux 2-1 et 2-2). La perception de la forme de l'imparfait pourrait donc être difficile pour les apprenant/es du français, raison pour laquelle ils choisissent d'exprimer l'imperfectif par les formes du présent.

4.2.3. Aspect lexical dans le corpus ESF – approche fonctionnelle

L'objectif de l'analyse des données du corpus ESF (Dietrich *et al.*, 1995) est d'étudier l'organisation des relations temporelles par les apprenant/es. Bien qu'inspiré par les travaux d'Andersen (Dietrich *et al.*, 1995 : 9) leur cadre d'analyse ne se base pas sur les mêmes catégories. Leur analyse des relations temporelles est basée sur les trois groupes de critères suivants (Noyau, 1990 : 103-104 ; Klein, 1995) :

- les intervalles temporels : moment de parole, moment de situation, moment en question (*topic time*), avant, après etc.

- l'aspect grammatical
- les propriétés temporelles inhérentes des procès (aspect lexical)

L'analyse des formes verbales produites par les apprenant/es Abdelmalek et Zahra, arabophones du Maroc, et Berta, hispanophone du Chili, a montré que les apprenant/es en début d'acquisition du français ont principalement recours à deux formes verbales de base :

- une forme *V-Ø* par exemple *travaille*, *porte*, *paye*, la forme que j'appelle la *forme courte* dans ce travail
- et une forme *V-[e]* par exemple *arrêTE*, *fermE*, *licenciE* (exemples de Noyau *et al.*, 1995 : 167), la forme que j'appelle la *forme longue* dans ce travail.

Les données de Zahra montrent un développement dans l'utilisation de ces deux formes que je résume ici en quatre étapes :

1. les deux formes sont produites sans valeurs distinctes (Noyau *et al.*, 1995 : 160)
2. les formes courtes et longues expriment une différenciation sémantique de la situation, la forme courte (ainsi que d'autres formes, *V-X*) ce qui est duratif et la forme longue ce qui est ponctuel (Noyau *et al.*, 1995 : 165)
3. les formes permettent de distinguer ce qui est \pm télique (dans leur vocabulaire \pm complet¹⁷). La distribution des formes n'est pas présentée dans le texte, mais les exemples laissent entendre que la forme courte est utilisée avec les verbes atéliques et la forme longue avec les verbes téliques (Noyau *et al.*, 1995 : 167-168)
4. la forme courte exprime ce qui est présent ou prospectif et la forme longue ce qui est passé ("V-e for completed thus past at some time", Noyau *et al.*, 1995 : 168)

Chez Abdelmalek, on observe des étapes similaires ; elles sont également présentées ici :

¹⁷ L'utilisation de la distinction \pm complet est un peu surprenante, le français ne faisant pas de distinction entre le parfait et l'aoriste à l'oral (voir Kihlstedt, 1998 : 28). Cette distinction n'est pas présentée de manière distincte dans le chapitre sur le français et reste difficile à comprendre.

1. on trouve une variation dans l'utilisation des deux formes sans valeur aspectuelle ou temporelle (Noyau *et al.*, 1995 : 172)
2. les deux formes apparaissent selon les traits temporels inhérents à la situation et on observe une tendance à la distribution des formes : les verbes à la forme longue apparaissent dans les situations dynamiques et à la forme courte dans les situations statiques (Noyau *et al.*, 1995 : 172-173)
3. d'un côté, la forme courte exprime les situations à un état (<John be ill>) et la forme longue les situations à deux états (<John become ill>)¹⁸ ; de l'autre côté, la forme courte apparaît dans les situations atéliques (*non-completed*), en cours et prospectives et la forme longue dans les situations téliques (*completed*) ou passées (Noyau *et al.*, 1995 : 174)
4. finalement, les deux formes alternent pour un même lexème verbale indiquant un contraste aspectuel : la forme longue pour le perfectif et la forme courte pour l'imperfectif (Noyau *et al.*, 1995 : 175)

L'explication proposée par Noyau *et al.* est que les apprenant/es sont guidés par différentes hypothèses successives dans leur recherche d'une fonction correspondant aux formes qu'ils ont découvertes :

Hi: temporal information encoded morphologically concerns the inherent temporal features of the situations

Ha: morphological temporal information differentiates aspectual perspectives on situations (perfective vs. imperfective, perhaps perfect)

Ht: verbal morphology encode temporal relations. (Noyau *et al.*, 1995 : 184)

Le lien entre forme et fonction se développe donc selon un itinéraire fixe à l'instar de l'hypothèse de l'aspect proposée par Andersen. Selon leur analyse, Zahra s'arrête à *Hi* et Abdelmalek à *Ha/t* (Noyau *et al.*, 1995 : 184).

Cette analyse fonctionnelle des données présente donc l'influence de l'aspect lexical sur la production des formes par les apprenant/es débutant/es ainsi qu'un itinéraire dans l'utilisation des formes courtes et longues allant de l'aspect lexical à l'aspect grammatical et finalement au temps. Selon Noyau, les traits de télicité (clôture) et de durée sont "les notions temporelles clé pour la

¹⁸ Exemples repris de la présentation du cadre d'analyse (Klein, 1995 : 20).

construction de la temporalité dans le discours par les apprenants” (Noyau, 1990 : 110).

L’acquisition précoce de moyens pour exprimer les premiers [les traits de télicité et de durée] suggère que les besoins déterminés par la structure conceptuelle du domaine guident la saisie par les apprenants des moyens correspondants dans la langue à apprendre (Noyau, 1990 : 111).

Les oppositions statique vs dynamique et des formes présent vs passé que l’on avait pu noter dans les études à partir de l’hypothèse de l’aspect sont donc également attestées dans les travaux à partir du corpus ESF, ce qui indique qu’il pourrait s’agir d’oppositions caractéristiques pour les apprenant/es du français LE/L2.

4.3. Résumé et discussion

Ce chapitre a tout d’abord présenté les principaux concepts impliqués dans celui de l’aspect lexical :

- l’aspect lexical concerne les propriétés inhérentes à la situation exprimée par le verbe
- les distinctions exprimées par l’aspect lexical sont considérées comme universelles et innées (Smith, 1991 ; Giacalone-Ramat, 2002)
- l’aspect lexical est lié à un certain nombre d’oppositions : statique – dynamique, duratif – ponctuel, atélique – télique
- à un niveau typologique (Sasse, 2002), l’aspect lexical et l’aspect grammatical sont liés au niveau de l’expression inhérente d’une limitation ou non :
 - l’imperfectif, passé ou non passé, est étroitement lié aux verbes statiques en raison de l’absence de limitation inhérente
 - le perfectif est étroitement lié aux verbes téliques dans le sens d’une limitation inhérente
 - les verbes d’activités ne sont pas liés à un aspect grammatical spécifique

Les études sur le français L1 ont montré que les sous-catégories \pm duratif et \pm résultatif permettent de faire des prédictions au niveau de l’utilisation de formes (Bonnotte & Fayol, 2001) et que les adultes préfèrent utiliser les formes

imperfectives avec les verbes statiques et d'activités et les formes du passé composé (du plus-que-parfait et du passé simple) avec les verbes téliques (Bonnotte & Fayol, 2001 ; Bonnotte, 2008).

Les études sur l'acquisition du français L1 ont montré que les enfants utilisent surtout le présent avec les verbes atéliques et le passé composé avec les verbes téliques (qui mènent à un résultat ou non ; Bronckart & Sinclair, 1973) et aussi, que les enfants produisent les verbes statiques au présent, et non à l'infinitif, dès leur plus jeune âge (Ferdinand, 1996).

Les études dans le cadre de l'hypothèse de l'aspect ainsi que celles à partir du corpus ESF dans une approche fonctionnelle ont montré que la production des formes par les apprenant/es du français LE/L2 suivait l'opposition statique vs dynamique, avec une utilisation des formes du présent avec les verbes statiques et les formes du passé composé avec les verbes dynamiques. La forme de l'imparfait est donc rarement utilisée par les apprenant/es. Le fait qu'elle soit homophone avec la forme du participe passé pour 90% des verbes de la langue française la rend peut-être peu saillante pour les apprenant/es du français LE/L2 contrairement aux autres langues romanes (*cf.* tableau 2-2).

Bien que le présent travail ne traite pas de la morphologie du passé, on pourrait formuler l'hypothèse que les formes produites par les apprenant/es soient influencées par l'aspect inhérent du lexème verbal quel que soit le contexte. L'influence de l'aspect lexical permettrait ainsi d'expliquer la variation de formes courtes et longues dans les contextes du présent et de l'infinitif, la forme courte étant produite avec les verbes statiques et la forme longue avec les verbes dynamiques.

La question de l'influence de la fréquence des formes dans l'input a été abordée dans plusieurs études sur l'hypothèse de l'aspect, notamment par Kaplan (1987), Kihlstedt (1998) et Labeau (2005) en particulier au niveau de la fréquence d'utilisation du passé composé. Mais l'utilisation des formes du présent pourrait elle aussi refléter la fréquence d'utilisation de ces formes dans l'input oral. On peut ainsi se demander si les apprenants ne sont pas simplement influencés par la fréquence des formes des différents verbes dans l'input, certains verbes étant plus fréquents au présent et d'autres au non-présent (*cf.* tableau 2-1) et que le regroupement de ceux-ci à partir de l'aspect lexical est un filtre ajouté par les linguistes. L'ambition de ce travail est de présenter des arguments qui permettent de discuter l'influence de l'aspect lexical par rapport à celle de la fréquence des formes des verbes à partir de deux hypothèses séparées.

5. Cadre théorique : perspective basée sur l'usage de la langue (*usage-based*)

Ce chapitre présente le cadre théorique utilisé pour ce travail. Il s'agit d'une approche basée sur l'usage de la langue qui considère que la langue produite par les apprenant/es est influencée par celle à laquelle ils sont exposés, l'input.

La première section présentera les principaux concepts de l'approche basée sur l'usage, dont la fréquence, alors que la deuxième section traitera des facteurs qui pourraient renforcer ou diminuer le rôle de la fréquence. La dernière section résumera les points essentiels de ce cadre théorique et présentera l'esquisse d'une hypothèse basée sur l'influence de la fréquence des formes.

5.1. Principaux concepts de la perspective basée sur l'usage de la langue

5.1.1. Expérience de la langue

Les tenants d'une approche basée sur l'usage de la langue postulent un lien étroit entre l'usage de la langue d'une part et la manière dont la langue est représentée en mémoire (Bybee, 2001; 2008; Kemmer & Barlow, 2000) : “[...] Usage-Based Theories of language recognize the impact of usage on the cognitive representation of language” (Bybee, 2008 : 216). Dans cette perspective, le système linguistique du locuteur est fondé sur les événements d'usage (*usage events*), à savoir les instances de production et de compréhension de la langue (Kemmer & Barlow, 2000 : viii). L'expérience de la langue est ainsi au centre de cette approche et la grammaire est vue comme l'organisation cognitive de l'expérience de la langue de chacun (Bybee, 2008 : 216 ; 2006 : 711).

From these studies there is now emerging a new theory of language in which grammar is viewed as the cognitive organization of one's experience with language (Bybee, 2006). In this view, as users of language experience tokens of language use, they categorize them at varying degrees of abstractness. This categorization process creates a vast

network of phonological, semantic and pragmatic associations that range over what has traditionally been designated as lexicon and grammar (Bybee, 2008 : 216-217)

5.1.2. Constructions

Dans cette approche, la langue est essentiellement vue comme constituée de constructions, allant du simple mot lexical (“*table, decide, pretty*”) à des constructions complètement abstraites (“*they gave him an award*”) en passant notamment par des syntagmes verbaux, nominaux etc. (“*VERB + Past Tense ; the + NOUN*”) et des expressions (*idioms*) avec un contenu lexical fixe (*go great guns*¹⁹) (voir Bybee, 2008 : 217 pour une liste complète). Selon Bybee (2008 : 217) un des avantages de l’utilisation de constructions comme unité grammaticale de base est qu’elle permet de voir un continuum du lexique à la grammaire. Reformulé dans mes propres termes : les constructions grammaticales sont stockées et représentées en mémoire selon les mêmes processus que les mots individuels.

L’expérience de la langue joue un rôle dans la formation de ces constructions ainsi que dans leur automatisisation. Les constructions sont des *chunks* dans le sens psycholinguistique du terme : “In Usage-Based Theory, constructions are chunks – neuromotor routines – with movable parts.” (Bybee, 2008 : 220 ; sur le *chunking* voir Ellis, 2001). La grammaire et le lexique sont ainsi constitués de routines linguistiques qui ont été développées avec le temps et qui sont utilisées de manière automatisée.

Les théories basées sur l’usage postulent une approche dynamique du système linguistique : le système mis en place peut être sujet à révision au fur et à mesure de l’expérience et des événements d’usage rencontrés.

5.1.3. Fréquence

En raison du rôle central de l’expérience, la fréquence d’utilisation est un facteur clé dans la construction des représentations mentales de la langue. La citation d’Ellis ci-dessous résume le processus d’acquisition d’une langue guidé par la fréquence.

In these usage-based perspectives, the acquisition of grammar is the piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them. Language learning is the associative learning of representations that reflect

¹⁹ Expression fixe signifiant *avoir du succès*.

the probabilities of occurrence of form-function mappings. Frequency is thus a key determinant of acquisition because “rules” of language, at all levels of analysis (from phonology, through syntax, to discourse), are structural regularities that emerge from learners’ lifetime analysis of the distributional characteristics of the language input. Learners have to *figure* language out. (Ellis, 2002 : 144 – son soulignement)

Une des principales contributions des travaux de Bybee est d’avoir montré l’importance de la distinction entre la fréquence d’occurrence (*token frequency*) et la fréquence type (*type frequency*). La *fréquence d’occurrence* correspond au nombre de fois qu’une unité (lettre, syllabe, mot, syntagme, phrase) apparaît dans un texte alors que la *fréquence type* correspond au nombre d’items lexicaux différents applicables à un même *pattern* (motif récurrent) ou à une même construction (Bybee & Thompson, 2000 ; Bybee 2008). “Only patterns of language have type frequency because this refers to how many distinct items are represented by the pattern” (Bybee, 2008 : 218). La fréquence type peut s’appliquer à différents éléments de la langue, dont voici quelques exemples mentionnés par Bybee (2008 : 218) :

- des séquences phonotactiques, par exemple le nombre de mots qui commencent par [sp] ou [sf]
- des *patterns* morphologiques tels que les combinaisons *racine + affixe*, par exemple la formation du passé en anglais de *know* en *knew* ou l’ajout du suffixe *-ed* pour former le passé des verbes réguliers en anglais
- des *patterns* syntaxiques ou des constructions, par exemple la construction ditransitive *He gave me the change* ne peut être utilisée qu’avec un petit nombre de verbes en anglais, alors que la construction *He gave the change to me* est possible avec un plus grand nombre de verbes.

Selon Bybee, les fréquences type et d’occurrence jouent un rôle important dans l’établissement et le maintien des représentations en mémoire. Les mots et les formes retenus et représentés en mémoire à long terme ne sont pas des unités isolées mais reliées entre elles par des sets de connexions lexicales plus ou moins fortes²⁰. “Words entered in the lexicon are related to other words via sets

²⁰ Pour cette raison, le modèle de Bybee (1985) est aussi appelé *modèle de réseau* (*network model*).

of lexical connections between identical and similar phonological and semantic features.” (Bybee, 1995 : 428)

Les connexions lexicales entre les mots diffèrent selon le genre de fréquence, type ou occurrence. Les mots avec une haute fréquence d’occurrence présentent une plus grande autonomie et donc une plus faible connexion avec d’autres mots (Bybee, 1995 : 428s ; 2008 : 218s). Plus un mot sera autonome plus son degré de *force lexicale* (*lexical strength*) sera élevé, selon la terminologie utilisée par Bybee. Les mots ayant une grande force lexicale sont plus facilement accessibles, présentent une autonomie qui les rend résistants au changement et sont enclins à l’indépendance sémantique. Les formes irrégulières des verbes, en particulier les formes supplétives (en français par exemple *va* ou *sont*) ont typiquement un fort degré de force lexicale (Bybee, 1995 : 429).

Si la fréquence d’occurrence entraîne l’autonomisation des mots, la fréquence type au contraire a pour effet de renforcer voire de généraliser des connexions établies. Les sets de mots qui ont des *patterns* similaires de connexions sémantiques et phonologiques se renforcent les uns les autres et créent des généralisations émergentes descriptibles comme *schémas*. Suivant la force et le degré d’ouverture (de spécificité) du schéma, il sera appliqué à des nouveaux items ou à des items dont les connexions ne sont pas connues (Bybee, 1995 : 430). Plus le schéma comprend de participants (fréquence type élevée) plus il sera productif (Bybee, 2001 : 120-121). La productivité d’un schéma est ainsi un effet important de la fréquence type car elle peut donner naissance à une catégorie selon le nombre d’items qui y participent.

L’étude de Bartning (1998) sur la production des verbes à la 3^e du singulier et du pluriel en français est une bonne illustration de la généralisation d’un schéma en raison de sa productivité. Bartning a en effet pu observer que des apprenant/es avancé/es du français LE/L2 produisaient les verbes irréguliers sur le modèle du paradigme des verbes en *-er* en utilisant la même forme à la 3^e du singulier et du pluriel, une forme courte (*il prend / ils prend, il met / ils met*) ou une forme longue (*il prende / ils prende, il mette / ils mette*), ne marquant donc pas l’opposition entre ces formes.

La fréquence type est ce qui différencie le modèle de Bybee de modèles postulant un traitement dual des formes (*dual processing*) (Bybee, 1995 : 427). Selon les tenants du *dual processing*, deux processus cognitifs distincts sont impliqués dans le maniement des formes régulières et irrégulières : les flexions régulières sont dérivées par l’application de règles symboliques, i.e. syntaxiques, alors que les flexions irrégulières sont représentées dans le lexique (*cf.* par exemple Pinker, 1998 : 221). Dans l’approche de Bybee, la différence entre les

formes se situe au niveau de leur fréquence. Pour simplifier on peut dire que les verbes irréguliers ont une fréquence *d'occurrence* élevée, alors que les verbes réguliers ont une fréquence *type* élevée. Mais il est important de noter que ce qui importe dans l'approche basée sur l'usage, c'est le nombre de verbes qui participent à un même schéma et non la régularité ou l'irrégularité du verbe concerné (Bybee, 1995).

5.1.4. Input et usage de la langue

Dans la perspective d'une théorie basée sur l'usage, *l'input* est considéré comme la langue à laquelle l'apprenant/e est exposé (Kemmer & Barlow, 2000 : xv ; Gass & Mackey, 2007 : 177). L'input comprend notamment des mots, des constructions, des phonèmes, des morphèmes, des régularités et des irrégularités. Il comprend donc d'une part la langue cible et ses fréquences de co-occurrence, mais il comprend aussi la propre production de l'apprenant/e. Le système linguistique du locuteur est en effet basé sur la compréhension comme sur la production de la langue par ce locuteur (Kemmer & Barlow, 2000 : viii). On peut même dire que le fait de produire du langage sensibilise l'apprenant/e à l'input (Gass & Mackey, 2002 ; 2007 ; VanPatten, 2002 : 762ss) :

In other words, the process of producing or struggling to produce output may sensitize learners to patterns and associations in future input, possibly making future input more salient. (Gass & Mackey, 2002 : 252)

Finalement l'input est fortement lié à l'environnement dans lequel l'apprenant/e fait l'expérience de la langue. Apprendre une langue en contexte formel, comme la classe de langue, entraîne un certain nombre de biais au niveau de l'input en général, d'une part par la méthode utilisée et d'autre part par la production des autres apprenant/es²¹.

On peut ainsi dire, qu'il y a trois sortes d'inputs :

1. l'input qui est inhérent à la langue cible, telle que la réalisation du perfectif et de l'imperfectif, le fait que les verbes réguliers en *-er* à l'oral n'aient que deux formes ; la langue cible en tant que telle, qui comprend des systématiquités, des associations prototypiques, par exemple le passé composé avec des verbes téliques ou le présent avec des verbes statiques

²¹ Cf. Ellis, 2003 : 72 : "[...] classroom environments for second or foreign language teaching can distort the patterns of exposure [...]". Cf. aussi Anderson (2007).

2. l'input auquel l'apprenant/e est réellement exposé, les sources concrètes d'occurrences et qui sont plus ou moins proches du prototype de la langue cible, par exemple en raison de biais propres au 'locuteur' (y compris l'apprenant/e lui-même/elle-même) ou au contexte, ce qu'on pourrait appeler des variétés de français
3. l'input tel que perçu par l'apprenant/e qui est lié à des mécanismes cognitifs qui facilitent ou inhibent le traitement de la langue à laquelle il est exposé

La différence entre la première sorte d'input et la deuxième peut être difficile à cerner étant donné que la langue est étroitement liée à l'usage qu'on en fait. Comme on le verra dans le chapitre 6 sur les hypothèses, cette différenciation d'input permettra de distinguer une hypothèse basée sur l'aspect lexical d'une hypothèse basée sur la fréquence des formes.

5.2. Autres facteurs en jeu

Comme on vient de le voir, la fréquence joue un rôle fondamental dans l'approche basée sur l'usage et elle contribue à l'acquisition d'une langue sur différents plans. Néanmoins, pour que la fréquence puisse avoir un effet, afin que l'input ne reste pas une série de sons, il est nécessaire de mentionner d'autres facteurs. L'*attention* telle que présentée par les psychologues semble jouer un rôle clé dans la transmission du matériel de la mémoire de travail en mémoire à long terme. Un autre facteur important est celui de *saillance* que l'on peut résumer par les chances qu'a un élément dans la structure cible d'obtenir l'attention nécessaire pour être retenu. Finalement le processus d'*intake* sera également présenté, car il est souvent mentionné lorsqu'on parle d'influence de l'input dans la littérature.

5.2.1. Attention

L'attention est un mécanisme cognitif nécessaire permettant de stocker en mémoire à long terme ce qu'un être humain perçoit, afin que cela ne reste pas seulement dans la mémoire de travail (Schmidt, 2001 : 16). Ainsi, l'attention joue un rôle important dans l'acquisition d'une LE/L2 :

[...] SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be. (Schmidt, 2001: 3-4)

Le terme psychologique *d'attention* a un corollaire considéré comme plus subjectif, mais plus fréquemment utilisé dans la littérature sur l'acquisition des langues, à savoir le processus de *noticing*. Dans l'approche de Schmidt, qui correspond bien, à mon avis, à une approche basée sur l'usage²², il est important de séparer le processus de *noticing* d'une conscience métalinguistique, “[...] the objects of attention and noticing are elements of the surface structure of utterances in the input – instances of language, rather than any abstract rules [...]” (Schmidt, 2001: 5). Ce qui compte ici concerne donc les éléments de surface, les occurrences concrètes de la langue.

En psychologie, l'attention est entre autres considérée comme étant limitée, sélective et essentielle pour apprendre (Schmidt, 2001 : 11-12).

Une difficulté propre aux apprenant/es débutant/es est qu'ils sont cognitivement surchargés et qu'ils ne peuvent pas faire attention à tous les éléments porteurs de sens en même temps (Schmidt, 2001: 7). Ils font donc des choix ou comme le dit Schmidt (2001 : 13) une analyse de coûts et bénéfices : les ressources de la mémoire de travail étant limitées il s'agit de bien les utiliser.

Limited attentional resources are directed first at those elements that carry message meaning, primarily lexicon, and only later, when the cost comes down, towards communicatively redundant formal features of language. (Schmidt, 2001 : 13)

VanPatten (2002, 2007) a traité la question de la gestion de ressources en formulant un certain nombre de principes cognitifs à partir de son modèle de traitement de l'input, *Input Processing*. Selon lui, “[...] the intersection of high communicative value and frequency should have a favorable effect on acquisition” (VanPatten, 2002 : 760)²³. En résumé, les apprenant/es centrent d'abord leur attention sur ce qu'ils considèrent comme prioritaire au niveau de la communication.

²² D'après Schmidt, le rôle de l'attention est avant tout souligné par les approches qui voient que les apprenants traitent une LE/L2 à partir de facteurs cognitifs généraux, tels que la saillance ou la fréquence, facteurs qui ne jouent pas ou peu de rôle dans des approches nativistes ou connexionnistes (Schmidt, 2001 : 6).

²³ Une haute valeur communicative comprend les caractéristiques [+valeur sémantique inhérente/-redondance] (VanPatten, 2002 : 759).

L'attention permet de contrebalancer la fréquence en raison de la limitation des ressources disponibles au traitement de l'input. Plus l'apprenant/e a d'expérience de la langue plus il pourra détacher son attention de ce qui a une haute valeur communicative pour découvrir d'autres éléments. Le processus d'attention pourrait ainsi expliquer des itinéraires d'acquisition comme celui de Bartning & Schlyter (2004 ; cf. section 3.2.).

5.2.2. Saillance

La saillance est directement liée au phénomène d'attention, dans la mesure où ce qui est saillant attire l'attention alors que ce qui ne l'est pas est ignoré ou omis. Goldschneider et DeKeyser (2001 : 22) définissent la saillance (perceptuelle) en termes de facilité à entendre ou à percevoir une structure donnée. Bardovi-Harlig (1987) définit la saillance en termes de disponibilité (*availability*) dans l'input. Comme le soulignent Goldschneider et DeKeyser (2001 : 22), la saillance est vue de deux manières : soit comme une caractéristique de l'input, soit en référence à d'autres facteurs qui font que certaines parties de l'input deviennent saillantes.

En raison du lien étroit entre attention et saillance on pourrait reprendre les propositions de VanPatten présentées ci-dessus et dire que les mots de contenu et les éléments qui ont une haute valeur communicative sont saillants. Par contre les marques grammaticales manquent de saillance (et sont redondantes). Selon Ellis (2008 : 381ss) c'est pour cette raison qu'elles sont difficiles à acquérir. Comme le dit Bybee (2008 : 231) certaines marques grammaticales restent dans la langue par convention, mais elles sont superflues pour les apprenant/es.

La prosodie et l'accentuation des mots et des phrases peuvent également contribuer à la saillance ou non de certains éléments (Klein 1989 ; pour une vue d'ensemble de critères de saillance voir Rast, 2008 : 146ss). Pour revenir aux marques grammaticales, les auxiliaires par exemple sont souvent monosyllabiques et inaccentués (Labelle, 2005 : 445). Carroll (2004) souligne le fait que les syllabes accentuées sont plus longues et plus stables dans leur prononciation que les syllabes non-accentuées. Ces syllabes accentuées seraient plus faciles à segmenter.

Le degré de saillance d'un mot dans une phrase peut ainsi contrebalancer les effets de fréquence. Un élément peut être fréquent dans la langue tout en étant peu saillant. Un tel élément sera moins facilement noté et retenu en mémoire à long terme qu'un élément fréquent et saillant.

5.2.3. Intake

Le terme d'*intake* provient de Corder (1967) et fait référence à la partie de l'input que les apprenant/es traitent cognitivement (VanPatten, 1996 : 7). Pour VanPatten (1996), l'*intake* est le résultat de ce qui a été traité dans l'input et ce qui sera transmis à un système de développement pour un traitement ultérieur d'accommodation ou de restructuration. Dans la définition de (2002) l'apport de la mémoire de travail est plus clairement mentionné.

Intake is that subset of filtered input that the learner actually processes and holds in working memory during on-line comprehension. (VanPatten, 2002 : 761)

En d'autres termes il s'agit de matériel qui est en bonne voie d'être retenu en mémoire à long terme. Ce qui correspond aussi au processus décrit par Carroll (2001). Carroll part de l'idée que les apprenant/es entendent du langage parlé dans des contextes qui ont du sens, mais que ce qu'ils entendent correspond avant tout à des événements acoustico-phonétiques qui stimulent le système de perception auditif (Carroll, 2001 : 8). Autrement dit, les apprenant/es perçoivent du bruit qu'ils doivent traiter. Plutôt que d'utiliser les termes d'input et d'*intake* Carroll reprend le vocabulaire de l'intelligence artificielle : elle parle de stimulus comme information qui entre dans les systèmes de perception puis de stimulus traité et représenté mentalement : "intake is a mental representation of a physical stimulus" (Carroll, 2001 : 8-9). En ce sens, l'*intake* apparaît comme un processus très proche de ce qui est décrit par l'attention.

Comme le dit Bybee, la question du stockage ou non d'une unité linguistique dépend de l'expérience de l'utilisateur de la langue en question (Bybee, 2006 : 717).

Bien qu'une théorie sur le traitement de l'input comme 'stimulus acoustique' soit nécessaire, elle est difficilement utilisable dans ce travail. Ainsi le processus d'*intake* sera abandonné au profit de celui de l'attention.

5.3. Résumé et discussion

Ce chapitre a présenté le cadre théorique qui sera utilisé pour l'étude de la variation des formes dans la production des apprenant/es. Les points les plus importants de cette théorie peuvent se résumer comme suit :

- la manière dont la langue est représentée en mémoire est liée à l'usage de la langue

- la grammaire et le lexique sont vus comme des routines linguistiques qui peuvent être revues au fur et à mesure de l'expérience de la langue à laquelle l'apprenant/e est exposé
- la fréquence d'occurrence et la fréquence type jouent un rôle central au niveau de la représentation des items en mémoire
- l'input est essentiellement la langue à laquelle l'apprenant/e est exposé, mais il comprend aussi le système de la langue cible à un niveau plus général ainsi que l'input tel que perçu par l'apprenant/e
- le processus cognitif de l'attention joue un rôle central au niveau des occurrences concrètes notées et retenues en mémoire
- la saillance joue un rôle dans la mesure où ce qui est facilement disponible dans l'input est plus facilement retenu en mémoire, à l'inverse de ce qui passe inaperçu

Ce cadre théorique permettra d'étudier la variation de formes à partir de l'influence de l'aspect inhérent comme de la fréquence des formes de surface. En effet, du moment où l'aspect lexical correspond à une faculté innée de distinguer les situations, il devrait être étroitement lié au processus d'attention d'une part et de saillance d'autre part, afin que les apprenant/es trouvent les formes exprimant les oppositions les plus importantes et les utilisent.

Une hypothèse à partir de la fréquence des formes peut se formuler ainsi : les apprenant/es sont sensibles à la fréquence des formes utilisées dans la langue à laquelle ils sont exposés, indépendamment du contexte syntaxique ou sémantique du verbe, du moins en tout début d'acquisition. La forme la plus fréquente d'un verbe sera plus solidement représentée en mémoire chez les apprenant/es et plus accessible en production. Ainsi, un verbe fréquent à la forme courte du présent dans l'input sera plus facilement produit à la forme courte par l'apprenant/e et un verbe fréquent à la forme longue dans l'input sera plus facilement produit à la forme longue par l'apprenant/e. Les verbes doivent ainsi pouvoir se placer sur un continuum allant des verbes très fréquents à la forme courte aux verbes très fréquents à la forme longue. Une telle hypothèse devrait permettre de contraster l'hypothèse sur l'influence de l'aspect lexical pour lequel on s'attend à l'expression d'une opposition. On devrait ainsi trouver des verbes statiques moins fréquents à la forme courte ou des verbes dynamiques moins fréquents à la forme longue dans l'input. Pour cette raison, il sera nécessaire de regarder les occurrences concrètes des verbes dans des sources de données d'input.

6. Hypothèses

L'objectif de ce chapitre est de présenter les deux hypothèses qui ont été évoquées à la fin des chapitres 4 et 5 en les rassemblant sous le cadre théorique pour ce travail (*cf.* chapitre 5).

Comme on l'a vu en introduction l'objectif du présent travail est de vérifier si l'on trouve une utilisation systématique des formes courtes (*parle*) et longues (*parle*) des verbes réguliers en *-er* dans les contextes syntaxiques du présent et de l'infinitif par les apprenant/es débutant/es. Pour cela j'ai choisi une approche théorique basée sur l'usage qui propose que la langue utilisée par les apprenant/es soit le fruit de l'expérience de la langue à laquelle ils sont exposés, l'input.

Comme on l'a vu au chapitre 2, le paradigme des verbes réguliers en *-er* à l'oral est formé de deux principales formes, la forme courte et la forme longue (*cf.* tableau 2-1). Les études antérieures qui ont étudié la production de ces formes à partir des catégories de la finitude (chapitre 3) ont confirmé la variation dans l'utilisation des formes dans les contextes finis et non-finis, comme on avait pu le voir dans l'exemple (1-1) de l'introduction. Pour cette raison, le présent travail ne sera pas centré sur la syntaxe mais sur l'étude d'un nombre précis de verbes et leur production par les apprenant/es.

Les études présentées au chapitre 4 ont montré que l'on pouvait trouver une certaine systématisme dans la production des formes du passé par les apprenant/es, à partir de l'aspect lexical des verbes (*cf.* tableau 4-1). Selon ces études, l'opposition statique *vs* dynamique semble saillante pour les apprenant/es du français L2. Mais comme je l'ai proposé à la fin du chapitre 4, la production des formes dans les contextes du présent et de l'infinitif pourrait elle aussi être influencée par l'aspect lexical inhérent. La première hypothèse sera donc formulée dans les termes suivants :

Hypothèse 1 – influence de l'aspect lexical

La production des formes courtes et longues des verbes réguliers en *-er* par les apprenant/es est influencée par l'aspect lexical des verbes : ils produisent la forme courte avec les verbes statiques et la forme longue avec les verbes dynamiques.

Comme on l'avait vu lors de la présentation des verbes en *-er* dans le tableau 2-1, on peut voir une distribution générale des formes courtes et longues, dans le

sens où la forme courte correspond au présent et la forme longue au non-présent. On peut ainsi se demander si l'opposition statique *vs* dynamique ne serait pas le reflet de la fréquence des différents verbes au présent *vs* non-présent, les verbes étant plus ou moins fréquemment utilisés dans ces deux contextes.

Ainsi selon l'hypothèse 2, la production des apprenant/es est le reflet de la fréquence d'occurrence de la forme courte et de la forme longue pour les verbes pris individuellement dans l'input. La forme la plus fréquente d'un verbe est la forme la plus accessible. Comme on l'a vu à la fin du chapitre 5, on peut supposer que la proportion des formes courtes et longues varie d'un verbe à l'autre selon sa fréquence d'utilisation au présent, au passé ou au futur, ce qui permet de formuler cette hypothèse à partir d'un continuum.

Hypothèse 2 – influence de la fréquence des formes de surface

La production des formes courtes et longues des verbes réguliers en *-er* par les apprenant/es est influencée par la fréquence d'occurrence de ces formes dans l'input. Les apprenant/es produisent la forme la plus fréquente du verbe dans l'input, les verbes étant répartis sur un continuum allant des verbes fréquents à la forme courte aux verbes fréquents à la forme longue.

Ces hypothèses ont uniquement été formulées pour les verbes réguliers en *-er*, mais on peut les transposer pour les verbes irréguliers. On s'attend ainsi à la forme du présent singulier à la place de la forme courte et une autre forme que le présent pour la forme longue, en particulier la forme de l'infinitif ou du participe passé.

Le chapitre 7 qui suit présentera les principaux choix qui ont été faits dans le but de vérifier ces deux hypothèses.

7. Principaux choix méthodologiques

L'objectif de ce chapitre est de présenter les principaux choix méthodologiques qui ont été faits pour vérifier les deux hypothèses formulées au chapitre précédent. La première section présentera le choix des sources de données pour l'input et les apprenant/es. La deuxième section introduira le choix des verbes pour l'analyse ainsi que leur classement à partir de l'aspect lexical. La dernière section présentera la manière de traiter les données à savoir la définition des formes et contextes, leur codage ainsi que les méthodes statistiques choisies.

7.1. Provenance des données empiriques

Les données empiriques de ce travail proviennent de trois sources différentes : des données d'input, des données d'apprenant/es de corpus et des données de participants à la partie expérimentale.

7.1.1. Données d'input

Tout d'abord, dans le but de travailler à partir d'occurrences concrètes au niveau de l'input pour l'hypothèse 2, j'ai choisi quatre sources de données d'input. Ces données ont été choisies parce qu'elles pourraient refléter les principales variétés de français auxquelles des apprenant/es débutant/es peuvent être exposés, même s'il ne s'agit pas de la même intensité pour chacune d'elles. Ces sources de données comportent trois sources d'input oral : les données de deux interlocutrices en conversation avec des apprenant/es débutant/es, les données d'enseignants en classe de français LE/L2 au lycée ainsi que les données de corpus de français parlé en France du DVD C-ORAL-ROM (2005), et une source de données écrites, celles de la méthode de français pour apprenant/es débutant/es suédophones, *Escalade 1* (Tillman *et al.*, 2001). Les données de cette méthode nous donnent indirectement accès à un usage de la langue en classe de débutants. Ces sources de données d'input sont présentées plus en détail dans la partie méthodologique du chapitre 8 sur la fréquence des formes dans l'input.

7.1.2. Données d'apprenant/es

Pour ce qui est des données d'apprenant/es, elles proviennent donc d'une part de corpus de conversation et d'autre part de la partie expérimentale.

Des données de corpus permettent d'avoir accès à des données de production libre relativement spontanées, même si les conversations sont souvent guidées par les interlocutrices. Ces données proviennent d'une part du corpus Thomas et d'autre part du corpus Lund (Schlyter).

a) Apprenant/es du corpus Thomas

Le corpus Thomas est un corpus pilote constitué de 20 apprenant/es adultes (moyenne d'âge 57 ans) apprenant le français dans le cadre de cours du soir, à raison de 90 minutes par semaine. Il a été constitué en 2003 et 2004 dans le but de travailler sur la morphologie verbale, notamment sur les formes du type *je parle* ou *je lire*²⁴.

La plupart des enregistrements sont d'une durée de 10 à 15 minutes. Quelques apprenant/es ont été enregistrés deux ou trois fois (une fois par semestre) durant 20 à 30 minutes. Ce corpus compte au total 25 conversations, transcrites et codées à l'aide du logiciel CLAN²⁵ (MacWhinney, 2000). Environ la moitié des apprenant/es avaient déjà appris le français à l'école dans leur jeunesse (faux débutants), l'autre moitié a commencé dans le cadre de ces cours du soir. Comme on peut le voir dans le tableau 7-1, les apprenant/es les plus débutants ont un semestre de cours derrière eux (un peu plus de 24 heures de cours) et les plus avancés quatre semestres (un peu plus de 96 heures). Tous les apprenant/es ont le suédois comme première langue, sauf une (TAM) qui a le norvégien, et ils ont tous l'anglais comme première langue étrangère. La plupart savent également l'allemand. Tous les apprenant/es ont une formation universitaire ou équivalente.

Le tableau 7-1 présente les noms de code des 20 apprenant/es du corpus Thomas avec à gauche les débutants et à droite les personnes qui avaient déjà des notions de français au moment de commencer les cours du soir. Le nombre de semestres concerne les cours du soir.

²⁴ Mon rôle est triple dans ce corpus : enseignante, interlocutrice et étudiante-chercheuse.

²⁵ Disponible sur le net : <http://childes.psy.cmu.edu/clan/> (page consultée le 13 août 2009).

Tableau 7-1 : Corpus Thomas : nombre de semestres par apprenant/e

débutants	semestres	faux débutants	semestres
CAR 1	1	BER 1	0
CAR 2	2	BER 2	1
CAT 1	2	DEN	0
CAT 2	3	JAC	0
CAT 3	4	MEL	0
FLO	1	TAM	0
FRE 1	1	RUT	0
FRE 2	2	JEA	1
PIE	2	MAR	1
RIC	2	JUL	3
LUC	4	SAN	3
SEB	4	SAR	4
		TAN	4

Les enregistrements contiennent des conversations que l'on peut qualifier de semi-spontanées, l'interlocutrice (INT-A) menant la conversation par un certain nombre de questions, dans le but de faire parler les apprenant/es (cf. section 8.1.1.). Les apprenant/es sont néanmoins libres de produire ce qu'ils souhaitent. Lors de la première série d'enregistrements, les apprenant/es savaient à l'avance qu'ils devraient se présenter et parler de leurs loisirs. Ils avaient tous déjà eu ce genre de conversations en classe. Pour les apprenant/es ayant été enregistrés une deuxième et une troisième fois, le cadre de discussion était plus libre mais toujours assez personnel. Les apprenant/es ayant plus de difficultés à s'exprimer ont pu choisir de parler à partir d'une image représentant une personne au début de la conversation. En général, les questions favorisaient plutôt l'expression du présent, moi, ici et maintenant.

Le niveau de français des apprenant/es *débutants* de ce corpus est estimé aux stades 1-2 de Bartning et Schlyter (2004) alors que les *faux débutants* sont estimés aux stades 2 et 3. Cette estimation est basée sur une analyse détaillée de quelques apprenant/es dans une étude antérieure (Thomas, 2003)²⁶. Les stades ont été présentés à la section 3.2. (voir aussi l'annexe 1).

On peut mentionner en passant, que les données des apprenant/es de ce corpus forment en quelque sorte le point de départ historique de ce travail. Les premières analyses de l'influence de l'input sur la production des apprenant/es ont en effet été faites à partir de ce corpus (Thomas, 2006). De nombreux choix, en particulier celui des verbes, découlent du matériel trouvé dans ce corpus.

²⁶ Il s'agit des apprenants CAR1, CAT1, FRE1, BER1 et JEA.

b) Karl et Lisa du corpus Lund (Schlyter)

Les données du corpus Thomas sont complétées par celles de deux apprenants du corpus Lund (Schlyter), Karl, trois enregistrements (sur cinq) et Lisa, quatre enregistrements. Ces apprenants ont été choisis parce qu'ils utilisent beaucoup la forme longue (*parlE*). Karl (23 ans) est un apprenant qui a essentiellement appris le français sur place en France, alors que Lisa (20 ans) a appris le français à l'école en Suède durant 6 ans, à raison de ± 3 heures par semaine en moyenne. Tous les enregistrements sont d'une durée d'environ 60 minutes et ils ont été menés à environ 2 mois d'intervalle. L'activité dominante est celle de la conversation, principalement sur la vie actuelle, passée et future de l'apprenant/e, mais il y a également un certain nombre de tâches (raconter un conte de fée, décrire un chemin). Avec Karl, les conversations traitent avant tout de son travail et de ses voyages passés ou futurs alors qu'avec Lisa, le thème principal est celui des études de français. Pour plus de détails sur ce corpus et les apprenant/es voir Granfeldt (2003 : 65-70), Bozier (2005 : 147) et Bartning & Schlyter (2004 : 284).

Selon l'analyse de Bartning et Schlyter (2004 : 294), les trois premiers enregistrements de Lisa (1 à 3) sont estimés au stade 2 et Lisa 4 au stade 3. Karl 1 est estimé aux stades 2-3 et Karl 2 et 3 au stade 3 (Karl 4 et 5 ne sont pas utilisés dans la présente étude car ils sont évalués au stade 4).

7.1.3. Données expérimentales

Dans le but d'obtenir des données plus ciblées pour vérifier les hypothèses, les données de corpus ont été complétées par des données expérimentales. Ces données proviennent de 33 apprenant/es et de 10 natifs (groupe contrôle). Les apprenant/es sont des lycéens (moyenne d'âge 17,7 ans) qui suivent des cours de français à un niveau débutant dans différentes classes de lycées de Lund et de Malmö. Les participant/es à la partie expérimentale sont présentés à la section 10.5.

Ces participant/es ont effectué les trois tâches suivantes :

- a) une tâche de production libre dans laquelle les apprenant/es racontent une histoire à partir d'une série d'images (chapitre 11)
- b) un test expérimental d'imitation, dans lequel les apprenant/es répètent à voix haute des phrases enregistrées (chapitre 12)
- c) un test expérimental de rappel libre, dans lequel les apprenant/es répètent à voix haute des séries de formes verbales (chapitre 13)

Les données d'apprenant/es proviennent donc d'une part de tâches en production libre, en conversation ou en monologue et d'autre part de deux tests expérimentaux. La validité de données expérimentales par rapport à des données de production libre est souvent discutée (*cf.* par exemple R.Ellis, 2008 ; Tarone, 1998). La citation ci-dessous résume bien la problématique.

[...] experimental elicitation may only tell us what learners can produce under conditions of experimental elicitation and may or may not reflect what they can do under more natural conditions of language use. (R.Ellis & Barkhuizen, 2005 : 36)

Comme le relève Hulstijn (1997), si l'utilisation de données expérimentales implique un certain degré d'abstraction, elle a l'immense avantage de permettre le contrôle d'un grand nombre de dimensions, du choix du test au traitement des données en passant par la consigne et les stimuli.

Dans le cas où les données obtenues dans les deux conditions devaient indiquer des tendances très différentes, la validité des données expérimentales sera discutée.

7.2. Choix des verbes

Dans ce travail, j'ai choisi de vérifier les hypothèses à partir d'un nombre limité de verbes.

7.2.1. Présentation des verbes choisis

Le choix des verbes pour cette étude s'est fait en plusieurs étapes. L'objectif principal a été de trouver des verbes utilisés dans la plupart des sources de données et en particulier chez les apprenant/es les plus débutants (corpus Thomas). J'ai choisi un nombre limité de verbes afin de pouvoir les comparer dans leur utilisation dans l'input et par les apprenant/es. J'ai ainsi choisi d'analyser 16 verbes réguliers en *-er* dans la partie des données de corpus et 12 dans la partie expérimentale (le choix des verbes pour la partie expérimentale est présenté à la section 10.3. Même si cette étude est avant tout centrée sur l'utilisation des verbes réguliers, il est intéressant d'analyser des données de verbes irréguliers à titre de contrôle. En effet, contrairement aux verbes réguliers en *-er* (chapitre 2) les formes non-finies et les formes du passé sont souvent distinctes à l'oral pour les verbes irréguliers. Il arrive par contre que les formes

du présent singulier et du participe passé soient les mêmes. La liste des verbes étudiés dans ce travail est présentée dans le tableau 7-2.

Tableau 7-2 : Listes des verbes étudiés dans ce travail

verbes réguliers input, apprenants et expérimental	verbes réguliers input et apprenants	verbes irréguliers input et apprenants
acheter	jouer	aller
adorer	étudier	connaître
aimer	travailler	comprendre
détester	voyager	croire
habiter		dire
manger		faire
parler		lire
penser		pouvoir
préférer		prendre
regarder		savoir
trouver		voir
visiter		vouloir

7.2.2. Classement des verbes selon leur aspect lexical

Pour pouvoir vérifier l'hypothèse 1, il est nécessaire d'établir l'aspect lexical des verbes choisis. Pour cela, il existe un certain nombre de tests (Vendler, 1957 ; Platzack, 1979 ; Vet, 1980 ; Robison, 1995). Des tests de classement permettent de réduire, mais non d'éliminer, les biais subjectifs des juges (Housen, 1994 : 282-283). Comme on l'a vu dans la présentation de l'aspect lexical (section 4.1.), le verbe représente une situation et on pourrait aller jusqu'à établir l'aspect lexical dans chaque contexte d'énonciation, en prenant en compte le contexte discursif (premier plan – arrière plan, dialogue, narration ou discours argumentatif), le contexte temporel et aspectuel de la situation à laquelle il est fait référence. Robison (1995) et Weist (2002) proposent que l'établissement de l'aspect lexical d'un verbe se fasse loin de tout contexte, avec les verbes à l'infinitif :

Tests such as these promote the classification of Aktionsart separately from viewpoint aspect and help to avoid a circular argument. (Weist 2002 : 36)

Une autre manière de faire est de demander à des juges de représenter les verbes à partir d'un certain nombre de codes graphiques, comme l'ont fait Bonnotte *et al.* (1993 ; Bonnotte, 2008).

L'hypothèse 1 est formulée à partir d'une opposition entre les verbes statiques vs dynamiques. Ces deux catégories de verbes ne sont pas toujours clairement distinctes, car les verbes dynamiques peuvent être utilisés dans un sens statique : "Many activities (and some accomplishments and achievements) have a 'derived' state sense." (Vendler, 1957 : 152). En outre, Vendler (1957 : 150-151) souligne le fait que les habitudes dans le sens large du terme ("occupations, dispositions, abilities, and so forth") sont des états. Il y a donc un certain champ de variation au niveau de l'établissement de l'aspect lexical des verbes²⁷.

Dans le présent travail, je chercherai, dans la mesure du possible, à classer les verbes en tant que lexèmes en essayant de capter leur aspect lexical prototypique.

Pour distinguer les verbes d'état des verbes dynamiques, trois tests sont proposés. En principe on ne peut pas répondre aux questions *que s'est-il passé ?* et *à quelle heure ?* en utilisant un verbe d'état (Vet, 1980 : 55-56) et l'utilisation de l'adverbe *lentement* est incompatible avec un sens statique, puisqu'il n'y a pas de processus à décrire (Platzack, 1979 : 93-94). L'exemple (7-1) présente ces tests avec d'autres verbes que les verbes utilisés dans ce travail. L'astérisque * indique que la réponse ne semble pas possible ou uniquement dans un contexte très spécifique. Dans cet exemple le verbe *admirer* apparaît comme un verbe statique (et le verbe *tomber* apparaît comme un verbe dynamique).

(7-1)	Que s'est-il passé ?	À quelle heure ?	Lentement
	Le verre est tombé	Le verre est tombé à 16h	Le verre est tombé lentement
	*Il a admiré l'acteur	*Il a admiré l'acteur à 16h	*Il a admiré l'acteur lentement

Comme on l'a vu au début du chapitre 4, un verbe d'état présente une situation stable qui existe sans apport d'énergie et qui se termine au moment où quelque chose change l'état (Bonnotte *et al.*, 1993 ; Andersen, 1991, 1993, 2002 ; Vendler, 1957). Les verbes dynamiques présentent des activités, des actions ou des événements qui demandent un apport d'énergie (volontaire ou non) à un moment donné du processus (Bonnotte *et al.*, 1993 ; Andersen, 1991, 1993, 2002 ; Vendler, 1957).

On peut en outre mentionner qu'il y a "incompatibilité absolue de la préposition *de* avec le passif des verbes dynamiques : *Ce plat a été cuisiné de ma femme" (Riegel *et al.*, 1994 : 437).

Le tableau 7-3 présente la répartition des verbes qui seront analysés dans ce travail à partir des outils et définitions présentés ci-dessus. Les verbes pour

²⁷ Cf. aussi le processus de *recatégorisation* (Sasse, 2002) ou de *coercion* (DeSwart, 1998).

lesquels il est difficile de trancher sur un aspect lexical prototypique ont été mis dans les deux catégories avec un point d'interrogation.

Tableau 7-3 : Répartition des verbes selon aspect lexical prototypique

verbes réguliers		verbes irréguliers	
statiques	dynamiques	statiques	dynamiques
adorer	acheter	connaître	aller
aimer	jouer	? comprendre	? comprendre
détester	étudier	croire	dire
habiter	manger	pouvoir	faire
penser	parler	savoir	lire
préférer	regarder	? voir	prendre
? trouver	travailler	vouloir	? voir
	? trouver		
	visiter		
	voyager		

a) Commentaires sur les verbes réguliers

Le fait qu'on trouve moins de verbes à double appartenance pour les verbes réguliers en *-er* n'est pas une caractéristique de ces verbes, mais le reflet d'un choix conscient des verbes pour ce travail. J'ai en effet choisi des verbes que l'on peut classer assez clairement selon aspect lexical prototypique.

Le verbe *trouver* est le seul verbe régulier à appartenir aux deux catégories d'aspect lexical. Il est en effet intéressant de prendre un exemple de verbe avec une double appartenance et de confronter les deux hypothèses. Le sens statique peut se définir comme étant celui dans le sens de *trouver* – à son avis comme dans *je trouve cette maison jolie* ainsi que celui de *se trouver quelque part* alors que le sens dynamique est celui de *trouver ce que l'on cherche*, un verbe dynamique télélique. Les données présenteront ce verbe sous trois formes : *trouver S* pour statique, *trouver D* pour dynamique et *trouver Σ* pour le verbe au total, à savoir S + D.

Le verbe *parler* est un verbe dynamique, mais comme on l'a vu plus haut, les verbes dynamiques ont aussi un sens statique. L'utilisation de ce verbe dans des énoncés comme *je parle allemand, français et anglais* renvoie plutôt à un sens statique. Une telle utilisation du verbe est probablement courante avec les apprenant/es LE/L2 débutants.

On peut noter que le choix des verbes statiques pour ce travail est assez particulier par rapport à d'autres études. Comme le dit Bardovi-Harlig (2000 : 236) les verbes statiques sont souvent dominés par les occurrences des verbes *être* et *avoir* et leurs équivalents. Il est vrai que les verbes auxiliaires et modaux sont tous statiques. Dans la liste des verbes choisis, le verbe *habiter* est le seul

verbe qui ne puisse pas être utilisé dans une fonction modalisante (*j'aime voyager, je pense partir*) et c'est d'ailleurs aussi le seul verbe qui ne décrive pas un état mental et qui représente un contenu concret. On pourrait ainsi dire que le verbe *habiter* est le verbe le moins statique du groupe.

Finalement on peut considérer que le nombre de verbes statiques différents est trop élevé par rapport à ce que l'on trouve naturellement dans la langue. D'un autre côté le fort déséquilibre entre le nombre de verbes statiques et le nombre de verbes dynamiques pose souvent problème pour l'interprétation des résultats (Bardovi-Harlig, 2002).

b) Commentaires sur les verbes irréguliers

La liste des verbes irréguliers comprend un plus grand nombre de verbes à double appartenance.

Le verbe *aller* dans son sens statique correspond à son utilisation en tant qu'auxiliaire ainsi que dans le sens *d'aller bien, mal* alors que le sens dynamique, télique correspond au sens de *se rendre à une destination*. Dans ce travail seul le sens dynamique sera analysé.

Le verbe *faire* est avant tout un verbe dynamique, mais il a un sens statique lorsqu'on parle de la météo, *il fait beau*, et en général dans son utilisation impersonnelle, par exemple dans les utilisations de *ça fait mal, ça fait longtemps*, etc.

Le verbe *comprendre* a une double appartenance qui est assez forte, car il est parfois classé comme verbe télique (Kihlstedt, 1998 : 81) et parfois comme verbe statique (Salaberry & Ayoun, 2005 : 9) dans la littérature. Pour ce verbe, il est peut-être difficile de séparer l'aspect grammatical de l'aspect lexical. Dans une utilisation au passé composé comme dans *j'ai compris le problème* le sens sera plutôt ponctuel alors que dans l'utilisation au présent, comme marqueur conversationnel ou pour introduire une subordonnée, *je comprends, je comprends que*, le sens est plutôt statique (de même pour l'utilisation de l'imparfait *je comprenais*). Au moment de tester le verbe, même à l'infinitif, le résultat est probablement différent selon la forme conjuguée à laquelle on pense. Mais, contrairement au verbe *trouver*, le sens du verbe *comprendre* reste proche dans les deux sens aspectuels : il s'agit toujours de <comprendre>.

Finalement, le verbe *voir* est un verbe proprement polysémique. Kihlstedt (1998 : 81-82) l'a mis dans toutes les catégories (cf. aussi Howard, 2004 : 327). Comme pour le verbe *comprendre* le sens statique du verbe *voir* correspond à l'utilisation du verbe comme marqueur conversationnel ou pour introduire une

subordonnée et on peut également dire que le sens sémantique <voir> reste présent dans toutes les utilisations de ce verbe.

La classification des verbes a été comparée à celle de Kihlstedt (1998 : 80-82) et elle correspond pour tous les verbes. Les verbes réguliers ont en outre été discutés avec Lisa Christensen qui travaille depuis plusieurs années sur l'aspect lexical en suédois (*cf.* Christensen, à paraître). Elle a accepté le classement des verbes sauf le verbe *visiter* (en suédois *besöka*) qu'elle considère comme un verbe statique, du moins en suédois. Le fait que l'on puisse utiliser le même verbe pour *rendre visite à qn* et *visiter un musée* en suédois joue peut-être un rôle.

L'aspect lexical des verbes a ainsi été établi à partir d'une réflexion métalinguistique. Le classement présenté dans le tableau 7-3 plus haut servira de base à l'hypothèse 1. Il va de soi qu'il devra être discuté au fur et à mesure de l'analyse des données.

7.3. Traitement des données

7.3.1. Formes verbales et contextes

L'analyse des données dans les sources de données d'input se fera à partir de toutes les formes *parle* et *parlE* présentées dans le tableau 2-1 pour les verbes réguliers en *-er* dans l'input. Pour les verbes irréguliers, il s'agira des formes correspondantes, à savoir, celles du présent singulier, 2^e et 3^e du pluriel, celles de l'imparfait singulier et 3^e du pluriel, de l'infinitif et du participe passé.

Pour les données d'apprenant/es, l'utilisation des formes des verbes en *-er* et les contextes au centre de ce travail ont été présentés dans le tableau 1-1 de l'introduction. Il s'agit donc d'analyser l'utilisation correcte et incorrecte de la forme courte du type *parle* qui correspond à la forme du présent singulier et 3^e du pluriel et de la forme longue du type *parlE* qui correspond à l'infinitif (*parler*), au participe passé (*parlé*), à l'imparfait singulier et 3^e du pluriel (*parlais*) ainsi que la 2^e du pluriel du présent (*parlez*) comme on l'a vu dans le tableau 2-1.

Le contexte du présent correspond aux contextes d'utilisations du présent en français alors que l'infinitif correspond au placement d'une forme verbale après une préposition (*pour parler*), après un verbe modal (*je veux parler*) ou semi

modal (*j'aime parler*), l'auxiliaire du futur proche (*je vais parler*) et dans les autres constructions demandant l'infinitif.

7.3.2. Codage des formes

Le codage des verbes dans les exemples et les tableaux pour les apprenant/es se fera à partir de la combinaison entre la forme produite et le contexte dans lequel la forme a été insérée. Pour plus de clarté sur les utilisations correctes et incorrectes, les codes de ces dernières seront précédés d'un astérisque*. Le tableau 7-4 présente les codes utilisés dans le chapitre 9 sur les données d'apprenant/es de corpus et ceux pour la tâche de production libre de la partie expérimentale dans le chapitre 11.

Tableau 7-4 : Codage de la production des verbes réguliers en –er

description	code	exemple
forme longue dans contexte de l'infinitif	LdsINF	<i>je veux parLE, pour parLE</i>
forme courte dans contexte de l'infinitif	*CdsINF	<i>je veux parle, pour parle</i>
forme courte dans contexte du présent	CdsPRE	<i>maintenant je parle</i>
forme longue dans contexte du présent	*LdsPRE	<i>maintenant je parLE</i>

Au chapitre 9, les verbes irréguliers seront analysés selon le même principe que les verbes réguliers. Les formes considérées dans les données d'apprenant/es seront la forme courte du présent singulier (*vois*), la forme de l'infinitif (*voir*) et celle du participe passé (*vu*) dans les contextes pour une forme du présent et de l'infinitif.

7.3.3. Format des données (logiciels)

Les transcriptions des corpus Thomas et Lund (Schlyter), interlocutrices et apprenant/es, ainsi que celles des enseignants de lycée ont été utilisées dans le format CHAT du logiciel CLAN (MacWhinney, 2000). Les principales conventions de transcription ont été mises dans la liste des abréviations après la table des matières au début de ce travail. La compilation des données a été faite à l'aide de différents programmes inclus dans le logiciel CLAN, dont COMBO qui permet de voir les formes en contextes.

Les données du DVD C-ORAL-ROM (2005) sont uniquement accessibles sur celui-ci. La compilation des données a été faite à l'aide du logiciel de concordance *contextes* fourni avec le DVD.

Les données expérimentales ont été enregistrées à l'aide d'un appareil numérique et elles ont été transcrites à l'aide d'un logiciel de traitement de texte

pour la tâche de narration libre et dans un tableur pour les tests d'imitation et de rappel. Les enregistrements ont été écoutés à l'aide du logiciel *iTunes*.

7.3.4. Méthodes statistiques

La mesure de comparaison entre les différentes sources de données sera le pourcentage de formes courtes. J'ai choisi cette mesure car la forme courte correspond à une seule fonction, celle du présent, contrairement à la forme longue comme on l'a vu au chapitre 2 et que le lien entre l'aspect lexical statique et cette forme a été attesté dans plusieurs études à partir de l'aspect lexical.

Les données provenant d'apprenant/es débutant/es et le nombre de verbes analysés étant relativement petit, le nombre d'occurrences est modeste et sujet à variation d'un verbe à l'autre.

Pour cette raison, les calculs statistiques se feront à partir de méthodes non paramétriques. De telles méthodes sont considérées comme ayant moins de force mais sont par contre plus robustes. Le risque de faire une erreur de type I est plus élevé avec de tels tests qu'avec des tests paramétriques. Cela signifie que l'on peut obtenir des différences significatives alors qu'il n'y en a pas (Heiman, 2006 : 370). Il est donc important de discuter les données de manière qualitative. Contrairement aux méthodes paramétriques qui effectuent les calculs à partir de moyennes, les méthodes non paramétriques utilisent les occurrences brutes. Ces méthodes peuvent également s'utiliser avec des pourcentages, en particulier dans le cas de comparaison d'échelles. Une telle comparaison sera en effet faite pour l'hypothèse 2, dans laquelle l'influence des formes de surface sera mesurée à partir d'un classement des verbes selon le pourcentage de formes courtes (continuum)

Pour la vérification de l'hypothèse 1 sur l'influence de l'aspect lexical, je me servirai du test de Mann-Whitney U qui est un test analogue au t-test. Ce test sera uniquement utilisé pour les verbes réguliers en *-er*. Pour les verbes irréguliers on a en effet deux verbes sur douze (*voir* et *comprendre*) qui sont à la fois des verbes statiques et des verbes dynamiques et pour lesquels il est difficile de séparer les deux sens.

Pour l'hypothèse 2, je me servirai du test de corrélation du r de Spearman pour la comparaison des échelles de pourcentage de formes courtes. Le coefficient de corrélation de rangs de Spearman, r , décrit la relation linéaire entre deux variables résumées par ordre de rang (Heiman, 2006 : 158ss). Si l'on obtient 1 cela signifie que la corrélation est totale alors qu'un 0 signifie qu'il n'y

a aucune corrélation. La corrélation est souvent vérifiée avec un graphique présentant les données et la ligne de corrélation parfaite (1). Ce test permettra donc de vérifier si l'échelle de pourcentage de formes courtes dans l'input et celles des apprenant/es sont corrélées au niveau du classement des formes les unes après les autres. On considère traditionnellement qu'une corrélation, significative ou non, d'au moins 30% est une bonne corrélation. La significativité des corrélations ne sera donc pas présentée dans les résultats.

PARTIE DONNÉES DE CORPUS

8. Données d'input

L'objectif de ce chapitre est d'obtenir des données d'input pour les verbes choisis pour ce travail afin de pouvoir établir le continuum de fréquence des formes qui servira de base à la vérification de l'hypothèse 2.

Ce chapitre comporte trois parties. La première partie présentera les choix méthodologiques, en particulier les différentes sources de données d'input. La partie empirique présentera d'abord l'étude des verbes réguliers en *-er* puis, celle des verbes irréguliers dans l'input. Ce chapitre se terminera sur une discussion et les prédictions concrètes pour l'hypothèse 2 comparées à celles de l'hypothèse 1.

8.1. Choix méthodologiques

8.1.1. Sources de données d'input

Le choix des sources de données d'input est motivé par les différents groupes d'apprenant/es dont les données sont analysées dans ce travail :

- des apprenant/es qui prennent des cours du soir de français (débutants du corpus Thomas, *cf.* section 7.1.2.)
- des apprenant/es qui ont appris le français dans leur jeunesse, qui ont passé beaucoup de temps en France pour les vacances et qui reprennent des cours du soir à un niveau débutant (faux-débutants du corpus Thomas)
- un apprenant qui a essentiellement appris le français sur place en France (Karl du corpus Lund – Schlyter, *cf.* section 7.1.2)
- une apprenante en première année de français à l'université (Lisa du corpus Lund – Schlyter)
- des lycéens (participants à la partie expérimentale, *cf.* section 10.5.)

Étant donné que ce travail est centré sur la production orale, j'ai tout d'abord choisi un certain nombre de sources de données de français parlé :

- les données d'interlocutrices en conversation avec des apprenant/es

- les données d'enseignants du français LE/L2 de lycée en Suède
- les données du corpus C-ORAL-ROM (2005)

Ces sources de données semblent couvrir la plupart des usages de la langue orale à laquelle les apprenant/es sont exposés : en effet la plupart des apprenant/es ont des conversations avec une personne plus compétente qu'eux (enseignants ou personnes qu'ils rencontrent en France), la plupart suivent des cours de français et la plupart ont accès à un français que l'on peut appeler standard, soit directement sur place en France, soit par le biais de matériel audio-visuel authentique ou non. Même si la proportion d'exposition à ces usages du français parlé peut varier d'un/e apprenant/e à l'autre, on peut dire que tous les apprenants sont potentiellement exposés à ces inputs. Ces sources de données correspondent d'ailleurs *grosso modo* aux trois variétés d'input décrites par R.Ellis, *foreigner talk*, *teacher talk*, et *native speaker usage* (R. Ellis, 1994 : 243ss).

Ces sources de données d'input oral ont été complétées par les données d'une méthode de français pour débutants suédophones, *Escalade 1* (Tillman *et al.*, 2001). Contrairement aux enfants apprenant leur(s) L1, les adultes ont directement accès à l'écrit. L'utilisation d'un livre de cours dès les premiers contacts avec la langue cible en tant qu'adulte est plutôt la règle que l'exception. Sans entrer dans les détails du traitement de l'écrit en LE/L2, les occurrences d'une méthode peuvent nous informer sur la fréquence des formes que les apprenant/es sont censés traiter lors de leurs premiers contacts avec la langue cible et indirectement sur l'usage probable de la langue en classe, par les enseignants comme par les apprenant/es.

Ces quatre sources de données sont présentées ci-après.

a) Interlocutrices en conversation avec des apprenant/es

Cette source de données comprend la production de deux interlocutrices natives du français en conversation avec des apprenant/es. Il s'agit d'une part de l'interlocutrice des apprenant/es du corpus Thomas (INT-A) et d'autre part de l'interlocutrice de Lisa du corpus Lund (INT-C) (*cf.* section 7.1.2.). Ces deux interlocutrices ont l'expérience de l'acquisition d'une ou plusieurs langues étrangères/secondes, notamment le suédois, ainsi que de l'enseignement du français LE/L2. Leurs données contiennent essentiellement des questions, le rôle des interlocutrices étant de faire parler les apprenant/es. De manière générale, la langue qu'elles utilisent est dénuée de constructions complexes tout en restant naturelle et spontanée. Leur utilisation de la langue française pourrait avoir une

influence directe sur l'apprenant avec lequel elles sont en train de discuter, au-delà des prises directes (*cf.* Bozier, 2005).

L'INT-A est également l'enseignante des apprenant/es du corpus Thomas (et l'auteur du présent travail). Sa manière d'utiliser les différentes formes verbales pourrait donc avoir une influence assez importante sur la production des apprenant/es du corpus Thomas.

Les données de l'INT-A sont caractérisées par le fait qu'elle vouvoie ses interlocuteurs et une faible variation de vocabulaire. L'INT-A s'était en effet tenue à une grille de conversation pour les 20 premiers enregistrements des apprenant/es²⁸. La principale référence temporelle est celle du présent (moi, ici et maintenant) même s'il y a aussi quelques références au passé et au futur. Des exemples de questions typiques posées par l'INT-A sont présentés en (8-1).

- (8-1) Est ce que vous pouvez vous présenter?
Qu'est ce que vous aimez faire quand vous êtes libre?
Est ce que vous jouez au golf, au bridge ?
Qu'est-ce que vous aimez comme films ?
Vous allez souvent en France?
Qu'est ce que vous faites quand vous êtes en France?

Les données de l'interlocutrice INT-C de l'apprenante Lisa du corpus Lund sont plus variées. Il s'agit de quatre conversations différentes, chacune d'une durée d'une heure. Les principaux thèmes discutés sont la vie actuelle, passée et future de l'apprenante, en particulier les études de français. L'apprenante a également été invitée à produire un certain nombre de tâches (raconter un conte, décrire un chemin, etc.). Comme le montrent les quelques exemples de questions typiques en (8-2), l'INT-C tutoie l'apprenante.

- (8-2) Et les autres matières, qu'est-ce que tu en penses ? les autres heu <cours?> sujets ?
oui les autres cours ?
Je crois que tu avais le choix entre le français et le suédois . qu'est-ce que tu as choisi alors ?
Hm et pour l'écriture ? est-ce que tu as trouvé des manières pour améliorer ton écriture ?

²⁸ Les questions étaient un peu plus variées lors de nouveaux enregistrements des mêmes apprenant/es.

b) Enseignants de français au lycée en Suède

Cette source de données comprend la production des enseignants de français de trois classes de lycées en Suède du corpus Flyman Mattsson (2003). Ces données ont été compilées à partir de 39 enregistrements de leçons de durées variées, pour un total d'environ 12 heures. Les enregistrements ont été faits dans trois classes à deux ans d'intervalle et les leçons enregistrées pour chaque année sont des leçons successives (Flyman Mattsson, 2003 : 34-35). Pour une des classes, il y a eu un changement d'enseignant entre la 1^{ère} et la 3^e année (enseignants LA). Trois enseignants ont le suédois comme langue maternelle et une enseignante est locutrice native du français (enseignante LC). Les élèves ont 16-17 ans en première année et 18-19 ans en troisième de français. Au moment de commencer les enregistrements, les élèves ont déjà eu 3 ans de français LE/L2. Le niveau de ces élèves est donc en principe plus avancé que celui des apprenant/es analysé/es dans le présent travail. Un résumé des données sur les enseignants est présenté dans le tableau 8-1 (pour plus d'informations sur ce corpus voir Flyman Mattsson 2003 : 33ss).

Les transcriptions transmises contiennent uniquement celles des *activités*, correction d'exercices, discussions générales ou sur un texte, traduction, grammaire (Flyman Mattsson, 2003 : 49-50). Comme on le verra la part de français et de suédois varie beaucoup d'un enseignant à l'autre. Les enregistrements d'exposés des élèves ainsi que de travail en groupes n'ont donc pas été utilisés ici, l'intérêt se situant au niveau de la production des enseignants.

Tableau 8-1 : Données sur les enseignants du corpus Flyman Mattsson

	classe A		classe B		classe C	
année	1 ^{ère}	3 ^e	1 ^{ère}	3 ^e	1 ^{ère}	3 ^e
sexe	homme	femme	homme	homme	femme	femme
code	LA	LA ²⁹	LB	LB	LC	LC
langue maternelle	suédois	suédois	suédois	suédois	français	français
nombre d'enregistrements ³⁰	8	4	7	8	5	7
et durée ³¹	180 min	28 min	142 min	126 min	149 min	114 min

Ces enseignants ont des manières très différentes d'utiliser et d'enseigner le français.

²⁹ Bien qu'il ne s'agisse pas du même enseignant, le code est le même, les données de l'enseignante de 3^e année étant peu nombreuses.

³⁰ Nombre d'enregistrements que j'ai eus à ma disposition.

³¹ Le nombre de minutes a été calculé d'après les données mentionnées dans les transcriptions.

L'enseignant LA de la 1^{ère} année est un enseignant que l'on pourrait qualifier de traditionnel, la forme semble l'emporter sur la communication. Il utilise très peu de mots français, en somme il parle essentiellement suédois (28 % des mots produits selon les données de Flyman Mattsson, 2003 : 77) et les verbes sont essentiellement présentés sous la forme des temps primitifs par exemple *parler-parlant- parlé- je parle- je parlai*. Cette manière d'utiliser les verbes correspond à environ un tiers des occurrences des verbes choisis pour ce travail. Cet enseignant utilise très peu de phrases complètes en français et les verbes sont avant tout produits dans l'objectif d'en enseigner les formes aux élèves comme le montre l'exemple (8-3) ci-dessous.

(8-3)	Travail sur les formes verbales (Classe A, 1 ^{ère} année – A3 ; ELC = élève)	Traduction (en <i>italique</i> les mots en français dans l'original)
*LA1:	parler infinitiv # infinitiv att tala ## är ni med att tala parlant # det är presens particip ## talande är det ofta man säger talande på franska tror ni parlant # vad tyckte Nadia?	<i>parler</i> infinitif # infinitif parler ## vous suivez parler <i>parlant</i> # c'est le participe présent ## parlant vous croyez qu'on dit souvent parlant en français <i>parlant</i> # que pense Nadia ?
*ELC:	jag tror inte det.	je ne crois pas
*LA1:	du tror inte det nej det här är inte ofta man säger parlant man säger det oftare än man säger talande på er franska på svenska men det är en annan historia parlé # det är perfekt particip # och det betyder jag kanske ska göra så # det betyder talat.	tu ne crois pas non ce n'est pas souvent qu'on dit <i>parlant</i> on le dit plus souvent qu'on dit parlant en euh français en suédois mais c'est une autre histoire <i>parlé</i> # c'est le participe passé # et cela signifie oui peut-être je vais faire comme ça # ça signifie parlé.

L'enseignante LA de la 3^e année parle très peu (1162 occurrences de mots au total) – les élèves font des exposés – mais essentiellement en français (95% des mots produits selon Flyman Mattsson, 2003 : 77). On peut ainsi conclure que les apprenant/es de cette classe ont peu l'occasion d'avoir accès à un flot de français parlé. L'enseignant de première année présente les formes verbales dans un flot de langue suédoise et l'enseignante de troisième année laisse les élèves parler. En somme l'input de ces élèves est essentiellement celui de leurs co-apprenant/es.

L'enseignant LB a la particularité de présenter la plupart des verbes avec un contexte et il communique en français avec les élèves (63 % de mots en français en 1^{ère} année et 83 % en 3^e selon Flyman Mattsson, 2003 : 77). Cet enseignant

cite rarement un verbe à l'infinitif sans immédiatement le faire suivre de la forme du présent singulier. Voici un exemple (8-4) pour cet enseignant.

(8-4) Activité de traduction (Classe B, 1 ^{ère} année – B7 ; ELJ = élève)	Traduction (en <i>italique</i> les mots en français dans l'original)
*LB1: un ami visst men det är ## det är ju lika rätt båda var det någon som valde nåt annat än faire des études man ja man kan säga+...	<i>un ami</i> bien mais c'est ## mais tous les deux sont corrects est-ce que quelqu'un avait choisi autre chose que <i>faire des études</i> on peut dire +...
*ELJ: qui étudie.	<i>qui étudie.</i>
*LB1: ja alltså istället för qui fait des études som bedriver studier vilket är riktigt så kan man skriva étudie med verbet étudier il étudie à Bruxelles mm www.	oui donc à la place de <i>qui fait des études</i> qui fait des études ce qui est correct on peut écrire <i>étudie</i> avec le verbe <i>étudier</i> il <i>étudie</i> à Bruxelles mm www.

Finalement l'enseignante LC est celle qui produit le plus haut pourcentage de mots en français en 1^{ère} année avec 67% des mots produits (91% en 3^e, Flyman Mattsson, 2003 : 77). Bien qu'elle parle essentiellement français, elle utilise également le suédois, notamment pour traduire ce qu'elle dit. Elle ne présente jamais les verbes au niveau des temps primitifs, mais elle utilise beaucoup la conjugaison des verbes au présent, qu'elle décline en entier. Voici un exemple (8-5) pour cette enseignante.

(8-5) Activité de comparaison de deux images (Classe C, 1 ^{ère} année – C4 ; ELI = élève)
*LC1: la dame avec un tricot un tricot orange avec un tricot orange # qu' est+ce qu' elle fait elle+...
*ELI: <i>lire</i> une er livre.
*LC1: elle lit un livre lit lire ah je lis tu lis il lit s s t elle lit un livre c' est vrai sur ici et puis là non # autre chose Mayser+...

La production des enseignants est donc assez variée et elle reflète probablement les principaux styles d'enseignement auxquels on peut être exposé en tant qu'élève en classe de langue.

Il serait intéressant d'analyser les données des élèves de ces classes dans le cadre d'un travail semblable à celui-ci pour voir l'impact de l'input des enseignants sur leur production (d'autres analyses sur l'influence de l'input des enseignants sur ces élèves ont été faites par Flyman Mattsson, 2003). On peut en effet se demander dans quelle mesure l'input utilisé en classe reflète une utilisation véritable de la langue. Du moment où les formes sont présentées de manière isolée dans un discours en suédois comme dans l'exemple (8-3) ou dans

un paradigme entier comme dans l'exemple (8-5) elles ne sont probablement pas perçues de la même manière que dans un flot continu de français. Les formes de ces verbes sont peut-être présentées de manière plus saillante, mais les élèves ne peuvent pas s'aider des constructions syntaxiques pour noter et mémoriser l'utilisation des formes verbales.

c) DVD C-ORAL-ROM (2005)

Les données du corpus de C-ORAL-ROM (2005) comprennent la transcription de 164 enregistrements de français parlé en France. Au total C-ORAL-ROM (2005) contient 300'000 mots provenant de différents corpus³². Même si le nombre de mots n'est pas très élevé, cette base de données a été préférée à celle de *Corpaix* de J.Véronis³³ (Véronis, 2000). En effet, la liste de *Corpaix* ne présente que les formes, sans contexte de phrase, ni étiquetage des formes, ce qui pose surtout un problème pour les formes homographes (par exemple *fait*, 3^e personne du présent singulier, participe passé et substantif). La plupart des enregistrements de C-ORAL-ROM datent de 2001 et 2002. On y trouve des monologues (par exemple un exposé), des conversations entre deux ou plusieurs personnes dans des contextes privés et publics, des extraits d'émissions télévisées ainsi que des conversations au téléphone (y compris avec des textes de machines) (C-ORAL-ROM, 2005 : 111-133). Pour des raisons pratiques, tous les enregistrements ont été pris en compte dans cette étude.

d) Manuel scolaire Escalade 1

Le manuel scolaire *Escalade 1* (Tillman *et al.*, 2001) comprend deux livres, un livre de textes et un livre d'exercices. Le livre de textes comporte 20 chapitres avec à chaque fois un texte à visée didactique, c'est-à-dire le texte étudié dans le chapitre concerné, et un document authentique en français avec sa traduction en suédois. Le livre d'exercices comprend des exercices sur le texte du chapitre, des exercices de vocabulaire, de grammaire (exercices à trous, traduction, conjugaison), de rédaction ainsi que des exercices auditifs et de communication par paires. Ce manuel suédois est destiné à des élèves débutants en français LE/L2 au lycée ou plus tard à l'âge adulte. Il est couramment utilisé au lycée

³² Aix-en-Provence 18 enregistrements, Corpus de Référence du Français parlé 31 enregistrements, corpus collecté pour le projet C-ORAL-ROM (2005) 115 enregistrements.

³³ *Corpaix* est accessible sur le site <http://sites.univ-provence.fr/veronis/data/freq-oral.txt> (page consultée le 22 août 2009).

8.1.2. Choix des verbes

Ce travail est basé sur l'étude d'un nombre limité de verbes qui ont été présentés à la section 7.2. Le choix des verbes pour ce travail a été fait à partir des verbes les plus utilisés par les apprenant/es de corpus, en particulier du corpus Thomas. Il était en effet important de choisir des verbes lexicaux que les apprenant/es les plus débutants utilisent en production libre.

Ce choix ne correspond pas vraiment à ce qui est produit par l'enseignant LA et l'INT-C chez lesquels on trouve peu de données pour plusieurs de ces verbes (cf. annexes 2 et 3).

Comme on l'a vu au chapitre 7, l'analyse des données de fréquence dans l'input est d'une part basée sur 16 verbes réguliers en *-er*. Pour ces verbes, l'analyse portera sur la proportion de formes courtes *parle* et longues *parlE* telles que présentées dans le tableau 2-1. Cette analyse sera complétée par les formes correspondantes de 12 verbes irréguliers (uniquement dans le corpus d'input oral).

8.1.3. Compilation des formes

La compilation des formes verbales choisies pour ce travail a été faite de différentes manières selon les sources de données.

La compilation des formes des interlocutrices natives et des enseignants a été réalisée à l'aide du programme COMBO du logiciel CLAN (MacWhinney, 2000). Ce logiciel présente les formes recherchées dans les énoncés. L'utilisation de ce programme a permis de corriger les éventuelles erreurs d'orthographe dans les transcriptions (par exemple *parler* à la place de *parlé*).

La compilation des formes du DVD C-ORAL-ROM (2005) a été faite à l'aide du logiciel de concordance *contextes* fourni avec le DVD. Ce logiciel permet la recherche détaillée des formes sur le corpus étiqueté. Les formes verbales sont présentées dans un format de concordance, présentant une dizaine de mots avant et après la forme recherchée. Ce logiciel permet également de demander une liste complète des occurrences des différentes formes du verbe choisi accompagnées de leur code pour le mode, le temps et la personne grammaticale.

Pour le manuel *Escalade 1*, les formes verbales des verbes en *-er*³⁴ ont été compilées manuellement dans les deux livres, textes et exercices. Pour le livre de textes, les occurrences ont été comptées dans les textes à visée didactique comme dans les textes authentiques. Dans le cahier d'exercice, les formes

³⁴ Les formes des verbes irréguliers n'ont pas été compilées dans ce manuel.

verbales des verbes en *-er* ont été comptées dans tous les exercices y compris dans le matériel supplémentaire (*extra*, par exemple une recette de cuisine) et les titres. Dans les tableaux de conjugaison, toutes les formes des paradigmes présentés ont été comptées, même les formes dans le titre. Dans la série d'exercices de prononciation à la fin du livre d'exercice, seuls les verbes donnés en français ont été comptés. Dans les exercices de dialogues par paires, chaque forme verbale a été comptée une fois. Aucune spéculation n'a été faite sur les formes qui pourraient être produites par les apprenant/es. En somme, toute forme des verbes choisis imprimée en français a été comptée dans le manuel.

8.2. Analyse empirique des données d'input

8.2.1. Verbes réguliers en *-er* dans l'input

L'objectif de cette analyse est d'établir la proportion de formes courtes et longues pour chaque verbe pour pouvoir placer les verbes sur une échelle allant des verbes fréquents à la forme courte aux verbes fréquents à la forme longue. Cette analyse sert donc à obtenir des données concrètes pour l'hypothèse 2 sur la fréquence des formes de surface.

Les formes prises en compte dans les calculs sont les formes courtes et longues telles que présentées dans le tableau 2-1, à savoir les formes du présent singulier et 3^e du pluriel (*parle, parles, parle, parlent*) représentées par la forme *parle*, celles de l'imparfait singulier et 3^e du pluriel (*parlais, parlais, parlait, parlaient*) représentées par la forme *parlais*, la forme de la 2^e du pluriel du présent *parlez*, du participe passé *parlé* et de l'infinitif *parler*.

Les seules formes exclues du calcul sont les occurrences dont la transcription ne permet pas de connaître le contexte visé³⁵.

Les occurrences d'autorépétition ont toutes été comptées (*on peut on peut dire* est compté comme deux occurrences de *peut*).

L'utilisation de formes de l'impératif est assez importante dans le manuel *Escalade 1*, surtout à la 2^e du pluriel avec 71 occurrences, ce qui correspond à environ 10% des occurrences du manuel (sans compter les verbes qui n'apparaissent qu'une fois). Il s'agit avant tout des titres d'exercices *communiquez, écoutez, prononcez* avec une occurrence par chapitre. On n'a par

³⁵ Par exemple *dans quelques minutes on en parler* (corpus Flyman Mattsson) ou *on pas trouver* (C-ORAL-ROM, 2005).

contre que six occurrences de la 2^e du singulier, toutes du verbe *regarder*. Chez les enseignants, on trouve aussi quelques occurrences de l'impératif, notamment neuf occurrences de *regardez*. Les occurrences de l'impératif n'entrent pas dans les calculs.

Les occurrences des formes courtes et longues correspondent à plus de 90% des occurrences pour la plupart des verbes dans C-ORAL-ROM (2005). Ce résultat confirme que la forme courte et la forme longue couvrent la plupart des occurrences en français parlé comme je l'avais suggéré au chapitre 2. Pour les verbes où l'on n'atteint pas le 90% des occurrences, il manque soit des formes du conditionnel, *aimerais* et *préférerais*, soit de l'impératif, *regardez* et *regarde*.

Pour ce qui est du manuel *Escalade 1*, outre les formes de l'impératif, on trouve également six occurrences de la 1^{ère} du pluriel du présent, dont trois avec le verbe *habiter*. Les formes choisies couvrent donc aussi la plupart des occurrences de ce manuel pour les verbes choisis. À noter qu'il n'y a pas d'occurrences de verbe en *-er* à l'imparfait dans *Escalade 1*.

Le tableau 8-2 ci-dessous présente les résultats de l'analyse de la fréquence de chaque forme, la forme courte et les quatre formes longues, des 16 verbes en *-er* dans les sept sources de données d'input (deux interlocutrices, trois enseignants, C-ORAL et *Escalade 1*). Les occurrences et pourcentages pour chaque source de données spécifique et chaque verbe se trouvent dans l'annexe 2. Les données du tableau 8-2 indiquent la somme des occurrences de chaque forme. Le pourcentage de fréquence de chaque forme (présenté entre parenthèses) a été obtenu en divisant le nombre total d'occurrences de la forme en question par la somme des formes courtes et longues produites pour ce verbe. La présentation des formes verbales dans le tableau 8-2 est différente de ce que l'on trouve dans l'annexe 2, l'objectif ici étant de pouvoir mettre en évidence la fréquence des différentes formes dans l'input pour chaque verbe individuel, alors que l'annexe sert à rapidement repérer les occurrences.

Le tableau 8-2 se lit horizontalement et verticalement. Les formes verbales ont été ordonnées horizontalement selon le nombre d'occurrences au total de chaque forme. Les formes les plus fréquentes se trouvent en position 1 à gauche et les moins fréquentes en position 5 à droite. Le classement vertical a été fait à partir du pourcentage de formes courtes (en gris clair) pour chaque verbe. Dans le cas d'un pourcentage identique, les verbes ont été classés selon le nombre d'occurrences de la forme courte. Ainsi les verbes dans le haut du tableau 8-2 ont un haut pourcentage de formes courtes et les derniers un faible pourcentage

de forme courte (et donc un haut pourcentage de formes longues). Pour mémoire, les formes du type *parle* représentent les occurrences des formes du présent singulier et 3^e du pluriel et celles du type *parlais* celles de l'imparfait singulier et 3^e du pluriel.

Tableau 8-2 : Fréquence des formes des verbes en –er dans l'input

1	2	3	4	5
adore 26 (81%)	adorer 3 (9%)	adorais 2 (6%)	adorez 1 (3%)	adoré 0 (0%)
pense 270 (73%)	pensais 30 (8%)	penser 28 (8%)	pensez 23 (6%)	pensé 17 (5%)
trouve S 131 (72%)	trouvé S 18 (10%)	trouvais S 17 (9%)	trouvez S 10 (5%)	trouver S 6 (3%)
habite 108 (72%)	habitais 20 (13%)	habitez 15 (10%)	habiter 6 (4%)	habité 1 (1%)
déteste 7 (70%)	détester 3 (30%)	détesté 0 (0%)	détestais 0 (0%)	détestez 0 (0%)
préfère 26 (58%)	préfèrez 10 (22%)	préféré 4 (9%)	préfèrais 3 (7%)	préférer 2 (4%)
aime 187 (55%)	aimez 85 (25%)	aimé 27 (8%)	aimer 22 (6%)	aimais 20 (6%)
trouve Σ 185 (47%)	trouvé Σ 102 (26%)	trouver Σ 71 (18%)	trouvais Σ 21 (5%)	trouvez Σ 14 (4%)
joue 72 (42%)	jouer 56 (33%)	jouez 21 (12%)	joué 16 (9%)	jouais 6 (4%)
regarde 86 (39%)	regarder 53 (24%)	regardez 41 (18%)	regardé 35 (16%)	regardais 8 (4%)
parle 220 (38%)	parler 174 (30%)	parlé 92 (16%)	parlais 49 (8%)	parlez 46 (8%)
travaille 98 (37%)	travailler 92 (34%)	travaillé 30 (11%)	travaillez 27 (10%)	travillais 20 (7%)
manger 61 (38%)	mange 47 (29%)	mangé 38 (24%)	mangez 11 (7%)	mangais 3 (2%)
trouvé D 84 (40%)	trouver D 65 (31%)	trouve D 54 (26%)	trouvez D 4 (2%)	trouvais D 4 (2%)
acheté 44 (40%)	acheter 38 (35%)	achète 16 (15%)	achetez 10 (9%)	achetais 2 (2%)
voyagé 18 (35%)	voyager 14 (27%)	voyagez 10 (19%)	voyage 7 (13%)	voyageais 3 (6%)
étudier 24 (35%)	étudié 21 (31%)	étudiez 15 (22%)	étudie 7 (10%)	étudiais 1 (1%)
visité 12 (48%)	visiter 10 (40%)	visite 2 (8%)	visitez 1 (4%)	visitais 0 (0%)

Légende : trouver S : sens statique ; D : sens dynamique ; Σ : sens statique + dynamique.

Le nombre d'occurrences des verbes est bas pour la plupart des verbes et tout particulièrement pour les verbes *adorer*, *détester* et *visiter*. Les résultats présentés indiquent donc avant tout des tendances.

La présentation des occurrences des différentes formes verbales dans le tableau 8-2 ci-dessus nous permet de constater que les verbes se laissent regrouper selon la fréquence des différentes formes analysées. On trouve ainsi quatre groupes de verbes :

1. Un groupe de verbes pour lesquels la forme courte est particulièrement fréquente avec 70% et plus d'occurrences de la forme courte. Il s'agit des verbes *adorer*, *penser*, *habiter*, *trouver* (statique) et *détester*.
2. Deux verbes pour lesquels la forme courte est plus fréquente que la forme longue mais à un pourcentage plus proche de 50% que le groupe précédent. Il s'agit des verbes *préférer* et *aimer*. Ce résultat est avant tout lié à la production de l'INT-A qui vouvoie les apprenant/es. Comme on le voit dans le tableau 2 en annexe, la forme de la 2^e du pluriel pour les verbes *aimer* et *préférer* a été produite à plus de 80% des occurrences par l'INT-A. Cela signifie que ces verbes sont fréquents au présent, mais pas toujours à la forme courte.
3. Un groupe de verbes pour lesquels la forme courte du présent singulier et la forme longue de l'infinitif se trouvent dans les deux premières positions. Pour ces verbes, le pourcentage de formes courtes a passé en dessous de 50%, mais la forme courte est la forme avec le plus grand nombre d'occurrences (sauf pour le verbe *manger*). On peut ainsi dire que ces verbes sont fréquents à la forme courte comme à la forme longue, donc qu'aucune forme ressort comme étant clairement la plus fréquente. Pour les verbes *trouver* (au total), *jouer*, *regarder*, *parler* et *travailler* la forme courte est plus fréquente que la forme de l'infinitif alors que c'est l'inverse pour le verbe *manger*. On peut ainsi dire que ce groupe de verbes est caractérisé par une plus grande variation dans l'utilisation des contextes syntaxiques pour une forme courte et longue, que les autres groupes de verbes. Comme on le voit dans l'annexe 2 ainsi que dans le tableau 8-3 plus loin dans le texte, le haut pourcentage de formes courtes avec ces verbes est avant tout influencé par les données du manuel *Escalade 1*. L'utilisation de la référence au présent, ici et maintenant, est en effet une des caractéristiques des premières utilisations d'une langue cible.

4. Finalement, les verbes dans le bas du tableau 8-2 forment un groupe de verbes pour lesquels la forme longue est particulièrement fréquente et où les deux formes les plus fréquentes en occurrence sont les formes non-finies de l'infinitif et du participe passé. Il s'agit des verbes *trouver* (*dynamique*), *acheter*, *voyager*, *étudier* et *visiter*.

Selon l'hypothèse 2, on s'attend donc à la production de la forme courte par les apprenant/es pour les verbes des groupes 1 et 2, ces verbes étant fréquents au présent. Le fait que les verbes du groupe 2 soient fréquents à la forme de la 2^e du pluriel en particulier chez l'INT-A, permettra de vérifier l'influence de cette forme dans les données des apprenant/es de corpus. Pour les verbes du groupe 3 on s'attend à plus de variation que pour les verbes des deux autres groupes. Ces verbes n'apparaissent pas clairement comme des verbes au présent ni au non-présent. On pourrait ainsi s'attendre à un pourcentage de formes courtes et longues plus proche de 50% chez les apprenant/es ou du moins une préférence moins marquée pour une des formes que pour les autres verbes. Finalement pour les verbes du groupe 4, on s'attend à l'utilisation de la forme longue par les apprenant/es, ces verbes étant fréquents à la forme longue dans l'input. Ce regroupement des verbes selon leur fréquence sera utilisé pour simplifier la discussion, mais comme mentionné au chapitre 7, les calculs statistiques seront faits à partir du continuum des verbes³⁶.

Comme on l'a vu dans la présentation et l'analyse des données, les sources de données d'input choisies présentent certaines spécificités. Le tableau 8-3 ci-dessous présente une comparaison des occurrences des 16 verbes pour les quatre groupes de sources de données. Les verbes sont classés dans le même ordre que dans le tableau 8-2 à savoir selon le pourcentage de formes courtes au total (TOT INP). Pour chaque source de données, on a d'abord le nombre d'occurrences du verbe au total puis le pourcentage de formes courtes. La dernière ligne présente le total des occurrences pour chaque source de données ainsi que le pourcentage de formes courtes (sans les données de *trouver* Σ).

³⁶ La différence de pourcentage de formes courtes entre les trois groupes de verbes – fréquents à la forme courte, à aucune des formes, à la forme longue – est significative (Mann-Whitney U, $p < .01$) dans l'input au total.

Tableau 8-3 : Comparaison des sources d'input

	PROFS		INT		C-ORAL		ESC		TOT INP	
	Σ occ	%C	Σ occ	%C	Σ occ	%C	Σ occ	%C	Σ occ	%C
adorer	0	-	0	-	11	73%	21	86%	32	81%
penser	31	52%	29	69%	307	76%	1	0%	368	73%
trouver S	10	40%	14	57%	153	75%	5	100%	182	72%
habiter	18	78%	34	74%	54	54%	44	91%	150	72%
détester	2	50%	0	-	4	100%	4	50%	10	70%
préférer	8	63%	6	0%	27	63%	4	100%	45	58%
aimer	26	77%	104	20%	142	58%	69	93%	341	55%
trouver Σ	51	14%	19	42%	286	52%	37	57%	393	47%
jouer	5	0%	33	33%	90	36	43	67%	171	42%
regarder	37	14%	39	13%	128	47%	19	84%	223	39%
parler	142	38%	90	29%	319	36%	30	80%	581	38%
travailler	35	31%	41	34%	171	33%	20	85%	267	37%
manger	75	32%	11	9%	62	21%	12	75%	160	29%
trouver D	41	7%	5	0%	133	26%	32	50%	211	26%
acheter	19	0%	5	40%	72	11%	14	43%	110	15%
voyager	4	0%	28	7%	16	25%	4	25%	52	13%
étudier	17	18%	33	6%	14	7%	4	25%	68	10%
visiter	2	0%	8	0%	13	15%	2	0%	25	8%
Total	472	34%	480	29%	1716	47%	328	77%	2996	46%

Légende : PROFS : enseignants de lycée ; INT : interlocutrices natives ; C-ORAL : C-ORAL-ROM (2005) ; ESC : manuel Escalade 1 ; TOT INP : input au total ; %C : pourcentage de formes courtes ; Σ occ : somme des occurrences du verbe pour la source de données concernée ; trouver S : sens statique ; D : sens dynamique ; Σ : sens statique + dynamique.

Si l'on regarde la dernière ligne du tableau 8-3, on peut constater que le pourcentage de formes courtes est particulièrement élevé pour le manuel *Escalade 1*. En effet, seuls les derniers verbes de la liste et le verbe *détester* ont 50% et moins de formes courtes. Cela implique que l'input adressé aux apprenant/es les plus débutants est fortement marqué par la fréquence de la forme courte. À l'inverse, les données des interlocutrices et des enseignants ont un faible pourcentage de formes courtes et seuls quelques verbes sont plus fréquents à la forme courte que longue.

Un calcul statistique de corrélation (r de Spearman) du pourcentage de formes courtes entre les différentes sources de données d'input indique les corrélations suivantes (sans les données de *trouver Σ*) :

Tableau 8-4 : Corrélation entre les pourcentage de formes courtes dans l'input

	PROFS	INT	C-ORAL	ESC	TOT INP
PROFS	1.000	.356	.721	.571	.772
INT	.356	1.000	.439	.212	.613
C-ORAL	.721	.439	1.000	.474	.925
ESC	.571	.212	.474	1.000	.556
TOT INP	.772	.613	.925	.556	1.000

Légende : voir tableau 8-3.

À l'exception de la corrélation entre les données du manuel *Escalade 1* et celles des interlocutrices, toutes les corrélations sont bonnes (au-delà de 35%). La corrélation entre les différentes sources de données et le TOTAL se situe entre .556 et .925. La corrélation entre le manuel *Escalade 1* et l'input au total est la plus faible. Un nouveau calcul mais cette fois sans les données du verbe *penser* pour lequel on a une seule occurrence dans *Escalade 1* donne une corrélation de .810. Ces résultats signifient que les pourcentages de formes courtes obtenus au total (TOT INP) sont suffisamment représentatifs pour pouvoir être utilisés comme base pour la comparaison avec les données des apprenant/es.

Si l'on regarde les données du tableau 8-2 du point de vue des catégories aspectuelles (hypothèse 1), on peut constater que les verbes statiques correspondent aux sept premiers verbes du tableau 8-2, des verbes qui sont tous plus fréquents à la forme courte que longue, alors que les verbes dynamiques sont tous plus fréquents à la forme longue que courte (pour la distribution des verbes selon l'aspect lexical voir section 7.2.2.). Les deux hypothèses se recoupent donc au niveau des verbes d'état alors que les verbes dynamiques se divisent en deux groupes au niveau de la fréquence des formes : les verbes fréquents aux deux formes et les verbes fréquents à la forme longue. C'est cette différence au niveau du pourcentage de formes courtes qui permettra de discuter les deux hypothèses dans les données d'apprenant/es.

On peut mentionner en passant que les données présentées ne montrent pas d'arguments en faveur d'une catégorisation plus fine des verbes dynamiques pour ce travail. En effet, les verbes téléiques *trouver* (*dynamique*) et *acheter* ne se distinguent pas des autres verbes dynamiques. Le groupe de verbes fréquents aux deux formes non-finies, infinitif et participe passé, n'appartiennent pas non plus à une catégorie aspectuelle propre. On y trouve en effet des verbes d'activité comme des verbes téléiques.

Un test statistique sur le pourcentage de formes courtes selon l'aspect lexical des verbes montre que la différence est significative pour toutes les sources de données sauf pour les interlocutrices, qui ont un haut pourcentage de formes longues en *-ez* avec les verbes statiques, et plus faible avec le manuel *Escalade 1*, qui a un haut pourcentage de formes courtes aussi avec les verbes dynamiques. Le tableau 8-5 présente ces résultats.

Tableau 8-5 : Test statistique pour l'aspect lexical

	PROFS	INT	C-ORAL	ESC	TOT INP
Mann-Whitney U	,000	13,000	,000	15,000	,000
Wilcoxon W	55,000	68,000	55,000	70,000	55,000
Z	-3,278	-1,475	-3,418	-1,957	-3,418
Asymp. Sig. (2-tailed)	,001	,140	,001	,050	,001

Légende : voir tableau 8-3.

La différence entre les deux hypothèses pourrait aussi se voir dans la production du verbe *trouver*. Comme on l'a vu dans la présentation des verbes à la section 7.2.2. et tout au long du présent chapitre, les données du verbe *trouver* sont présentées de trois manières: d'une part le sens statique (*à mon avis, se trouver quelque part*), d'autre part le sens dynamique (*trouver ce que l'on cherche*) et finalement dans les deux sens pris ensemble.

Une analyse des occurrences du verbe *trouver* dans les différentes sources de données orales montre que l'utilisation la plus fréquente du sens statique, à *son avis*, dans C-ORAL-ROM (2005) est *je trouve* (*je trouve que ..., ce tableau est intéressant je trouve, je trouve ça drôle* ; 110 occurrences de *trouver à son avis*, dont 82 au présent et 68 avec le sujet *je*). Le sens statique sert donc avant tout de verbe de pensée et de marqueur discursif. Chez les enseignants et les interlocutrices, le sens statique du verbe est moins fréquent à la forme courte. Leurs données ne sont en effet pas réellement des données de conversation libre, dans laquelle chacun exprime son avis. L'INT-A produit six questions avec *vous trouvez* (exemple 8-6) et l'enseignant LA a une longue séquence où il cherche des élèves qui ont *trouvé le texte intéressant, ennuyeux*, etc. (*quelqu'un qui a trouvé le texte ennuyeux ## pourquoi # pourquoi est+ce que c'est ennuyeux?* A3, 3^e) ce qui donne cinq occurrences au participe passé.

- (8-6) *INT: vous trouvez que c' est difficile d' apprendre le français ?
 *CAT2: oui , difficile , très très très difficile

Les données du tableau 8-2 ont montré une distinction claire au niveau des formes associées aux deux sens du verbe *trouver*, la forme courte avec le sens statique et la forme longue (participe passé) avec le sens dynamique. Les données du verbe au total (Σ) reflètent cette distribution de manière quasiment parfaite en présentant la forme courte en première position, mais à seulement 47%, et la forme longue du participe passé en deuxième position. Dans le manuel *Escalade 1* ainsi que chez les enseignants, le sens dynamique est plus fréquent en nombre d'occurrences que le sens statique (cf. tableau 8-3), mais au niveau des occurrences au total, le sens dynamique est seulement légèrement plus fréquent que le sens statique. Selon l'hypothèse 1, on s'attend à la production de la forme courte avec le sens statique et la forme longue avec le sens dynamique par les apprenant/es et pour l'hypothèse 2, on s'attend à ce que ce verbe soit produit selon la fréquence des formes au total, donc comme un verbe fréquent aux deux formes.

8.2.2. Verbes irréguliers dans l'input oral

L'objectif de cette section est de présenter la fréquence des formes de verbes irréguliers dans le corpus d'input *oral* de ce travail (cette étude ne comprend donc pas de données du manuel *Escalade 1*) pour pouvoir vérifier l'hypothèse 2 au niveau des verbes irréguliers sur les données des apprenant/es de corpus. Comme on l'a vu dans l'introduction à ce chapitre, il est intéressant de regarder l'influence des formes des verbes irréguliers de l'input sur la production des apprenant/es car la plupart des formes verbales sont distinctes à l'oral. On trouve donc une meilleure correspondance entre forme et contexte dans l'input qu'avec les verbes en *-er*.

Les formes compilées pour les verbes irréguliers sont les mêmes que pour les verbes en *-er* – toutes les formes du présent sauf celle de la 1^{ère} du pluriel, les formes de l'imparfait singulier et 3^e du pluriel, de l'infinitif et du participe passé – sauf que la 3^e du pluriel du présent a été comptée à part si elle n'était pas homophone avec les formes du singulier (*prennent* par exemple). Les formes homophones et homographes de la 3^e singulier du présent et du participe passé (*fait*, *dit*) ont été comptées séparément.

Les données du verbe *aller* concernent uniquement le sens lexical du verbe, à savoir toutes les utilisations du type *je vais à Paris* ou des expressions comme *ça allait vite*. Les formes de l'infinitif citées par les enseignants comme dans

l'énoncé *il existe des verbes irréguliers comme aller*³⁷ (enseignant LA) ont également été comptées comme lexicales. Les utilisations du verbe comme auxiliaire (*je vais travailler*), dans le sens *ça va* (y compris *un style qui leur va bien ; ça vous va bien ; tout va bien*) et finalement les constructions modales à l'imparfait comme dans *j'allais dire une bêtise, j'allais leur parler*, n'entrent donc pas dans les calculs mais les occurrences sont présentées dans l'annexe 3 sous *aller otr*.

Les formes verbales dans les expressions *en fait, point de vue, c'est-à-dire*³⁸ ainsi que la forme *connu* comme adjectif³⁹ n'ont pas été comptées. L'occurrence du verbe *dire* dans l'énoncé *mm que que dîtes le groupe trois* (LA, A8-1^{ère}) et du verbe *faire* dans *quel temps a-t-il faisait* (LB, B10-3^e) du corpus des enseignants n'ont pas été comptées, la transcription ne permettant pas de savoir quel était le contexte visé.

Le tableau 8-6 ci-dessous présente les résultats de l'analyse des 12 verbes irréguliers dans les sources de données d'input oral. Une présentation détaillée des occurrences pour chaque source de données se trouve dans l'annexe 3. Les données du tableau 8-6 indiquent la somme des occurrences de chaque forme. Le pourcentage de fréquence de chaque forme (présenté entre parenthèses) a été obtenu en divisant le nombre total d'occurrences de la forme en question par la somme des formes retenues dans le calcul pour ce verbe.

Les formes verbales ont été ordonnées horizontalement selon le nombre d'occurrences de chaque forme au total. Les formes les plus fréquentes se trouvent en position 1 à gauche et les moins fréquentes en position 5 à droite. Le classement vertical a été fait à partir du pourcentage de formes du présent singulier (en gris clair) pour chaque verbe.

Pour alléger le tableau 8-6, les formes du présent singulier et celles l'imparfait (singulier et 3^e du pluriel) sont représentées par une seule forme, la 1^{ère} du singulier (par exemple *crois* et *croyais*).

³⁷ Traduction de *det finns ju oregelbundna typ aller också #*.

³⁸ Ces formes apparaissent en effet comme des occurrences des verbes au présent, au participe passé et à l'infinitif respectivement dans C-ORAL-ROM (2005).

³⁹ La forme de l'énoncé *vous vous êtes connus* (INT-A) n'a pas été comptée non plus.

Tableau 8-6 : Fréquence des formes des verbes irréguliers dans l'input

1	2	3	4	5	6
crois/nt 204 (80%)	croire 22 (9%)	croyais 13 (5%)	cru 9 (4%)	croyez 6 (2%)	
sais 646 (70%)	savez 103 (11%)	savoir 99 (11%)	savais 43 (5%)	savent 21 (2%)	su 12 (1%)
peux 782 (65%)	pouvez 118 (10%)	pu 105 (9%)	pouvoir 76 (6%)	pouvais 71 (6%)	peuvent 48 (4%)
veux 417 (61%)	voulais 98 (14%)	voulez 93 (14%)	voulu 37 (5%)	veulent 27 (4%)	vouloir 10 (1%)
vois/nt 374 (42%)	voir 267 (30%)	vu 164 (18%)	voyez 42 (5%)	voyais 41 (5%)	
connais 109 (40%)	connaissez 51 (19%)	connaître 39 (14%)	connu 39 (14%)	connaissais 26 (9%)	connaissent 10 (4%)
comprends 64 (33%)	comprendre 58 (30%)	compris 42 (22%)	comprenez 21 (11%)	comprennent 5 (3%)	comprenais 2 (1%)
faire 892 (35%)	fais 746(29%)	fait 496 (20%)	faisais 172 (7%)	faites 130 (5%)	font 103 (4%)
dire 584 (36%)	dis 449 (28%)	dit 389 (24%)	disais 102 (6%)	dites 51(3%)	disent 32 (2%)
prendre 128 (30%)	pris 128 (30%)	prends 110 (26%)	prenez 21 (5%)	prenais 19 (5%)	prennent 14 (3%)
aller lex 177 (32%)	vais lex 137 (25%)	allé lex 99 (18%)	allez lex 63 (11%)	allais lex 47 (9%)	vont lex 28 (5%)
lire 62 (35%)	lisez 45(25%)	lu 43 (24%)	lis 22 (12%)	lisent 4 (2%)	lisais 2 (1%)

En général, le nombre d'occurrences pour les verbes irréguliers est plus élevé que celui des verbes réguliers, ce qui correspond à une différence typique entre verbes réguliers et irréguliers comme on l'a vu au chapitre 5. On peut aussi voir qu'il y a plus de variation au niveau des formes qui arrivent en deuxième position que pour les verbes réguliers en *-er*. Les deux formes les plus fréquentes sont néanmoins les formes du présent singulier et de l'infinitif, qui se partagent les deux premières places pour sept des douze verbes.

Si l'on regarde le nombre d'occurrences et le pourcentage des formes en première position, on obtient trois groupes de verbes :

1. Les verbes dans le haut du tableau 8-6 dont la forme du présent singulier est la plus fréquente avec un pourcentage supérieur à 60%. Il s'agit des verbes *croire*, *savoir*, *pouvoir* et *vouloir*.
2. Les verbes au milieu du tableau 8-6 dont la forme du présent singulier a le plus d'occurrences mais avec un pourcentage en dessous de 50%. Il s'agit des verbes *voir*, *connaître*, *comprendre*.

3. Les verbes dans le bas du tableau 8-6 dont la forme de l'infinitif est la plus fréquente avec un pourcentage supérieur à 60%. Il s'agit des verbes *faire*, *dire*, *prendre*, *aller*⁴⁰ et *lire*. Pour le verbe *prendre* les deux formes non-finies, infinitif et participe passé se partagent la première position et c'est le seul verbe dont les deux formes non-finies sont les plus fréquentes.

Selon l'hypothèse 2, on s'attend, chez les apprenant/es, à la production de la forme du présent pour les verbes du groupe 1, ces verbes étant fréquents à cette forme dans l'input et à la forme de l'infinitif pour les verbes du groupe 3, cette forme étant la plus fréquente dans l'input. Pour les verbes du groupe 2, on s'attend à plus de variation dans la production des formes que pour les autres groupes de verbes.

Si l'on regarde les données d'input selon l'aspect lexical des verbes (hypothèse 1), on peut voir que les verbes statiques *croire*, *savoir*, *pouvoir*, *vouloir* et *connaître* sont associés à la forme du présent et que les verbes dynamiques *faire*, *dire*, *prendre*, *aller* et *lire* sont associés à la forme de l'infinitif dans l'input présenté ici. On peut néanmoins noter que le verbe *connaître* est moins fréquent à la forme courte que les autres verbes statiques, mais qu'il est fréquent au présent, la forme arrivant en deuxième position étant la 2^e du pluriel.

Les verbes *croire*, *savoir*, *vouloir*, mais aussi *voir* et *comprendre* sont souvent utilisés comme marqueurs discursifs ou dans des expressions toutes faites telles que *si tu veux*, *je veux dire*, *ça veut dire*, *si vous voulez*, *je crois (que)*, *je (ne) sais pas/plus*, *tu sais*, *vous savez* dans C-ORAL-ROM (2005)⁴¹. Pour le verbe *voir*, les expressions comme *tu vois* correspondent à plus de 60% des formes *vois* dans C-ORAL-ROM (2005). Ainsi les verbes *voir* et *comprendre* sont plus fréquents dans leur utilisation statique que dynamique dans l'input présenté ici.

On peut aussi mentionner que le sens statique du verbe *faire* dans son emploi impersonnel comme dans les expressions liées à la météo ou dans des expressions du type *ça fait*, *ça se fait* ou *ce qui fait que* correspond à un tiers des occurrences de *fait* au présent dans C-ORAL-ROM (2005).

Le fait que les cinq verbes dynamiques dans le bas du tableau 8-6 sont plus fréquents à la forme de l'infinitif signifie que les verbes dynamiques ne sont pas

⁴⁰ Les autres utilisations du verbe (auxiliaire, *aller bien*, etc. cf. plus haut) montent par contre à 77% de formes du présent singulier (cf. annexe 3 sous *aller otr*)

⁴¹ Sur l'utilisation des verbes comme marqueurs discursifs voir par exemple Dostie (2004).

particulièrement fréquents à la forme du participe passé dans cet input, même pour les verbes qui ont aussi un sens télique comme *voir*, *comprendre*, *dire* et *prendre*. Si l'on regarde les occurrences de ces verbes dans C-ORAL-ROM (2005) uniquement, on trouve les mêmes formes que dans le tableau 8-6, sauf pour le verbe *prendre* où la forme la plus fréquente est celle du participe passé. Les formes de l'infinitif et du présent semblent donc plus fréquentes à l'oral.

8.3. Résumé et prédictions

L'analyse de la fréquence des différentes formes des verbes réguliers en *-er* présentées dans le chapitre 2 (tableau 2-1) dans les différentes sources de données d'input choisies, nous a permis d'obtenir un continuum de fréquence à partir du pourcentage de formes courtes des verbes. L'analyse a permis de regrouper les verbes par similarité de fréquence : des verbes fréquents à la forme courte, des verbes fréquents aux deux formes et des verbes fréquents à la forme longue. Ce regroupement est une manière simplifiée de présenter le continuum de fréquence dans l'input. Selon l'hypothèse 2, on s'attend donc à retrouver chez les apprenant/es, la forme courte pour les verbes fréquents à la forme courte dans l'input et la forme longue pour les verbes fréquents à la forme longue dans l'input. Pour les verbes fréquents aux deux formes, on s'attend à une préférence moins claire pour la forme longue ou la forme courte que pour les autres verbes chez les apprenant/es.

Bien que les différentes sources de données présentent une bonne corrélation, certaines spécificités sont apparues. Ces spécificités pourraient avoir une influence sur la production des apprenant/es selon l'hypothèse 2.

- Le pourcentage de formes courtes est apparu comme particulièrement élevé avec la plupart des verbes du manuel *Escalade 1*. Cela signifie que les apprenant/es qui utilisent cette méthode sont exposés à un input massif de la forme courte. On pourrait ainsi s'attendre à un pourcentage de formes courtes plus élevé que dans l'input, tout en suivant le continuum des verbes les uns par rapport aux autres.
- Le pourcentage de formes longues est particulièrement élevé dans les données de l'interlocutrice INT-A, en raison des occurrences de la 2^e du pluriel du présent (vouvoiement). Si cette forme a un effet de fréquence sur la production des apprenant/es (hypothèse 2) on devrait trouver un plus haut pourcentage de formes longues avec les

verbes *aimer* et *préférer* qu'avec les autres verbes fréquents à la forme courte dans les données des apprenant/es de corpus (corpus Thomas).

En comparant le pourcentage de formes courtes et l'aspect lexical des verbes, on a pu constater que la fréquence dans l'input et l'aspect lexical s'avèrent être étroitement liés. Les verbes plus fréquents à la forme courte que longue sont en effet tous des verbes statiques alors que les verbes plus fréquents à la forme longue que courte sont tous des verbes dynamiques dans l'input présenté ici. Les deux hypothèses pourraient donc être difficiles à vérifier sur les données d'apprenant/es. La seule manière de les distinguer sera de voir si les verbes dynamiques sont préférés à la forme longue comme attendu par l'hypothèse 1 ou si le groupe de verbes dynamiques moins clairement fréquent à la forme longue dans l'input est produit avec plus de variation par les apprenant/es. Par variation, j'entends ici moins clairement à la forme courte ou à la forme longue que les autres verbes.

Le tableau 8-7 présente les prédictions pour les verbes selon les deux hypothèses. L'aspect lexical des verbes a été présenté à la section 7.2.2. et les données d'input dans le tableau 8-2.

Tableau 8-7 : Formes attendues chez les apprenant/es – verbes en -er

	hypothèse 1 aspect lexical	hypothèse 2 fréquence de surface
adorer	C	C
penser	C	C
trouver S	C	-
habiter	C	C
détester	C	C
préférer	C	C
aimer	C	C
<i>trouver</i> Σ	-	C/L
jouer	L	C/L
regarder	L	C/L
parler	L	C/L
travailler	L	C/L
manger	L	C/L
trouver D	L	-
acheter	L	L
voyager	L	L
étudier	L	L
visiter	L	L

Comme mentionné plus haut, il n'y a pas lieu de réviser le nombre de catégories d'aspect lexical à partir de ces données, les deux verbes téliques, *trouver* (*dynamique*) et *acheter* se trouvent dans le groupe de verbes clairement fréquents à la forme longue, mais ensemble avec des verbes d'activités.

L'analyse des 12 verbes irréguliers a montré les mêmes tendances que pour les verbes réguliers en *-er*. Les verbes se répartissent sur un continuum et présentent trois groupes de fréquence : les verbes fréquents à la forme du présent singulier, les verbes fréquents à l'infinitif et les verbes fréquents aux deux formes. Selon l'hypothèse 2, on s'attend à retrouver les mêmes préférences au niveau des formes pour les différents verbes chez les apprenant/es.

Les données ont également montré que les verbes fréquents à la forme du présent (singulier) étaient des verbes statiques et que les verbes fréquents à l'infinitif étaient des verbes dynamiques. On a aussi pu noter l'absence de fréquence de la forme du participe passé dans cet input. Ce résultat pourrait être lié au fait que la plupart des verbes ont aussi un sens statique, en particulier dans leur utilisation en tant que marqueur conversationnel. La distribution des formes dans cet input reflèterait donc des spécificités de l'utilisation des verbes en français parlé. Pour les verbes *voir* et *comprendre* cette utilisation l'emporte sur le sens dynamique. Le fait que ces verbes ont à la fois un aspect lexical statique et dynamique rend la distinction entre les deux hypothèses encore plus difficile que pour les verbes réguliers.

Le tableau 8-8 présente les prédictions pour les verbes selon les deux hypothèses. L'aspect lexical des verbes a été présenté à la section 7.2.2. et les données d'input dans le tableau 8-6.

Tableau 8-8 : Formes attendues chez les apprenant/es – verbes irréguliers

	hypothèse 1 aspect lexical	hypothèse 2 fréquence de surface
croire	pré sg	pré sg
savoir	pré sg	pré sg
pouvoir	pré sg	pré sg
vouloir	pré sg	pré sg
voir	pré sg ou inf/ppé	variation
connaître	pré sg	variation
comprendre	pré sg ou inf/ppé	variation
faire	inf/ppé	inf
dire	inf/ppé	inf
prendre	inf/ppé	inf
aller lex	inf/ppé	inf
lire	inf/ppé	inf

Maintenant que les données d'input pour l'hypothèse 2 ont pu être précisées, il est temps de vérifier les hypothèses sur des données d'apprenant/es.

9. Données d'apprenant/es de corpus

Ce chapitre est le premier chapitre d'analyse à partir de données d'apprenant/es⁴². Le travail sur corpus est une méthode classique pour l'étude de données d'apprenant/es de LE/L2. Les données sont considérées comme reflétant une utilisation libre et spontanée de la langue par les apprenant/es, même si les conversations sont souvent structurées par les besoins de la recherche.

Ce chapitre comporte une première section de présentation des choix méthodologiques suivie de la présentation des résultats, d'abord pour les verbes réguliers en *-er* puis pour les verbes irréguliers. Il se terminera sur une synthèse et une discussion des résultats, d'une part par rapport à l'input et d'autre part par rapport aux possibilités et limitations propres aux données de corpus.

9.1. Choix méthodologiques

9.1.1. Présentation des apprenant/es

Les données de la production des apprenant/es proviennent du corpus Thomas, 20 apprenant/es et du corpus Lund (Schlyter), 2 apprenant/es. Les apprenant/es de ces corpus ont été présentés à la section 7.1.2.

9.1.2. Choix des verbes

Les verbes analysés dans ce chapitre sont les mêmes verbes, 16 réguliers et 12 irréguliers, qu'au chapitre précédent sur l'input. Comme mentionné au chapitre 7, les verbes choisis pour ce travail sont les verbes les plus utilisés par les apprenant/es de corpus, en particulier du corpus Thomas, qui comprend les apprenant/es les plus débutants. Ces verbes sont moins bien représentés dans les données du corpus Lund.

⁴² Ce chapitre est une version remaniée d'une étude exploratoire sur la variation des formes en vue de la création de la partie expérimentale.

9.1.3. Calcul des occurrences

Les transcriptions ont été traitées à l'aide du logiciel CLAN (MacWhinney, 2000). Les principales conventions de transcription ont été mises dans la liste des abréviations au début de ce travail.

Les calculs sont faits à partir du nombre d'occurrences de chaque verbe dans les contextes du présent et de l'infinitif. La principale difficulté dans l'analyse de données d'apprenant/es de corpus est d'identifier le contexte visé par l'apprenant/e, la forme produite n'étant pas toujours un indice fiable.

Comme mentionné au chapitre 7, les contextes suivants ont été considérés comme contextes pour une forme du présent :

- a) tous les contextes dans lesquels on attend une forme du présent en français.
- b) lorsque l'apprenant/e produit des formes du présent dans un récit du passé et s'y tient. Dans les cas où l'apprenant/e varie les formes du présent et du passé dans un récit du passé, elles ont toutes été considérées comme produites dans un contexte du passé et n'ont donc pas été retenues ici.
- c) L'utilisation de formes du présent pour exprimer un futur, même avec un adverbe temporel (*demain je travaille à la maison* – exemple fictif).

Un contexte pour l'infinitif correspond au placement d'une forme verbale après une préposition (*pour parler*), après un verbe modal (*je veux parler*) ou semi modal (*j'aime parler*), l'auxiliaire du futur proche (*je vais parler*) et dans les autres constructions demandant l'infinitif, y compris dans une fonction de sujet (*jouer au golf n'est pas compliqué*).

Comme dans la plupart des travaux sur corpus, mais peut-être encore plus dans des énoncés de débutants, il faut compter avec un grand nombre de contextes ambigus et une certaine marge d'erreurs dans les interprétations.

Les principes généraux de compilation des formes sont les suivantes :

- Les quelques occurrences des formes de la 1^{ère} du pluriel et de la 2^e du pluriel n'ont pas été comptées ni celles d'autres temps ou modes (futur simple, conditionnel, etc.).
- L'accord sujet-verbe n'a pas été considéré. Ainsi les formes *je parle* et *nous parle* ont toutes deux été codées comme une forme courte.

- Les répétitions ou imitations des formes d'abord exprimées par l'interlocutrice n'ont pas été comptées, du moment où elles n'ont pas été changées par l'apprenant/e.
- Lorsque les apprenant/es ont répété plusieurs fois la même forme verbale, seule la dernière a été calculée. De même lorsqu'il y a un changement entre la première forme et la deuxième forme du même verbe, seule la deuxième a été comptée, les apprenant/es comme les natifs devant avoir une chance de se corriger⁴³. Cette manière de faire est discutable dans la mesure où la première forme produite est la forme que l'on peut considérer comme la plus aisément accessible. Mais globalement, le nombre d'erreurs est équivalent si l'on compte les premières plutôt que les deuxièmes occurrences ce qui signifie que les tendances trouvées ici auraient été les mêmes avec le calcul de premières occurrences.
- Les erreurs de vocabulaire comme l'utilisation de *venir* pour *aller* ; *travailler* pour *voyager* n'ont pas été prises en considération : les formes des verbes ont été comptées ensemble avec les autres formes du même verbe, donc *venir* avec *venir* etc.

9.2. Résultats des apprenant/es de corpus

9.2.1. Verbes réguliers en *-er*

L'objectif de cette analyse est donc de vérifier si la production des apprenant/es est influencée par l'aspect lexical des verbes (hypothèse 1) ou si la production des apprenant/es est plutôt influencée par la fréquence des formes de surface dans l'input (hypothèse 2). Les hypothèses ont été formulées au chapitre 6 et ont été précisées par les données concrètes des verbes à la fin du chapitre 8.

L'analyse présentera d'abord l'utilisation des formes dans les contextes du présent et de l'infinitif puis une comparaison des pourcentages de formes courtes chez les apprenant/es et dans l'input.

Le tableau 9-1 présente les résultats de la production des formes courtes (*parle*) et longues (*parle*) dans les contextes du présent et de l'infinitif chez les apprenant/es de corpus. Les verbes sont présentés dans le même ordre que dans le tableau 8-2 à savoir dans l'ordre décroissant de la fréquence de la forme

⁴³ Les formes comptées dans ces cas se trouvent donc *après* les signes [/] ou [//].

courte en moyenne dans l'input. Les verbes ont été regroupés selon les trois groupes de fréquences établis à partir de l'input, les verbes fréquents à la forme courte, les verbes fréquents aux deux formes et les verbes fréquents à la forme longue dans l'input. Pour mémoire le premier groupe comprend uniquement des verbes statiques et les deux autres des verbes dynamiques selon le classement établi au chapitre 7 (section 7.2.2.).

Les deux premières colonnes $\Sigma C+L$ présentent la somme des occurrences des formes courtes et longues produites dans les contextes de l'infinitif et du présent pour chaque verbe puis le nombre de formes courtes au total ΣC .

Les colonnes du milieu présentent l'utilisation correcte de la forme longue dans le contexte pour une forme de l'infinitif puis la sur-utilisation de la forme courte dans ce contexte, suivi de son pourcentage $\%*C$. Ce pourcentage a été obtenu en divisant le nombre de $*CdsINF$ (*je veux parle*) par la somme des occurrences produites dans un contexte pour l'infinitif (INF). Finalement les colonnes de droite présentent l'utilisation correcte de la forme courte dans le contexte pour une forme du présent puis la sur-utilisation de la forme longue dans ce contexte, suivi de son pourcentage $\%*L$ obtenu en divisant le nombre de $*LdsPRE$ (*je parle*) par la somme des contextes pour une forme du présent (PRE). Les numéros à gauche renvoient au regroupement des verbes à partir de l'input (section 8.2.1.).

Tableau 9-1 : Production des verbes en –er par les apprenant/es de corpus

forme	ΣC+L	ΣC	LdsINF	*CdsINF	%*C	CdsPRE	*LdsPRE	%*L
adorer	3	3	-	-	-	3	0	0%
penser	28	28	0	1	100%	27	0	0%
(trouver S)	1	1	-	-	-	1	0	0%
habiter	66	61	2	3	60%	58	3	5%
(détester)	1	1	-	-	-	1	0	0%
préférer	8	8	-	-	-	8	0	0%
aimer	49	49	-	-	-	49	0	0%
trouver Σ	7	2	4	0	0%	2	1	33%
jouer	11	2	3	0	0%	2	6	75%
regarder	18	9	4	1	20%	8	5	38%
parler	113	82	24	5	17%	77	7	8%
travailler	57	41	4	5	56%	36	12	25%
manger	27	17	9	1	10%	16	1	6%
trouver D	6	1	4	0	0%	1	1	50%
acheter	6	3	2	0	0%	3	1	25%
voyager	11	3	6	0	0%	3	2	40%
étudier	10	0	6	0	0%	0	4	100%
visiter	9	0	8	0	0%	0	1	100%

Légende : C : forme courte ; L : forme longue ; LdsINF : forme longue dans contexte infinitif ; *CdsINF : forme courte dans contexte infinitif ; CdsPRE : forme courte dans contexte présent ; *LdsPRE : forme longue dans contexte présent ; trouver S : sens statique ; D : sens dynamique ; Σ : sens statique + dynamique.

Les premiers verbes dans le haut du tableau 9-1, de *adorer* à *aimer* des verbes statiques fréquents à la forme courte dans l'input, ont presque uniquement été produits à la forme courte et dans un contexte pour le présent par les apprenant/es. Le seul verbe pour lequel on trouve un peu plus de variation est le verbe *habiter*, qui a aussi été produit dans le contexte de l'infinitif, mais trois fois sur cinq à la forme courte comme illustré dans (9-1). Bien que l'interlocutrice utilise le verbe *habiter* à la forme longue, l'apprenante ne la reprend pas.

- (9-1) *INT: mais vous n'allez pas *habiter* en France?
 *JUL: peut être en quatre ans.
 *INT: et quelle est la région que vous préférez en France?
 *JUL: la sud.
 *INT: le sud.
 *JUL: euh **nous allons habite** en euh Perpignan ou euh +... [*CdsINF]

La production des verbes dans le haut du tableau 9-1 correspond donc à ce qui est attendu par l'hypothèse 1 sur l'aspect lexical et à l'hypothèse 2 sur la fréquence des formes de surface.

On peut aussi noter qu'aucune forme longue n'a été produite avec les verbes *aimer* et *préférer*. Ce résultat indique que la fréquence de la forme de la 2^e du pluriel comme seule forme longue n'a pas d'influence sur la production des apprenant/es. Cela implique que les apprenant/es pourraient faire une distinction entre les verbes au présent vs ceux qui ne le sont pas.

Pour ce qui est des verbes dynamiques et fréquents aux deux formes dans l'input (de *trouver* Σ à *manger*), on trouve plus de variation dans la production de ces formes par les apprenant/es que pour les autres verbes. Ces verbes, en particulier *regarder*, *parler* et *travailler* ont été utilisés dans les deux contextes, du présent et de l'infinitif. La forme produite est le plus souvent correcte, mais on voit aussi des erreurs. L'exemple (9-2) présente une telle variation.

- (9-2) *INT: vous *regardez* la télévision?
 *MAR: oui un peu.
 *INT: tous les jours?
 *MAR: oui euh à **je regarde** à le soir. [*LdsPRE]
 *INT: quelles émissions est ce que vous aimez *regarder*, qu'est ce que vous *regardez*?
 *MAR: **je regarde** euh nyheterna@x [= les nouvelles]. [CdsPRE]

L'exemple (9-3) montre un contexte d'utilisation du verbe *parler* qu'on peut considérer comme étant assez typique pour des apprenant/es de LE/L2, c'est-à-dire une utilisation plutôt statique de *parler une langue*. Cette utilisation se fait surtout avec la forme courte du présent.

- (9-3) *INT: eh est-ce que tu penses que tu as plus de facilités que d'autres personnes pour apprendre le français ? que c'est plus facile pour certaines personnes d'apprendre les langues ?
 *LIS4: oui , je crois . parce que eh mes parents et ma soeur eh # ne # **parl(ent)** eh # a des difficultés de **parle** eh des langues aussi . [CdsPRE] [*CdsINF]
 *INT: mh . à ton avis , d'où elles viennent ces difficultés , à cause de quoi , pourquoi est-ce qu'ils ont des difficultés ?
 *LIS4: parce que mon père **parle** anglais eh # très mal , hehe . [CdsPRE]
 www
 *INT: mais est-ce que ça vient parce qu'il n'est pas , jenesaispas , est-ce que c'est un problème physique ? est-ce que c'est un problème eh # ?
 *LIS4: oui # eh c'est une problème # dans la tête , je crois . mais , ma mère eh elle aime de la des langues .
 *INT: mh
 *LIS4: elle eh # apprendre italien maintenant.
 *INT: mh .
 *LIS4: et elle **parle** français , anglais et allemagne . [CdsPRE]

L'exemple (9-4) présente finalement deux utilisations de la forme courte avec le verbe *travailler* par deux apprenantes différentes alors que l'exemple (9-5) l'apprenante SAN présente sa famille en utilisant systématiquement la forme longue avec ce verbe, alors que le verbe *habiter* est produit correctement à la fin de l'extrait.

(9-4) *CAT3: je aime <travaillons> [/] euh **travail** au jardin .[*CdsINF]

*RUT: euh j' aime de <travailler> [/] **travail** dans le jardin .[*CdsINF]

(9-5) *SAN: et euh ma mon fils **travail** à banque. [*LdsPRE]

*INT: hm.

*SAN: et ma sa son sa femme **travail** avec euh www@x. [*LdsPRE]

*INT: lahhh, les trucs du marché, les études de marché?

*SAN: oui euh elle **travail** à www, une banque suédoise. [*LdsPRE]

*SAN: mais ma fille **travail** à www. [*LdsPRE]

*INT: ah bon.

*SAN: et ils **habitent** à une maison très merveilleuse. [CdsPRE]

La production de ce deuxième groupe de verbes parle plus en faveur de l'hypothèse 2 que de l'hypothèse 1, dans le sens où l'on trouve plutôt une variation des formes qu'un recours clair à la forme longue, comme attendu pour les verbes dynamiques. La forme longue est sur-utilisée dans les contextes du présent, mais la forme courte est aussi sur-utilisée dans les contextes de l'infinitif, ce qui n'est pas attendu par l'hypothèse sur l'influence de l'aspect lexical.

Finalement pour le groupe de verbes fréquents à la forme longue dans l'input, on peut constater que ces verbes ont plus souvent été produits dans des contextes pour une forme de l'infinitif que du présent. On peut ensuite constater que les verbes ont essentiellement été produits à la forme longue, correctement pour l'infinitif mais en sur-utilisation dans les contextes pour une forme du présent. Pour ces verbes, on ne trouve aucune sur-utilisation de la forme courte dans les contextes de l'infinitif.

L'exemple (9-6) montre une utilisation du verbe *acheter* dans un contexte du présent et l'exemple (9-7) montre la variation dans l'utilisation de la forme longue du verbe *voyager*.

- (9-6) *INT: et quand vous allez en France, vous parlez français aussi?
 *SEB: euh non non.
 *INT: non?
 *INT: avec les les +/.
 *SEB: avec les oui.
 *SEB: un peu un peu euh euh euh quand **je achetE** à la kiosque. [*LdsPRE]
- (9-7) *FRE2: **je vais voyager** à Sao+Paolo +/. [LdsINF]
 *INT: hm.
 *FRE2: +, et Rio+de+Janeiro.
 *INT: ah oui.
 *FRE2: le congrès est à Sao Paolo.
 *INT: hm.
 *FRE2: <ma> [/] mon vacances est à Rio+de+Janeiro.
 *INT: ah oui, ah voilà. vous allez présenter un papier aussi?
 *FRE2: ah oui.
 *INT: oui.
 *FRE2: je parle euh quinze minutes.
 *INT: quinze minutes ah oui.
 *FRE2: <je voyage> [/] **je voyagE** pour quinze heures +/. [*LdsPRE]

La production des verbes de ce dernier groupe correspond donc à ce qui est attendu par l'hypothèse 1 sur l'aspect lexical et par l'hypothèse 2 sur la fréquence des formes de surface, les verbes étant produits à la forme longue.

La mesure de vérification des hypothèses étant le pourcentage de formes courtes (*cf.* section 7.3.4.), le tableau 9-2 présente les verbes sur une échelle de pourcentage de formes courtes pour les apprenant/es et l'input. Pour les apprenant/es, ce pourcentage a été calculé à partir du nombre de formes courtes divisé par la somme des formes courtes et longues du tableau 9-1 ci-dessus. Les pourcentages de l'input sont repris des tableaux 8-2 et 8-3 et les lignes délimitent les trois groupes de fréquence (*cf.* tableau 8-7).

Tableau 9-2 : Pourcentages de formes courtes : apprenant/es de corpus et input

apprenant/es	ΣC+L	% C	input	ΣC+L	% C
adorer	3	100%	adorer	32	81%
penser	28	100%	penser	368	73%
aimer	49	100%	trouver S	182	72%
préférer	8	100%	habiter	150	72%
habiter	66	92%	détester	10	70%
parler	113	73%	préférer	45	58%
travailler	57	72%	aimer	341	55%
manger	27	63%	trouver Σ	393	47%
regarder	18	50%	jouer	171	42%
acheter	6	50%	regarder	223	39%
trouver Σ	7	29%	parler	581	38%
voyager	11	27%	travailler	267	37%
jouer	11	18%	manger	160	29%
trouverD	6	17%	trouver D	211	26%
étudier	10	0%	acheter	110	15%
visiter	9	0%	voyager	52	13%
(détester)	1	(100%)	étudier	68	10%
(trouver S)	1	(100%)	visiter	25	8%

Légende : C : forme courte ; L : forme longue ; trouver S : sens statique ; D : sens dynamique ; Σ : sens statique + dynamique.

On peut voir que le pourcentage de formes courtes monte à 90% et plus chez les apprenant/es pour les verbes qui sont attendus à la forme courte par les deux hypothèses. Comme on l'a vu plus haut, les verbes *aimer* et *préférer* ont uniquement été produits à la forme courte, on n'observe donc pas d'influence de la fréquence de la 2^e du pluriel par les apprenant/es. Pour les verbes attendus à la forme longue par les deux hypothèses, on voit que les prédictions se réalisent bien pour les verbes *étudier*, *visiter*, *trouver* (*dynamique*) et *voyager* mais pas pour le verbe *acheter* pour lequel on a 50% de formes courtes et longues chez les apprenant/es. On aurait peut-être obtenu un résultat différent avec des contextes pour le passé avec ce verbe télique. Finalement on peut voir que les verbes *parler*, *travailler*, *manger* et *regarder* se regroupent aussi sur le continuum des apprenant/es, avec une préférence nette pour la forme courte avec les verbes *parler*, *travailler* et *manger* alors que l'on trouve 50% de formes longues et courtes pour le verbe *regarder*. Comme mentionné dans la présentation des verbes au chapitre 7 et illustré dans l'exemple (9-3), le verbe *parler* est peut-être surtout utilisé dans le sens de *parler une langue* avec les apprenant/es, une utilisation qui est plus fréquente au présent qu'aux autres temps et formes. La fréquence de l'utilisation statique du verbe l'emporte donc sur l'utilisation prototypique, dynamique. On peut finalement noter que le verbe *jouer* est plus fréquent à la forme longue que ne le prédit la fréquence des

formes. Malgré les différences pour les verbes *acheter* et *penser*, les similarités entre le classement des verbes dans l'input et chez les apprenant/es parlent en faveur d'une influence de la fréquence des formes dans l'input.

Les données pour le verbe *trouver* sont malheureusement peu nombreuses. On n'a qu'une seule occurrence du sens statique ("tu ne sais pas où se trouve moi", Karl 2). Le sens dynamique semble néanmoins plus facilement utilisé que le sens statique.

D'un point de vue qualitatif l'hypothèse 1 ne se vérifie donc pas dans la mesure où les apprenant/es produisent volontiers la forme courte avec des verbes dynamiques, comme pour les verbes *parler*, *travailler* et *manger*. D'un point de vue statistique, le pourcentage de formes courtes est significativement plus élevé avec les verbes statiques que les verbes dynamiques (Mann-Whitney U, $Z=-3,496$; $p<.001$). Cela signifie que la forme courte est nettement plus fréquente avec les verbes statiques que dynamiques. On peut néanmoins dire que la production des apprenant/es semble plutôt suivre la fréquence des formes dans l'input, en raison du haut pourcentage de formes courtes avec un grand nombre de verbes dynamiques, ainsi que la sur-utilisation de la forme courte dans les contextes de l'infinitif. En effet, comme on l'a vu dans le tableau 9-2, on retrouve les trois groupes de fréquences de l'input, même si on trouve des différences comme pour les verbes *acheter* et *jouer*. Comme le montrent les données de corrélation (r de Spearman ; sans *trouver* Σ) du tableau 9-3 ci-dessous, la corrélation entre les données des apprenant/es de corpus et celles de l'input est bonne, même si elle est un peu plus faible avec les interlocutrices, dont les données sont caractérisées par la forme longue et avec le manuel *Escalade 1*, dont les données sont caractérisées par la forme courte.

Tableau 9-3 : Corrélations entre input et apprenant/es de corpus

	PROFS	INT	C-ORAL	ESC	TOT INP	APPR
PROFS	1.000	.356	.721	.571	.772	.846
INT	.356	1.000	.439	.212	.613	.525
C-ORAL	.721	.439	1.000	.474	.925	.870
ESC	.571	.212	.474	1.000	.556	.598
TOT INP	.772	.613	.925	.556	1.000	.862
APPR	.846	.525	.870	.598	.862	1.000

Légende : PROFS : enseignants ; INT : interlocutrices ; C-ORAL : C-ORAL-ROM (2005) ; ESC : manuel *Escalade 1* ; TOT INP : input au total ; APPR : apprenant/es de corpus.

En résumé, cette première analyse de données d'apprenant/es semble indiquer que les apprenant/es sont plutôt influencés par la fréquence des formes dans l'input que par l'aspect lexical des verbes dans leur production des formes.

9.2.2. Verbes irréguliers

Les hypothèses ont avant tout été formulées pour les verbes réguliers, mais elles ont été transposées pour les verbes irréguliers à la fin du chapitre 6. Les prédictions concrètes pour chaque verbe à partir de l'input ont été résumées à la fin du chapitre 8, dans le tableau 8-8.

Bien que les hypothèses proposent que la forme du participe passé puisse être utilisée dans les contextes du présent et de l'infinitif, on n'a en fait que deux occurrences claires de sur-utilisation de la forme du participe passé avec la forme *vu* dans un contexte pour l'infinitif. Étant donné qu'il semble s'agir d'un cas unique, les formes du participe passé ne seront pas traitées dans cette partie.

Pour les verbes *faire* et *dire* la forme du présent singulier et celle du participe passé sont homophones. Elles ont été comptées comme formes du présent singulier dans les contextes du présent et de l'infinitif présentés ici.

Le tableau 9-4 ci-dessous présente les résultats pour les verbes irréguliers dans les données des apprenant/es de corpus. Les premières colonnes présentent la somme des occurrences des formes du présent singulier (*vois*) et de l'infinitif (*voir*) qui ont été produites dans les contextes pour une forme de l'infinitif ou du présent puis la somme des formes du présent singulier (pré sg).

Les colonnes du milieu présentent l'utilisation correcte de la forme de l'infinitif dans le contexte pour une telle forme puis la sur-utilisation de la forme du présent singulier, suivi de son pourcentage $\%*pré\ sg$. Ce pourcentage a été obtenu en divisant le nombre de $*prédsINF$ (*je vais vois*) par la somme des occurrences produites dans un contexte pour l'infinitif (INF). Finalement les colonnes de droite présentent l'utilisation correcte de la forme du présent singulier dans le contexte pour une forme du présent puis la sur-utilisation de la forme de l'infinitif dans ce contexte, suivi de son pourcentage $\%*inf$ qui a été obtenu en divisant le nombre de $*infdsPRE$ (*je voir*) par la somme des contextes pour une forme du présent (PRE). Les verbes sont présentés selon le continuum de pourcentage de formes du présent singulier du tableau 8-8 de l'input.

Tableau 9-4 : Production des verbes irréguliers par les apprenant/es de corpus

verbes	Σ occ	Σ pré sg	inf ds INF	*pré ds INF	%*pré	pré ds PRE	*inf ds PRE	%*inf
croire	141	141	0	0	-	141	0	0%
savoir	32	29	3	2	40%	27	0	0%
pouvoir	37	37	0	0	-	37	0	0%
vouloir	28	28	0	0	-	28	0	0%
voir	24	7	13	0	0%	7	4	36%
connaître	28	23	2	1	33%	22	3	12%
comprendre	28	22	6	0	0%	22	0	0%
faire	135	46	57	2	3%	44	32	42%
dire	52	35	17	2	11%	33	0	0%
prendre	37	29	4	1	20%	28	4	13%
aller lex	86	31	25	2	7%	29	30	51%
lire	52	23	20	3	13%	20	9	31%

Légende : Σ occ : somme occurrences comptées ; Σ pré sg : somme formes du présent singulier ; infdsINF : forme de l'infinitif dans contexte infinitif ; *prédsINF : forme du présent dans contexte infinitif ; prédsPRE : forme du présent dans contexte présent ; *infdsPRE : forme de l'infinitif dans contexte présent.

On peut voir que les verbes dans le haut du tableau 9-4, les verbes fréquents à la forme du présent singulier dans l'input ont essentiellement été produits à la forme du présent singulier et dans le contexte du présent, à l'instar de ce que l'on avait vu pour les verbes réguliers en *-er*. Le verbe *savoir* est le seul verbe à avoir été produit dans le contexte de l'infinitif mais 2 fois sur 5 à la forme du présent. Ces occurrences proviennent de Lisa 2 et 4 dans l'expression "difficile de sait".

La production des verbes de ce premier groupe correspond donc aux prédictions de l'hypothèse 1 sur l'aspect lexical et de l'hypothèse 2 sur la fréquence des formes de surface.

Les verbes *voir*, *connaître* et *comprendre* devraient être produits à la forme de l'infinitif selon l'hypothèse 1 et de manière variée selon l'hypothèse 2. On peut voir que ces verbes ont plus souvent été produits dans les deux contextes que les verbes dans le haut du tableau 9-4. Le verbe *comprendre* a toujours été produit correctement, alors que l'on trouve un peu plus de variation pour le verbe *connaître*. Pour le verbe *voir* on a aussi une sur-utilisation de la forme de l'infinitif dans le contexte du présent. On peut mentionner que le verbe *voir* a surtout été utilisé dans le sens de *voir qc*, donc dynamique, comme dans l'exemple (9-8) et l'utilisation du verbe *connaître* comme dans l'exemple (9-9) m'apparaît plutôt comme statique. Comme on peut le voir dans l'exemple (9-10) le verbe *comprendre* est surtout utilisé dans le sens de *comprendre le français*,

qui pourrait être une utilisation statique et Lisa produit quelques expressions où le verbe est utilisé comme marqueur conversationnel “tu comprends ?”.

(9-8) *LIS4: **je** eh # tous le temps **voir** dans eh le dictionnaire . mais eh # # ils ? <>
eux parlent # ” ba ”. [*infdsPRE]

(9-9) *A propos d'un cours de grammaire en suédois*

*LIS2: donc c'est ma ma langue # maternelle

*INT: mh .

*LIS2: mais **je ne connaître pas** tout. [*infdsPRE]

www

*LIS2: mais je (ne #) pas beaucoup grammaire.

*INT: mh .

*LIS2: eh le grammaire tu " eller " **je connaître** eh. [*infdsPRE]

*LIS2: je j'apprenD (-dre?) à au français ou anglais .

*INT: ouais .

*LIS2: mais , c'est assez important eh **de connaître**. [infdsINF]

(9-10) *KAR3: il faut que je **comprends** quelque chose pour **comprendre** la [/] la
totalité. [prédsPRE], [infdsINF]

*INT: oui c'est ça.mm.

*INT: mais les actualités si tu sais déjà ce qui se passe?

*KAR3: oui je peux **comprendre**. [infdsINF]

Pour ces verbes, on peut dire que la production des apprenant/es semble plutôt suivre les prédictions de l'hypothèse 1. En effet, bien que l'on trouve un peu plus de variation que pour les premiers verbes, il semble que les verbes *connaître* et *comprendre* soient plutôt produits dans le sens statique et donc à la forme du présent et le verbe *voir* dans le sens dynamique et donc plus souvent à l'infinitif (voir aussi tableau 9-5 sur le pourcentage de formes du présent ci-dessous). Pour ces verbes, l'hypothèse 2 sur la fréquence de surface se vérifie donc plus faiblement.

Finalement le dernier groupe de verbes a été produit dans les deux contextes, sauf le verbe *prendre* qui a surtout été utilisé au présent. Les verbes *faire* et *aller* sont fréquemment utilisés à la forme de l'infinitif pour exprimer le présent avec près de 50% de sur-utilisation. Les exemples (9-11) et (9-12) présentent de telles utilisations.

(9-11) *INT: qu' est ce que **vous aimez faire** quand vous êtes libre ?

*DEN: **je fais** de jardinage et le natation et **je faire** de du vélo , oui # comme ça .
[prédsPRE] [*infdsPRE]

- (9-12) *PIE: alors cet été **nous allE** à euhm en France euh non, Normandie en Normandie. [*infdsPRE]
 *INT: en Normandie.
 *PIE: en Normandie cet été.
 *INT: vous allez aussi goûter du vin là bas?

La production de ces verbes semble suivre les prédictions des deux hypothèses, même si le résultat est moins clair que pour les verbes réguliers en *-er* fréquents à l'infinitif dans l'input. Les verbes sont utilisés de manière productive dans les contextes pour l'infinitif et la forme de l'infinitif est sur-utilisée dans le contexte du présent. Mais on trouve aussi la sur-utilisation de la forme du présent dans les contextes de l'infinitif ce qui n'est pas attendu par les hypothèses. On dirait qu'ils sont plutôt produits comme des verbes pour lesquels on attend une variation des formes chez les apprenant/es. Cette variation correspond peut-être à la variation que l'on trouve au niveau de l'aspect lexical de ces verbes. Les verbes *faire*, *dire* et *aller* ont en effet aussi un sens statique (cf. section 7.2.2.). On peut néanmoins se demander si des apprenant/es débutant/es sont sensibles à la polysémie des verbes. Les études à partir de l'aspect lexical présentées au chapitre 4, semblent plutôt suggérer que les apprenant/es soient sensibles à l'aspect lexical prototypique d'un verbe. Ainsi la variation dans la production des apprenant/es pourrait plutôt refléter la variation dans l'usage de ces verbes dans l'input et la variation de la fréquence des formes que cet usage implique.

Afin de pouvoir observer et comparer le pourcentage de formes du présent singulier des différents verbes, le tableau 9-5 présente d'un côté la somme des occurrences et le pourcentage de formes du présent (nombre de formes du présent produites divisé par le nombre d'occurrences) chez les apprenant/es et dans l'input. Les verbes ont été classés selon le pourcentage de formes du présent singulier dans chacune des sources de données.

Tableau 9-5 : Pourcentages de formes du présent: apprenant/es et input

apprenant/es	Σocc	%pré sg	input	Σocc	%pré sg
vouloir	28	100%	croire	254	80%
pouvoir	37	100%	savoir	924	70%
croire	141	100%	pouvoir	1200	65%
savoir	32	91%	vouloir	682	61%
connaître	28	82%	voir	888	42%
comprendre	28	79%	connaître	274	40%
prendre	37	78%	comprendre	192	33%
dire	52	67%	faire	2539	29%
lire	52	44%	dire	1607	28%
aller lex	86	36%	prendre	420	26%
faire	135	34%	aller lex	551	25%
voir	26	27%	lire	178	12%

Au niveau de l'hypothèse 1, on peut voir que tous les verbes statiques arrivent dans le haut du tableau 9-5 chez les apprenant/es et que le verbe *comprendre* est aussi produit comme tel par les apprenant/es. Dans le bas du tableau 9-5, on retrouve tous les verbes dynamiques. Les seuls verbes qui ne semblent pas suivre l'aspect lexical sont les verbes *prendre* et *dire* dont la fréquence de la forme courte est surprenante, au niveau des deux hypothèses. On peut néanmoins dire que la production des verbes par les apprenant/es de corpus semble suivre l'aspect lexical.

Pour ce qui est de l'hypothèse 2, on trouve une correspondance de fréquence pour un grand nombre de verbes, mais les prédictions au niveau des formes ne se vérifient que faiblement. Les verbes *prendre* et *dire* ont par exemple été préférés à la forme courte par les apprenant/es, alors que ces verbes étaient fréquents à la forme de l'infinitif dans l'input. À l'inverse, le verbe *voir*, fréquent au présent dans l'input, a été préféré à l'infinitif. On trouve néanmoins une corrélation tout à fait satisfaisante avec un r de Spearman de .683.

Les deux hypothèses semblent difficiles à séparer pour les verbes irréguliers bien que les formes des verbes soient plus clairement distinctes. En effet, le fait que plusieurs des verbes irréguliers fréquemment utilisés par les apprenant/es, et donc choisis pour ce travail, varient au niveau de l'aspect lexical demanderait une analyse beaucoup plus détaillée de chaque énoncé.

9.3. Résumé et discussion

L'analyse de la production des formes courtes (*parle*) et longues (*parlE*) dans les contextes du présent et de l'infinitif de 16 verbes réguliers en *-er* a montré le résultat suivant :

- les verbes fréquents à la forme courte dans l'input, les verbes statiques, ont essentiellement été produits à la forme courte et presque uniquement dans le contexte du présent
- les verbes également fréquents aux deux formes dans l'input, des verbes dynamiques, ont présenté une certaine variation dans l'utilisation des formes et des contextes, les apprenant/es produisant les formes courtes et longues dans les deux contextes, parfois correctement parfois non ; les apprenant/es n'ont donc pas systématiquement produit ces verbes à la forme longue, comme attendu par l'hypothèse sur l'aspect lexical
- les verbes fréquents à la forme longue dans l'input, des verbes dynamiques, ont essentiellement été produits à la forme longue avant tout dans les contextes de l'infinitif, mais aussi dans les contextes du présent

Même si la différence du pourcentage de formes courte est significative au niveau des verbes statiques vs dynamiques, ces résultats parlent en faveur d'une influence de la fréquence des formes dans l'input sur la production des apprenant/es comme prévu par l'hypothèse 2 (cf. tableau 9-2 ; le taux de corrélation entre les données de l'input au total et des apprenant/es était de .862). En effet, bien que les verbes statiques aient été produits à la forme courte, on n'a pas pu observer une préférence de la forme longue avec les verbes dynamiques comme attendu par l'hypothèse 1 sur l'influence de l'aspect lexical. Le haut pourcentage de formes courtes avec un grand nombre de verbes dynamiques pourrait être le reflet de la fréquence du présent dans la langue utilisée avec les débutants comme on l'a vu dans les données du manuel *Escalade 1* au chapitre 8. Une autre possibilité est que les verbes dynamiques soient fréquemment utilisés pour décrire des *habitudes* par les apprenant/es comme on l'a vu dans les exemples (9-2) et (9-3), une utilisation plutôt statique (cf. section 7.2.2.). Dans ce cas l'utilisation de la forme courte avec ces verbes serait influencée par la *fréquence* de l'aspect lexical statique de ces verbes, en d'autres termes par la fréquence de la forme courte dans l'input comme attendu

par l'hypothèse 2. L'influence de l'aspect lexical implique en effet la production de la forme qui correspond au sens prototypique du verbe.

Le fait que l'on ne trouve pas de forme longue avec les verbes *aimer* et *préférer*, qui sont fréquents à la forme de la 2^e du pluriel dans l'input des apprenant/es du corpus Thomas, pourrait parler en faveur d'une sensibilité des apprenant/es aux verbes fréquents au présent *vs* ceux qui ne le sont pas. L'utilisation de la forme longue pourrait ainsi être associée à l'utilisation d'autres fonctions que le présent.

L'analyse de la production de la forme du présent singulier (*vois*) et de l'infinitif (*voir*) de 12 verbes irréguliers dans les contextes du présent et de l'infinitif a montré un résultat différent (section 9.2.2). Les verbes semblent être produits selon deux schémas : on trouve des verbes produits à la forme du présent et dans le contexte du présent d'une part et des verbes produits dans les deux contextes avec une sur-utilisation des formes plus ou moins importante. La production des verbes irréguliers semble plutôt suivre une distribution aspectuelle comme attendu par l'hypothèse 1. Le sens statique des verbes a été exprimé par une forme du présent alors qu'on trouve plus de variation pour les verbes dynamiques. Mais si l'on considère que le sens prototypique des verbes *prendre* et *dire* est plutôt dynamique, on voit pourtant que ces verbes n'ont pas été produits comme attendu par l'hypothèse 1.

En somme, l'analyse des verbes irréguliers à partir des deux hypothèses formulées pour ce travail, et sans les contextes pour le passé, est encore plus difficile que celle des verbes réguliers en *-er*. Le fait que les formes soient distinctes ne semble donc pas apporter plus de clarté au travail.

Même si on trouve une distribution des formes du présent et de l'infinitif entre les deux catégories de verbes dans l'input, ce n'est peut-être que dans les contextes du passé que l'on pourrait trancher entre l'influence de l'aspect lexical et celui de la fréquence. Pour ces raisons, les verbes irréguliers ne seront pas testés dans la partie expérimentale.

Cette analyse de données d'apprenant/es sur corpus nous a donc permis d'observer un certain nombre de tendances au niveau de l'influence de l'aspect lexical et de la fréquence des formes de l'input dans la production des formes dans les contextes du présent et de l'infinitif. Les données semblent indiquer que la fréquence des formes dans l'input influence l'utilisation des formes par les apprenant/es. D'un autre côté, on a vu que les verbes fréquents à la forme courte, qui sont aussi des verbes statiques, n'ont pratiquement pas été produits

dans un autre contexte que celui du présent. On n'a donc pas pu observer la production de ces verbes dans les contextes de l'infinitif et voir si les apprenant/es avaient recours à la forme la plus fréquente dans l'input là-aussi. Pour cette raison, deux tests expérimentaux ont été construits, dans lesquels les apprenant/es doivent traiter les deux formes dans les deux contextes pour tous les verbes de test.

PARTIE EXPERIMENTALE

10. Présentation de la partie expérimentale

10.1. Introduction

L'objectif de cette partie expérimentale est de vérifier les deux hypothèses formulées au chapitre 6 à partir de données obtenues de manière ciblée en condition expérimentale. L'analyse des données des apprenant/es de corpus a montré un certain nombre de tendances sur la question de l'influence de l'aspect lexical et de la fréquence des formes de surface, mais elles n'ont pas permis de vérifier les hypothèses sur tous les verbes. En effet, la production des apprenant/es a révélé une absence de variation des formes et des contextes pour les verbes statiques et fréquents à la forme courte dans l'input, ne permettant pas de savoir si la forme produite était la forme la plus accessible.

Cette partie expérimentale a été construite dans le but d'observer le traitement par les apprenant/es de la forme courte (*parle*) et de la forme longue (*parlE*) des verbes réguliers en *-er* dans les contextes du présent et de l'infinitif en utilisant des tests d'élicitation ciblés, un test d'imitation de phrases et un test de rappel de séries de formes verbales.

La partie expérimentale comprend quatre chapitres. Le présent chapitre sera consacré à une présentation générale de la partie expérimentale à savoir la présentation du déroulement complet de l'expérience soumis aux apprenant/es, le choix des verbes, les règles éthiques suivies ainsi que la présentation des participant/es. Les chapitres suivants seront consacrés à l'analyse des trois principales tâches, la tâche de production libre au chapitre 11, le test d'imitation au chapitre 12 et finalement le test de rappel libre au chapitre 13.

10.2. Déroulement de l'expérience

Le choix des éléments constituant la partie expérimentale s'est fait en plusieurs étapes révisées au fur et à mesure de l'évaluation de tests pilotes. Un facteur important outre la construction adaptée au niveau des apprenant/es débutant/es, a été le facteur temps. Pour éviter que les participant/es ne s'épuisent et pour augmenter les chances d'obtenir un nombre suffisant de participant/es, l'expérience ne devait pas dépasser 60 minutes y compris les consignes et

feedbacks. La priorité a ainsi été mise sur des tests ciblés sur la question de recherche.

Le cœur de cette partie expérimentale est formé des deux tests psycholinguistiques, le test *d'imitation* (*elicited imitation*) et le test de *rappel libre* (*free recall*) qui seront présentés en détail dans leur chapitre respectif. On peut tout de même déjà dire que ces deux méthodes ont été choisies

- pour leur proximité avec la production orale : les apprenant/es traitent et répètent les stimuli
- pour leur complémentarité : dans le test d'imitation, les verbes sont présentés dans une phrase alors que dans le rappel libre ils sont présentés dans des séries de verbes.

Ces tests permettent donc d'observer la production des formes verbales avec et sans le contexte de la phrase.

Dans le but de vérifier la validité des données des deux tests expérimentaux avec celles de données en production libre, une tâche de narration à partir de la série d'images *Le voyage en Italie* a été incluse dans la présente partie. Le fait que ces données proviennent toutes des mêmes participant/es garantira la qualité des corrélations. Comme on le verra au chapitre 11, la tâche de production libre servira aussi de test de niveau pour éviter que les données d'apprenant/es trop avancés soient incluses dans les résultats.

Le déroulement complet de l'expérience comprend huit moments présentés dans le tableau 10-1 ci-dessous. Chaque moment était précédé de consignes et suivi d'un moment de discussion sur le déroulement de l'activité passée.

Tableau 10-1 : Déroulement de l'expérience pour les apprenant/es

élément de test	durée prévue
1 Introduction générale (règles éthiques suivies)	3'
2 Activité d'échauffement : se présenter	5'
3 Test d'imitation	20'
4 Tâche de production libre : narration du <i>Voyage en Italie</i>	10'
5 Test de rappel libre	15'
6 Traduction des verbes	3'
7 Questionnaire sur le/la participant/e	2'
8 Consentement, bon cadeau et remerciement	2'

L'expérience s'est déroulée de la même manière pour le groupe contrôle de natifs mais sans les tâches de traduction et de présentation.

La tâche de traduction des verbes avait pour objectif de vérifier combien de verbes les apprenant/es semblaient connaître. Les verbes étaient présentés

oralement aux deux formes, courte et longue, et l'apprenant/e disait oralement la/les traductions qu'il savait. Ces données ne seront pas présentées à part dans ce travail.

10.3. Sélection des verbes pour la partie expérimentale

Comme mentionné au chapitre 7, le nombre de verbes pour la partie expérimentale a été réduit à 12 verbes réguliers en *-er*. Le nombre final de verbes a été fortement contraint par le nombre limité de verbes statiques disponibles, l'objectif était en effet de travailler à partir du même nombre de verbes statiques et dynamiques, l'aspect lexical étant une des variables de base. Les verbes dynamiques ont ensuite été choisis en cherchant à obtenir un équilibre avec les verbes statiques au niveau du nombre de syllabes et des particularités telles que le changement de voyelle entre la forme courte et longue comme dans *préférer* – *préfère*. Sur les 16 verbes réguliers analysés aux chapitres 8 et 9, quatre verbes dynamiques ont donc été éliminés :

- Le verbe *travailler* en raison de son utilisation dans le sens *voyager* (cf. *travel* en anglais) par les apprenant/es ainsi qu'en raison de la fréquence du substantif *travail*.
- Les verbes *voyager* et *étudier* ont été éliminés parce que j'ai préféré choisir les verbes *visiter* et *acheter* parmi les verbes fréquents à la forme longue dans l'input (cf. tableaux 8-2 et 8-3). Comme on l'a vu, le verbe *visiter* a uniquement été produit à la forme longue par les apprenant/es alors que pour le verbe *acheter* on a vu une plus grande variation (cf. tableau 9-2). Ces deux verbes sont en outre plus faciles à prononcer pour des suédophones.
- Finalement le verbe *jouer* n'a pas été retenu parce que les autres verbes choisis étaient plus faciles à prononcer et plus courants chez les débutants.

Le verbe *trouver* a été sélectionné d'office en raison de sa double appartenance aspectuelle, statique dans le sens de *à son avis* (et *se trouver*) et dynamique dans le sens de *trouver ce que l'on cherche*. Le nombre d'occurrences du verbe étant plus élevé pour le sens dynamique chez les enseignants et le manuel *Escalade 1* ainsi que chez les apprenant/es de corpus, il a été compté comme verbe dynamique (cf. tableau 8-3).

Le tableau 10-2 présente les 12 verbes de tests regroupés par le nombre de syllabes de la forme longue et de la forme courte, les verbes plus courts dans le haut du tableau et les verbes plus longs dans le bas. Les verbes pour lesquels il y a un changement de prononciation à l'intérieur du verbe entre la forme courte et longue sont signalés sous + *changement*.

Tableau 10-2 : Verbes de test de la partie expérimentale

	statiques	dynamiques
2 – 1 syllabes	aimer – aime penser – pense	manger – mange parler – parle trouver – trouve acheter – achète
+ changement		
3 – 2 syllabes	adorer – adore habiter – habite	regarder – regarde visiter – visite
+ changement	détester – déteste préférer – préfère	

10.4. Présentation des règles éthiques suivies

Les règles éthiques suivies pour le déroulement de cette expérience correspondent aux règles utilisées pour les expériences effectuées dans le laboratoire expérimental du Centre de Langues et Littérature de l'université de Lund (*Humanistlaboratoriet*). Ces règles se répartissent sur trois moments : le recrutement, durant et après l'expérience.

Au moment du recrutement, les apprenant/es ont notamment été informés sur l'objectif de l'expérience : “en apprendre plus sur comment font les débutants alors qu'ils ont peu de moyens à leur disposition” ainsi que sur le genre de tâches qui les attendraient. Ils ont également été informés sur la liberté de participation et sur la possibilité d'interrompre l'expérience à tout moment ainsi que sur la rémunération : un bon de 100 couronnes suédoises (environ 10 euros) pour une librairie.

Ces informations ont été répétées au moment de l'introduction au début de l'expérience avec quelques informations sur le traitement du matériel enregistré et les risques par rapport à l'intégrité de la personne.

À la fin de l'expérience, l'apprenant/e et l'expérimentatrice ont signé une feuille de consentement résumant les points éthiques les plus importants.

10.5. Présentation des participant/es

10.5.1. Apprenant/es

Au total 33 apprenant/es ont participé à la partie expérimentale. Une personne avait d'abord participé comme pilote de la version finale des tests, les autres sont toutes des élèves de lycée, cinq lycées de Lund et un de Malmö pour un total de 10 classes de français différentes. La moyenne d'âge est de 17,7 ans. Le nombre de filles (26) est nettement plus élevé que celui de garçons (7). Au moment des tests 17 participant/es se trouvaient au deuxième semestre du niveau scolaire de français, *Steg1* et 16 au niveau *Steg2*, qui correspond au quatrième semestre de français. Un semestre correspond environ à 40 heures de français. Certaines personnes avaient déjà fait du français avant le lycée, d'autres ont choisi de commencer le français à un certain moment de leur cursus de lycée. Le tableau 10-3 présente les principales informations sur les participant/es, le sexe, l'âge, la/les L1 (langues parlées avant l'âge de 4 ans), le nombre de semestres de français, séjour en pays francophone ou non ainsi que les autres langues parlées ou étudiées (autre l'anglais).

Tableau 10-3 : Présentation des apprenant/es

part	H/F	âge	L1	smstr fra	steg	séjour France	autres langues que l'anglais
P01	H	18	suédois, cantonnais	4	2	non	
P02	H	20	arabe (Irak)	4	2	oui	suédois, danois
P03	F	18	suédois	4	2	oui	espagnol- 1 an
P04	F	17	suédois, hongrois	6	2	non	allemand- 1 an
P05	H	18	suédois	4	2	oui	espagnol- 1 smstr
P06	F	17	suédois, espagnol	2	1	oui	allemand- 1 smstr
P07	F	17	suédois, danois	2	1	oui	espagnol- 4 ans
P08	F	18	suédois	4	2	oui	espagnol- 4 ans
P09	F	18	suédois	4	2	non	allemand- 2 ans
P10	F	16	suédois, perse	2	1	oui	espagnol- 4 ans
P11	F	19	suédois	4	2	oui	allemand- 6 ans, latin- 1 an
P12	F	17	suédois	2	1	oui	allemand- 3 ans
P13	H	18	suédois	4	2	oui	
P14	F	18	suédois	4	2	oui	allemand- 3 ans, espagnol- 3 ans
P15	F	17	suédois	2	1	oui	allemand- 1 an, espagnol- 3 ans
P16	H	18	suédois	4	2	non	allemand- 4 ans
P17	F	18	suédois, mandarin	9	2	oui	espagnol- 4 ans
P18	F	20	chinois	6	3	non	suédois
P19	F	18	suédois	4	2	oui	allemand- 4 ans, norvégien
P20	F	18	suédois	6	1	oui	allemand- 5 ans, latin- 2 ans, japonais- 2 ans
P21	F	17	suédois	8	1	non	allemand- 1 an

part	H/F	âge	L1	smstr fra	steg	séjour France	autres langues que l'anglais
P22	F	18	suédois	4	2	non	allemand- 6 ans
P23	F	18	suédois	2	1	non	allemand- 6 ans, espagnol-1 an
P24	H	17	suédois	2	1	non	allemand- 4 ans, chinois et arabe autodidacte
P25	F	18	suédois, français	7	1	oui	italien- 1 an
P26	F	18	suédois	4	2	non	allemand- 5 ans
P27	H	17	suédois	2	1	oui	espagnol- 2 ans, danois- 2 ans
P28	F	18	suédois, hongrois	2	1	oui	italien- 2 ans, latin- 1 an, espagnol- 3 ans
P29	F	17	suédois	2	1	oui	allemand- 3 ans
P30	F	19	suédois, punjabi	2	1	non	allemand- 3 ans, espagnol- 1 an, danois- 2 ans
P31	F	18	suédois	2	1	non	espagnol- 5 ans, latin- 1 an
P32	F	17	suédois, hongrois	10	1	oui	italien
P33	F	17	suédois	2	1	oui	espagnol- 3 ans

Légende : part : participant/es ; H/F : homme / femme ; smstr fra : nombre de semestres de français ; Steg : niveau scolaire de français.

Tous les participant/es ont le suédois comme L1 sauf P02 et P18. Ces personnes ont néanmoins appris le français en Suède, ce qui n'est pas le cas pour P04, qui a fait une première année de français en Grande-Bretagne. Tous les participant/es ont un bon niveau d'anglais, qui est la première langue étrangère en Suède (enseignée dès l'âge de 10 ans).

Plus de la moitié des participant/es ont passé une semaine et plus en pays francophone, plus ou moins récemment. La participante P25 a vécu à Paris durant les 3 premières années de sa vie (mais elle a oublié le français ensuite) et les participantes P08 et P14 ont un frère ou une sœur à Paris à qui elles rendent visite régulièrement.

Les enregistrements ont été faits dans une salle du lycée de la personne enregistrée sauf trois qui ont été faits dans une salle à l'université. Il y a donc un certain niveau de bruit pour la plupart des enregistrements. Ce bruit fait partie du quotidien des participant/es et n'a pas forcément un effet néfaste sur leur production. Chaque personne a été enregistrée individuellement en présence de l'expérimentatrice.

Bien que l'échantillon des participant/es présente un certain nombre de différences, celles-ci ne semblent pas assez importantes pour éliminer des participant/es d'entrée.

Comme on le verra plus loin, le vocabulaire des tests a essentiellement été choisi à partir du manuel pour débutants *Escalade 1* (Tillman *et al.*, 2001 ; cf. section 8.1.1.). La tâche de traduction avait montré que la méthode *Génial 1*

avait un vocabulaire assez différent de la méthode *Escalade 1*. Afin de voir le possible impact des méthodes sur la production des apprenant/es de la partie expérimentale, le tableau 10-4 présente les méthodes de français utilisées.

Tableau 10-4 : Méthodes de français utilisées dans les classes de lycées

méthode	participant/es
Escalade 1 (Tillman <i>et al.</i> , 2001)	P06,P07, P10, P12, P21, P33
Escalade 2 (Tillman <i>et al.</i> , 2002)	P22, P26
Escalade B (Tillman <i>et al.</i> , 1997)	P03, P04, P05, P08, P09, P16
Génial 1 (Sanner & Wennberg, 2002)	P24, P27, P28, P29, P30, P31, P32
Un air de Paris (Persson & Lindfors, 2002)	P15, P20, P23, P25
Un air de France (Persson & Lindfors, 2002)	P11, P13, P14, P17, P19
A Propos 2 (Öman & Hedman-Ekman, 1990)	P01, P02
?	P18

10.5.2. Groupe contrôle de natifs du français

Le groupe contrôle est constitué de 10 personnes, six femmes et quatre hommes, ayant le français comme L1. Elles proviennent de différentes régions francophones : Belgique, Côte d'Ivoire, France, Martinique, Maroc et Québec. Six d'entre elles sont des étudiant/es d'échange, et quatre sont établies en Suède. La moyenne d'âge est de 26 ans (entre 21 et 37 ans). Le premier enregistrement a été fait de manière très spontanée, la personne ne restant que quelques jours à Lund. Pour des raisons techniques, les données de la Nat10 ont uniquement été retenues pour la tâche de production libre.

11. Tâche de production libre

L'analyse des données de la partie expérimentale s'ouvre sur celles de la tâche de production libre. En effet, ces données servent d'une part à dégager les principales caractéristiques de la production des apprenant/es de la partie expérimentale et d'autre part à vérifier le niveau des apprenant/es. C'est pour cette raison que cette analyse a été placée en premier, afin de pouvoir, le cas échéant, écarter les données d'apprenant/es trop avancés. Comme on l'a vu dans le tableau 10-1 sur le déroulement de l'expérience, cette tâche vient entre le test d'imitation et le test de rappel.

Ce chapitre commencera par une présentation générale de la tâche à la section 11.1. Les principes de traitement des données seront présentés à la section 11.2. La section 11.3. sera consacrée à l'évaluation du niveau des apprenant/es à partir des récits obtenus. La section 11.4 présentera la production des verbes réguliers en *-er* ainsi que la vérification des hypothèses sur les verbes de test produits. Le chapitre se terminera par un résumé.

11.1. Présentation de la tâche de production libre

La tâche de production libre consistait à raconter une histoire à partir de la série d'images "*Le voyage en Italie*" en maximum 10 minutes. Cette série d'images, en couleurs, a été créée et utilisée pour une tâche d'élicitation du pluriel et des adjectifs pour des apprenant/es de différents niveaux, y compris des débutants (Ågren, 2008). Cette série d'images donne en outre l'occasion de produire la plupart des verbes de test. Elle est reproduite dans la figure 11-1 ci-dessous.

Cette tâche a été introduite dans l'expérience dans un double objectif. Premièrement, il s'agissait de pouvoir évaluer le niveau des apprenant/es, i.e. de vérifier qu'ils se trouvent à un niveau qui corresponde aux stades 1 à 3 de Bartning et Schlyter (2004) comme les apprenant/es de corpus (chapitre 9). Ces stades ont été présentés à la section 3.2. et sont cités en détail dans l'annexe 1. Deuxièmement, cette tâche de narration sert à obtenir des données de production libre par les apprenant/es de la partie expérimentale. De telles données permettront d'établir des corrélations avec les données des deux tests expérimentaux puis de discuter les similarités et différences obtenues. Ces

données de production libres pourront également être comparées à celles présentées au chapitre 9 (apprenant/es de corpus), qui proviennent d'un groupe moins homogène d'apprenant/es que ceux de la partie expérimentale.

La consigne donnée, "Raconte cette histoire à une personne qui ne voit pas les images. Qu'est-ce qu'il se passe durant le voyage en Italie ?", devait encourager les apprenant/es à raconter l'histoire au présent, de manière descriptive afin d'éviter les contextes du passé. À la différence des données d'apprenant/es de corpus, les apprenant/es produisent en principe un monologue. On peut mentionner en passant, que la plupart des apprenant/es ont trouvé cette tâche extrêmement difficile, la plus difficile de toutes, malgré la consigne *fais comme tu peux*.

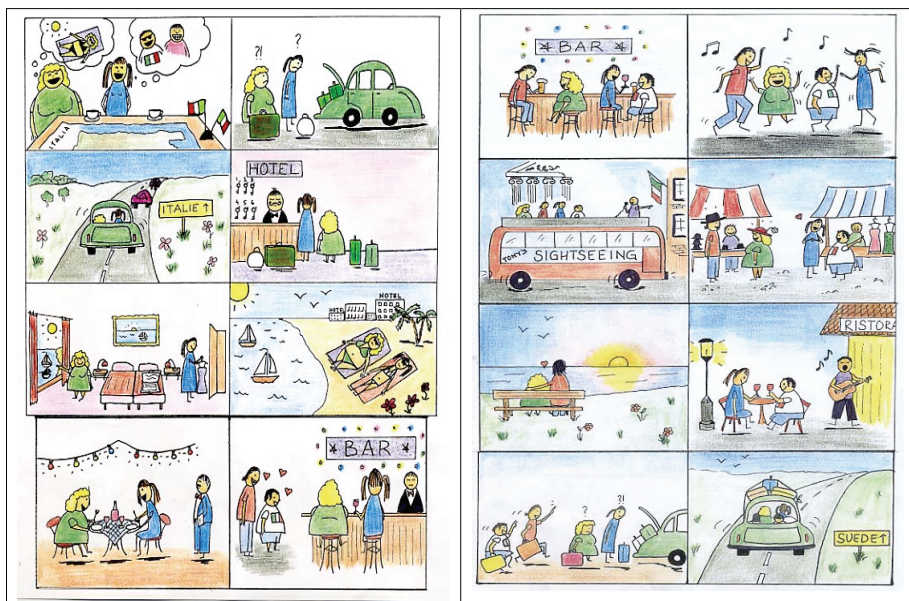


Figure 11-1 : Série d'images "Le voyage en Italie"

11.2. Traitement des données

Les 33 récits du *Voyage en Italie* ont été transcrits en format texte à l'aide du logiciel audio *iTunes*. Au total le matériel comprend environ 3 heures d'enregistrements pour 4790 mots comptés (les apprenant/es ont mis 5'15 minutes en moyenne pour leur récit).

Les enregistrements ont été écoutés au moins trois fois, deux fois au moment de la première transcription du matériel (un/e participant/e après l'autre) et une autre fois lors de la réécoute (par tâche). Les enregistrements de P04, P09, P11, P19, P24 ont également été écoutés et transcrits par une autre personne principalement dans le but de pouvoir trancher sur des formes verbales difficiles à percevoir (par exemple *fait* vs *faire*). Les propositions de cette personne pour ces verbes ont donc toutes été acceptées.

Les 10 récits des natifs (groupe contrôle) sont d'une durée moyenne de 3'16 minutes (de 1'48 à 5'35 minutes) pour un total d'environ 30 minutes et de 4049 mots.

Les transcriptions ont été faites selon les principes suivants :

- La narration étant un monologue de l'apprenant/e, les interventions de l'expérimentatrice sont rares et brèves. Elles n'ont pas été transcrites et elles n'ont pas été enlevées du calcul de durée de la narration sauf pour P25 (20 secondes).
- L'accent a été mis sur la transcription des mots produits en français, mais quelques mots étrangers ont aussi été transcrits.
- Les marqueurs conversationnels *euh*, *ah* et *ja* n'ont pas été transcrits. Par contre les *oui* ainsi que les *je ne sais pas* (en un mot *jenesaispas*) et les *peut-être* ont été transcrits puisqu'il s'agit de mots français.
- Les pauses prolongées n'ont pas été notées, mais elles ont parfois été indiquées par une virgule, d'autres signes étant comptés comme mots dans le logiciel de traitement de texte.
- En principe toutes les utilisations de la forme longue des verbes en *-er* dans un contexte pour une forme courte ont été transcrites avec un *-E* final comme dans *elle voyagE*. Lorsque le /r/ final d'une forme verbale a été prononcé, il a été transcrit : dans l'exemple *le voiture voyager*, le verbe a été prononcé *voyagére*. Lorsque le verbe a été utilisé correctement dans un contexte syntaxique pour l'infinitif, il a également été transcrit avec le *-er* final (*va voyager*).

Les questions de transcriptions liées à la morphologie silencieuse des marques du pluriel seront traitées et discutées à la section 11.3.2.

11.3. Évaluation du niveau des apprenant/es

11.3.1. Choix des méthodes

Les critères d'évaluation du niveau d'apprenant/es L2 sont nombreux comme en témoigne notamment le travail de compilation des méthodes de Wolfe-Quintero *et al.* (1998). Les critères d'évaluation de niveau se regroupent en trois principales catégories de mesures : la fluidité, la complexité grammaticale et lexicale et l'exactitude⁴⁴ (*accuracy*) (*cf.* Wolfe-Quintero *et al.*, 1998, pour l'évaluation de plus d'une centaine de mesures). Pour la présente étude, j'ai choisi deux mesures utilisables avec des données de production orale d'apprenant/es débutant/es tout en étant peu coûteuses en temps et nombre de manipulations. J'ai ainsi fait le choix de calculer le débit et d'évaluer le niveau des participant/es à l'aide du logiciel d'analyse automatique *Direkt Profil 2.2.* (Granfeldt, Nugues *et al.*, 2005)⁴⁵.

La mesure du débit semble être une mesure fiable (Iwashita *et al.*, 2008) et facile à établir, même si différentes méthodes sont présentées dans la littérature comme on peut le voir ci-après :

- dans Iwashita *et al.* (2008) le débit (*speech rate*) a été calculé en divisant le nombre total de syllabes produites sur un échantillon par la durée de l'échantillon en secondes, sans les données d'auto-reformulation (*repair*) ni les pauses de trois secondes et plus.
- dans Mehnert (1998) le débit est calculé à partir du nombre de syllabes d'un échantillon divisé par sa durée en secondes avec ou sans les éléments répétés, reformulés, remplacés
- Freed *et al.* (2004) calculent le nombre de mots/minute sans les faux départs, répétitions et éléments de réparation (seuls les mots de la construction réparée sont comptés)
- Towell *et al.* (1996) font la distinction entre la mesure du débit de parole (*speaking rate*) où ils divisent le nombre total de syllabes produites sur la totalité du temps passé et la mesure du débit d'articulation (*articulation rate*) dans laquelle les pauses ne sont pas comptées

⁴⁴ *Cf.* Defays *et al.* (2007 : 45) qui proposent de traduire *accuracy* par *exactitude* pour désigner à la fois la correction et la précision.

⁴⁵ *Direkt Profil* est accessible sur <http://profil.sol.lu.se:8080/profil/logon.jsp> (page consultée le 22 août 2009).

Le logiciel *Direkt Profil* (Granfeldt, Nugues *et al.*, 2005) est un outil d'évaluation de niveau basé sur les stades de Bartning et Schlyter (2004), qui, comme on l'a vu à la section 3.2., ont mis en évidence un itinéraire détaillé de l'acquisition du français par des suédophones (*cf.* aussi annexe 1). Ces stades ont déjà été utilisés pour évaluer le niveau des apprenant/es de corpus du chapitre 9. Ce logiciel effectue une analyse à partir de 142 attributs, essentiellement les membres de la phrase nominale et de la phrase verbale, ainsi que des données statistiques, telles que le nombre d'occurrences, le nombre de types de mots, le nombre de phrases, le nombre de mots pour chaque catégorie grammaticale et finalement une analyse du vocabulaire. Une des fenêtres d'output permet de voir directement en couleur sur le texte introduit, le nombre de syntagmes nominaux et verbaux retenu, alors qu'une autre fenêtre offre un aperçu des résultats en occurrences et pourcentages ainsi qu'un stade. Le résultat en *stade* est donné par trois algorithmes différents. Une évaluation de la précision de chacun des trois classificateurs a montré un résultat nuancé et le classificateur appelé *SVM* semble donner les meilleurs résultats pour les trois premiers stades, les stades qui nous intéressent ici (Granfeldt & Nugues, 2007).

L'évaluation du niveau des apprenant/es de la partie expérimentale se fera ainsi à partir de mesures appartenant aux trois principales catégories de mesure, la fluidité avec le débit ainsi que la complexité et l'exactitude avec *Direkt Profil*. L'utilisation d'un logiciel d'analyse automatique pour l'évaluation du niveau des apprenant/es présente en outre l'avantage de limiter les facteurs de subjectivité, même si, comme on le verra ci-après, l'utilisation de ce logiciel avec des transcriptions de données orales pose un certain nombre de problèmes, en particulier au niveau du traitement de la morphologie silencieuse.

11.3.2. Procédure

a) Traitement des transcriptions par *Direkt Profil* 2.2

L'utilisation de *Direkt Profil* (Granfeldt, Nugues *et al.*, 2005), un logiciel conçu pour l'évaluation du niveau de textes écrits, avec des transcriptions de production orale pose un problème au niveau de la morphologie silencieuse. Les principaux choix à faire ont des implications au niveau de l'accord sujet – verbe *il(s) parle(nt)* et du syntagme nominal *l[e] voiture(s)*, *l[e] italien(s)*. Comme on peut le voir dans la figure 11-1 au début du chapitre, la plupart des événements et actions du *Voyage en Italie* impliquent plusieurs personnes. En cas de doute,

la transcription pourrait se faire en se basant sur le nombre de personnes sur les images.

Ainsi, pour ce qui est du syntagme verbal on pourrait ajouter un /s/ à tous les pronoms personnels référant à des sujets au pluriel. Selon Ågren (2008, 152), un tel marquage est précoce en français LE à l'écrit, avec 87,5% des contextes marqués au stade 1 et 97% au stade 2. Mais, le problème se situe au niveau de la forme verbale : on ne peut pas savoir quelles sont les formes verbales qui prennent un *-nt* ou non. En effet, le pourcentage de marquage du pluriel sur les verbes réguliers n'est que de 20% au stade 1. Il monte à 54% au stade 2 et à 75% au stade 3 à l'écrit (Ågren 2008, 175). Pour cette raison, j'ai choisi de ne pas transcrire les marques du pluriel inaudibles au niveau du syntagme verbal sauf sur les pronoms précédant les formes verbales *ont*, *vont*, *font*, *sont* et, le cas échéant, d'autres verbes qui sont clairement à la forme de la 3^e du pluriel comme *prennent* ou *sortent*. L'énoncé /il parl/ n'a par contre pas été transcrit par *ils parlent* parce qu'on ne peut savoir si le *-nt* aurait été produit ou non, ni par *ils parle*, ce qui induit une erreur d'accord qui n'existe pas à l'oral, mais par *il parle*. Les /s/ ont par contre été transcrits lorsqu'ils étaient audibles. Par exemple dans une série comme *elle elle elles* le /s/ du troisième *elle* a été prononcé par l'apprenant/e⁴⁶.

Pour ce qui est du syntagme nominal, les /s/ ont été transcrits s'ils étaient audibles ainsi que sur les noms précédés d'un déterminant du pluriel comme *deux*, *les* ou *des*. La recherche sur l'écrit semble en effet indiquer que le pluriel sur les noms se fait dès le début de l'acquisition du français LE/L2 (78% au stade 1, 87% au stade 2, 92% au stade 3 selon Ågren, 2008 : 144).

Si l'utilisation de *Direkt Profil* présente de nombreux avantages, il est important d'être conscient de ses faiblesses, dont voici les principales :

- le stade obtenu avec un texte de moins de 100 mots est moins fiable
- *Direkt Profil* a de la difficulté à traiter les noms propres. Les phrases verbales avec un nom propre comme sujet ne sont pas prises en compte. Pour vérifier le possible impact de ce fait, un pronom personnel a été ajouté dans la transcription soumise au logiciel afin que la forme verbale soit prise en compte

⁴⁶ Étant donné que l'étude ne concerne pas le pluriel, on peut trouver des inconsistances dans les transcriptions.

- *Direkt Profil* ne prend pas les verbes *faire* et *prendre* et rarement le verbe *aller*⁴⁷
- *Direkt Profil* ne compte pas les verbes sans sujet et les verbes corrects à l'infinitif ont rarement été pris en compte
- *Direkt Profil* considère une forme longue se terminant par un *-E* comme une forme courte du présent singulier. Ainsi dans *elle voyagE*, la forme verbale est comptée comme *voyage*. Le pourcentage d'accord sujet – verbe correct pour les participant/es de cette étude pourrait alors être plus élevé qu'en réalité. Mais, cette différence pourrait compenser les accords potentiellement corrects au pluriel (*les filles aime*). Pour les participant/es avec un haut taux de formes longues, le stade a été contrôlé en remplaçant les formes en *-E* par l'infinitif

L'analyse des transcriptions par *Direkt Profil* a été vérifiée *online* pour chaque texte entré à partir des soulignements en couleur présentant le traitement des éléments de la phrase, en particulier de la phrase nominale et de la phrase verbale.

b) Calcul du débit

Pour ce qui est du calcul du débit, il existe de nombreuses manières de le calculer, comme on l'a vu à la section précédente. Dans le présent travail, le débit a été calculé de la manière la plus simple possible en divisant le nombre de mots comptés par la durée totale en secondes. Dans le tableau 11-1, les données ont été converties en mots/minutes afin de faciliter la lecture et la représentation du résultat.

La durée a été calculée en nombre de secondes entre le début du premier mot et la fin du dernier mot. Le temps passé à produire des mots qui n'ont pas été transcrits ni comptés (en particulier des mots en suédois) ainsi que les rares interventions de l'expérimentatrice entrent donc aussi dans le calcul de la durée. Les résultats pour le débit pourraient ainsi être plus bas que la normale. Mais, le nombre d'interventions en suédois ou de sollicitations diminuent lorsque le niveau de l'apprenant/e augmente étant donné que le cadre invite à la production d'un monologue. Le temps d'hésitation ou les *euh* précédant le premier mot ou un silence après le dernier mot n'ont pas été pris en compte.

⁴⁷ Il semble que la forme *aller* ou *allE* après un sujet n'est prise que lorsqu'il y a une production correcte de la forme de l'infinitif dans le texte comme dans *Marie va aller en Italie. (...) Il allE dans une bar...* (P02).

Le nombre de mots a été calculé à l'aide d'un logiciel de traitement de textes. Les mots étrangers *shopping, sightseeing, sol, ristorante, room, view, playa, drink, mar* sont en principe les seuls mots étrangers qui ont été comptés comme mots. Les mots étrangers qui n'ont pas été comptés ont été surlignés en gris dans les transcriptions. Un certain nombre d'hispanismes a été laissé tel quel, notamment l'utilisation des pronoms et déterminants.⁴⁸

11.3.3. Résultats de l'évaluation du niveau des apprenant/es

Le tableau 11-1 présente le résultat de l'évaluation du niveau des 33 apprenant/es de la partie expérimentale. Il présente le stade attribué, le nombre de mots produits transcrits, la durée en seconde puis le nombre de mots/minute. Pour avoir une idée rapide du débit des apprenant/es, ils ont été répartis en trois zones de débit selon le nombre de mots/minute : le tiers le plus lent a été considéré comme étant dans la zone de débit 1, les intermédiaires dans la zone de débit 2 et les plus rapides dans la zone de débit 3. Le tableau indique aussi les transcriptions comprenant moins de 100 mots. Les participant/es ont été classés selon le stade puis selon le nombre de mots/minutes. Les données des natifs sont présentées à la fin du tableau.

Tableau 11-1 : Classement des participant/es par stade et nombre de mots/minutes

part	stade final	nb mots	nb sec	mots /min	zone débit	<100 mots	commentaire
P15	1	54	490	7	1	+	
P30	1	63	325	12	1	+	
P10	1	49	226	13	1	+	20 secondes enlevées ⁴⁹
P27	1	73	294	15	1	+	
P05	1	23	71	19	1	+	
P33	1	94	275	21	1	+	
P25	1	70	204	21	1	+	20 secondes enlevées ⁵⁰
P28	1	107	218	29	2		
P31	1	98	186	32	3	+	
Σ		631	2289	17			

⁴⁸ Les mots non connus suivants ont aussi été comptés comme mots : *disteur* (se distraire ? P03), *il avi* (habite ? a vit ? P07), *plakau* (place ? P17).

⁴⁹ Elles correspondent à un long silence entre le dernier et l'avant-dernier mot.

⁵⁰ Interventions de l'expérimentatrice.

part	stade final	nb mots	nb sec	mots /min	zone débit	<100 mots	commentaire
P20	2	71	228	19	1	+	DProf: stade 3 mais <100 mots
P29	2	68	204	20	1	+	DProf: stade 3 mais <100 mots
P21	2	163	468	21	1		
P12	2	90	247	22	1	+	
P19	2	146	367	24	2		DProf: stade 3
P07	2	89	217	25	2	+	DProf: stade 3 mais <100 mots
P13	2	194	443	26	2		
P23	2	27	61	27	2	+	
P09	2	175	391	27	2		
P03	2	273	488	34	3		
P04	2	156	274	34	3		
P24	2	183	305	36	3		
P01	2	174	259	40	3		
P16	2	140	193	44	3		
Σ		1949	4145	28			
P26	3	99	244	24	2	+	
P06	3	160	387	25	2		
P22	3	303	670	27	2		
P08	3	155	330	28	2		
P18	3	206	422	29	2		
P32	3	215	399	32	3		
P17	3	456	810	34	3		DProf: stade 4
P02	3	119	141	51	3		
P14	3	159	181	53	3		
P11	3	338	378	54	3		
Σ		2210	3962	33			
N01	N	418	182	138			
N02	N	470	214	132			
N03	N	237	130	109			
N04	N	518	301	103			
N05	N	394	206	115			
N06	N	202	108	112			
N07	N	379	193	118			
N08	N	840	335	150			
N09	N	288	159	109			
N10	N	303	134	136			
Σ		4049	1962	124			

Légende : nb mots : nombre de mots comptés ; nb sec : durée en secondes ; mots /min : nombre de mots minutes (débit) ; zone débit : 1 – tiers de participant/es le plus lent, 2 – tiers intermédiaire, 3 – tiers le plus rapide ; <100 mots : le + indique que moins de 100 mots ont été produits ; DProf : Direkt Profil.

On peut tout d'abord constater qu'on ne trouve pas d'apprenant/e à un stade au-delà du stade 3, ce qui signifie que les 33 participant/es seront retenus pour la

suite des analyses. Certains classements demandent un bref commentaire présenté ci-dessous.

La participante P17 a été évaluée au stade 4 par *Direkt Profil*, ce qui pourrait indiquer qu'elle est trop avancée. La transcription de sa narration est reproduite en (11-1). On peut voir que P17 a un très bon niveau, mais qu'elle n'est pas assez avancée pour être placée au stade 4⁵¹. A ce stade, on ne devrait en effet plus trouver de formes de l'infinitif dans les contextes pour le présent comme c'est le cas ici avec *aller* et *prendre* mais aussi avec des verbes en *-er*.

- (11-1) (P17 – Stade 3) Lisa et Marta. Marta est la fille blonde et Lisa est la fille **azul**. Marta **pense, pense** elle, elle **veut** beaucoup être, elle **veut** beaucoup est près de du soleil. Lisa **veut raconte** le le garçons d'Italie parce qu'il elle **trouve** qu'ils sont très belles. Lisa et Marta **prende** les les les portef les portef et c'est trop qu'il. Marta et Lisa **vont** à Italie. Et après trois trois jours, elle **arrive** à l'hôtel près de la mer et soleil. Elle elle est très très gaie et parce qu'il c'est soleil et la mer est près de l'hôtel. Elle elles **vont** à la playa et c'est beaucoup de fleurs rouges et c'est beaucoup de oiseaux et et il elle **l'aime** beaucoup. Le soir première en l'Italie Lisa et Marta **mange** au le restaurant. Elle **aime** aussi une bar. Elle **mange** le spécialité de l'Italie le spaghetti avec bolognaise. Après elle **célèbre** avec une verre de vin, vigne non vin vin. Après deux garçons, Ben qui un peu grosse et Tony, une un garçon trop trop **slim slimant**. Marta et Lisa a il **dit** bonjour et Ben et Tony **prendre** une une verre. Et c'est très enchanté. Après le drink et un peu de chatte, le **danse** un peu. Et et ils ils ont très très drôle. Et c'est le un la nouveau musique de Justin Timberlake qui était la chanson favorite de le. Après la jour après cet soir Tony Ben Marta et Lisa **veut, vont** à Tony's sightseeing en l'Italie et c'est il ils a, ils avons **vu** beaucoup de maisons et plakau très fameuse. Dans la midi de la jour Tony Marta Lisa et Ben **alle**, allons, allons, **alle**, ils ont à une market et c'est beaucoup de, beaucoup de trucs pour la tête et robe. Très beau de robe. Et Ben la **dit** que la robe rouge est très très beau pour Lisa et Lisa le **dit** que c'est elle, elle **l'aime** beaucoup aussi. Et Tony le **achète** de la robe à Lisa. La la soir final, il, Tony et Marta et Ben et Lisa **vont** à place différente parce qu'ils vont, **voulent veulent** être sole seuls. Marta et Tony a une vue la soleil et les oiseaux. Et Lisa et Ben **prende** une drink à une restaurant avec une avec un garçon qu'il le **joue** de la guitare et **chante** pour le. La jour après le soir final elle elle est très désolé et parce qu'il, Marta et Lisa, va **aller, retourner** à Suède, donc Ben et Tony **coupe coupe coupe** à Marta et Lisa, parce qu'ils aussi **voulent aller** à Suède avec Lisa, Marta. Après Lisa Marta Tony et Ben **alle** à Suède. Et **va** à elle la maison d'elle et après c'est fini.

Pour P19 *Direkt Profil* a proposé un classement au stade 3 mais elle a été mise au stade 2 en raison du manque d'éléments présents dans les autres textes du stade 3, tels que des connecteurs autres que *et*, des adverbes de temps ou encore un verbe lexical à changement de racine au pluriel (p.ex. *ils sortent*). La transcription de son récit est reproduite en (11-2).

⁵¹ L'algorithme *SVM* indique un stade 4 et *LTM* un stade 3.

- (11-2) (P19 – stade 2) Les dames **pense** à l'Italie et une **pense** à le soleil et et une **pense** à les les hommes et le monde à Italie. Elle, elle, elles ont beaucoup de choses. Elle elle **vais**, elle **va** au Italie avec sé auto. Elle **arrive** au une hôtel. Et il elles ont une chambre avec une belle vue. Elle elle **voit** la mer et le soleil. Elle **reste** à la plage. Et elle **fait** très beau. Elle **mange** à une restaurant et elle **boit** du vin. Et elle **vais**, elles **vont** à une bar, et elle **rencontre** deux hommes. Elle **parle** avec les hommes et elle, elle **pense** le autobus et **voit** le la ville. Elles **vont** au marché et elle elle **voit** le soleil. Avec une homme. Et et une dame **va** au restaurant avec une homme. Et et elle **va** au en Suède avec les hommes.

Les participant/es P28 et P31 ont peut être été classés en dessous de leur niveau (au stade 1 au lieu de 2), mais elles ont toutes deux ont un taux d'omission du sujet assez élevé, un phénomène auquel *Direkt Profil* pourrait être sensible puisqu'il est assez rare en français LE (Granfeldt, 2003 : 116). Elles ont été laissées au stade 1 pour éviter d'avoir un trop grand nombre de personnes au stade 2. L'exemple (11-3) présente la production de P28.

- (11-3) (P28 – Stade 1) Deux femmes **pensE** en en Italie. On se **paque** le voiture et **travaille** à Italie de l'Italie. **ArrivE** à l'hôtel, et et je voudrais **réserver** un un chambre de deux nuits. Le chambre est en la mer, de, d'un lit. De⁵² **mange** un restaurant et et et **travaille** à un bar. En el bar **parlE** avec, avec deux hommes hommes et **dansE** à le musique. Aujourd'hui **travaille** à une bus de Tony sightseeing et **regardE** le. Le femme à le bleu kläddning **aime** il garçon d'Italie. La femme et le garçon a une glace de vin et **écoutE** à le musique et et. Et **travaille** à le Suède.

La décision finale du niveau des participant/es après la considération des questions discutées ci-dessus donne 9 participant/es au stade 1, 14 participant/es au stade 2 et 10 participant/es au stade 3.

La corrélation entre le stade et le nombre de mots s'élève à .666 $p < 0.01$ (r de Spearman) et à .582 $p < 0.01$ entre le stade et le nombre de mots/minutes. Ces corrélations valident le regroupement des apprenant/es par stade. Un calcul statistique au niveau du nombre de mots et du nombre de mot/minute selon les stades (tableau 11-2) indique une différence significative entre les stades 1 et 2, mais pas entre les stades 2 et 3.

⁵² Le *de* est ici une espèce de sujet. Beaucoup de participant/es avaient de la peine à trouver les pronoms *ils* ou *elles*.

Tableau 11-2 : Tests statistiques pour le regroupement des participant/es

	nb mots	nb mots /min
Stade 1-2, N=23 :		
Mann-Whitney U	23,000	23,000
Wilcoxon W	68,000	68,000
Z	-2,520	-2,520
Asymp. Sig. (2-tailed)	0,012	0,012
Stade 2-3, N=24 :		
Mann-Whitney U	39,000	41,000
Wilcoxon W	144,000	146,000
Z	-1,815	-1,698
Asymp. Sig. (2-tailed)	0,069	0,089
Stade 1-3 N=19 :		
Mann-Whitney U	1,000	10,000
Wilcoxon W	46,000	55,000
Z	-3,593	-2,858
Asymp. Sig. (2-tailed)	0,000	0,004

Si l'on compare les données de débits avec les résultats obtenus par Granfeldt (2007 : 93) pour des apprenant/es de stade 3 et par Hilton (2006) pour des apprenant/es débutant/es de l'anglais (A1 et A2 du *Cadre commun européen de référence pour les langues*) on peut constater que le débit obtenu dans la présente étude est plus bas que les leurs. Ce résultat était attendu à cause de la manière de calculer le débit (*cf.* plus haut). Sans les mots incomplets, sans les mots non français, et sans les mots répétés (*all meaningful non repeated French words*) Granfeldt arrive en effet à 62,4 mots/minute en moyenne pour les apprenant/es indiqués au stade 3 par *Direkt Profil*, alors que Hilton arrive à 65,3 mots/minute, tandis que les apprenant/es du stade 3 de la présente étude ne produisent que 33 mots/minute. La principale différence, outre la manière de calculer, est le nombre d'années passées à étudier la langue cible, environ quatre ans chez Granfeldt et sept ans chez Hilton. On pourrait donc penser que les apprenant/es du présent travail ont été surévalués au stade 3 comparé à ces travaux. D'un autre côté, le fait de devoir parler à partir d'images pourrait ralentir le débit par rapport à une production plus libre.

Pour les natifs, le débit moyen obtenu est de 122 mots/minute (de 103 à 150), ce qui correspond au débit normal pour des natifs (Bonin, 2002 : 44).

Dans le but de vérifier la validité de l'évaluation des niveaux six textes, tirés au sort, deux par stade (sans P11 et P17 qui avaient déjà été discutés), ont été soumis à trois juges experts pour une évaluation manuelle. À noter que le stade indiqué par *Direkt Profil* pour le récit 2 diffère du stade final. *Direkt Profil* avait indiqué le stade 3. Le tableau 11-3 présente les différentes évaluations de

niveau. Même si le juge 3 est plus sévère dans ses critères de classement, le stade établi pour ces six textes semble valable.

Tableau 11-3 : Évaluation manuelle du stade de six récits par trois juges

récit	stade	juge1	juge2	juge3
1	3	2-3	2	1
2	2	2-3	3	2
3	3	3	2	2
4	1	1	1	1
5	1	1	1	1
6	2	2	2	2

11.4. Résultats de la production libre

Cette section présente une analyse de la production des verbes de la tâche de production libre pour les verbes réguliers en *-er*. L'objectif de cette analyse est de dégager les principales caractéristiques de ce groupe d'apprenant/es. Cette analyse se fera d'abord au niveau de la production des formes courtes et longues dans les contextes du présent et de l'infinitif pour tous les verbes en *-er*. Les deux hypothèses formulées pour ce travail seront ensuite vérifiées sur les verbes de tests produits dans cette tâche.

11.4.1. Calcul des occurrences

Les formes analysées dans cette tâche sont les mêmes que pour les apprenant/es de corpus du chapitre 9, à savoir les formes courtes (*parle*) et longues (*parle*) des verbes en *-er* dans les contextes pour une forme du présent singulier et 3^e du pluriel et dans les contextes pour une forme de l'infinitif (et le cas échéant du passé).

Les 9 occurrences de la 1^{ère} et les 3 occurrences de la 2^e du pluriel n'entrent pas dans les calculs ni les 12 occurrences de formes se terminant par les sons *-ant(e)* ou *-onte*, même si la 3^e du pluriel pourrait être visée par prononciation du *-ent* comme dans *la fille* et *le garçon mangeant*.

Les formes verbales ont été classées selon la forme visée au niveau syntaxique, présent et infinitif indépendamment de l'accord sujet verbe (cf. section 7.3).

Les quelques occurrences de formes du passé ont été comptées sous contexte pour l'infinitif, dans la mesure où ces formes ont uniquement été reconnues comme telles en raison de l'auxiliaire qui les précède. Toutes les formes longues

sans auxiliaire ont été considérées comme une forme longue dans un contexte du présent. Ce problème se pose avant tout pour les données de P11 dont voici un extrait en (11-4). Les formes *voyagé* et *demandé* ont été mises avec les LdsINF alors que les occurrences de *arrivE* ont été mises dans *LdsPRE même si la forme visées par l'apprenante pourrait être un passé.

- (11-4) (P11 – Stade 3) Deux femmes a **voyagé** à Italie. Elle a **voyagé** avec son auto, son auto vert. C'est c'est une route très long. Et mais quand il **arrivE**, elle **arrivE**, le portier dit, il a, elle a **demandé** de la son chambre et son clé.

Le codage des formes verbales est le même que celui pour les apprenant/es de corpus (présenté dans le tableau 7-4).

11.4.2. Distribution des formes et contextes chez les participant/es

Cette section présente les caractéristiques de la production des verbes réguliers en *-er*. Les données sont présentées en détail pour chaque participant/e, y compris du groupe contrôle de natifs dans l'annexe 4.

Cette section ne comporte pas d'analyse détaillée des verbes irréguliers, mais on peut tout de même mentionner que le contexte pour une forme du présent s'élève à 93% pour les verbes irréguliers dans ces récits (sans les occurrences de *être* et *avoir*, ni de *aller* comme auxiliaire) et que le pourcentage de formes de l'infinitif dans le contexte du présent s'élève à 25%. Sur les 46 occurrences d'une forme de l'infinitif dans un contexte pour le présent, 17 proviennent du verbe *aller* et 15 du verbe *faire*.

La figure 11-2 présente la distribution des contextes du présent et de l'infinitif produits pour les verbes réguliers en *-er* par les apprenant/es ainsi que par les natifs (N).

Les pourcentages ont été obtenus en divisant le nombre de formes produites dans le contexte respectif par la somme des occurrences dans les deux contextes pris ensemble. Toutes les données sont des occurrences brutes.

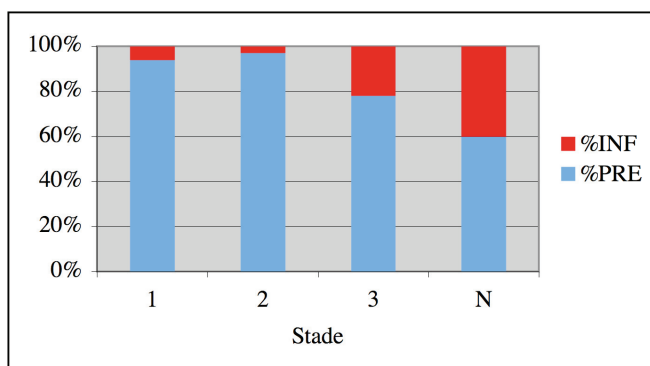


Figure 11-2 : Contextes du présent et de l'infinitif – tâche de narration

La production de contextes pour une forme du présent est donc massive pour ces apprenant/es montant jusqu'à 97% au stade 2 (94% au stade 1) et la production de contextes pour les formes longues, infinitif (ou passé) n'apparaît réellement qu'au stade 3. On ne trouve que quatre apprenant/es qui ont produit des contextes pour une forme longue aux stades 1 et 2 (P25 et P28 au stade 1, P13 et P16 au stade 2) pour un total de 7 occurrences.

Chez les natifs, la proportion de contextes pour le présent et l'infinitif est plus proche que chez les apprenant/es, mais le pourcentage de contextes pour le présent l'emporte nettement avec 59% des occurrences en moyenne. Cela signifie que la préférence pour le présent est inhérente à la tâche (et cohérente avec la consigne). Un exemple de récit de natif (Nat 09) est reproduit en (11-5).

- (11-5) (Nat 09) La famille Dupont **part** en Italie, la mère **a** envie de se **prélasser** au soleil et la et la jeune **a** envie de **rencontrer** des des garçons italiens. Bon lorsqu'ils **font** lorsqu'ils lorsqu'ils **tentent** de **mettre** les bagages dans la voiture, ils se **rendent** compte qu'ils **n'ont** pas fin que elles elles **n'ont** pas assez de place, c'est c'est seulement la mère et la fille en fait. Bon après après lorsque les problèmes **sont** réglés ils décid ils **partent** vers l'Italie, ils **arrivent** à l'hôtel, donc ils ont **réservé** une chambre, fin je **dis** chaque fois ils mais en fait c'est elles. Après elles **s'installent** dans leur chambre qui **a** une vue sur la mer, la maman d'ailleurs **a** l'air assez assez contente. Elles vont elles vont naturellement se **prélasser** sur la plage directement. Le soir elles **vont** elles **vont** au restaurant et et ensuite elles vont **prendre** un cocktail au bar. Là elles ont **été repérées** par deux jeunes beaux garçons, un qui **s'intéresse** à la mère et l'autre à la fille et ensuite elles **partent danser** dans les clubs. On ne **sait** pas ce qu'il se **pass**e la nuit mais le lendemain **partent faire** un tour de sightseeing, elles, avec avec avec les garçons qu'elles ont **rencontrés** dans le club et les les garçons qui leur **font découvrir** un peu le les marchés. Après ils **regardent** le coucher de soleil, enfin la fille la fille **prend** un verre avec son ami pendant que pendant que la maman **regarde** le coucher **observe** le coucher de soleil

avec avec son nouveau copain et ensuite il **est** temps de **se dire** au revoir et de **rentrer** au pays, la Suède, en fait et et voilà.

La figure 11-3 présente la production des formes courtes et longues dans le contexte pour une forme courte du présent. Les pourcentages ont été obtenus en divisant le nombre de formes courtes (C) et longues (*L) par la somme des occurrences dans le contexte du présent.

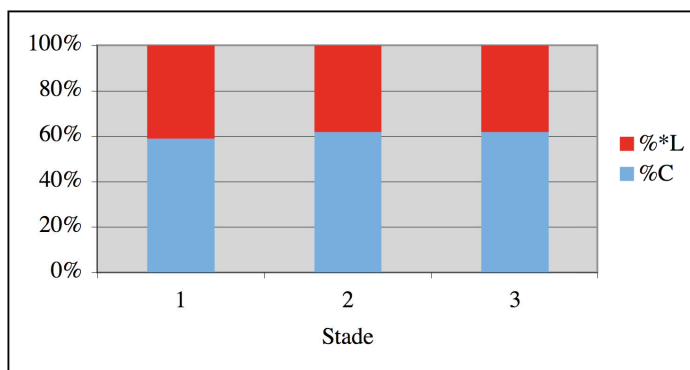


Figure 11-3 : Formes dans les contextes pour le présent – tâche de narration

On peut voir que les apprenant/es des trois stades ont produit près de 60% de formes courtes et 40% de formes longues dans le contexte du présent. On voit donc une légère préférence pour la forme courte, mais comme l'ont souligné les études à partir des contextes de la finitude présentées au chapitre 3, présentées de cette manière, les formes courtes et longues semblent être utilisées de manière aléatoire.

Bien que les pourcentages soient les mêmes pour les trois stades, il y a beaucoup de différences individuelles comme on peut le voir dans l'annexe 4. On peut surtout noter que 5 des 14 apprenant/es du stade 2 ont produit 0% de formes longues. Ces apprenant/es ont donc une production très correcte, dans la mesure où ils produisent la forme attendue par le contexte du présent. Voici un exemple (11-6) d'une telle production des verbes en *-er*.

- (11-6) (P07 – Stade 2) Marie **va** en vacances à Italie. La voiture est trop petite (...) Au hôtel il **avi**, au il **habite** il est rouge. Et en Italie c'est bien chaleur. Et en la soir, il **mangue** et oui il **mangue** beau. Et il **va** au boîte et en mercredi il est un sightseeing en Rome. Il **marche** et et **aller** au la vert. Et Marie est a un petit ami. Et vendredi ils ils **aller** au Suède.

Au stade 3, tous les apprenant/es ont produit les deux formes, le pourcentage de formes longues variant entre 8% et 73%. Plusieurs apprenant/es de ce stade sont donc moins corrects que ceux du stade 2. Un exemple du stade 3 est présenté en (11-7) plus bas.

Finalement la figure 11-4 présente la production des formes dans les rares contextes de l'infinitif, la forme courte *C et la forme longue L. On peut voir que les apprenant/es produisent surtout la forme longue dans ce contexte, mais on a quelques occurrences de la forme courte. Ce contexte est donc plus rapidement et plus souvent correct comme le soulignaient Bartning & Schlyter (2004) et Prévost (2004 ; 2007) (*cf.* chapitre 3).

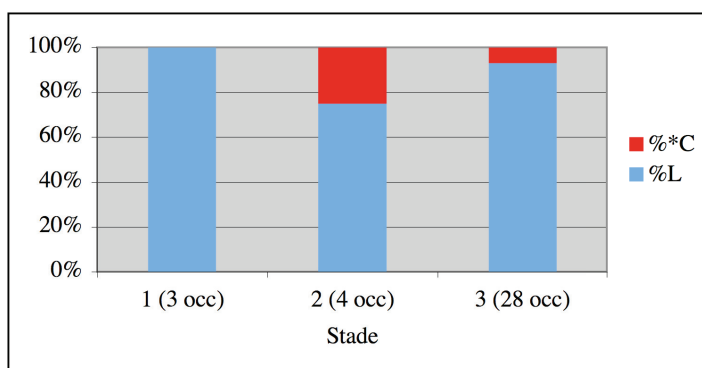


Figure 11-4 : Formes dans les contextes pour l'infinitif – tâche de narration

Six des dix apprenant/es du stade 3 ont produits des contextes pour la forme longue, ce qui signifie que ces apprenant/es commencent à se familiariser avec cette construction. Le fait qu'ils produisent de tels contextes signifie aussi que la fréquence de cette construction dans l'input a augmenté chez ces apprenant/es ou du moins son traitement. Cette situation pourrait expliquer le haut pourcentage de formes longues dans les contextes pour le présent comme on l'a vu plus haut. Le pourcentage de formes longues dans le contexte pour une forme courte est en effet plus élevé pour ces apprenant/es (sauf P17) que pour ceux qui se limitent au contexte du présent (*cf.* annexe 4). L'exemple (11-7) présente un extrait de la production d'une apprenante de ce stade (*cf.* aussi (1-1)).

- (11-7) (P22 – Stade 3) La mère et la non. C'est un pair soeur qui **rêve** en Italie. Il **veut** venu à en Italie. Et mais il a beaucoup de bagages et, et il **faut faire rester** quelque bagage à la maison. Il, ils **vont** en Italie dans avec auto et il elle **arrive** à la à l'hôtel pour, et il elle. Il **parle** avec le réceptionniste dans l'hôtel pour **rentrer** le clé pour sa chambre. (...) Et dans la nuit, elle **mange** dans un restaurant et elle **boire** du vin. Et depuis il elle **visite** un bar et là il est de italiens charmants et il. Les italiens adorant les soeurs. Les italiens parlant avec les soeurs, il a, et il **danse**, oui il **danse**. Et le lendemain, les Italiens et les soeurs faisant sightseeing dans, tour de sightseeing et elle voyant, le village. Il elle **vient** en. Il **vient** dans place qui est là on va **trouver** beaucoup de choses qui elle pouvant **acheter**, par exemple, les robes etcetra. La soeur avec de cheveux blonds **regardE** avec son italien le soleil dans le soir. Et la soeur avec de cheveux noirs **boire** du vin avec son italien et il **écoute** un homme qui **joue** du, **joue** un instrument instrument instrument en restaurant. Mais elle **trouve** c'est un bon vacances et elle va **rentrer** dans sa maison dans, son auto. Et mais le italien **veut** avec elle **continuer**, il **veut visiter** elle maison. Et et enfin, les soeurs et l'italien **va**, non, **vont** à la Suède.

En résumé, les données des apprenant/es de la partie expérimentale semblent surtout utiliser le contexte du présent, mais on trouve la variation caractéristique des apprenant/es débutant/es dans leur réalisation de ce contexte avec un pourcentage élevé de formes longues. Les contextes pour une forme de l'infinitif n'apparaissent qu'au stade 3 et ils sont le plus souvent produits correctement. Ce résultat rejoint ce que d'autres ont pu constater, que le contexte de l'infinitif pose moins de problèmes que le contexte du présent. Comme on l'a vu au chapitre 3, Bartning & Schlyter (2004) utilisent le critère de pourcentage de formes courtes dans le contexte du présent.

11.4.3. Vérification des hypothèses

Dans le but de vérifier l'hypothèse 1 sur l'influence de l'aspect lexical dans la production des verbes par les apprenant/es, le tableau 11-4 présente les données des 12 verbes de test regroupés selon leur aspect lexical (*cf.* section 7.2.2. pour la discussion sur l'aspect lexical des verbes). Les occurrences sont présentées pour chaque stade et au total (Σ) pour les apprenant/es. Les données du groupe contrôle de natifs sont présentées à droite. Le nombre de formes courtes *C* (*parle*) et de formes longues *L* (*parle*) a été obtenu en additionnant le nombre de formes produites indépendamment du contexte dans lequel elles ont été produites. Le tableau présente également le pourcentage de formes courtes, %C. À noter que les verbes *détester* et *préférer* n'ont pas été produits par les apprenant/es dans cette tâche.

L'objectif de l'analyse est de vérifier si les verbes statiques sont plus fréquents à la forme courte et les verbes dynamiques à la forme longue.

Tableau 11-4 : Production des verbes selon l'aspect lexical – tâche de narration

	stade 1			stade 2			stade 3			Σapprenant/es			natifs		
	C	L	%C	C	L	%C	C	L	%C	C	L	%C	C	L	%C
adorer	1	0	100%	2	0	100%	1	1	50%	4	1	80%	0	0	-
aimer	2	0	100%	9	0	100%	9	0	100%	20	0	100%	1	0	100%
détester	0	0	-	0	0	-	0	0	-	0	0	-	0	0	-
habiter	0	0	-	2	0	100%	0	0	-	2	0	100%	1	0	100%
penser	0	2	0%	8	0	100%	7	0	100%	15	2	88%	6	1	86%
préférer	0	0	-	0	0	-	0	0	-	0	0	-	0	1	0%
trouver S	0	0	-	0	0	-	2	0	100%	2	0	100%	1	0	100%
<i>statique</i>	3	2	60%	21	0	100%	19	1	95%	43	3	93%	9	2	82%
acheter	0	1	0%	2	3	40%	2	7	22%	4	11	27%	1	5	17%
manger	4	4	50%	12	7	63%	10	4	71%	26	15	63%	0	5	0%
parler	1	3	25%	8	3	73%	5	3	63%	14	9	61%	4	3	57%
regarder	2	2	50%	6	4	60%	0	2	0%	8	8	50%	7	4	64%
trouver D	1	0	100%	0	0	-	1	2	33%	2	2	50%	5	1	83%
visiter	0	0	-	1	1	50%	0	3	0%	1	4	20%	0	5	0%
<i>dynamique</i>	8	10	44%	29	18	62%	18	21	46%	55	49	53%	17	23	43%

Légende : C : forme courte ; L : forme longue ; Σ : somme ; trouver S : sens statique ; D : sens dynamique.

À quelques rares exceptions près, les verbes statiques ont été produits à la forme courte comme attendu par l'hypothèse 1. On peut voir que le pourcentage de formes courtes est plus bas au stade 1 qu'aux autres stades où l'on est proche de 100%.

Pour les verbes dynamiques, on peut voir que les apprenant/es des stades 1 et 3 ont produit plus de formes longues que courtes pour ces verbes, alors qu'au stade 2 la forme courte l'emporte. La forme longue est utilisée avec la plupart des verbes dynamiques du stade 1 avec 50% et plus de formes longues, mais il y a très peu d'occurrences. Au stade 2 seuls les verbes *acheter* et *visiter* ne sont pas plus souvent produits à la forme courte. Au stade 3, seuls *manger* et *parler* n'ont pas été préférés à la forme longue.

Chez les natifs, la forme longue l'emporte aussi pour les verbes dynamiques, mais on peut voir que plusieurs verbes ont été produits à la forme courte par les natifs, ce qui implique que le haut pourcentage de formes courtes pourrait aussi être dû à la tâche. Seuls les verbes *acheter*, *manger* et *visiter* ont été préférés à la forme longue par les natifs.

Pour revenir aux apprenant/es, on peut dire que l'hypothèse 1 se vérifie assez bien sur ces données. La forme courte a été préférée avec les verbes statiques et la forme longue a été préférée avec un grand nombre de verbes dynamiques. Un calcul statistique sur le pourcentage de formes courtes au total pour les apprenant/es selon l'aspect lexical, sans les verbes *préférer* et *détester*, qui n'ont

pas été produits par les apprenant/es, a montré que la différence était significative (Mann-Whitney U, $Z=-2,770$ $p=.004$). On peut par contre voir que la différence de pourcentage au niveau des verbes en tant que groupe est la plus faible au stade 1 et correspond le mieux à l'hypothèse au stade 3. Un tel résultat indique que l'influence de l'aspect lexical pourrait se mettre en place qu'après une certaine expérience de la langue.

Dans le but de vérifier l'hypothèse 2, sur l'influence de la fréquence des formes dans l'input, le tableau 11-5 ci-dessous présente le pourcentage de formes courtes pour les 12 verbes de test, classés selon le continuum de pourcentage de formes courtes. Le tableau 11-5 ne présente que les pourcentages au total, mais les pourcentages pour les différents stades peuvent être lus dans le tableau 11-4 ci-dessus. Pour mémoire, les données de l'input ont été analysées au chapitre 8 et les données présentées ci-dessous proviennent des tableaux 8-2 et 8-3. Les prédictions pour chaque verbe à partir de l'influence de la fréquence de surface ont été présentées dans le tableau 8-7.

Les données des verbes *détester* et *préférer* ont été mises à part étant donné que ces verbes n'ont pas été produits par les apprenant/es.

Tableau 11-5 : Pourcentages de formes courtes – tâche de narration et input

apprenant/es	Σ C+L	% C	input	Σ C+L	% C
aimer	20	100%	adorer	32	81%
habiter	2	100%	penser	368	73%
trouver S	2	100%	trouver S	182	72%
penser	17	88%	habiter	150	72%
adorer	5	80%	aimer	341	55%
<i>trouver Σ</i>	6	67%	<i>trouver Σ</i>	393	47%
manger	41	63%	regarder	223	39%
parler	23	61%	parler	581	38%
regarder	16	50%	manger	160	29%
trouver D	4	50%	trouver D	211	26%
acheter	15	27%	acheter	110	15%
visiter	5	20%	visiter	25	8%
<i>détester</i>	0		<i>détester</i>	10	70%
<i>préférer</i>	0		<i>préférer</i>	45	58%

Légende : C : forme courte ; L : forme longue ; trouver S : sens statique ; D : sens dynamique ; Σ : sens statique + dynamique.

La comparaison des échelles de pourcentage de formes courtes entre les apprenant/es de la partie expérimentale et l'input montre que les prédictions formulées au niveau de l'influence de la fréquence de surface correspondent à ce qui est produit par les apprenant/es. Du moment où plusieurs verbes n'ont que

peu d'occurrences, ces données montrent surtout des tendances. Les verbes *aimer*, *habiter*, *penser* et *adorer* ont été produits massivement à la forme courte alors que les verbes *acheter* et *visiter* ont plus souvent été produits à la forme longue. Le verbe *trouver dynamique* n'a par contre pas été produit comme un verbe fréquent à la forme longue. Les verbes *trouver* (au total) *manger*, *parler*, *regarder* présentent plus de variation que les autres, ce qui correspond à ce qui est attendu par l'hypothèse 2, même si les pourcentages de formes courtes sont plus élevés que dans l'input. Comme on l'a vu ces apprenant/es ont une préférence pour le contexte du présent, qui reflète aussi la fréquence de la forme courte dans le manuel *Escalade 1*. La corrélation entre les données de la tâche de narration et les données d'input au total est de .797 (r de Spearman), ce qui est une excellente corrélation.

Les données obtenues en production libre à partir de la narration du *Voyage en Italie* parlent donc en faveur des deux hypothèses, mais l'hypothèse 2 semble mieux expliquer la production des apprenant/es du moment où l'on retrouve les trois groupes de fréquences de l'input : les verbes statiques, fréquents à la forme courte dans l'input sont produits à la forme courte à un haut pourcentage par les apprenant/es. Les verbes dynamiques, fréquents aux deux formes dans l'input sont produits à un pourcentage plus faible de formes courtes alors que les verbes dynamiques fréquents à la forme longue sont clairement produits à la forme longue par les apprenant/es.

11.5. Résumé des données de la tâche de narration

Le traitement des données de la tâche de production libre, *Le voyage en Italie*, a servi à la vérification du niveau des apprenant/es, à l'analyse des principales caractéristiques de la production des apprenant/es de la partie expérimentale et finalement à la vérification des hypothèses sur les verbes de tests produits dans cette tâche.

La vérification du niveau des apprenant/es a été faite à partir de deux mesures : d'une part l'évaluation du stade morphosyntaxique d'après Bartning et Schlyter (2004) à l'aide du logiciel *Direkt Profil* (Granfeldt, Nugues *et al.*, 2005) et d'autre part d'après la mesure du débit en nombre de mots par minute. Aucun apprenant n'a été évalué à un stade supérieur au stade 3, ce qui signifie que les données des 33 participant/es seront analysées dans les sections qui suivent. Le regroupement des apprenant/es a donné 9 participant/es au stade 1, 14 au stade 2 et 10 au stade 3, qui est le stade le plus avancé. Ces deux mesures

ont indiqué une bonne corrélation, même si les différences entre les stades 2 et 3 semblent faibles. Malgré cela, cette répartition des apprenant/es sera gardée pour la suite dans le but de voir si on trouve des caractéristiques selon le niveau plus ou moins débutant des apprenant/es.

L'analyse de la production de tous les verbes réguliers en *-er* dans la tâche de production libre a montré que les apprenant/es de la partie expérimentale semblent surtout utiliser le contexte du présent. Dans ce contexte, on trouve environ 40% de formes longues ce qui indique que l'on retrouve ici la variation caractéristique des apprenant/es débutant/es. On trouve néanmoins beaucoup de variation individuelle (cf. annexe 4). Les contextes pour une forme de l'infinitif n'apparaissent qu'au stade 3 ici et sont le plus souvent produits correctement.

Finalement, la vérification des deux hypothèses sur les verbes de test produits a montré

- que les verbes statiques, fréquents à la forme courte dans l'input, étaient surtout produits à la forme courte par les apprenant/es, comme attendu par les deux hypothèses
- que les verbes dynamiques, fréquents aux deux formes dans l'input étaient préférés à la forme courte, à un pourcentage plus faible de formes courtes que les verbes statiques, ce qui n'est pas attendu pour les verbes dynamiques selon l'hypothèse 1 (influence de l'aspect lexical), mais par l'hypothèse 2 (influence de la fréquence des formes)
- que les verbes dynamiques fréquents à la forme longue dans l'input étaient clairement préférés à la forme longue par les apprenant/es (sauf le verbe *trouver dynamique*) ce qui parle en faveur des deux hypothèses.
- que l'on pouvait constater une possible influence de l'aspect lexical au stade 3

Bien que l'on trouve davantage de formes longues avec les verbes dynamiques qu'avec les verbes statiques (la différence au niveau du pourcentage de formes courtes entre les deux catégories de verbes était significative), une comparaison des échelles de fréquence de forme courtes parle plutôt en faveur d'une influence de la fréquence des formes que de l'aspect lexical avec un taux de corrélation de .797. Le fait que les verbes dynamiques moins fréquents à la forme longue dans l'input aient été préférés à la forme courte par les apprenant/es parle également en faveur de l'hypothèse 2 sur l'influence de la fréquence des formes dans l'input.

12. Test d'imitation

Ce chapitre présente l'analyse des données du test d'imitation. Ce test psycholinguistique permet de tester le traitement et la production des verbes dans des phrases. Les 12 verbes de tests ont été soumis aux apprenant/es dans les contextes du présent et de l'infinitif, chaque verbe une fois dans chaque contexte. L'objectif de ce test est de voir si les apprenant/es produisent la forme qui correspond à l'aspect lexical du verbe (hypothèse 1) ou à la forme la plus fréquente dans l'input (hypothèse 2) indépendamment du contexte syntaxique du stimulus et de l'output de l'apprenant/e.

Ce chapitre comporte tout d'abord une présentation du test (section 12.1) puis de son déroulement et de sa construction (section 12.2). La section 12.3. sera consacrée au traitement des données et la section 12.4. aux résultats. Le chapitre se terminera par un résumé.

12.1. Présentation du test d'imitation

Dans le test d'imitation sollicitée (*elicited imitation*), le participant entend une phrase qu'il doit ensuite répéter le plus exactement possible. Ce test a été utilisé dans le cadre de recherches sur le développement de la langue aussi bien pour des locuteurs L1 que L2 (Naiman, 1974 ; Markman *et al.*, 1975 ; Gallimore & Tharp, 1981 ; Munnich *et al.*, 1994 ; R.Ellis, 2005 ; Erlam, 2006). Le principal avantage de ce test est sa proximité avec le mode de production orale tout en étant ciblé sur les structures à éliciter.

This test is a more practical alternative to communicative tests of free language production in that it allows for the testing of learners' knowledge of specific language structures. (Erlam, 2006 : 467)

Cette méthode permet en quelque sorte de forcer la rencontre entre l'apprenant/e et des structures qu'il cherchera peut-être à éviter en production spontanée sachant qu'elles lui posent problème (Naiman, 1974 : 3). Elle permet ainsi une analyse des structures produites mais aussi de celles qui sont omises (non imitées), ce qui est un avantage sur des données de production libre.

Le point le plus discuté depuis le début de l'utilisation de ce test est sa nature ou plus exactement la nature de l'output obtenu (*cf.* e.a. Naiman, 1974 : 3 ; Vinther, 2002 : 54-55). Le participant a-t-il simplement imité la phrase stimulus comme une chaîne de sons ou l'a-t-il comprise et re-produite avec ses propres moyens ? Les résultats de nombreuses études semblent indiquer que le test d'imitation implique plusieurs niveaux de traitement : les systèmes de compréhension et de production du langage ainsi que la mémoire (Bley-Vroman & Chaudron, 1994 : 247). Ce test semble engager un processus dit *reconstructif* durant lequel le participant décode la phrase perçue pour ensuite la re-encoder avec sa propre grammaire.

In reconstruction, the subject hears the sentence and, to the extent that the sentence is comprehended, reconstructs the meaning using his or her own grammar. (Bley-Vroman & Chaudron, 1994 : 246)

Selon Naiman (1974 : 4) le participant doit utiliser les règles de son propre système de production pour re-encoder la phrase, les mêmes qu'il utilise en production spontanée. Pour qu'une phrase puisse être imitée correctement, il ne suffit pas qu'elle soit comprise, il faut que le participant ait suffisamment de grammaire et de mémoire pour la ré-encoder. Ce test permet ainsi de comparer la grammaire de la phrase stimulus avec celle du participant/e (Bley-Vroman & Chaudron, 1994 : 245). Naiman (1974 : 23) propose quatre possibles scénarios :

- a) la structure du stimulus n'est pas présente dans la grammaire et il n'y aura pas d'imitation correcte
- b) une forme alternative est présente dans la grammaire et sera produite à la place de la forme stimulus
- c) la structure commence à se stabiliser et elle sera parfois imitée correctement
- d) la structure est présente et elle est imitée correctement

Comme le souligne Vinther dans la citation ci-dessous, les outputs les plus intéressants sont ceux qui ne correspondent pas complètement au stimulus, car ils reflètent le niveau linguistique du participant/e.

Whether this is due to a failure to memorize the content (possibly because of the amount of attention invested in decoding) or to the fact that the subject has not mastered the necessary formal system to recode the meaning, this output provides the most interesting

F[oreign]L[anguage]A[cquisition] research data, since it reflects the subject's level of language proficiency. (Vinther, 2002 : 58)

Pour que le test donne des résultats intéressants, il est donc important de tenir compte de trois principaux facteurs : la longueur et la structure des phrases stimulus, le niveau grammatical de l'apprenant/e ainsi que la capacité de la mémoire de travail et du stock phonologique de celui-ci.

Pour éviter une répétition par cœur (*rote imitation*), la longueur de la phrase stimulus doit dépasser la capacité (l'empan) de la mémoire de travail et du stock phonologique du participant tout en étant suffisamment courte pour pouvoir être traitée (Naiman, 1974 : 25). Comme le soulignent Bley-Vroman et Chaudron (1994 : 251) la longueur de la phrase stimulus est subjective : elle dépend du participant/e. Ainsi un apprenant complètement débutant risque de traiter le stimulus à un niveau purement phonologique comme une série de non-mots ou à un niveau lexical s'il réussit à reconnaître un certain nombre de mots, alors que l'apprenant plus avancé pourra rappeler une plus grande partie de la phrase stimulus par le processus de regroupement des mots, *chunking*. C'est à ce niveau que l'on trouve la relation entre grammaire et mémoire.

If the grammar can reduce the number of units so that they can fit into short-term memory, then the sentence has a chance of being repeated accurately. (Bley-Vroman & Chaudron, 1994 : 248)

De ce point de vue, la grammaire permet donc le regroupement des membres de la phrase ce qui facilite la mémorisation puis l'imitation de celle-ci. Les résultats de Markman *et al.* (1975) soulignent plus généralement le rôle de l'expérience de la langue cible, notamment au niveau du vocabulaire. Si la grammaire permet d'assembler les constituants, la connaissance du vocabulaire facilite le décodage et la mémorisation du contenu sémantique de la phrase (*cf.* aussi Ellis, 2001 : 48). Bref, une meilleure connaissance de la langue cible permet de mieux organiser le matériel de la phrase stimulus et augmente les imitations correctes.

La mesure de capacité de la mémoire de travail ainsi que du stock phonologique (empan) des apprenant/es pourrait donner des informations intéressantes et augmenter la validité des résultats obtenus. Pour des raisons de contraintes de temps au niveau de la durée totale de l'expérience, de tels tests n'ont finalement pas été inclus. Plusieurs recherches semblent indiquer qu'un bon empan facilite l'acquisition d'une langue étrangère notamment au niveau du vocabulaire (Gathercole, 1995 ; Gathercole & Baddeley, 1993 ; Service, 1992 ; Baddeley *et al.*, 1988).

On pourrait ainsi s'attendre à ce que les personnes évaluées comme étant à un niveau de français LE/L2 plus avancé réussissent mieux le test d'imitation que les personnes à un niveau moins avancé, en raison de l'interaction entre grammaire et mémoire. Un bon empan augmente les possibilités de noter et de stocker des mots isolés, mais aussi des constructions grammaticales (syntaxe) qui à leur tour aident à la mémorisation de matériel linguistique.

Repetition of sequences in the phonological loop allows their consolidation in phonological L[ong] T[erm] M[emory]. (Ellis, 2001 : 41)

En résumé, le test d'imitation sollicitée est une bonne alternative aux données en production libre et il présente de nombreux avantages : il implique les mêmes mécanismes qu'en production orale, il permet d'étudier la grammaire des participant/es de manière ciblée et surtout il permet d'observer le traitement de l'input en direct. En outre, des études récentes (R.Ellis, 2005 ; Erlam, 2006 ; R.Ellis, 2008) ont montré que les résultats de ce test étaient comparables à des données obtenues en production libre.

12.2. Déroulement et construction du test d'imitation

Le test d'imitation utilisé dans ce travail suit le déroulement suivant⁵³ :

1. Un point d'exclamation apparaît au milieu de l'écran durant 2 secondes suivi du son de la phrase enregistrée.
2. Dès la fin du son, les chiffres 1-2-3-4-5-6-7-8-9 apparaissent en haut de l'écran, durant 4 secondes. À ce moment-là, le participant doit compter une fois de 1 à 9 en français (tâche de suppression articulatoire).
3. Lorsque les chiffres disparaissent, une image sensée rappeler le contenu de la phrase apparaît ainsi qu'une série de traits et c'est le moment où le participant doit rappeler la phrase (ou ce dont il/elle se souvient).
4. Lorsqu'il a terminé, le participant doit cliquer pour aller à la phrase suivante (*self paced*).

⁵³ Le test a été soumis aux apprenant/es sous la forme de diapositives du logiciel *Microsoft PowerPoint*.

Cette section présentera le matériel utilisé dans ce test à commencer par les phrases puis la tâche de suppression articulatoire, les images indices et finalement les consignes.

12.2.1. Construction du matériel sonore (phrases stimulus)

La version du test d'imitation construite pour la présente recherche comporte 27 phrases de test et 4 phrases d'exercice. Deux phrases ont été créées pour chaque verbe, une avec le verbe à la forme courte du présent et une avec le verbe à la forme longue de l'infinitif. Pour le verbe *trouver* on a un total de cinq phrases : une phrase au présent et une à l'infinitif par sens aspectuel – statique et dynamique – puis une phrase supplémentaire au passé composé dans le sens dynamique. La raison de cette unique phrase au passé composé est liée aux résultats d'une série de tests pilotes aux cours de laquelle le verbe *trouver* avait soit été omis soit remplacé par un autre verbe. Comme les données sur la fréquence des formes dans l'input avaient montré que la forme du participe passé était la forme la plus fréquente pour le sens dynamique de ce verbe (cf. tableau 8-2 dans le chapitre sur l'input) une phrase au passé composé a été ajoutée pour voir si ce verbe était plus facilement rappelé au passé ou si le verbe était difficile à imiter quel que soit le contexte. Les données pour cette phrase différente des autres ne seront pas retenues si elle ne donne pas de résultat particulier.

Les 27 phrases de test sont présentées dans le tableau 12-1 dans l'ordre alphabétique des verbes suivies des quatre phrases d'exercice (en italique). Ce tableau présente également le nombre de syllabes (*syll*) et de mots pour chaque phrase ainsi que la durée de l'enregistrement en secondes (calculées par le logiciel *Praat*).

Tableau 12-1 : Phrases du test d'imitation par ordre alphabétique des verbes

n°	début	sujet	aux	verbe	fin	syll	mots	durée"
1	Au supermarché	Carl		achète	un melon	11	6	3,153
2	Aujourd'hui	Christine	veut	acheter	une robe	10	6	3,405
3	Je crois que	Marie		adore	les cafés	10	7	2,351
4	Elle croit que	Carl	va	adorer	ce dessert	11	8	2,627
5	En hiver	Bruno		aime	le cinéma	10	6	2,782
6	En été	Marie	va	aimer	la Suède	11	7	3,092
7	Il croit que	Marie		déteste	la musique	10	7	2,922
8	Elle croit que	Carl	va	détester	Paris	10	7	2,712
9	En France	Carl		habite	au bord de la mer	10	9	3,236
10	En mars	Marie	veut	habiter	à Paris	11	7	2,914
11	A midi	Christine		mange	au restaurant	10	6	3,018
12	Maintenant	Bruno	veut	manger	une salade	10	6	3,145

n°	début	sujet	aux	verbe	fin	syll	mots	durée"
13	Au travail	Carl		parle	avec une collègue	10	7	3,487
14	Maintenant	Marie	veut	parler	avec son chef	11	7	3,085
15	Le matin	Christine		pense	à sa famille	10	7	3,302
16	Aujourd'hui	Carl	veut	penser	aux vacances	10	6	3,253
17	A Paris	Christine		préfère	le métro	10	6	3,124
18	A Londres	Bruno	va	préférer	le bus	10	7	3,198
19	Maintenant	Carl		regarde	la télévision	10	5	3,15
20	Samedi	Christine	va	regarder	un film	10	6	3,565
21	Aujourd'hui	Marie		trouve	qu'il fait très froid	10	7	3,448
22	A Nice	Carl	va	trouver	que la mer est belle	11	10	3,469
23	Au marché	Bruno		trouve	des belles tomates	10	7	3,484
24	J'espère que	Christine	va	trouver	un travail	11	7	3,373
25	Je vois que	Bruno	a	trouvé	ses clés	10	8	2,626
26	A Paris	Marie		visite	la tour Eiffel	11	7	3,337
27	En avril	Carl	veut	visiter	l'Espagne	10	6	3,459
<i>ex a</i>	<i>Le dimanche</i>	<i>Christine</i>		<i>fait</i>	<i>beaucoup de sport</i>	<i>10</i>	<i>7</i>	<i>3,406</i>
<i>ex b</i>	<i>A trois heures</i>	<i>Carl</i>	<i>va</i>	<i>faire</i>	<i>de la musique</i>	<i>10</i>	<i>9</i>	<i>3,781</i>
<i>ex c</i>	<i>A Copenhague</i>	<i>Bruno</i>		<i>prend</i>	<i>un taxi</i>	<i>10</i>	<i>6</i>	<i>3,168</i>
<i>ex d</i>	<i>Le week-end</i>	<i>Marie</i>	<i>veut</i>	<i>lire</i>	<i>des romans</i>	<i>10</i>	<i>7</i>	<i>3,049</i>

Le nombre de syllabes a été maintenu le plus constant possible avec 19 phrases à 10 syllabes et 8 phrases à 11 syllabes pour une moyenne de 10,30 syllabes par phrase et un écart type de 0,47⁵⁴. Les recherches psycholinguistiques semblent en effet indiquer que la syllabe est une unité de segmentation fondamentale dans le traitement de la parole, du moins en français⁵⁵ (Gineste & Le Ny, 2002 : 27-31 ; Segui & Ferrand, 2002 ; Billières & Spanghero-Gaillard, 2004).

Le nombre de mots n'a par contre pas pu être tenu constant : les phrases contiennent entre 5 et 10 mots avec une moyenne de 6,85 mots et un écart type de 1,03⁵⁶. Selon Munnich *et al.* (1994 : 233), c'est la structure de la phrase stimulus plutôt que le nombre de mots qui détermine les résultats, du moment où le nombre de syllabes est identique.

La durée moyenne des enregistrements est de 3,14 secondes y compris une plage de silence avant (\pm 0,20 secondes) et après (\pm 0,30 secondes) chaque

⁵⁴ Les syllabes ont été comptées à partir de la prononciation orale et sans l'expression des *e* muets. Ainsi *parle* correspond à une syllabe et *acheter* à deux.

⁵⁵ La segmentation par syllabes pourrait jouer un rôle moins important en anglais et dans les langues germaniques (Segui & Ferrand, 2002).

⁵⁶ Les mots se terminant par une apostrophe ont été comptés ensemble avec le mot suivant (*aujourd'hui* = 1 mot, *qu'il* = 1 mot) ainsi que les mots avec un trait d'union (*week-end* = un mot).

phrase. Ces silences ont été laissés pour éviter que la mise en route et la fin de la plage de son n'interfère avec la phrase stimulus.

Pour ce qui est de la structure, les phrases comprennent cinq parties : un début, un sujet, un auxiliaire pour les phrases avec le verbe testé à l'infinitif, le verbe de test et une fin. La forme verbale de test se trouve toujours au milieu de la phrase pour éviter au maximum la simple imitation de l'élément central du test. En effet, les données de Naiman (1974 : 25) montrent que les éléments en début de phrase sont significativement mieux rappelés que les autres et les éléments en fin de phrase mieux que ceux au milieu (effets sériels de primauté et de récence ; Ganoac'h & Larigauderie, 2000 : 27). La position au milieu de la phrase est donc la plus difficile à rappeler.

Les phrases ont été construites avec une structure aussi simple que possible tout en restant naturelles et sans tomber dans des phrases toutes faites. Les phrases obtenues sont probablement un peu plus difficiles que celles des livres de cours où le sujet est souvent le premier élément de la phrase (surtout dans les exercices de production par l'étudiant/e).

Les tests d'imitations comprennent souvent un certain nombre de phrases agrammaticales ou déviantes (*cf.* Markman *et al.*, 1975 ; Munnich *et al.*, 1994 ; Erlam, 2006), mais ce n'est pas le cas ici. En effet, la présente étude ne portant que sur deux structures cible, à savoir le verbe à la forme courte du présent et à la forme longue de l'infinitif, l'agrammaticalité aurait dû s'exprimer sur la forme verbale ce qui pourrait attirer l'attention du participant/e sur la forme.

Les débuts de phrases comprennent des expressions de lieux (8 fois), des expressions de temps aussi neutres que possible pour éviter la référence au passé (12 fois) et finalement des syntagmes avec un verbe de pensée (6 fois). Les syntagmes comprenant un verbe de pensée introduisent les phrases avec les formes *adore*, *adorer*, *déteste* et *détester*, des verbes statiques, mais aussi les phrases avec *trouvé* et *trouver* dans le sens dynamique. Le recours à de telles constructions est motivé par la difficulté de trouver des débuts de phrases accessibles au niveau du vocabulaire pour les verbes statiques à l'infinitif. Étant donné que les formes longues des verbes statiques sont plus rares dans l'input et donc moins aisées à produire selon l'hypothèse formulée, la même structure syntaxique a été utilisée pour la forme courte du verbe concerné. Ainsi le niveau de difficulté lié à la subordination de la forme verbale de test sera semblable pour les deux formes. Selon Clahsen & Felser (2006) les apprenant/es L2 semblent sous-utiliser l'information syntaxique durant le traitement. Selon eux, les apprenant/es semblent plutôt centrer leur attention sur la structure prédicat – arguments et les informations sémantiques et pragmatiques. Si tel est le cas, on

ne devrait donc pas trouver de différences liées à la structure syntaxique des phrases.

Toutes les phrases à l'infinitif ont été construites avec un auxiliaire. Il n'y a donc pas de constructions prépositionnelles. L'utilisation de phrases avec l'auxiliaire *aller* induit une dimension temporelle prospective plus ou moins explicite qui n'a pas pu être évitée.

Malgré les efforts pour minimiser les différences entre les phrases de test, on peut constater que les phrases avec le verbe *trouver* accumulent les particularités: le verbe de la phrase 24 est en subordination, alors que dans les phrases 21 et 22 les verbes introduisent une subordonnée. Ces particularités s'ajoutent à la difficulté du double sens du verbe, statique (*à son avis*) et dynamique (*trouver ce que l'on cherche*). On peut donc faire l'hypothèse que les phrases avec le verbe *trouver* seront généralement moins bien rappelées.

Les quatre phrases d'exercice (en italique dans le tableau 12-1) ont été créées selon les mêmes principes que les phrases de test, sauf que les verbes utilisés sont des verbes irréguliers. Les débuts et les fins de phrases sont différents de ce que l'on trouve dans les phrases de test, mais les prénoms et les auxiliaires sont les mêmes.

Comme on l'a vu dans la présentation du test d'imitation, la familiarité avec le vocabulaire utilisé dans les phrases peut avoir une influence sur la capacité à les reproduire. Pour cela, la présence des mots utilisés a été contrôlée dans le manuel pour débutants *Escalade 1* (Tillman *et al.*, 2001) en consultant les listes alphabétiques français-suédois et suédois-français à la fin du livre, ainsi que les planches de vocabulaire illustrées et les pages de grammaires. Sur la centaine de mots différents utilisés dans les phrases, une quinzaine n'a pas été trouvée dans *Escalade 1*. Il s'agit de mots qui sont identiques en suédois (*collègue, chef, bus, film, télévision, et taxi*), des formes verbales *crois, croit, espère, vois*, du pronom démonstratif *ce* et du substantif *marché*. La forme *veut* est utilisée dans des constructions, mais le verbe *vouloir* n'est pas présenté en tant que tel. La plupart des mots utilisés dans les phrases de test pourraient donc être connus des apprenant/es débutant/es. Le fait que les verbes utilisés en début de phrase (*crois, croit, espère, vois*) ne se trouvent pas dans le manuel *Escalade 1* pourrait ajouter une difficulté supplémentaire au rappel de ces phrases. Ainsi les phrases 3, 4, 7, 8, 24 et 25 (verbes *adorer, détester et trouver*) risquent d'être moins bien rappelées que les autres.

Les 27 phrases ont été présentées en suivant l'ordre de quatre listes différentes afin de minimiser les effets de liste.

Pour la première liste – la liste A – les phrases ont été mises dans un ordre aléatoire à partir des numéros de phrases du tableau 12-1 ci-dessus. Cet ordre a été corrigé pour éviter que les deux formes verbales du même verbe se suivent, ainsi que pour éviter d'avoir plus de deux fois de suite une forme longue ou courte, un verbe statique ou dynamique. L'unique phrase au passé composé (n° 25, *a trouvé*) a été mise en dernier sur toutes les listes, afin de ne pas suggérer la possibilité de contextes pour le passé pour les autres phrases.

Dans le but d'alterner la première forme du verbe rencontrée, une liste B en miroir de la liste A a été créée. Dans cette liste B la phrase avec la forme courte et celle avec la forme longue du même verbe ont été échangées.

Finalement, les listes C et D ont été créées en inversant les 13 premières et les 13 dernières phrases de la liste A respectivement B dans le but de réduire les effets de série (donc sans compter la phrase au passé composé qui est toujours en dernier).

Les phrases ont été enregistrées par une Française originaire de Nantes, âgée de 20 ans, qui, au moment de l'enregistrement final, avait passé environ 9 mois en Suède. Il s'agit de la même personne qui avait enregistré le matériel pour deux séries de tests pilotes. L'enregistrement a été fait dans une salle sans écho et avec un enregistrement numérique (à l'aide du logiciel *Audacity*). Le script d'enregistrement a été fait à partir des listes A et B du test d'imitation, recopiées puis mélangées manuellement pour éviter que les phrases soient lues dans un même ordre. Les listes d'enregistrement ont été faites de telle sorte qu'il n'y ait pas deux fois la même phrase en fin de page et en général qu'il n'y ait pas deux phrases de suite avec le même verbe (ni même à une forme différente). Chaque phrase a été enregistrée trois fois, à raison de trois tours successifs comprenant toutes les phrases. La consigne donnée était de regarder la phrase puis de la dire par cœur au micro afin que les phrases ressemblent le plus possible à du français parlé et le moins possible à du français lu.

Le matériel sonore a ensuite été traité avec le logiciel *Praat*. Les critères qui ont servi de base à la sélection de la meilleure version de chaque phrase sont les suivants :

- bonne articulation de la phrase, rythme plutôt lent
- forme verbale bien perceptible
- pas de bruits de langue, ni trop de « ss », pas de voix qui s'étouffe
- intonation naturelle
- bonne tonalité, ni trop fort ni trop doucement

12.2.2. Tâche de suppression articulatoire

Pour éviter le plus possible que les phrases ne soient simplement imitées par cœur et donc pour augmenter les chances d'un traitement complet et reconstitutif des phrases stimulus, une tâche de suppression articulatoire a été insérée dans la construction du test d'imitation. Selon le modèle de mémoire de Baddeley (Baddeley, 2000), c'est la *boucle phonologique* (*phonological loop*) qui joue un rôle essentiel dans le maintien temporaire des éléments en mémoire (*stock phonologique*) lors du traitement de la langue parlée. Cette boucle phonologique est formée de deux composants (Gathercole & Baddeley, 1993 : 8 ; Olive & Piolat, 2005 : 376) :

- le *stock phonologique* qui maintient temporairement les représentations sous forme de code phonologique.
- le mécanisme d'*articulation subvocale* qui rafraîchit les représentations en voie de détérioration (*decaying*) dans le stock phonologique et par là permet de maintenir les éléments en mémoire

Une des fonctions de la boucle phonologique serait d'aider à l'apprentissage de nouveaux mots, le séjour de nouveaux mots dans la boucle phonologique aidant à la construction de représentations à long terme (Baddeley *et al.*, 1998 : 158).

D'après les recherches menées par Baddeley et ses collègues il est possible d'empêcher le travail du mécanisme d'articulation subvocale en demandant au participant/e de répéter du matériel hors de propos, par exemple une série de chiffres. Par cette technique de suppression articulatoire (*articulatory suppression*) le contenu du stock phonologique ne sera pas rafraîchi et décroîtra rapidement (Gathercole & Baddeley, 1993 : 9). La capacité du stock phonologique est estimée à environ 2 secondes (Cowan *et al.*, 2003 : 74 ; Speciale *et al.*, 2004 : 294), raison pour laquelle il est important que la tâche de suppression dure au moins deux secondes (McDade *et al.*, 1982, proposent quant à eux un délai de 3 secondes pour une tâche d'imitation).

Dans les expériences de Cowan *et al.* (2003), les participant/es devaient par exemple compter de 1 à 8 en 1,5 secondes, dans Fournet *et al.* (2003) lire un certain nombre de chiffres à haute voix durant 2, 8 ou 24 secondes, dans Mobayyen & de Almeida (2005) compter à l'envers durant 9 secondes et dans Poirier & Saint-Aubin (1995) répéter le mot *mathématiques* aussi durant le son. L'utilisation des chiffres dans une tâche de suppression articulatoire semble donc courante. Pour cette raison, j'ai choisi de demander aux participant/es de compter une fois de 1 à 9 en français, les chiffres, 1-2-3-4-5-6-7-8-9, étant

présentées à l'écran durant 4 secondes. Des tests pilotes ont montré que les apprenant/es mettaient facilement plus de quatre secondes pour compter de 1 à 9, même après exercice. Pour cette raison, j'ai demandé aux participant/es de compter jusqu'à ce qu'apparaisse l'image qui aide au rappel du contenu de la phrase, même s'ils n'étaient pas arrivés à 9, l'important étant l'articulation de matériel autre que celui des phrases durant 4 secondes.

12.2.3. Images indices

Le rôle des images dans le déroulement du test d'imitation est de donner une piste au participant/e pour le rappel du contenu de la phrase entendue, mais aussi de centrer son attention sur le contenu plutôt que sur la forme (Erlam, 2006 : 471).

Étant donné que ce test est centré sur les verbes, les images devaient le plus possible inclure le sens du verbe. Pour certains verbes comme *acheter* ou *manger*, de telles images sont possibles contrairement à *aimer* ou *adorer* pour lesquels il est presque impossible de trouver une bonne image. Par exemple, pour la phrase *En été Marie va aimer la Suède* l'image présente une jolie maison suédoise avec ciel bleu et drapeau suédois.

Le rôle de l'image étant d'aider au rappel après que le participant a calculé de 1 à 9, il était important qu'elle ne prenne pas trop d'attention cognitive. J'ai donc préféré prendre le risque que l'image n'aide pas, plutôt que de demander implicitement au participant/e de *comprendre ce que l'expérimentatrice a bien voulu dire*. Pour cette raison, il y a toujours une seule image par phrase, même pour les phrases avec le verbe *penser* (*à sa famille, aux vacances*).

Les images ne devaient pas non plus laisser transparaître les facteurs qui font l'objet de cette recherche, à savoir statique vs dynamique et présent vs infinitif/futur. Par exemple, les images pour les deux phrases avec le verbe *acheter* montrent une personne en train de considérer l'achat de l'objet (*melon, robe*). L'utilisation (respectivement la sur-utilisation) des deux formes est donc possible dans les deux phases. De même, la phrase *En hiver Bruno aime le cinéma* est montrée par un homme content d'être au cinéma et la phrase *Samedi Christine va regarder un film* une femme au cinéma.

Toutes les images ont été trouvées dans une base de données de photos destinées à la vente sur internet, raison pour laquelle elles ne sont pas reproduites ici.

12.2.4. Consignes

Les consignes ont été données oralement en suédois par l'expérimentatrice. Il s'agissait d'une description du déroulement tel qu'il a été présenté au début de cette section 12.2. Les trois moments visuels, le point d'exclamation, les chiffres et l'image, étaient présentés sur papier l'un après l'autre. Les participant/es étant des débutants, la priorité a été mise sur une consigne qui encourage avant tout à la production de matériel plutôt qu'une répétition exacte : "là, tu répètes la phrase que tu avais entendue. Si tu ne te souviens pas de tout, dis les mots dont tu te souviens, même si ce ne sont que des bouts de la phrase."

12.3. Traitement des données

12.3.1. Conventions de transcription

La transcription des phrases produites dans le cadre du test d'imitation a été faite dans un tableur avec une attention particulière à la production des verbes de test. Les enregistrements ont été écoutés au moins trois fois, deux fois au moment de la première transcription du matériel (un/e participant/e après l'autre) et une autre fois lors de la réécoute (par tâche). Les enregistrements de P11, P16, P19, P21 et P22 ont également été écoutés et transcrits par une autre personne. Aucune différence majeure n'a été détectée au niveau des verbes de test.

Les phrases produites par les participant/es ont été réparties dans des colonnes comprenant les informations suivantes : le stade, le code participant/e, le numéro de la phrase, le début, le sujet, le verbe modal, le verbe de test, la fin et le codage. Voici un exemple en (12-1) avec le stimulus (*stim*) et la production de P06 pour la phrase n° 8. Le codage sera présenté à la section suivante.

(12-1)	Stade	Part	n°	Début	Sujet	mod	verbe	fin	Code
			<i>Stim</i>	8	Elle croit que	Carl	va	détester	Paris
	3	P06	8	Elle croit que	Carl	va	détester	à Paris	L>L

Les éléments produits par les apprenant/es ont été mis le plus possible sous le stimulus tout en gardant l'ordre des mots produits. L'orthographe du stimulus a été gardée même si cela posait parfois problème comme pour les phrases 13 et 24 au niveau de la transcription de *travail* /*travaille* (exemple 12-2). Étant donné que le verbe *travailler* n'est pas un verbe de test, le choix effectué aura uniquement un effet sur le calcul du nombre de phrases sans verbes dans le cas

où *travail* a été considéré comme un substantif (six fois avec la phrase n° 24, une fois avec la phrase n° 13 et une fois avec la n° 16).

(12-2)	Stim 13	Au travail	Carl	parle	avec une collègue	
2	P24 13		Carl	travaille		C>øtr
	Stim 24	J'espère que	Christine	va trouver	un travail	
2	P20 24		Christine	-	travail	C>øtr

La colonne *verbe* ne contient que des verbes reconnaissables en français. Les verbes d'une autre langue, les verbes incompréhensibles ou les verbes ne rappelant aucune forme en français ont été mis dans une autre colonne. Seules les formes de la colonne *verbe* entrent dans les calculs. Si plusieurs formes verbales ont été produites, elles ont été mises dans une autre colonne. En principe la colonne *verbe* contient la première forme verbale produite, sauf si une des formes du verbe stimulus a été produite plus tard dans la phrase comme dans les trois cas ci-dessous (exemple 12-3) :

(12-3)	Part	n°	Début	Sujet	mod	verbe	fin	
	Stim 7		Il croit que	Marie		déteste	la musique	
2	P04 7		Il trouve il	Marie		déteste	la musique	C>C
	Stim 11		A midi	Christine		mange	au restaurant	
1	P28 11			Marié	aime	mange	à la restaurant	C>C
	Stim 20		Samedi	Christine	va	regarder	un film	
3	P26 20		A samedi Carina achète	Carina		regardE	une film	L>L

Lorsque le /r/ de l'infinitif a été prononcé, il a été transcrit par un *-R* (*mangeR* a été prononcé *mangéRe*). Les formes longues produites après un auxiliaire ont été transcrites avec un *-er* ou un *-é* final, alors que les autres ont été transcrites avec un *-E*. Les formes du verbe *regarder* prononcées *regardeu* ont été transcrites *regarde*. Les séries de syllabes incompréhensibles ont été transcrites par des xxx.

Finalement, certain/es participant/es (par exemple P24) ont produits de nombreux sons *hé*, ce qui ne facilite pas toujours la transcription. Dans le même ordre de difficulté, toutes les phrases avec les prénoms *Christine* et *Marie* suivi du mot *a* ont été écoutées soigneusement afin de distinguer les *Christina* (0 occurrences), *Maria* (2 occurrences) des *Christine a* (7 occurrences) ou *Marie a* (3 occurrences).

Dans trois cas, une autre forme que la courte ou la longue a été produite. Cette forme a soit été comptée comme autre verbe soit elle a été ignorée parce qu'une autre forme à compter a été produite dans la même phrase.

Les *va* et autres potentiels auxiliaires ont été mis dans la colonne *verbe* s'il n'y avait pas d'autre verbe, même si *va* ou *veut* était dans le stimulus, et codés comme autre verbe⁵⁷.

12.3.2. Codage des phrases

Les formes verbales dans la colonne *verbe* ont été codées en comparant le stimulus et l'output du même verbe indépendamment du contexte syntaxique. Ce codage permet de voir le nombre de fois où une forme a été répétée, remplacée par l'autre forme du même verbe ou évitée par la production d'un autre verbe ou rien. Le codage est donc centré sur les *formes* et non sur les contextes dans lesquels ces formes ont été produites. Les deux hypothèses pour ce travail prédisent en effet une production des formes indépendamment du contexte. Le tableau 12-2 présente le codage des formes de la colonne *verbe*.

Tableau 12-2 : Codage des formes verbales dans le test d'imitation

stimulus	output	code
forme courte	forme courte	C>C
forme courte	forme longue	C>L
forme courte	autre verbe ou rien	C>øtr
forme longue	forme longue	L>L
forme longue	forme courte	L>C
forme longue	autre verbe ou rien	L>øtr

Ces codes sont utilisés dans la présentation des exemples et dans les tableaux en annexe : l'annexe 5 sur les données par participant/e et l'annexe 6 sur les données par verbe.

12.4. Résultats du test d'imitation

Les résultats seront présentés dans quatre sections différentes : la première section sera consacrée aux outputs produits sans verbes, dans le but d'explorer la validité du test et de la répartition des apprenant/es dans les stades 1 à 3 de Bartning & Schlyter (2004). Les deux sections suivantes serviront à la vérification des deux hypothèses, tout d'abord celle sur l'aspect lexical et ensuite l'hypothèse 2 sur la fréquence des formes. La dernière section présentera une analyse qualitative de quelques exemples.

⁵⁷ 13 fois *va*, 3 fois *vais*, 1 fois *veut*, 3 fois *a*, 7 fois *est*.

Au total 889 phrases stimulus ont été traitées par les participant/es. Deux participant/es, P05 et P33 du stade 1, ont raté le son d'un stimulus chacun. En moyenne, ce test, y compris les exercices, a duré 10'52 minutes (entre 7'56 – P05 et 15'08 – P15 ; durée calculée du premier au dernier *enter*). Sur les 889 stimuli traités, 310 outputs (35%) ont été produits sans verbe dont 52 sans aucun mot. La distribution de ces outputs sans verbes entre les trois stades sera présentée à la section suivante. On trouve ensuite 143 phrases d'outputs (16%) avec un autre verbe que le verbe du stimulus, 54 avec un autre verbe de test (par exemple *adore* pour *aime*) et 89 avec des verbes comme *travailler*, *faire*, *prendre*, *être* etc. Seules les 435 phrases avec le même verbe dans le stimulus et l'output seront analysées dans ce travail. Ces phrases correspondent à 49% des phrases stimulus traitées par les apprenant/es.

L'exemple (12-4) présente les différentes sortes d'outputs mentionnés avec l'exemple du stimulus n° 22 : un output sans mots (P25), un output sans verbe (P33), un output avec un autre verbe (P18), un output avec un autre verbe mais de test (P21) et finalement deux outputs avec le même verbe que le verbe du stimulus (P26 et P19).

(12-4)	Stim	22	A Nice	Carl	va	trouver	que la mer est belle	
1	P25	22	-	-	-	-	-	L>øtr
1	P33	22	En	-	-	-	-	L>øtr
3	P18	22		Carl	écrit	Carl écrit la mer (...)		L>øtr
2	P21	22	A Nice	Carl	visitE	la mer		L>øtr
3	P26	22		Carl	trouve	qui elle est belle		L>C
2	P19	22		Carl	va	trouver	xx la mer est belle	L>L

Les natifs ont mis 8'26 minutes en moyenne pour faire le test et ils ont imité toutes les formes verbales correctement⁵⁸.

12.4.1. Outputs sans verbes

Comme on l'a vu dans la présentation du test à la section 12.1., un des avantages du test d'imitation est que l'on peut également analyser le matériel qui n'a pas été imité par les apprenant/es. Ainsi, cette section sera consacrée aux 52 outputs sans mots ainsi qu'aux 310 outputs sans verbe. Leur analyse d'un point de vue quantitatif nous informera sur la validité du test utilisé ici ainsi que sur la

⁵⁸ Les 20 changements effectués par les natifs concernent quelques débuts et fin de phrases (*marché* pour *supermarché*, *de belles tomates* pour *des belles tomates*) quelques utilisations de *va* pour *veut* et quelques changements de prénom.

répartition des apprenant/es en trois stades (cf. section 11.3.). Le test d'imitation impliquant un traitement complet de la phrase stimulus par les participant/es (*processus restructif*), on s'attend à ce que les apprenant/es plus avancés, en l'occurrence les apprenant/es du stade 3, rappellent plus de matériel que les apprenant/es moins avancés, du stade 1. Comme on l'a vu, l'expérience de la langue cible permet aux apprenant/es avancé/es de reconnaître le vocabulaire et de s'aider de la grammaire pour regrouper les mots par le processus de *chunking* (Naiman, 1974 ; Markman *et al.*, 1975 ; Bley-Vroman & Chaudron, 1994).

Le tableau 12-3 présente le nombre d'outputs sans verbes (y compris les outputs sans mots) et les outputs sans aucun mot pour les trois stades et pour les 27 phrases stimulus. Le pourcentage entre parenthèses a été calculé en divisant le nombre d'occurrences d'outputs sans verbe, respectivement sans mots, par le nombre de phrases stimulus traitées pour chaque stade.

Tableau 12-3 : *Outputs sans verbes et sans mots par stade*

	Σstimuli traités (n=27)	outputs sans verbe (y.c. sans mots)	outputs sans mots
stade 1 (9 part.)	241 ⁵⁹	143 (59%)	23 (10%)
stade 2 (14 part.)	378	135 (36%)	22 (6%)
stade 3 (10 part.)	270	32 (12%)	7 (3%)
Σapprenant/es	889	310	52

Le nombre d'outputs sans mots et les outputs sans verbes (y compris les outputs sans mots) diminuent en occurrences et surtout en pourcentage au fur et à mesure des stades. Le stade est corrélé au nombre d'outputs sans verbes (*r* de Spearman à .714) et aux outputs sans mots (*r* de Spearman à .416). La différence entre les stades 1 et 2 est plus faible qu'entre les stades 2 et 3, mais elle est significative au niveau des outputs sans verbes comme le montrent les résultats du test statistique dans le tableau 12-4⁶⁰.

Tableau 12-4 : *Test statistique pour les outputs sans verbes par stade*

	stades 1-2	stades 2-3	stades 1-3
Mann-Whitney U	27,500	18,500	4,000
Wilcoxon W	132,500	73,500	59,000
Z	-2,246	-3,034	-3,374
Asymp. Sig. (2-tailed)	,025	,002	,001

⁵⁹ Deux participant/es ont raté le son d'un stimulus.

⁶⁰ Par contre les différences au niveau des outputs sans *mots* sont faibles, on ne trouve qu'une faible différence entre les stades 1 et 3 ($p < .05$).

Ces résultats indiquent que les différences entre les stades dans ce test sont différentes de celles trouvées à partir des données de production libre, où la différence était plus forte entre les stades 1 et 2 que 2 et 3 (tableau 11-2).

Ces résultats valident le test tel qu'utilisé ici ainsi que la répartition en stades : le nombre de verbes rappelés augmente au fur et à mesure des stades, ce qui signifie aussi que les apprenant/es ont de plus en plus d'expérience de la langue au fur et à mesure des stades.

12.4.2. Vérification de l'hypothèse 1 sur l'influence de l'aspect lexical

Selon l'hypothèse 1, les apprenant/es devraient produire la forme courte avec les verbes statiques et la forme longue avec les verbes dynamiques. Le tableau 12-5 ci-dessous présente les résultats du test d'imitation. Les verbes ont été regroupés selon l'aspect lexical (*cf.* section 7.2.2.), statique dans le haut du tableau et dynamique dans le bas. Les données des deux verbes *trouver* ont été regroupées d'après le sens du stimulus. Les occurrences ont été calculées d'après le nombre de formes produites pour le verbe en question, indépendamment de la forme attendue par le contexte de la phrase stimulus ou de l'output. L'analyse est donc centrée sur les *formes* produites. Les occurrences correspondent à la production codée C>C, C>L, L>L et L>C (*cf.* annexe 6). Pour les apprenant/es le nombre de stimuli par verbe est de 66 au total sauf pour les verbes *habiter* et *trouver statique* qui n'en ont que 65 en raison de deux sons ratés. Le tableau 12-5 présente aussi les données des natifs, qui ont donc produit tous les verbes comme dans les stimuli. Les données de la phrase avec le verbe *trouver* au participe passé n'entrent pas dans les calculs. L'annexe 6 présente les résultats détaillés au niveau du rapport entre stimulus et output pour chaque verbe ainsi que les données de la phrase avec *trouvé* (et l'annexe 5 pour chaque participant/e)

Tableau 12-5 : Production des verbes selon l'aspect lexical – test d'imitation

	stade 1			stade 2			stade 3			Σ apprenant/es			natifs		
	C	L	%C	C	L	%C	C	L	%C	C	L	%C	C	L	%C
adorer	1	0	100%	3	1	75%	5	1	83%	9	2	82%	9	9	50%
aimer	4	0	100%	8	1	89%	10	0	100%	22	1	96%	9	9	50%
détester	4	2	67%	10	4	71%	11	5	69%	25	11	69%	9	9	50%
habiter	7	0	100%	15	3	83%	11	3	79%	33	6	85%	9	9	50%
penser	0	1	0%	4	0	100%	9	2	82%	13	3	81%	9	9	50%
préférer	9	4	69%	14	6	70%	10	7	59%	33	17	66%	9	9	50%
trouver S	0	1	0%	3	4	43%	11	1	92%	14	6	70%	9	9	50%
<i>statique</i>	25	8	76%	57	19	75%	67	19	78%	149	46	76%	63	63	50%
acheter	0	6	0%	4	11	27%	4	10	29%	8	27	23%	9	9	50%
manger	8	0	100%	6	9	40%	11	3	79%	25	12	68%	9	9	50%
parler	2	7	22%	14	3	82%	9	7	56%	25	17	60%	9	9	50%
regarder	8	4	67%	17	5	77%	11	5	69%	36	14	72%	9	9	50%
trouver D	1	0	100%	8	1	89%	5	5	50%	14	6	70%	9	9	50%
visiter	3	4	43%	10	12	45%	5	13	28%	18	29	38%	9	9	50%
<i>dynamique</i>	22	21	51%	59	41	59%	45	43	51%	126	105	55%	54	54	50%

Légende : C : forme courte ; L : forme longue, Σ : somme ; trouver S : sens statique ; D : sens dynamique.

On peut tout d'abord constater que les apprenant/es n'ont pas toujours produit le verbe à la même forme que dans le stimulus, contrairement aux natifs.

On peut voir que les verbes statiques ont surtout été produits à la forme courte à quelques exceptions près, comme le verbe *penser* au stade 1 (1 seule occurrence) et le verbe *trouver statique* qui a été préféré à la forme longue aux stades 1 et 2. Les verbes statiques ont donc été produits comme attendu par l'hypothèse 1 à tous les stades.

Pour les verbes dynamiques, on peut constater que le pourcentage de formes courtes l'emporte, mais le pourcentage de formes longues est beaucoup plus élevé que pour les verbes statiques à tous les stades, même si la différence est plus faible au stade 2. Les verbes *acheter* et *visiter* ont été préférés à la forme longue à tous les stades. Le verbe *parler* a également été préféré à la forme longue au stade 1 et le verbe *manger* au stade 2. Les verbes dynamiques n'ont donc pas tous été préférés à la forme longue contrairement à ce qui était attendu par l'hypothèse 1.

Ces résultats ressemblent globalement à ce que l'on avait pu voir dans la tâche de production libre (chapitre 11) mais avec des différences plus faibles entre les deux groupes de verbes. Le fait de présenter les verbes statiques dans les deux contextes syntaxiques a en effet augmenté le pourcentage de formes longues pour ces verbes.

La différence du pourcentage de formes courtes entre les verbes statiques et dynamiques n'est significative qu'au stade 3 (Mann-Whitney U, $Z = -2,435$, $p = .015$) et au total (Mann-Whitney U, $Z = -2,074$, $p = .038$). Ce résultat rejoint aussi ce que l'on avait pu constater dans l'analyse des données de la production libre (chapitre 11) et implique que l'influence de l'aspect lexical ne se ferait qu'après une certaine expérience de la langue cible, ici chez les apprenant/es du stade le plus avancé.

12.4.3. Vérification de l'hypothèse 2 sur la fréquence des formes

Selon l'hypothèse 2, les apprenant/es produisent la forme la plus fréquente du verbe dans l'input. Le tableau 12-6 présente une comparaison des échelles de pourcentage de formes courtes entre les apprenant/es selon les stades et l'input. Les données de l'input proviennent du tableau 8-2 et les lignes délimitent les regroupements de fréquence des formes (cf. tableau 8-7). Les verbes ont à chaque fois été classés par ordre décroissant de formes courtes. Les données du verbe *trouver* correspondent au total des deux sens. Les données indiquées sous Σ correspondent à la somme des formes courtes et longues produites pour le verbe concerné (cf. tableau 12-5 ci-dessus).

Tableau 12-6 : Pourcentages de formes courtes – test d'imitation et input

stade 1	Σ	%C	stade 2	Σ	%C	stade 3	Σ	%C	input	Σ	%C
adorer	1	100%	penser	4	100%	aimer	10	100%	adorer	32	81%
aimer	4	100%	aimer	9	89%	adorer	6	83%	penser	368	73%
habiter	7	100%	habiter	18	83%	penser	11	82%	habiter	150	72%
manger	8	100%	parler	17	82%	habiter	14	79%	détester	10	70%
préférer	13	69%	regarder	22	77%	manger	14	79%	préférer	45	58%
détester	6	67%	adorer	4	75%	trouver	22	73%	aimer	341	55%
regarder	12	67%	détester	14	71%	détester	16	69%	trouver	393	47%
trouver	2	50%	préférer	20	70%	regarder	16	69%	regarder	223	39%
visiter	7	43%	trouver	16	69%	préférer	17	59%	parler	581	38%
parler	9	22%	visiter	22	45%	parler	16	56%	manger	160	29%
penser	1	0%	manger	15	40%	acheter	14	29%	acheter	110	15%
acheter	6	0%	acheter	15	27%	visiter	18	28%	visiter	25	8%

Légende : C : forme courte ; Σ : somme des formes courtes et longues.

Le tableau 12-6 montre que les échelles de fréquence se ressemblent beaucoup entre les apprenant/es et l'input. Le continuum du stade 1 diffère sur quatre verbes, le verbe *manger* est clairement préféré à la forme courte et le verbe *détester* l'est un peu moins que dans l'input. Le verbe *visiter* a plutôt été produit aux deux formes et le verbe *penser* à la forme longue. Au stade 2 on trouve surtout des verbes à la forme courte sauf les verbes *visiter* et *acheter*. Les verbes

parler et *regarder* sont un peu plus fréquents à la forme courte que prévu, mais le verbe *trouver* est au milieu. Finalement, au stade 3, comme chez les apprenant/es au total, les verbes *détester* et *préférer* sont plus bas dans le continuum que dans l'input.

On peut donc voir un certain nombre de différences, qui semblent être plus fortes dans les premiers stades qu'au stade 3, mais les données semblent indiquer que les apprenant/es produisent plutôt la forme la plus fréquente dans l'input comme prévu par l'hypothèse 2.

Un calcul de corrélation avec l'input montre cette progression avec un r de Spearman de .063 pour le stade 1⁶¹, .344 au stade 2 et finalement .760 au stade 3 et .679 au total. La figure 12-1 présente l'absence de corrélation entre l'input et les données du stade 1. Le placement des verbes se fait horizontalement selon le pourcentage de formes courtes dans l'input et verticalement selon le pourcentage de formes courtes du stade 1 du test d'imitation. La ligne illustre une corrélation parfaite, c'est-à-dire lorsque le pourcentage de formes courtes est le même pour les deux sources de données. Ainsi, plus les verbes sont proches de la ligne de corrélation plus ils sont corrélés et plus ils sont éloignés, moins ils sont corrélés au niveau du pourcentage de formes courtes dans les deux sources de données.

⁶¹ La corrélation avec *trouver* au total et les sans deux sens monte à .088, ce qui ne fait pas de différence.

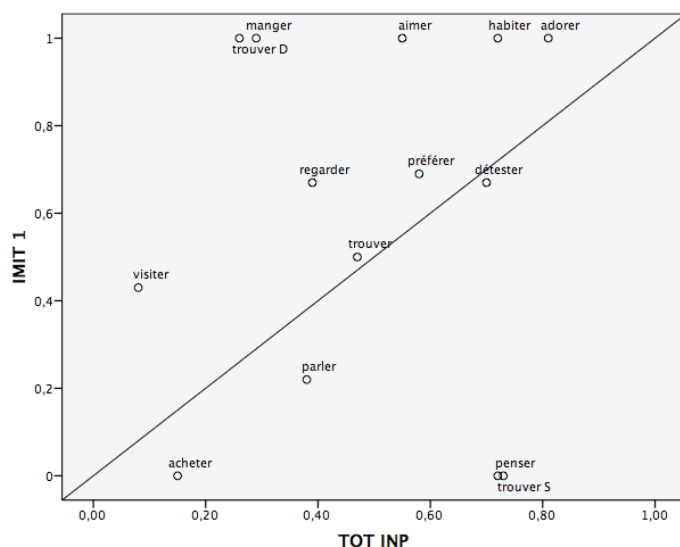


Figure 12-1 : Test d'imitation – corrélation entre le stade 1 et l'input

On constate que les verbes *détester*, *préférer*, *trouver* (au total) sont les mieux corrélés, alors que les verbes *manger* et *trouver* (dynamique), 100% à la forme courte, puis *penser* et *trouver* (statique), 0% à la forme courte, ont été produits de manière complètement différente par les apprenant/es du stade 1 que dans l'input. Les verbes *penser* et *trouver* sont les verbes les moins bien connus par les apprenant/es et il n'y a qu'une seule occurrence de chaque (cf. annexe 6). Sans les données de ces deux verbes, on arrive à une corrélation de .639, ce qui signifie que l'on observe une influence de l'input au niveau des verbes connus par les apprenant/es du stade 1, donc chez les apprenant/es les plus débutants.

On peut en outre noter que le rôle de la syntaxe des stimuli joue un rôle secondaire au niveau de la difficulté d'imitation. En effet, on s'attendait à un rappel plus faible des phrases des verbes *adorer*, *détester* et *trouver*, ces verbes étant présentés en subordonnée et avec un vocabulaire moins connu des apprenant/es. Les données de pourcentage d'outputs pour ces verbes au total (%outp dans l'annexe 6) montrent que le verbe *adorer* a été le plus faiblement rappelé, plus faiblement que le verbe *penser*, qui est le verbe le moins connu des apprenant/es. Pour le verbe *détester* le pourcentage monte à 55% de formes rappelées avec une forme du verbe et pour les verbes *trouver* on arrive à un tiers. La connaissance et la reconnaissance des verbes semblent néanmoins avoir

une plus forte influence que la présentation du verbe en subordonnée et introduite par un verbe peu connu.

12.4.4. Analyse qualitative de quelques exemples

Dans le but d'illustrer plus concrètement la vérification des deux hypothèses ainsi que les différences entre les apprenant/es des trois stades, les pages qui suivent seront consacrées à la présentation de quelques exemples de traitement de phrases de stimulus au présent et à l'infinitif pour le même verbe. Ces exemples permettront de discuter les deux hypothèses de manière plus qualitative. Pour ces exemples, j'ai choisi le verbe *habiter*, un verbe statique et fréquent à la forme courte dans l'input et le verbe *acheter*, un verbe dynamique fréquent à la forme longue dans l'input, ainsi que le verbe *parler*, un verbe dynamique dont la fréquence des formes courtes et longues est proche dans l'input. Pour le verbe *habiter* on s'attend à la production de la forme courte et pour le verbe *acheter* la forme longue pour les deux hypothèses. Pour le verbe *parler* on s'attend à la forme longue selon l'hypothèse 1, puisqu'il s'agit d'un verbe dynamique et à une variation des formes selon l'hypothèse 2 sur la fréquence des formes.

L'exemple (12-5) présente les 21 outputs pour le verbe *habiter* à la forme courte dans le stimulus.

(12-5)	Stim	9	En France	Carl	habite	au bord de la mer	
1	P15	9		Carl	habite	près du mer	C>C
1	P28	9	En France	Carl	habite	de la mer	C>C
1	P30	9	En France	Carl	habite	au nãgonting la mer	C>C
1	P31	9	A la France	quatre mille	habite	dans la mer la mer	C>C
2	P01	9	En a France	Catherine	habite	de la mer	C>C
2	P04	9	En va France	Carl	habite	dans le bord de la mer	C>C
2	P07	9	En France		habite	au la mer	C>C
2	P12	9	En France	Carl	habite	au bord de la mer	C>C
2	P13	9	D'abord	Carl	habite	avec sa mer	C>C
2	P19	9	En France	il	habite	au bord de la mer	C>C
2	P20	9	En France	Carl	habite	en le mer	C>C
2	P24	9	A la France	Carl	habite	près de la mer	C>C
2	P16	9	En France	Carl	habitE	la mer	C>L

3	P02	9	En France	Carl	habite	de la nâgonting mer	C>C
3	P06	9	En	Carl	habite	bord de la mer	C>C
3	P08	9	En Nice	Carl	habite	au bord de la mer	C>C
3	P11	9	En France	Carl	habite	d'abord la mer	C>C
3	P14	9	En France	Paul	habite	à bord de la mer	C>C
3	P18	9	En France	Carl Carl	habite	à près (...)	C>C
3	P22	9	En à Nice	Carl	habite	dans le bord de la mer	C>C
3	P26	9	En France	Carl	habite	à bord de la mer	C>C

On constate que tous les participant/es sauf une (P16, stade 2) ont produit la forme courte du présent comme dans la phrase stimulus, ce qui correspond à ce qui est attendu pour un verbe statique et un verbe fréquent à la forme courte dans l'input.

L'exemple (12-6) présente les 18 outputs pour le verbe *habiter* à la forme longue dans le stimulus.

(12-6)	Stim	10	En mars	Marie	veut	habiter	à Paris	
1	P10	10				habite	à Paris	L>C
1	P15	10	En mars	nâgon		habite	de Paris	L>C
1	P33	10	En mars			habite	à Paris	L>C
2	P03	10	En verdi	Marie		habite	à Paris	L>C
2	P07	10	En mars	Carine		habite	à Paris	L>C
2	P12	10	En mars	elle		habite	en Paris	L>C
2	P21	10	En mars	j'		habite	à Paris l'habite à Paris	L>C
2	P23	10	En mars	Marie		habite	à Paris	L>C
2	P24	10	A mars	Carl		habite	à Paris	L>C
2	P19	10	En mars	Marie	vais	habite	à Paris	L>C
2	P01	10	En mars	Marie	habi	habitE	en Paris	L>L
2	P16	10	En mars	Marie		habitE	à Paris	L>L
3	P32	10	Marie en mars	Marie		habite	en Paris	L>C
3	P06	10	En mars	Carine	va	habite	à Paris	L>C
3	P14	10	En mars	Marie	veut	habite	à Paris	L>C
3	P08	10	En mars	Carl		habitE	a visitE à Paris	L>L
3	P11	10	En mars	Christine	a	habité	à la Paris	L>L
3	P22	10	En mars	Marie	va	habiter	au à Paris	L>L

Les participant/es du stade 1, ainsi qu'un grand nombre de participant/es du stade 2, n'ont pas reproduit la forme longue du stimulus, mais la forme courte, la forme la plus fréquente dans l'input, la forme attendue pour un verbe statique. Ces participant/es ont en outre transformé la phrase stimulus à l'infinitif en phrase au présent.

Cinq participant/es, dont quatre au stade 3, ont réussi à reproduire le contexte syntaxique du stimulus, mais seulement deux ont réussi à produire une forme longue après l'auxiliaire. Il s'agit de P11 dont la stratégie de production du passé a déjà été notée en production libre et de P22 qui a un bon niveau de français et probablement une bonne mémoire (elle a en effet rappelé 90% des formes verbales comme dans le stimulus). La production de la forme courte après l'auxiliaire par les trois autres apprenant/es du stade 3 correspond à ce qui est attendu par les deux hypothèses.

Finalement, on trouve trois occurrences de la forme stimulus produite dans une phrase sans auxiliaire. Pour P01 (stade 2) et P08 (stade 3) on peut penser qu'ils ont pu retenir la forme longue du verbe, mais pas l'auxiliaire car ces deux apprenants ont produit la forme *habite* après le stimulus à la forme courte (exemple 12-5). Par contre P16 a produit la forme longue dans les deux cas, ce qui donne à penser qu'elle ne s'appuie pas sur l'aspect lexical du verbe *habiter*, mais que la forme longue est la plus accessible pour elle.

On peut néanmoins résumer en disant que le verbe *habiter* a été produit conformément aux deux hypothèses.

L'exemple (12-7) présente les 21 outputs pour le verbe *acheter* à la forme longue dans le stimulus.

(12-7)	Stim	2	Aujourd'hui	Christine	veut	acheter	une robe	
1	P05	2				achetE	un robe	L>L
1	P25	2	Aujourd'hui			achetE	une robe	L>L
1	P33	2		On		achetE	une robe	L>L
1	P10	2		Christi		achetE	un robe	L>L
2	P09	2	Aujourd'hui	Christine		achète	un robe	L>C
2	P01	2	Aujourd'hui	Marie		achetE	de la robe	L>L
2	P04	2	Aujourd'hui	Christine		achetE	une robe	L>L
2	P12	2		Christine		achetE	une robe	L>L
2	P13	2		Caroline		achetE	une robe	L>L
2	P16	2	Aujourd'hui	Marie		achetE	une robe	L>L
2	P21	2	Aujourd'hui	Christi		achetE	une robe	L>L
2	P19	2	Aujourd'hui	Christine	va	acheter	une robe	L>L

3	P18	2	Aujourd'hui	Catherine (...)	acheteR	une robe	L>L
3	P14	2	En demi	Christine	veut achète	un une robe	L>C
3	P02	2	Aujourd'hui	Christine	va acheter	une robe	L>L
3	P06	2	En lundi	Christine	va acheteR	une robe	L>L
3	P08	2		Catherine	va acheter	une robe	L>L
3	P11	2	Aujourd'hui	Christine	a acheté	une robe	L>L
3	P17	2	Aujourd'hui	Christine	veut acheter	une robe	L>L
3	P22	2	A hm	Christi	va acheter	un un robe	L>L
3	P26	2	Aujourd'hui	Christine	va acheter	une robe	L>L

On peut voir que tous les apprenant/es, sauf deux, ont reproduit la forme longue du stimulus, ce qui est attendu pour un verbe dynamique fréquent à la forme longue dans l'input. P09 (stade 2) a changé le stimulus en phrase au présent et P14 (stade 3) a produit la forme courte après un auxiliaire ce qui est très inattendu. On peut voir que les apprenant/es du stade 1 ont systématiquement produit la forme longue et que les apprenant/es des stades 1 et 2 produisent cette forme dans une construction sans auxiliaire. Ces données confirment l'absence de saillance de l'auxiliaire chez les apprenant/e/es les plus débutants (*cf.* section 5.2.) et l'attention sur le contenu de <acheter une robe>. Ce n'est qu'au stade 3 que les apprenant/es produisent la forme longue dans une construction pour une telle forme. On peut également noter que le nombre d'outputs pour la forme longue est plus grand que pour le verbe *habiter* présenté à la forme longue, de même que le nombre de phrases produites à l'infinitif est plus élevé.

L'exemple (12-8) présente les 14 outputs pour le verbe *acheter* à la forme courte dans le stimulus.

(12-8)	Stim	1	Au supermarché	Carl	achète	un melon	
1	P10	1	En supermarché		achetE	un	C>L
1	P15	1	A supermarché	Carl	achetE		C>L
2	P07	1	Au		achetE	nej marché supermarché	C>L
2	P16	1	A supermarché	Carl	achetE	un melon	C>L
2	P20	1	Au supermarché	Carl	achetE		C>L
2	P21	1	En supermarché	Carl	achetE		C>L
2	P12	1	En supermarché	Carl	achète	une vélo	C>C
2	P13	1	En supermarché	Carl	achète	une melon	C>C
2	P19	1	Au supermarché	Carl	achète	une melon	C>C
3	P02	1	Au marché	Carl	achetE	un melon	C>L
3	P11	1	A supermarché	Carl	a acheté	une melon	C>L
3	P08	1	A supermarché	Carl	achète	un melon	C>C
3	P22	1	Au marché	Carl	achète	un	C>C
3	P26	1		Carl	achète	une melon	C>C

Sur les 14 outputs, 8 ont été produits à la forme longue (57%), la forme attendue pour un verbe dynamique fréquent à la forme longue, tout en gardant la structure syntaxique du présent. La participante P11 est la seule à avoir transformé la phrase stimulus au présent en phrase au passé, faisant ainsi correspondre forme et structure syntaxique.

On peut noter que le nombre d'outputs est assez faible par rapport aux autres exemples, ce qui donne à penser que le verbe *acheter* n'a pas souvent été reconnu à la forme courte, la forme la moins fréquente.

Le pourcentage de formes courtes dans ces outputs s'élève à 43%, ce qui est plus élevé que les 28% de formes longues du verbe *habiter* présenté à la forme la moins fréquente. Un tel résultat pourrait s'expliquer par le fait que la construction syntaxique du présent est bien connue et activement utilisée par les participant/es. La fréquence de la construction pourrait ainsi aider à la mémorisation temporaire de la forme courte.

En résumé, le verbe *acheter* a été produit comme attendu par les deux hypothèses.

Les verbes *habiter* et *acheter* ont donc été produits comme attendus par les hypothèses. On a pu noter que le nombre d'apprenant/es qui pouvaient imiter une phrase à l'infinitif était plus élevé avec le verbe *acheter* que *habiter*. Cela signifie que la fréquence de la construction dans laquelle le verbe est le plus souvent présenté peut aussi jouer un rôle. Un tel effet de fréquence correspond à ce que Bybee (2008) appelle la fréquence type, la fréquence des *patterns* comme on l'a vu à la section 5.1.3.

L'exemple (12-9) présente les 23 outputs pour le verbe *parler* à la forme la longue dans le stimulus.

(12-9)	Stim	14	Maintenant	Marie	veut parler	avec son chef	
1	P10	14			parlE	avec un Français	L>L
1	P31	14			parlE	en su chef	L>L
1	P05	14	Maintenant	Marie	parlE	avec sa chef	L>L
1	P28	14	Marlé	Marie	parlE	français	L>L
1	P30	14	Nâgonting	Marie	parlE	au chef, maintenant...	L>L

2	P03	14	Marie	parle	avec sa chef	L>C
2	P20	14	Marie	parle	avec son chef	L>C
2	P24	14	Catherine	parle	avec elle chef	L>C
2	P12	14	Maintenant elle	parle	avec son chef	L>C
2	P04	14	Maintenant Marie	parle	de la de la chef	L>C
2	P13	14	Maintenant Marie	parle	avec sa chef	L>C
2	P19	14	Maintenant Marie	parle	avec son chef	L>C
2	P09	14	Un tnant Catr	parlE	avec lui chef	L>L
2	P16	14	Maintenant Carl	parlE	pour son chef	L>L
2	P21	14	En main temps Cari	parlE	au chef	L>L
3	P06	14	Maintenant Carine	parleR	avec su chef sa chef	L>L
3	P02	14	Maintenant Marie	a parle	avec son chef	L>C
3	P14	14	Maintenant Marie	veut parle	avec son chef	L>C
3	P26	14	Maintenant Marie	va parle	avec son chef	L>C
3	P08	14	Maintenant Carine	va parler	avec son chef	L>L
3	P11	14	Maintenant Marie	veut parler	avec son chef	L>L
3	P17	14	En maintenant Marie	veut parler	avec sa son chef	L>L
3	P22	14	Maintenant Marie	va parler	avec son chef	L>L

Les outputs pour le verbe *parler* présentent ici différentes facettes. Les participant/es du stade 1 ont tous produit le verbe à la forme longue comme attendu par l'hypothèse sur l'influence de l'aspect lexical. La production du verbe au stade 2 ne montre par contre pas d'influence de l'aspect lexical : bien que le stimulus soit à la forme longue, la plupart des apprenant/es de ce stade ont produit le verbe à la forme courte. Ce résultat parle pour une influence de la fréquence des formes, même si le pourcentage de formes courtes est plus élevé que dans l'input au total. Il correspond par contre au pourcentage de formes courtes dans *Escalade 1*, où le verbe *parler* est à la forme courte à 80% (cf. tableau 8-3).

Les participant/es du stade 3 réussissent pour la plupart à imiter la structure syntaxique du stimulus (sauf P06), or trois apprenant/es sur sept utilisent la forme courte. Le fait que ces apprenant/es du stade 3 ne produisent pas tous la forme longue parle également pour une influence de la fréquence des formes.

Finalement l'exemple (12-10) présente les 19 outputs pour le verbe *parler* à la forme courte dans le stimulus.

(12-10)	Stim	13	Au travail	Carl	parle	avec une collègue	
1	P31	13			parlE	dans une copain	C>L
1	P28	13	Au travail	Carl	parlE		C>L
1	P25	13	Au travail	Carl	parle	avec une collègue	C>C
1	P30	13	Au travail	hmhm	parle	nâgonting	C>C

2	P03	13			parle	avec une collègue	C>C
2	P04	13	A travail	Carl	parle	à une collègue	C>C
2	P07	13	Au travail	nâgon	parle		C>C
2	P12	13	Au travail	Carl	parle	avec une collègue	C>C
2	P13	13		Carl	parle	avec	C>C
2	P19	13	Au travail	Carl	parle	avec une collègue	C>C
2	P29	13	Au travail	Carl	parle	un collègue	C>C
3	P11	13	Au travail	Carl	parlE	avec un collègue	C>L
3	P17	13	Au travail	Carl	parlE	à une femme femme	C>L
3	P02	13	Au travail	Carl	parle	avec un collègue	C>C
3	P06	13	En travail	Carl	parle	avec une	C>C
3	P08	13	Au travail	Carl	parle	avec une collègue	C>C
3	P14	13	Au travail	Carl	parle	avec un collègue	C>C
3	P22	13	Au travail		parle	Paul parle avec un collègue	C>C
3	P26	13	Au travail	Carl	parle	au une collègue	C>C

Bien que le verbe *parler* soit un verbe dynamique et soit un peu plus fréquent à la forme longue, le nombre d'outputs à la forme longue ne sont que de 21% ici, ce qui est loin du résultat du verbe *acheter* avec le stimulus à la forme courte (57%). Ce résultat confirme celui des outputs avec le stimulus à la forme longue pour le verbe *parler* : la forme courte semble être la plus facilement produite par les apprenant/es, ce qui est plutôt le fruit de la fréquence des formes que de l'influence de l'aspect lexical.

Cette présentation d'exemples d'outputs nous a donc permis de voir que les verbes *habiter* et *acheter* étaient produits de préférence à la forme correspondant à l'aspect lexical et à la forme la plus fréquente dans l'input, indépendamment de la forme présentée en stimulus. On a aussi pu constater que les verbes sont produits de manière systématique à une forme indépendamment du contexte syntaxique créé par les apprenant/es. Ces résultats confirment le lien étroit entre les deux hypothèses lorsque la forme de l'aspect lexical correspond à la forme la plus fréquente dans l'input. Par contre, l'analyse des outputs du verbe *parler* a montré une plus grande variation des formes. Les apprenant/es du stade 1 ont préféré la forme longue, ce qui pourrait donner l'impression que leur production suit l'aspect lexical. Mais contrairement aux verbes *habiter* et *acheter*, on trouve les deux formes dans l'output après la forme courte dans le stimulus et l'apprenante P30 a même produit le verbe aux deux formes. Les apprenant/es des stades 2 et 3 ont aussi produit les deux formes, mais avec une préférence pour la forme courte. La plupart des apprenant/es du stade 3 ont su imiter la structure de l'infinitif, mais plusieurs l'ont produite avec une forme courte, ce qui n'est pas attendu par l'hypothèse sur l'influence de l'aspect lexical. Les

résultats pour ce verbe parlent donc plutôt en faveur d'une influence de la fréquence des formes dans l'input.

12.5. Résumé des résultats du test d'imitation

Le test d'imitation a permis d'observer la production des verbes par les apprenant/es dans un nombre limité de phrases données en stimulus, les verbes étant présentés le même nombre de fois.

Sur les 889 phrases stimulus traitées par les apprenant/es, 435 ont été retenues pour l'analyse, à savoir celles qui comportaient le verbe stimulus. Un calcul des outputs sans verbe a montré que le nombre de ces outputs diminuait avec l'augmentation des stades. Ce résultat a été interprété comme une forme de validation du test d'imitation et de la répartition en stades.

L'analyse a été faite au niveau des *formes* produites, indépendamment des contextes syntaxiques créés par les apprenant/es. La production des formes pour les 12 verbes de test a montré que

- les apprenant/es n'ont pas toujours imité la forme présentée en stimulus
- les verbes statiques, fréquents à la forme courte, ont été préférés à la forme longue par les apprenant/es, à tous les stades
- les verbes dynamiques au total ont plus souvent été produits à la forme longue que les verbes statiques, mais que ce pourcentage n'est jamais monté au-dessus de 50% ; les verbes dynamiques n'ont donc pas été préférés à la forme longue comme attendu par l'hypothèse 1
- les verbes *acheter* et *visiter*, les deux verbes dynamiques fréquents à la forme longue dans l'input, ont été préférés à la forme longue par les apprenant/es à tous les stades
- le pourcentage de formes courtes entre les deux catégories de verbes statiques et dynamiques ne s'est avéré significativement différent qu'au stade 3, ce qui ressemble à ce que l'on avait vu pour la tâche de production libre ; il se peut donc que l'influence de l'aspect lexical apparaisse au stade 3
- la corrélation au niveau des échelles de pourcentages de formes courtes entre l'input et les apprenant/es était proche de .700 pour les stades 2 et 3, alors que la corrélation était pratiquement nulle au stade 1. Cette absence de corrélation était essentiellement due à la différence de production des verbes *manger* et *trouver dynamique* ainsi que *penser* et

trouver statique (cf. figure 12-1). Un calcul de corrélation sans les verbes *penser* et *trouver*, qui sont les verbes les moins bien connus des apprenant/es, donne néanmoins un taux de corrélation de plus .600. Ces résultats parlent en faveur de l'hypothèse 2 sur l'influence de la fréquence des formes, même au stade 1.

Les données globales du test d'imitation parlent donc en faveur de l'hypothèse 2 sur la fréquence des formes dans l'input. Elles semblent aussi indiquer qu'une certaine connaissance des verbes soit nécessaire pour pouvoir obtenir un effet de fréquence des formes comme le montrent les données du stade 1. Pour les verbes connus des apprenant/es, l'effet de la fréquence semble tangible rapidement alors que l'influence de l'aspect lexical ne commence à se voir qu'au stade 3. On pourrait considérer que l'aspect lexical est un fruit de la fréquence type, les apprenants du stade 3 ayant eu suffisamment d'expérience pour pouvoir reconnaître les formes correspondant au sens prototypique des verbes.

Une analyse détaillée des outputs des verbes *habiter*, *acheter* et *parler* avec le stimulus au présent et à l'infinitif ont montré que

- les verbes *habiter* et *acheter* ont été produits conformément aux deux hypothèses, donc à la forme attendue selon l'aspect lexical et la fréquence dans l'input
- le verbe *parler* avait été produit avec un peu plus de variation, avec une préférence pour la forme courte aux stades 2 et 3, ce qui n'est pas attendu par l'hypothèse sur l'influence de l'aspect lexical, mais qui pourrait s'expliquer par la fréquence de la forme courte dans l'input pour ce verbe.

Ces outputs ont également permis de voir que les apprenant/es du stade 1 ont avant tout produit la forme verbale et un peu de matériel supplémentaire, alors que les apprenant/es du stade 2 ont le plus souvent produit des phrases au présent, quelle que soit la construction syntaxique du stimulus. Les apprenant/es du stade 3 ont souvent réussi à imiter la construction du stimulus, même lorsqu'elle était à l'infinitif. On a donc pu observer un développement de la syntaxe des apprenant/es entre ces trois stades. Selon Bybee (2008 : 221), la découverte des constructions est elle aussi le fruit de la fréquence type. On peut en effet voir la construction d'une phrase à l'infinitif comme suivant le *pattern* auxiliaire + verbe. La production des formes selon l'aspect lexical des verbes et

le développement de la syntaxe semblent donc se faire simultanément chez les apprenant/es au stade 3, tous deux étant le fruit d'une certaine expérience de la langue cible, alors que l'influence de la fréquence des formes pourrait avoir un effet dès que l'apprenant/e a eu l'occasion de traiter quelques occurrences d'un verbe, donc dès le stade 1.

13. Test de rappel

Ce chapitre présente l'analyse des données du test de rappel. Ce test psycholinguistique permet de vérifier si les apprenant/es ont plus facilement accès à une forme qu'à une autre, lorsqu'elle est présentée hors contexte syntaxique. Contrairement aux autres données, on exclut ici l'influence du contexte de la phrase de production. L'objectif de ce test est ainsi de voir si les apprenants ont plus facilement accès à la forme qui correspond à l'aspect lexical ou à la forme la plus fréquente du verbe, quelle que soit la forme présentée dans le stimulus.

Ce chapitre présente d'abord le test de rappel libre à la section 13.1. puis sa construction à la section 13.2. La section 13.3. présentera le traitement des données et la section 13.4. les résultats. Ce chapitre se terminera par un résumé.

13.1. Présentation du test de rappel

Le test de rappel utilisé ici est inspiré du test de *rappel sériel*, dans lequel le participant entend une série d'items, mots ou chiffres, qu'il doit ensuite rappeler dans le même ordre. Ce test est avant tout utilisé dans le cadre de la recherche sur le fonctionnement et les déficits de la mémoire, notamment pour mesurer la capacité de la mémoire de travail (par exemple rappel de chiffres) et du stock phonologique (par exemple rappel de non-mots). Dans la recherche sur l'acquisition des langues, il est surtout utilisé en corrélation avec d'autres tâches, notamment pour mettre en évidence les apports des différents éléments de la mémoire (*cf.* par exemple Gathercole, 1995 ; Baddeley *et al.*, 1998 ; Gathercole *et al.*, 1999).

D'après Ganoac'h & Larigauderie (2000 : 160), le rappel libre, sans la contrainte d'un rappel dans l'ordre, est un outil pour tester la présence ou non d'un contenu en mémoire de travail (activation d'une représentation).

La forme libre est plus classiquement utilisée comme outil d'investigation des contenus en mémoire de travail : on recueille, comme observable de la présence ou absence d'une information en mémoire de travail, sa présence ou son absence dans le rappel. (Ganoac'h & Larigauderie, 2000 : 161)

En d'autres termes, l'hypothèse générale à partir de ce test est donc que l'on considère que la forme rappelée par l'apprenant/e est la forme la plus accessible, la forme activée en mémoire.

La principale différence entre le test de rappel et le test d'imitation est que l'apprenant/e n'a pas d'indice syntaxique (présence d'un auxiliaire ou non) pour l'utilisation d'une forme ou d'une autre. La production des verbes dans ce test est donc différente de ce que l'on trouve en production libre, mais selon Allen & Hulme (2006), le test de rappel (sériel) implique les mécanismes habituels de production (et de compréhension) du langage.

In order to repeat a spoken list of words, a participant must be able to perceive the words presented, and subsequently produce them. More radically, it might be argued that limits to serial recall performance may reflect (at least in part) variations in the efficiency of language processing mechanisms, because the representations underlying recall presumably are the same representations as those underlying the perception, comprehension, and production of speech. (Allen & Hulme, 2006 : 64-65)

La construction des séries du test de rappel libre doit être faite en tenant compte d'un certain nombre de facteurs. Les caractéristiques des items constituant une série peuvent en effet avoir une influence sur le rappel de la série entière.

- Les séries mixtes seraient mieux rappelées au total que des séries pures (Saint-Aubin & LeBlanc, 2005). Comme on le verra à la section suivante sur la construction du test de rappel, chaque série de cinq verbes comportera au moins deux formes courtes, deux formes longues, deux verbes considérés comme statiques et deux comme dynamiques.
- Les mots concrets sont mieux rappelés que les mots abstraits, car la représentation sémantique joue un rôle dans la performance du rappel libre (Allen & Hulme, 2006). Ce facteur pourrait avoir une influence sur le rappel des verbes statiques, qui seraient moins bien rappelés que les verbes dynamiques. En effet, à l'exception du verbe *habiter*, les verbes statiques représentent plutôt des concepts abstraits.
- Les mots courts sont mieux rappelés que les longs (Cowan *et al.*, 2003 ; Farrell, 2006). Ainsi, la forme courte du présent serait mieux rappelée que la forme longue de l'infinitif.
- Finalement, les mots fréquents sont mieux rappelés que les mots peu fréquents (Allen & Hulme, 2006 ; Saint-Aubin & LeBlanc, 2005). Selon ce facteur, un verbe fréquent dans l'input serait mieux rappelé qu'un

verbe peu fréquent. Ce facteur n'a pas été considéré lors de la construction de chaque série, le petit nombre de verbes utilisés ne permettant pas beaucoup de marge de manœuvre.

Ces facteurs pourraient tous contribuer à un moins bon rappel des verbes statiques à la forme longue puisqu'elles sont abstraites, longues et peu fréquentes.

Finalement, le test de rappel libre implique la capacité de la mémoire de travail d'une manière similaire à ce qui a été présenté pour le test d'imitation. Un facteur important ici est celui des effets de primauté et de récence, selon lesquels les premiers items sont mieux rappelés que les autres et les derniers mieux que ceux du milieu (Ganoac'h & Larigauderie, 2000 : 27). Pour cette raison, les items de tests doivent être présentés au milieu des séries, même s'il n'y a pas de contrainte de rappel dans l'ordre.

13.2. Construction du test de rappel

Le test de rappel libre utilisé ici comporte 24 séries de 5 formes verbales. Les séries sont constituées de formes verbales à la forme courte (*parle*) et à la forme longue (*parlE*) provenant des 12 verbes de tests ainsi que de 12 verbes de remplissage. Les participant/es entendent une série à la fois et doivent la répéter directement après l'avoir entendue en entier, sans tâche de suppression articulatoire et sans indice aidant au rappel. Le nombre d'items qu'un sujet peut rappeler assez facilement dans un test de rappel sériel (donc avec la contrainte d'un rappel des items dans le même ordre que l'ordre de présentation) s'élève à 7 ± 1 item. Il semble que ce chiffre descend à 5 ± 1 en LE/L2 selon le niveau d'après Ganoac'h & Larigauderie (2000 : 236). Le choix de cinq items par séries devrait donc correspondre à la situation des participant/es⁶².

13.2.1. Construction des séries de formes verbales

Chaque série du test de rappel comporte donc cinq formes verbales, trois formes de test au milieu (positions 2, 3 et 4) et deux formes de remplissage, une au début (position 1) et une à la fin (position 5). Chaque forme verbale des 12 verbes de test (la forme courte et la forme longue d'un même verbe) apparaît

⁶² Des tests pilotes avaient montré que la tâche était trop facile avec quatre formes verbales.

une fois dans chaque position sérielle n° 2, 3 et 4, ce qui donne un résultat de 24 séries.

La distribution des formes verbales de test dans les séries a été faite de manière aléatoire. Cette répartition a été corrigée afin

- qu'il n'y ait pas deux fois le même verbe dans la même série,
- qu'il y ait au moins un verbe dynamique et un verbe d'état dans les positions 2, 3 et 4,
- qu'il y ait au moins une forme courte et une forme longue dans les positions 2, 3 et 4,
- que les verbes n'apparaissent pas toujours à la même forme en premier dans les trois positions ; chaque verbe est présenté d'abord au moins une fois à la forme courte et une fois à la forme longue.

Les positions 1 et 5 des séries ont été complétées avec 12 autres verbes : *admirer, aider, arriver, demander, désirer, écouter, jouer, passer, proposer, rester, tomber, travailler*. Ces verbes de remplissage ont été choisis soit par leur présence dans la production des apprenant/es de corpus les plus débutants, soit par leur présence dans une séquence de tâches autour d'un film qui avait été prévue au départ. Tous les verbes sauf *admirer* et *proposer* se trouvent dans le manuel pour débutants *Escalade 1* (Tillman *et al.*, 2001 ; cf. section 8.1.1.). Ces deux verbes devraient néanmoins être reconnus par les apprenant/es puisqu'ils existent en anglais. Les verbes *admirer* et *désirer* ont été considérés comme des verbes statiques et les autres comme verbes dynamiques.

Chaque forme des verbes de remplissage a été mise une fois en position 1 et une fois en position 5. La répartition de ces 24 formes verbales a été faite de manière aléatoire. Le résultat a été corrigé de manière à ce qu'il n'y ait pas plus de deux verbes statiques dans une même série, que le même verbe n'apparaisse pas en position 1 et 5 de la même série et que le dernier verbe d'une série ne soit pas le même que le premier de la série suivante.

Finalement, chaque série contient donc au moins deux formes courtes et deux longues ainsi que deux verbes d'état et deux verbes dynamiques.

Le tableau 13-1 ci-dessous présente la liste A avec les formes verbales ainsi que la durée des séries en secondes⁶³.

⁶³ Contrairement au test d'imitation, le nombre de syllabes par série n'a pas été calculé, car la présence des différentes formes et différents sens aspectuels a été considérée comme prioritaire et la marge de manœuvre faible.

Tableau 13-1 : Liste A du test de rappel libre

	pos 1	pos 2	pos 3	pos 4	pos 5	durée (sec)
1	désirer	regarder	déteste	mange	proposer	8,351
2	admire	préférer	trouve	manger	joue	8,12
3	proposer	adore	aime	trouver	passé	8,242
4	admirer	trouve	manger	adorer	aide	8,026
5	arrive	aime	habiter	regarder	tombe	8,019
6	reste	acheter	adorer	parle	passer	8,318
7	désire	parler	visite	adore	tomber	8,095
8	écoute	aimer	acheter	préfère	arrive	8,223
9	passé	penser	parle	habite	travailler	8,235
10	demander	habite	détester	visite	aider	7,871
11	rester	pense	regarde	aimer	travaille	8,074
12	aider	visiter	pense	trouve	écouter	8,204
13	joue	détester	visiter	achète	admire	8,233
14	jouer	achète	habite	parler	arriver	8,451
15	demande	manger	regarder	aime	admirer	8,179
16	aide	trouver	préfère	penser	demander	8,288
17	tomber	déteste	mange	visiter	désire	8,31
18	travaille	préfère	aimer	acheter	reste	8,26
19	arriver	mange	préférer	habiter	écoute	8,212
20	propose	adorer	parler	regarde	désirer	8,299
21	travailler	visite	achète	détester	jouer	8,129
22	écouter	regarde	penser	déteste	rester	8,155
23	passer	parle	adore	préférer	demande	8,18
24	tombe	habiter	trouver	pense	propose	8,323

Dans le but d'éviter les effets de série, une liste B, en miroir de la liste A, a aussi été créée : dans chaque série, la forme courte a été remplacée par la forme longue du même verbe et vice-versa la forme longue par la courte du même verbe. Finalement, des listes C et D ont été créées en inversant l'ordre du bloc des 12 premières et 12 dernières séries des listes A et B liste (13-24 puis 1-12).

Quatre séries d'exercices ont été construites en suivant les mêmes règles de construction que pour les séries de test. Elles sont présentées dans le tableau 13-2. Pour ces séries, les verbes suivants ont été choisis : *chercher*, *étudier*, *tourner*, *contrôler*, *fermer*, *bloquer*, *continuer*, *essayer* ainsi que deux verbes statiques *estimer* et *respecter*. La plupart de ces verbes étaient présents dans la séquence film qui a été enlevée. Les verbes *tourner*, *fermer*, *bloquer*, *estimer* et *respecter* ne se trouvent pas dans *Escalade 1* ce qui fait que les exercices sont probablement un peu plus difficiles que les séries de test.

Tableau 13-2 : Séries d'exercices

ex 1	tourne	contrôler	bloque	respecter	étudier
ex 2	estime	tourner	ferme	bloquer	cherche
ex 3	continuer	étudie	essayer	contrôle	chercher
ex 4	essaie	fermer	respecte	continue	estimer

Contrairement au test d'imitation, le test de rappel libre utilisé ici ne comprend pas de tâche de suppression articulatoire. Les tests pilotes effectués semblaient indiquer que la charge cognitive du nombre de formes à rappeler soit suffisamment élevée.

Les formes verbales ont été enregistrées dans les mêmes conditions et lors de la même session que les phrases du test d'imitation (section 12.2.1.). Dans le but d'obtenir des formes verbales prononcées de manière aussi proche que possible de formes prononcées dans le contexte d'une phrase, chaque forme a été insérée dans une phrase simple dans le script d'enregistrement. La locutrice devait a) regarder la phrase sur papier, b) la dire par cœur au micro, c) répéter la forme verbale après avoir dit la phrase. Les verbes de tests ont été enregistrés trois fois et les verbes de remplissage deux fois.

Le matériel sonore obtenu a été traité avec le logiciel *Praat*. La forme verbale ressemblant le plus à celle produite dans la phrase a été choisie⁶⁴. Parfois, la clarté de la prononciation (articulation ou lenteur) de la forme l'a emporté sur la proximité de la production en phrase. L'objectif était de fournir aux participant/es des formes aussi compréhensibles que possible. Les bruits de langue ou de respiration ont été enlevés le plus possible sauf lorsque la respiration ou autre bruit semblait directement faire partie de la forme prononcée.

L'insertion des 5 formes verbales dans un même fichier sonore s'est faite à partir d'un fichier sonore de silence dans lequel les 5 formes ont été introduites les unes après les autres. La durée de *onset* à *onset* des formes verbales (début de la forme 1 au début de la forme 2, etc.) d'une série s'élève à 1,7 secondes (entre 1,68 et 1,73 secondes). Des tests pilotes avec une durée de 2 secondes de *onset* à *onset* avait semblé trop longue. Les séries de la liste A ont une durée moyenne de 8,20 secondes (écart type de 0,126 secondes) et celles de la liste B une durée moyenne de 8,19 secondes (écart type de 0,118 secondes). Chaque série commence et fini par un moment de silence d'environ 0,3 secondes.

⁶⁴ J'ai donc choisi de ne pas découper les formes verbales dans les phrases, afin d'obtenir une forme verbale entière sans lien avec ce qui précède ou qui suit.

13.2.2. Déroulement et consignes

Le test de rappel suivait le déroulement décrit ci-dessous :

- un point d'exclamation apparaît durant deux secondes, suivi du son
- une seconde après la fin du son apparaissent cinq étoiles * * * * * et c'est le moment où le participant doit rappeler la série
- lorsqu'il a terminé, le participant doit cliquer pour aller à la fiche suivante (*self-paced*).

Les consignes ont été données oralement et en suédois par l'expérimentatrice comme pour le test d'imitation. Il s'agissait d'une description du déroulement présenté ci-dessus. La consigne du rappel était formulée de la manière suivante : "là, tu essaies de répéter les cinq verbes, à la même forme que tu les avais entendus". La consigne devait encourager la production de matériel, même si ce n'est que des parties de mots, dans l'espoir que le participant termine la production d'une forme verbale une fois qu'il l'avait entamée : "si tu ne te souviens pas de tout, dis autant que tu peux, même si ce n'est que des bouts de formes verbales."

13.3. Conventions de transcriptions et codage

La transcription des données du rappel libre a été faite dans un tableur. Tous les enregistrements de cette tâche ont été écoutés au moins deux fois, une fois au moment de la transcription (un/e participant/e à la fois) et une fois par tâche.

Les formes verbales ont été codées d'après un codage similaire à celui du test d'imitation, un codage qui permette la comparaison entre la forme du stimulus et de l'output. Pour faciliter la lecture des transcriptions, ce codage a aussi été marqué directement sur les formes verbales comme le montre le tableau 13-3 ci-dessous. La transcription des verbes de remplissages (position 1 et 5 des séries) a été faite selon les mêmes règles que pour les verbes de test, sauf que l'omission d'une forme a été transcrite par un \emptyset .

Tableau 13-3 : Codage des formes verbales dans le test de rappel

stimulus	output	codage	transcription
forme courte	forme courte	C>C	pense
forme courte	forme longue	C>L	pensE ⁶⁵
forme courte	autre verbe	C>øtr	<i>danse, dansE</i>
forme courte	omission	C>øtr	pense
forme longue	forme longue	L>L	penser
forme longue	forme courte	L>C	pense
forme longue	autre verbe	L>øtr	<i>dansE, danse</i>
forme longue	omission	L>øtr	pensE
verbe de remplissage	omission	-	ø

Étant donné que les formes ont été rappelées dans l'ordre choisi par les participant/es, le placement des formes par rapport aux stimuli est un choix à faire au moment de la transcription. Ici, le plus de formes possible ont été mises sous les stimuli qui semblaient correspondre. Par exemple, la forme *danse* a été mise sous le stimulus *penser*, du moment où ce verbe-ci n'a pas été produit⁶⁶. Les formes qui n'ont pas été reconnues comme faisant partie de la série en question ainsi que les mêmes verbes répétés (à la même forme ou non) ont été mis dans une colonne *V en plus*. Lorsque deux formes du même verbe ont été produites, c'est la première qui a été mise sous le stimulus, comme dans l'exemple (13-1) dans lequel *achète* a été produit avant *achetE*. Cet exemple permet également d'illustrer la manière de transcrire la production d'un/e apprenant/e.

(13-1) ----- positions -----
 --- verbes de test ---

Part	n°	1	2	3	4	5	V en plus
<i>Stim</i>	14	<i>jouer</i>	<i>achète</i>	<i>habite</i>	<i>parler</i>	<i>arriver</i>	
P01	14	jouer	achète	habite	parle	arrive	achetE
codes			C>C	C>øtr	L>C		

⁶⁵ La forme se termine par un -E même si un r final a été prononcé.

⁶⁶ La consigne *rappelle le plus de matériel possible* avait pour objectif d'encourager les participant/es à produire un maximum d'éléments, mais seuls les verbes de tests produits en entier sont intéressants pour l'analyse. Le placement des formes produites dans les différentes colonnes a donc été fait à partir des mots et non des syllabes.

Lorsqu'un verbe a été produit en deux temps par exemple *visite#E*, *achète#E*, où le # correspond à une pause, seule la forme finale a été transcrite, donc *visitE* et *achetE* pour les exemples cités.

La question du degré de sévérité pour les verbes considérés comme *autre verbe* (transcrits en italique) est discutable. En principe, les erreurs de prononciation ont été acceptées, par exemple *aïmer* pour *aimer* le /i/ étant présent dans la forme verbale ou *rigarde* pour *regarde* en raison de la possible influence de la prononciation du mot en anglais. Par contre, lorsqu'un autre verbe existant a été produit, comme *danser* pour *penser*, il a été considéré comme *autre verbe*, même si la production de cet *autre verbe* est le fruit du hasard comme dans *adhérer* pour *adorer*. Le tableau 13-4 ci-dessous présente les variantes qui ont été acceptées et comptées comme *autre verbe*. Le principal argument pour compter une forme comme *autre verbe* est que la forme a été considérée comme trop loin de la forme d'origine. Sans indication, la forme n'a été produite qu'une fois.

Tableau 13-4 : Formes verbales discutables dans le rappel libre

verbe	accepté (prononciation acceptable)	compté comme <i>autre verbe</i>
acheter		<i>échetE</i> (2 fois), <i>acht</i> , <i>asstE</i> (2 fois), <i>assète</i>
aimer	<i>aïmE</i> (3 fois), <i>amE</i> (5 fois), <i>ame</i>	<i>admE</i> , <i>aïmier</i>
adorer	<i>j'adore</i> : cette forme a été acceptée parce qu'elle indique clairement que la forme stimulus courte a été traitée et non pas simplement répétée.	<i>adorant</i> (2 fois), <i>adérE</i> , <i>adère</i> , <i>admange</i> , <i>padorE</i> , <i>abeurE</i> , <i>amorE</i>
détester	<i>détetetedétester</i> (2 fois), <i>dédétester</i> : la forme verbale a finalement été prononcée et elle semble avoir été visée dès le début.	<i>destétE</i> , <i>dédastE</i> , <i>estE</i> , <i>dégestE</i> , <i>testE</i> (4 fois), <i>destE</i> , <i>déchétester</i>
habiter	<i>habit</i> : la prononciation de la consonne finale est douteuse – peut-être seulement difficile à percevoir.	<i>invitE</i> , <i>adite</i> , <i>admite</i>
parler	<i>pèrle</i>	<i>pranler</i>
penser	<i>banse</i> : la forme a été acceptée en raison du doute de la prononciation de la première consonne.	<i>danser</i> (9 fois), <i>danse</i> (4 fois), <i>français</i> (2 fois), <i>prende</i> (4 fois), <i>passer</i>
préférer	<i>préfe</i> (2 fois) : la prononciation de la consonne finale est douteuse – peut-être seulement difficile à percevoir. <i>perférE</i> , <i>prévère</i>	<i>préparE</i> , <i>défère</i> , <i>prépère</i> , <i>privère</i> , <i>premier</i>
manger	<i>mangne</i> , <i>manguE</i>	<i>dumange</i> , <i>marche</i> , <i>demange</i> (2 fois)
regarder	<i>rigarde</i> (4 fois) : prononciation influencée de l'anglais.	

verbe	accepté (prononciation acceptable)	compté comme <i>autre verbe</i>
trouver	<i>trou</i> (4 fois) : la prononciation de la consonne finale de <i>trouve</i> est douteuse – peut-être seulement difficile à percevoir. <i>trovE</i> : l'erreur a été perçue comme une erreur de prononciation.	<i>atrouE</i>
visiter		<i>vester</i>

Les enregistrements de P03, P07, P09, P18, P23 et P27 ont également été écoutés et transcrits par une autre personne. Dans trois cas, il n'a pas été possible de savoir quelle était la forme produite par le participant. Il s'agit des occurrences de *préfère* ou *préférE* chez P23 (série 8) ainsi que chez P27 (série 23) classées comme C>øtr et une occurrence de *mange* ou *mangE* de P23 (série 15) classée comme L>øtr.

13.4. Résultats du test de rappel

Chaque participant/e a traité 24 séries pour un total de 792 séries. Les résultats sont ainsi basés sur le traitement de 2376 formes stimulus des verbes de test ce qui correspond à 198 occurrences de chaque verbe ou 99 de chaque forme. En moyenne, ce test, y compris les exercices, a duré 11'01 minutes pour les participant/es (entre 7'54 – P31 et 16'10 – P24 ; durée calculée du premier au dernier *enter*). Sur les 2376 formes verbales de test dans le stimulus, les participant/es en ont rappelé 1444 à la même forme ou à l'autre forme du même verbe, ce qui correspond à 61%. Le pourcentage de rappel est donc supérieur à celui du test d'imitation dans lequel seulement 49% des outputs contenaient le verbe stimulus (*cf.* section 12.4.).

Les natifs ont mis 9'18 minutes en moyenne et ils ont rappelé 91% des formes en moyenne. Pour des raisons techniques, les données de Nat10 n'ont pas été retenues pour ce test. Trois des neuf natifs ont rappelé tous les verbes comme dans le stimulus.

13.4.1. Vérification de l'hypothèse 1 sur l'influence de l'aspect lexical

Selon l'hypothèse 1, les apprenant/es produisent la forme courte avec les verbes statiques et la forme longue avec les verbes dynamiques. Le tableau 13-5 ci-dessous présente les résultats du test de rappel. Les verbes ont été regroupés selon l'aspect lexical (*cf.* section 7.2.2.), statique dans le haut du tableau et

dynamique dans le bas. Les données du verbe *trouver* ont été mises à part étant donné qu'il n'y a pas de contexte pour déterminer le sens aspectuel. Les occurrences du tableau 13-5 correspondent à la production codée C>C, C>L, L>L et L>C (cf. tableau 13-3). Pour les apprenant/es, le nombre de stimuli par verbe est de 198 et pour les natifs de 54 au total. L'annexe 8 présente les résultats détaillés au niveau du rapport entre stimulus et output pour chaque verbe (et l'annexe 7 pour chaque participant/e).

Tableau 13-5 : Production des verbes selon l'aspect lexical – test de rappel

	stade 1			stade 2			stade 3			Σ apprenants			natifs		
	C	L	%C	C	L	%C	C	L	%C	C	L	%C	C	L	%C
adorer	15	20	43%	23	19	55%	21	23	48%	59	62	49%	28	19	60%
aimer	24	11	69%	31	21	60%	26	19	58%	81	51	61%	27	21	56%
détester	20	8	71%	21	23	48%	30	16	65%	71	47	60%	27	21	56%
habiter	23	16	59%	22	29	43%	20	20	50%	65	65	50%	23	25	48%
penser	12	3	80%	22	11	67%	20	14	59%	54	28	66%	26	27	49%
préférer	9	22	29%	13	29	31%	18	27	40%	40	78	34%	28	24	54%
<i>statique</i>	103	80	56%	132	132	50%	135	119	53%	370	331	53%	159	137	54%
acheter	8	10	44%	19	29	40%	8	22	27%	35	61	36%	25	22	53%
manger	20	15	57%	26	25	51%	29	16	64%	75	56	57%	29	22	57%
parler	18	21	46%	27	19	59%	18	23	44%	63	63	50%	27	22	55%
regarder	20	18	53%	10	41	20%	16	24	40%	46	83	36%	23	24	49%
visiter	14	23	38%	25	30	45%	16	22	42%	55	75	42%	24	27	47%
<i>dynamique</i>	80	87	48%	107	144	43%	87	107	45%	274	338	45%	128	117	52%
trouver	23	11	68%	42	17	71%	24	14	63%	89	42	68%	26	23	53%

Légende : C : forme courte ; L : forme longue.

Les données du tableau 13-5 révèlent que les données obtenues dans ce test sont différentes de ce que l'on a obtenu dans les deux autres tâches (chapitres 11 sur la tâche de production libre et 12 sur le test d'imitation). Les données montrent que la plupart des pourcentages sont proches de 50% ce qui correspond à ce que l'on trouve dans les stimuli. Les données de l'annexe 7 montrent en outre que le taux de correspondance entre stimulus et output est de 70% et plus pour la plupart des participant/es.

La forme courte a été préférée avec les verbes statiques aux stades 1 et 3, mais pas au stade 2. Au total, les verbes statiques *adorer*, *habiter* et *préférer* n'ont pas été préférés à la forme courte et le verbe *préférer* a toujours été mieux rappelé à la forme longue que courte. On ne retrouve pas ici la préférence claire de l'utilisation de la forme courte avec les verbes statiques que l'on avait pu voir dans les autres tâches.

Par contre la forme longue a été préférée avec les verbes dynamiques et à tous les stades. Seul le verbe *manger* n'a jamais été préféré à la forme longue. Le verbe *trouver* a également été préféré à la forme courte.

A première vue l'hypothèse 1 sur l'influence de l'aspect lexical semble mieux se vérifier avec ces données qu'avec celles en production libre ou dans le test d'imitation. En même temps, ce qui donne l'impression d'un meilleur résultat est l'augmentation du pourcentage de formes longues en particulier avec les verbes dynamiques. On pourrait formuler l'hypothèse que la forme de l'infinitif est plus facilement accessible dans ce test. En effet, le test statistique ne montre aucune différence significative au niveau du pourcentage de formes courtes selon l'aspect lexical des verbes. L'hypothèse sur l'influence de l'aspect lexical ne semble donc pas se vérifier sur ces données.

Ce résultat signifie que le facteur "longueur des mots", selon lequel la forme courte serait plus facilement rappelée que la longue, (Cowan *et al.*, 2003 ; Farrell, 2006) n'a eu que peu d'effets sur la production des apprenant/es ici, par contre sur celle des natifs, qui ont produit plus de formes courtes que de longues. En outre, les verbes statiques n'ont pas été mieux rappelés que les dynamiques au total : les verbes concrets n'ont donc pas été mieux rappelés que les abstraits, comme l'avaient proposé Allen & Hulme (2006). Par contre, les mots plus fréquents ont été mieux rappelés que les moins fréquents (Allen & Hulme, 2006 ; Saint-Aubin & LeBlanc, 2005). Les verbes peu connus des apprenant/es, comme *penser*, ont moins souvent été rappelés, alors que des verbes comme *habiter*, *parler* ou *regarder* ont été rappelés à un haut pourcentage dès le stade 1 (*cf.* annexe 8). Ce facteur rejoint celui de l'expérience de la langue cible, plus l'apprenant est débutant moins il a de vocabulaire.

13.4.2. Vérification de l'hypothèse 2 sur la fréquence des formes

Selon l'hypothèse 2, les apprenant/es produisent la forme la plus fréquente du verbe dans l'input. Le tableau 13-6 présente une comparaison des échelles de pourcentages de formes courtes entre les apprenant/es selon les stades d'abord, puis au total et finalement chez les natifs (groupe contrôle) et dans l'input. Les données de l'input proviennent du tableau 8-2 et les lignes délimitent les regroupements de fréquence des formes (tableau 8-7). Les verbes ont à chaque fois été classés par ordre décroissant de formes courtes. Les données indiquées sous Σ correspondent à la somme des formes courtes et longues produites pour le verbe concerné (*cf.* tableau 13-5 ci-dessus).

Tableau 13-6 : Pourcentages de formes courtes – test de rappel et input

stade 1	Σ	%C	stade 2	Σ	%C	stade 3	Σ	%C
penser	15	80%	trouver	59	71%	détester	46	65%
détester	28	71%	penser	33	67%	manger	45	64%
aimer	35	69%	aimer	52	60%	trouver	38	63%
trouver	34	68%	parler	46	59%	penser	34	59%
habiter	39	59%	adorer	42	55%	aimer	45	58%
manger	35	57%	manger	51	51%	habiter	40	50%
regarder	38	53%	détester	44	48%	adorer	44	48%
parler	39	46%	visiter	55	45%	parler	41	44%
acheter	18	44%	habiter	51	43%	visiter	38	42%
adorer	35	43%	acheter	48	40%	préférer	45	40%
visiter	37	38%	préférer	42	31%	regarder	40	40%
préférer	31	29%	regarder	51	20%	acheter	30	27%

Σapprenants	Σ	%C	natifs	Σ	%C	input	Σ	%C
trouver	131	68%	adorer	47	60%	adorer	32	81%
penser	82	66%	manger	51	57%	penser	368	73%
aimer	132	61%	aimer	48	56%	habiter	150	72%
détester	118	60%	détester	48	56%	détester	10	70%
manger	131	57%	parler	49	55%	préférer	45	58%
habiter	130	50%	préférer	52	54%	aimer	341	55%
parler	126	50%	trouver	49	53%	trouver	393	47%
adorer	121	49%	acheter	47	53%	regarder	223	39%
visiter	130	42%	penser	53	49%	parler	581	38%
acheter	96	36%	regarder	47	49%	manger	160	29%
regarder	129	36%	habiter	48	48%	acheter	110	15%
préférer	118	34%	visiter	51	47%	visiter	25	8%

Légende : Σ : somme des formes courtes et longues ; C : forme courte.

Comme on l'a vu à la section précédente, le verbe *préférer* a plus souvent été rappelé à la forme longue, alors que le verbe *trouver* a surtout été rappelé à la forme courte. La corrélation des rangs du r de Spearman indique une corrélation entre l'input et le stade 1 de .371, avec le stade 2 de .245, avec le stade 3 de .347, avec les apprenant/es au total de .291 et avec les natifs de .218. Les données des stades 1 et 3 sont celles qui corrélaient le mieux avec l'input. Le résultat pour le stade 1 est donc différent de ce que l'on avait vu dans les autres tâches. L'hypothèse 2 sur l'influence de la fréquence des formes dans l'input ne se vérifie donc pas au total, mais avec les apprenant/es des stades 1 et 3.

La figure 13-1 présente la corrélation entre l'input au total et les données du rappel au total pour les apprenant/es. Le placement des verbes se fait horizontalement selon le pourcentage de formes courtes dans l'input et verticalement selon le pourcentage de formes courtes dans le test de rappel au total. La ligne illustre une corrélation parfaite, c'est-à-dire lorsque le

pourcentage de formes courtes est le même pour les deux sources de données. Ainsi, plus les verbes sont proches de la ligne de corrélation plus ils sont corrélés et plus ils sont éloignés moins ils sont corrélés au niveau du pourcentage de formes courtes dans les deux sources de données.

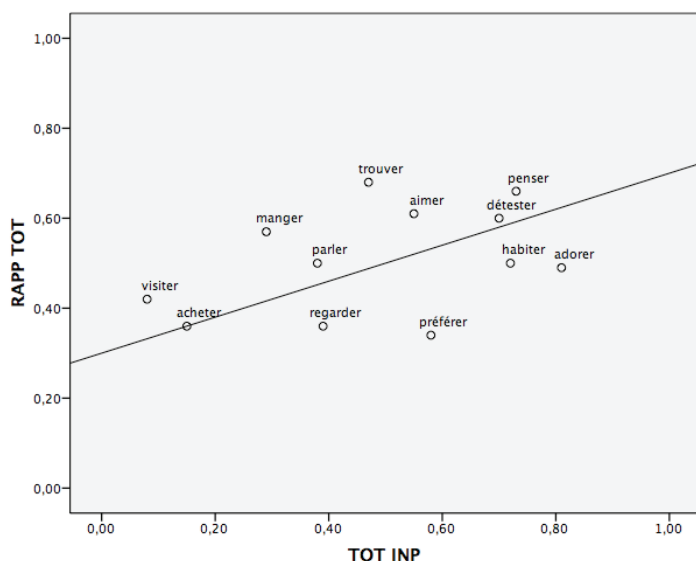


Figure 13-1 : Corrélation input et données du test de rappel au total

On constate que l'absence de corrélation est due au fait que la fréquence des verbes est inversée pour plusieurs verbes, *trouver* et *manger* ont été préférés à la forme courte chez les apprenant/es, alors que les verbes *préférer* et *adorer* ont été préférés à la forme longue. On voit aussi que les verbes sont regroupés autour de 50% dans le test de rappel alors que les données de l'input s'étalent sur toute la largeur.

13.5. Résumé des résultats du test de rappel

La vérification de l'hypothèse 1 sur l'aspect lexical a montré que les verbes statiques étaient plus souvent rappelés à la forme courte et les verbes dynamiques à la forme longue à tous les stades. À première vue, l'hypothèse semblait mieux se vérifier que dans les autres tâches. Dans ce test-ci, un certain

nombre de verbes dynamiques ont été préférés à la forme longue. En même temps, les verbes statiques ont moins souvent été produits à la forme courte que dans les autres tâches, ce qui a annulé la différence entre les deux groupes de verbes.

L'analyse des verbes à partir de l'hypothèse 2 sur la fréquence des formes a montré que la corrélation au niveau du pourcentage de formes courtes était plus faible que dans les autres tâches, même si elle était satisfaisante pour les stades 1 et 3. Ce résultat pourrait tout de même parler en faveur d'une influence de la fréquence des formes. En effet comme tous les verbes ont été présentés exactement le même nombre de fois et que le pourcentage de correspondance entre les stimuli et les outputs est particulièrement élevé dans ce test, le fait que l'on obtienne des corrélations pourrait parler en faveur d'une telle interprétation.

Les résultats pour ce test de rappel semblent indiquer que les deux formes verbales sont stockées et accessibles chez les apprenant/es pour la plupart des verbes. Les préférences d'utilisations sont donc plus visibles dans la production de phrases comme dans le test d'imitation ou en production libre que dans ce test. Cette différence pourrait être liée au fait que la charge cognitive dans la production de phrases est plus forte que dans la production de mots isolés.

Lors de la présentation du test de rappel (section 13.1), un certain nombre de facteurs pouvant influencer le résultat avaient été mentionnés. Les résultats obtenus semblent indiquer que les facteurs de longueur des mots ainsi que celui de mots concrets *vs* abstrait ne semblaient pas jouer de rôle important dans le rappel des formes par les apprenant/es. Par contre, la fréquence du verbe, dans le sens du niveau de connaissance du verbe, semblait jouer un rôle, le verbe *penser* peu connu des apprenant/es ayant été plus rarement rappelé que des verbes comme *habiter*, *parler* ou *regarder* plus fréquents dans l'input adressé aux apprenant/es.

SYNTHÈSE

14. Synthèse finale

14.1. Introduction

L'objectif de ce travail était d'étudier la variation des formes verbales dans la production orale d'apprenant/es en début d'acquisition du français LE/L2. Comme le montre l'exemple (14-1), repris de l'introduction, les formes courtes du type *parle* et longues du type *parle*⁶⁷ des verbes réguliers en *-er* sont parfois produites correctement dans les contextes du présent et de l'infinitif, parfois non.

- (14-1) (P08 – stade 3) C'est deux, c'est deux personnes que va **va voyageE** à en Italie. (...) Et elle fait du soleil en le, oui. En la nuit il elle elle **mange** en un restaurant. Et l'italienne, et deux deux hommes **aime** les filles. Et les hommes **va parle** avec les filles et boire le vin et la bière. Et ils il, il **dansE** ensemble. (...)

La question posée dans ce travail était de voir si l'on pouvait trouver une certaine systématisme dans la production des formes verbales par les apprenant/es.

Le phénomène de la variation des formes a été étudié depuis de nombreuses années à partir de différentes approches, notamment au niveau de la finitude et de la syntaxe (*cf.* chapitre 3). Ces études ont montré que les apprenant/es ont recours à des formes de base (Klein & Perdue, 1997) ou des formes invariables (Bartning & Schlyter, 2004). Les études dans une approche générative et en particulier dans le cadre de la *Missing Surface Inflection Hypothesis* (Prévost & White, 2000 a et b ; Herschensohn, 2001 ; Prévost, 2004, 2007) ont en outre montré que le développement de la morphologie verbale se faisait indépendamment de la syntaxe.

Il existe un grand nombre de recherches qui ont proposé que les apprenant/es de LE/L2 soient influencés par l'aspect lexical dans leur production de formes de la langue cible, en particulier dans l'expression du passé (*cf.* chapitre 4). L'aspect lexical est vu comme une distinction cognitive et perceptuelle qui permet à l'être humain d'organiser sa manière de voir le monde (Smith, 1991). Les principaux concepts liés à l'aspect lexical ont été résumés dans le tableau 4-1. Selon cette approche, les apprenant/es adultes en début d'acquisition

⁶⁷ Par convention l'archiphonème *-E* représente les sons [e] et [ɛ].

sélectionnent les formes verbales qui sont le plus proche de l'aspect lexical prototypique du verbe (*cf.* principes cognitifs dans Andersen, 1990).

Les études sur l'acquisition du français LE/L2 sur le passé, dans le cadre de l'hypothèse de l'aspect (Andersen, 1991, 1993, 2002 ; Kaplan, 1987 ; Bergström, 1997 ; Kihlstedt, 1998 ; Duperron, 2003 ; Howard, 2004 ; Labeau, 2005) ou dans le cadre du projet ESF (Dietrich *et al.*, 1995 ; Noyau, 1990 ; Noyau *et al.*, 1995) ont montré que la production des formes par les apprenant/es semblait avant tout suivre l'opposition statique *vs* dynamique, avec une utilisation préférentielle du présent avec les verbes statiques et du passé composé avec les verbes dynamiques. Après une discussion de l'influence de l'aspect lexical à la fin du chapitre 4, une première hypothèse a ainsi été développée à partir de l'influence de l'aspect lexical pour les contextes du présent et de l'infinitif (*cf.* aussi chapitre 6).

(14-2) Hypothèse 1 – influence de l'aspect lexical

La production des formes courtes et longues des verbes réguliers en *-er* par les apprenant/es est influencée par l'aspect lexical des verbes : ils produisent la forme courte avec les verbes statiques et la forme longue avec les verbes dynamiques.

Cette première hypothèse a été contrastée par une autre hypothèse proposant que la production des apprenant/es reflète la fréquence des formes des verbes individuels dans l'input (pour la motivation de cette hypothèse voir chapitre 5 sur le cadre théorique et 6 sur les hypothèses).

(14-3) Hypothèse 2 – influence de la fréquence des formes de surface

La production des formes courtes et longues des verbes réguliers en *-er* par les apprenant/es est influencée par la fréquence d'occurrence de ces formes dans l'input. Les apprenant/es produisent la forme la plus fréquente du verbe dans l'input, les verbes étant répartis sur un continuum allant des verbes fréquents à la forme courte aux verbes fréquents à la forme longue.

Ces hypothèses ont été testées sur un nombre restreint de verbes présentés au chapitre 7 (section 7.2). L'aspect lexical des verbes choisis a été discuté à la section 7.2.2. Dans le but d'obtenir des données concrètes pour l'hypothèse 2, ces verbes ont d'abord été étudiés dans des sources de données concrètes d'input. Ils ont ensuite été analysés dans des données d'apprenant/es de corpus puis dans le cadre d'une étude expérimentale.

Les principaux résultats de ces différentes parties seront présentés dans les sections qui suivent.

14.2. Résultats des données de l'input et prédictions

L'étude des verbes dans l'input dans le but de pouvoir établir le continuum de fréquence des formes pour la vérification de l'hypothèse 2 a été présentée au chapitre 8. Les sources de données d'input ont été choisies pour refléter les principales variétés de français auxquelles des apprenant/es débutant/es peuvent être exposés, même si la proportion d'exposition peut varier (*cf.* section 8.1.1.). Ces sources de données comportent trois sources d'input oral : les données de deux interlocutrices en conversation avec des apprenant/es débutant/es, les données d'enseignants en classe de français LE/L2 au lycée ainsi que les données de corpus de français parlé en France du DVD C-ORAL-ROM (2005), et une source de données écrites, celles du manuel de français pour apprenant/es débutant/es suédophones, *Escalade 1* (Tillman *et al.*, 2001). Les données de ce manuel nous donnent indirectement accès à un usage de la langue en classe de débutants.

L'analyse de 16 verbes réguliers en *-er* à partir de la fréquence des différentes formes couvertes par les formes du type *parle* et *parlE* (tableau 2-1) ainsi que de 12 verbes irréguliers, à titre de comparaison, a permis d'établir un continuum basé sur le pourcentage de formes courtes (tableau 8-2), respectivement du présent pour les verbes irréguliers (tableau 8-6). L'analyse de la fréquence des formes a montré que l'on pouvait regrouper les verbes par similarité de fréquence :

- les verbes fréquents à la forme courte du présent
- les verbes fréquents à la forme longue (infinitif et participe passé)
- et les verbes fréquents aux deux formes (présent et infinitif) ; la forme courte étant la plus fréquente en occurrences et la forme longue en pourcentage.

Ce regroupement a été utilisé pour simplifier la discussion, mais les corrélations ont été faites à partir du continuum en entier.

L'analyse de la fréquence des formes des verbes réguliers en *-er* a également révélé les différences et similitudes au niveau des sources d'input (tableau 8-3) : les verbes du manuel *Escalade 1* étaient pour la plupart fréquents à la forme courte, alors que le pourcentage de formes longues était particulièrement élevé chez les interlocutrices, notamment en raison de la fréquence de la 2^e du pluriel du présent (vouvoiement de l'INT-A).

La comparaison du pourcentage de formes courtes et de l'aspect lexical des verbes au niveau de l'input a révélé que les verbes fréquents à la forme courte

étaient tous des verbes statiques et que les verbes plus fréquents à la forme longue que courte (au niveau du pourcentage des formes) étaient tous des verbes dynamiques. Ce résultat a montré qu'il serait difficile de vérifier les deux hypothèses sur les données d'apprenant/es, la fréquence dans l'input et l'aspect lexical s'avérant être étroitement liés. La seule manière de les distinguer dans la production des apprenant/es était de regarder si tous les verbes dynamiques étaient produits de préférence à la forme longue comme attendu par l'hypothèse 1 ou si l'on pouvait voir une plus grande variation pour le groupe de verbes moins clairement fréquents à la forme longue dans l'input, comme attendu par l'hypothèse 2. Le tableau 14-1 présente d'une part les pourcentages de formes courtes dans l'input et d'autre par les prédictions concrètes pour les verbes individuels.

Tableau 14-1 : Input et prédictions pour la production des apprenant/es

	% formes courtes dans l'input	hypothèse 1 fréquence	hypothèse 2 aspect lexical
adorer	81%	C	C
penser	73%	C	C
trouver S	72%	-	C
habiter	72%	C	C
détester	70%	C	C
préférer	58%	C	C
aimer	55%	C	C
trouver Σ	47%	C/L	-
jouer	42%	C/L	L
regarder	39%	C/L	L
parler	38%	C/L	L
travailler	37%	C/L	L
manger	29%	C/L	L
trouver D	26%	-	L
acheter	15%	L	L
voyager	13%	L	L
étudier	10%	L	L
visiter	8%	L	L

Légende : trouver S : sens statique ; D : sens dynamique ; Σ : trouver lexème ; C : forme courte ; L : forme longue.

L'analyse de 12 verbes irréguliers a été présentée dans la section 8.2.2 et les prédictions ont été résumées dans le tableau 8-8. L'analyse a montré des tendances similaires à celles des verbes réguliers en *-er*. Elle a aussi révélé que les verbes dynamiques choisis ici n'étaient pas particulièrement fréquents au participe passé, mais plutôt à la forme de l'infinitif.

14.3. Résultats des données d'apprenant/es de corpus

D'une part, les données des apprenant/es de corpus provenaient de 25 enregistrements du corpus Thomas (Thomas, 2006) comprenant des conversations d'une durée de 10 à 30 minutes avec 20 apprenant/es suédophones (entre 30 et 70 ans) qui suivent des cours du soir de français. D'autre part, les données provenaient de sept enregistrements de 60 minutes du corpus Lund (Schlyter), trois enregistrements de l'apprenant Karl (23 ans) et quatre de l'apprenante Lisa (20 ans), le premier ayant appris le français en France, la seconde à l'école en Suède. Tous ces apprenants ont commencé à apprendre le français à l'âge adulte. Leur niveau de français se situe entre les stades 1 et 3 de Bartning & Schlyter (2004, *cf.* section 3.2. et annexe 1 ; *cf.* section 7.1.2. pour la présentation des apprenant/es).

L'analyse de la production des formes courtes (*parle*) et longues (*parle*) dans les contextes du présent et de l'infinitif des 16 verbes réguliers en *-er* par les apprenant/es a montré que :

- les verbes fréquents à la forme courte dans l'input, les verbes statiques, ont surtout été produits à la forme courte et presque uniquement dans le contexte du présent ; la fréquence de la 2^e du pluriel du présent des interlocutrices n'a pas influencé la production de ces verbes par les apprenant/es
- les verbes qu'on peut considérer comme fréquents aux deux formes dans l'input, des verbes dynamiques, ont été produits à la forme courte et à la forme longue dans les contextes du présent et de l'infinitif, parfois correctement, parfois non
- les verbes fréquents à la forme longue dans l'input, des verbes dynamiques, ont surtout été produits à la forme longue, avant tout dans les contextes de l'infinitif, mais aussi dans les contextes du présent.

Une analyse statistique a montré une différence significative au niveau du pourcentage de formes courtes entre les verbes statiques et les verbes dynamiques et une corrélation de .862 entre le pourcentage de formes courtes chez les apprenant/es et dans l'input au total. Les deux hypothèses semblaient donc se confirmer. Or, les verbes dynamiques n'ont pas tous été préférés à la forme longue. La différence significative entre les deux groupes aspectuels de verbes provient probablement du fait que les verbes statiques ont presque uniquement été produits à la forme courte (92% à 100%). On a pu voir que les

verbes dynamiques *parler, travailler, manger* et *regarder* avaient été produits avec un plus faible pourcentage de formes courtes que les verbes statiques, mais avec une préférence pour la forme courte. Ces résultats parlent donc plutôt en faveur de l'hypothèse 2 sur l'influence de la fréquence des formes.

L'analyse des verbes irréguliers chez les apprenant/es de corpus n'a pas permis de voir si l'une des hypothèses se vérifiait mieux que l'autre. A première vue, les apprenant/es semblaient suivre l'aspect lexical, mais on n'a pas pu voir que les verbes dynamiques étaient tous préférés à une autre forme que la forme du présent.

14.4. Résultats de la partie expérimentale

Les données des apprenant/es de corpus ont été complétées par des données expérimentales dans le but de vérifier les deux hypothèses de manière plus ciblée, en contrôlant les données à traiter par les apprenant/es. Le fait que les verbes statiques et fréquents à la forme courte aient presque uniquement été produits à la forme courte et dans le contexte du présent par les apprenant/es de corpus n'a en effet pas permis de savoir comment ces verbes seraient produits dans un contexte pour l'infinitif.

La partie expérimentale a été consacrée à l'étude de 12 verbes réguliers en *-er* (cf. tableau 10-2). Les verbes *jouer, travailler, voyager* et *étudier* n'ont pas été repris dans cette partie (cf. section 10.3. sur le choix des verbes).

Les données proviennent de 33 apprenant/es et de 10 natifs (groupe contrôle). Les apprenant/es sont des lycéens (moyenne d'âge 17,7 ans) qui suivent des cours de français à un niveau débutant dans différentes classes de lycées de Lund et de Malmö (cf. section 10.5.).

La production de ces verbes a été analysée dans trois tâches différentes :

- Une tâche de production libre (tâche de narration) qui consistait à raconter une histoire à partir de la série d'images *Le voyage en Italie* en maximum 10 minutes (cf. chapitre 11).
- Un test d'imitation (*elicited imitation*) (cf. chapitre 12) dans lequel le participant entend une phrase qu'il doit ensuite répéter le plus exactement possible. Les 12 verbes de test ont été présentés aux apprenant/es dans des phrases au présent et à l'infinitif, chaque verbe étant présenté une fois dans chaque contexte. Les 27 phrases de test ont été présentées dans le tableau 12-1.

- Un test de rappel libre (*cf.* chapitre 13) dans lequel les apprenant/es répètent à voix haute des séries de cinq formes verbales courtes et longues enregistrées. Les 24 séries ont été présentées dans le tableau 13-1.

La tâche de production libre a servi à vérifier le niveau des apprenant/es (*cf.* section 11.3.). Cette vérification a été faite à partir des stades de Bartning & Schlyter (2004 ; *cf.* section 3.2. et annexe 1) à l'aide du logiciel *Direkt Profil* (Granfeldt, Nugues *et al.*, 2005). Aucun apprenant n'a été évalué à un stade supérieur au stade 3 et le regroupement des apprenant/es a donné 9 participant/es au stade 1, 14 au stade 2 et 10 au stade 3, qui est le stade le plus avancé dans ce travail.

Une analyse des caractéristiques de la production des verbes réguliers en *-er* dans la tâche de narration par ces apprenant/es (section 11.4.2.) a montré

- une nette préférence pour la production de contextes du présent
- que la production de contextes pour l'infinitif n'apparaissait pratiquement qu'au stade 3 (*cf.* figure 11-2)
- que le contexte du présent était exprimé dans 40% des cas par une forme longue du type *je parle*, à tous les stades (*cf.* figure 11-3)
- que le contexte de l'infinitif était le plus souvent produit correctement avec une forme longue (*cf.* figure 11-4).

On a donc retrouvé la variation caractéristique des débutants chez ces apprenant/es. Comme l'avaient noté Bartning & Schlyter (2004) et Prévost (2004, 2007), la forme courte est essentiellement produite dans les contextes du présent alors que la forme longue est utilisée dans les deux contextes syntaxiques.

Mais, ces données présentent beaucoup de différences individuelles. Ainsi, pour les apprenant/es du stade 2, le haut pourcentage de formes longues dans les contextes du présent (40%) était le fait de la moitié des apprenant/es de ce stade, l'autre moitié ayant produit des phrases au présent et à la forme courte. Pour le stade 3, on a pu voir que les apprenant/es, qui avaient produit des contextes pour une forme de l'infinitif, avaient un taux plus élevé de formes longues dans les contextes du présent (*cf.* section 11.4.2. et annexe 4). Ces résultats suggèrent que la fréquence de la construction de l'infinitif dans l'input a augmenté pour les apprenant/es du stade 3 ou du moins le traitement de cette construction, ce qui

pourrait parler en faveur d'une influence de la fréquence dans l'input (*cf.* section 5.1.4).

L'analyse des 12 verbes de test a été faite de deux manières dans toutes les données de la partie expérimentale : d'une part au niveau de l'aspect lexical des verbes (hypothèse 1) et d'autre part au niveau de la corrélation des échelles de pourcentages de formes courtes (hypothèse 2). Les résultats de chaque partie ont été résumés à la fin des chapitres 11, 12 et 13. Dans cette synthèse, les deux hypothèses seront discutées à partir des mêmes tableaux : un premier tableau, 14-2, avec les résultats au total et un deuxième tableau, 14-3, avec les résultats par stade.

Dans le but d'obtenir une vue d'ensemble des différents résultats, le tableau 14-2 présente à gauche des données de l'input, l'input au total ainsi que les données du manuel pour débutants *Escalade 1* (qui sont également incluses dans l'input au total). Les colonnes de droite présentent d'abord les données des apprenant/es de corpus puis les données de la partie expérimentale (les données proviennent des tableaux 11-4, 12-5 et 13-5). Seuls les 12 verbes de tests sont présentés dans ce tableau. Les verbes sont présentés selon le continuum de pourcentage de formes courtes dans l'input au total (Σ input) comme dans le tableau 14-1. Les lignes indiquent les regroupements au niveau de la fréquence dans l'input (*cf.* plus haut section 14.2.).

Finalement, les deux dernières rangées du tableau 14-2 présentent les résultats de calculs statistiques. Les données de la ligne *SvsD* indiquent la probabilité que la différence de pourcentage entre les verbes statiques et dynamiques soit le fait du hasard selon le test de Mann-Whitney U. Les données de la ligne *Corr Σ inp* indiquent le taux de corrélation entre les données de la colonne concernée et l'input au total selon le test de Spearman (sans les données de *trouver Σ* , sauf pour le test de rappel). Ce test compare les rangs, c'est-à-dire les échelles de pourcentages de formes courtes. Le traitement statistique des données a été présenté à la section 7.3.4.

Tableau 14-2 : Pourcentages de formes courtes par tâche au total

	input		apprenant/es			
	ESC	ΣINP	CORP	ΣNARR	ΣIMIT	ΣRAPP
adorer	86%	81%	100%	80%	82%	49%
penser	0%	73%	100%	88%	81%	66%
habiter	91%	72%	92%	100%	85%	50%
trouver S	100%	72%	100%	100%	70%	-
détester	50%	70%	100%	-	69%	60%
préférer	100%	58%	100%	-	66%	34%
aimer	93%	55%	100%	100%	96%	61%
trouver Σ	57%	47%	29%	67%	70%	68%
regarder	84%	39%	50%	50%	72%	36%
parler	80%	38%	73%	61%	60%	50%
manger	75%	29%	63%	63%	68%	57%
trouver D	50%	26%	17%	50%	70%	-
acheter	43%	15%	50%	27%	23%	36%
visiter	0%	8%	0%	20%	38%	42%
SvsD <i>p</i>	.05	.001	.001	.004	.038	ns
Corr Σinp <i>r</i>	.556	1	.862	.797	.679	.291

Légende : trouver S : sens statique ; trouver D : sens dynamique ; trouver Σ : trouver comme lexème ; SvsD : statique vs dynamique ; Corr Σinp : corrélation avec ΣINP ; ESC : manuel Escalade 1 ; ΣINP : input au total ; CORP : apprenant/es de corpus ; ΣNARR: tâche de narration ; ΣIMIT : tâche d'imitation ; ΣRAPP : tâche de rappel ; ns : non-significative ($p > .05$).

Avant de commenter les résultats des tests expérimentaux, il convient de rappeler que les formes courtes et longues ont été présentées le même nombre de fois pour tous les verbes dans le test d'imitation et de rappel. Toute déviance de 50% est donc intéressante.

Au niveau des verbes individuels, on peut voir que les apprenant/es ont surtout produit la forme courte avec les verbes *adorer*, *aimer*, *détester*, *habiter*, *penser* et *trouver* (statique), qui sont des verbes statiques et des verbes fréquents à la forme courte dans l'input. Les données du test de rappel montrent généralement davantage de valeurs proches de 50%, de sorte que les 49% pour *adorer* et 50% pour *habiter* ne correspondent pas clairement à une forme courte ; ces valeurs n'indiquent cependant pas non plus clairement une forme longue. Le verbe *préférer* est le verbe statique avec les plus basses valeurs de formes courtes dans le test d'imitation et de rappel. Malgré ces différences, la production de ces verbes semble suivre les deux hypothèses d'un point de vue qualitatif.

Les deux verbes dynamiques fréquents à la forme longue dans l'input, *acheter* et *visiter* ont été préférés à la forme longue dans la partie expérimentale, alors que le verbe *acheter* avait été produit à 50% à chaque forme par les apprenant/es

de corpus. On peut tout de même considérer que les verbes ont aussi été produits selon les deux hypothèses.

Finalement, les verbes dynamiques pour lesquels on trouve plus de variation dans l'input, les verbes *trouver* (au total), *parler* et *manger*, ont été préférés à la forme courte alors que la production du verbe *regarder* varie selon la tâche. La production de ces verbes semble donc plutôt suivre la fréquence de formes dans l'input que l'aspect lexical, ce qui est un résultat en faveur de l'hypothèse 2.

Les valeurs (en pourcent) de formes courtes des verbes entre les différentes tâches de la partie expérimentale sont fortement corrélées avec un r de .763 entre la narration et l'imitation, de .747 entre la narration et le rappel et finalement de .525 entre le rappel et l'imitation. Ces corrélations indiquent que les apprenant/es ont produits les verbes à une proportion de formes courtes similaire dans les différentes tâches, ce qui confirme l'analyse qualitative.

Statistiquement, on voit que les deux hypothèses se vérifient de moins en moins bien : elles se vérifient le mieux dans la tâche de production libre (ainsi que dans les données des apprenant/es de corpus) et le moins bien dans le test de rappel. Les données du test de rappel sont en effet caractérisées par un haut degré de formes longues en comparaison avec les autres tâches, la plupart des verbes ayant été produits à un pourcentage proche de 50%. Ce résultat est contraire à ce qui est attendu pour ce test où la forme courte devrait être plus facilement produite (Cowan *et al.*, 2003 ; Farrell, 2006), ce qui a été le cas pour les natifs comme on l'a vu dans le tableau 13-5. Dans ce test, l'influence du contexte de la phrase a été écartée. On peut interpréter les résultats du test de rappel en disant que les deux formes verbales sont stockées en mémoire chez les apprenant/es, mais que la production des phrases implique une plus forte charge cognitive. Ce n'est qu'au moment de produire des phrases que les apprenant/es ont recours à la forme la plus accessible.

Les données de la partie expérimentale ont encore une fois montré combien les deux hypothèses étaient liées, l'influence de l'aspect lexical est difficile à distinguer de la fréquence pour les verbes statiques et les verbes dynamiques *acheter* et *visiter*. Par contre, la production des verbes dynamiques, *regarder*, *parler* et *manger*, moins fréquents à la forme longue dans l'input, semble influencée par la fréquence des formes dans l'input : ces verbes n'ont pas été préférés à la forme longue par les apprenant/es comme attendu pour les verbes dynamiques selon l'hypothèse 1. La production de ces verbes semble proche de la fréquence du manuel *Escalade 1*, ces verbes étant à la forme courte à près de 80% dans ce manuel, comme on le voit dans le tableau 14-2. Les données du

verbe *trouver* montrent aussi une préférence pour la forme courte, même lorsqu'il est produit dans un contexte dynamique.

La possibilité que les verbes *regarder*, *parler* et *manger* soient plus souvent utilisés comme des verbes statiques en classe de langue, par exemple pour décrire les habitudes de chacun, a été évoquée plusieurs fois. Mais, le fait que ces verbes soient préférés à la forme courte parle en faveur d'une influence de la fréquence et non d'une influence de l'aspect lexical. En effet, d'après Andersen (1990, 2002, voir aussi Noyau, 1990), les apprenant/es adultes cherchent dans la langue cible les formes qui correspondent le mieux à l'aspect lexical – la forme prototypique, pertinente, congruente – du verbe. Les verbes *regarder*, *parler* et *manger* sont clairement des verbes d'activités et donc des verbes dynamiques.

Les données au total semblent donc plutôt parler en faveur de l'influence de la fréquence des formes sur la production des apprenant/es.

On peut maintenant se demander si l'influence de l'aspect lexical ou de la fréquence des formes sont visibles dès le début de l'acquisition du français. À cet effet, le tableau 14-3 présente les pourcentages de formes courtes de manière similaire au tableau 14-2, avec la différence que les données de la partie expérimentale sont présentées par stades, le stade 1 étant le plus débutant et le stade 3 le plus avancé (Bartning & Schlyter, 2004 ; cf. section 3.2. et l'annexe 1). Cette fois-ci, les verbes ont été regroupés par aspect lexical et les données du verbe *trouver* au total (comme lexème) ont été mises à la fin. Le nombre de données par stade étant faible pour la tâche de narration, aucun calcul statistique n'a été fait. (Le tableau 14-3 se trouve sur la page suivante.)

Tableau 14-3 : Pourcentages de formes courtes par aspect lexical et par stades

input			apprenant/es									
			corpus	narration			imitation			rappel		
ESC	ΣINP			1	2	3	1	2	3	1	2	3
adorer	86%	81%	100%	100%	100%	50%	100%	75%	83%	43%	55%	48%
aimer	93%	55%	100%	100%	100%	100%	100%	89%	100%	69%	60%	58%
détester	50%	70%	100%	-	-	-	67%	71%	69%	71%	48%	65%
habiter	91%	72%	92%	-	100%	-	100%	83%	79%	59%	43%	50%
penser	0%	73%	100%	0%	100%	100%	0%	100%	82%	80%	67%	59%
préférer	100%	58%	100%	-	-	-	69%	70%	59%	29%	31%	40%
trouver S	100%	72%	100%	-	-	100%	0%	43%	92%	-	-	-
<i>statique</i>			97%	60%	100%	95%	76%	75%	78%	56%	50%	53%
acheter	43%	15%	50%	0%	40%	22%	0%	27%	29%	44%	40%	27%
manger	75%	29%	63%	50%	63%	71%	100%	40%	79%	57%	51%	64%
parler	80%	38%	73%	25%	73%	63%	22%	82%	56%	46%	59%	44%
regarder	84%	39%	50%	50%	60%	0%	67%	77%	69%	53%	20%	40%
trouver D	50%	26%	17%	100%	-	33%	100%	89%	50%	-	-	-
visiter	0%	8%	0%	-	50%	0%	43%	45%	28%	38%	45%	42%
<i>dynamique</i>			63%	44%	62%	46%	51%	59%	51%	48%	43%	45%
trouver Σ	57%	47%	29%	100%	-	60%	50%	69%	73%	68%	71%	63%
SvsD <i>p</i>	.05	.001	.001	-	-	-	ns	ns	.015	ns	ns	ns
Corr Σinp <i>r</i>	.556	1	.862	-	-	-	.063	.344	.760	.371	.245	.347

Légende : voir tableau 14-2.

Les données du tableau 14-3 montrent que

- la différence entre les verbes statiques et dynamiques au niveau des pourcentages de formes courtes est significative uniquement au stade 3 dans le test d'imitation. La différence à ce stade semble aussi la plus claire pour les données de la tâche de narration, avec 95% de formes courtes avec les verbes statiques et 46% avec les verbes dynamiques.
- la différence entre les verbes statiques et dynamiques semble la plus faible pour les apprenant/es du stade 2 qui ont produit un grand nombre de verbes à la forme courte ;
- les apprenant/es du stade 1 semblent préférer la forme courte avec les verbes statiques et la forme longue avec les verbes dynamiques, bien que l'on note beaucoup de valeurs extrêmes, 0% et 100%.

L'influence de l'aspect lexical (hypothèse 1) ne semble donc pas, selon les données analysées ici, se refléter dans la production des apprenant/es les plus débutants, comme l'ont proposé Perdue *et al.* (2002 ; cf. section 4.2). Il semble que l'influence de l'aspect lexical commence seulement à se voir au stade 3, donc après que les apprenant/es ont eu une certaine expérience de la langue cible

et ainsi l'occasion de noter les formes correspondant à l'aspect lexical prototypique des verbes.

L'influence de la fréquence des formes dans l'input (hypothèse 2) peut notamment se voir au niveau du taux de corrélation obtenu (*Corr Σ_{inp}* dans le tableau 14-3).

- Cette corrélation est toujours bonne au stade 3 et surtout avec le test d'imitation. Ce résultat pourrait indiquer que l'expérience permet aux apprenant/es d'être plus proches de ce que l'on trouve dans la langue cible.
- La corrélation pour le stade 1 dans le test d'imitation est extrêmement faible. Comme on l'a vu dans la figure 12-1, l'absence de corrélation est surtout due à la production des verbes *manger*, *trouver* (*dynamique*), *penser* et *trouver* (*statique*). Les verbes *trouver* et *penser* sont des verbes peu connus des apprenant/es comme l'avait montré la tâche de traduction des verbes (*cf.* section 10.2). La corrélation avec l'input au total sans les verbes *trouver* et *penser* monte à .639, ce qui signifie que l'influence de la fréquence des formes est visible chez les apprenant/es les plus débutants pour les verbes dont ils ont une certaine connaissance.
- Finalement, la corrélation pour les apprenant/es du stade 2 au niveau du test d'imitation est relativement faible. Plusieurs verbes dynamiques ont en effet été produits à la forme courte à un haut pourcentage. Ce résultat peut s'interpréter comme étant le fruit de la fréquence répétée de la forme courte dans la langue à laquelle ils sont exposés, langue qui pourrait principalement correspondre aux données du manuel *Escalade 1* (*cf.* section 5.1.4. sur l'input et l'usage de la langue).

Pour terminer cette synthèse sur la partie expérimentale, voici une présentation condensée de l'analyse de quelques outputs du test d'imitation (section 12.4.4.).

L'exemple (14-4) présente les deux stimuli pour le verbe *habiter* avec les outputs d'une apprenant/e de chaque stade, la même personne pour les deux outputs (pour le codage voir section 12.3.2.).

(14-4)	Stim	9	En France	Carl	habite	au bord de la mer	
1	P15	9		Carl	habite	près du mer	C>C
2	P24	9	A la France	Carl	habite	près de la mer	C>C
3	P06	9	En	Carl	habite	bord de la mer	C>C
	Stim	10	En mars	Marie	veut	habiter	à Paris
1	P15	10	En mars	nâgon	habite	de Paris	L>C
2	P24	10	A mars	Carl	habite	à Paris	L>C
3	P06	10	En mars	Carine	va	habite	à Paris
							L>C

On constate que la forme courte a été produite à tous les stades et que la seule différence réside dans le fait que l'apprenante du stade 3 a aussi su produire la construction syntaxique de l'infinitif. Ce résultat parle en faveur des deux hypothèses.

L'exemple (14-5) ci-après présente les stimuli et outputs pour le verbe *acheter*.

(14-5)	Stim	2	Aujourd'hui	Christine	veut	acheter	une robe	
1	P10	2		Christi		achetE	un robe	L>L
2	P16	2	Aujourd'hui	Marie		achetE	une robe	L>L
3	P08	2		Catherine	va	acheter	une robe	L>L
	Stim	1	Au supermarché	Carl		achète	un melon	
1	P10	1	En supermarché			achetE	un	C>L
2	P16	1	A supermarché	Carl		achetE	un melon	C>L
3	P08	1	A supermarché	Carl		achète	un melon	C>C

Cet exemple montre que les apprenant/es des stades 1 et 2 ont systématiquement produit la forme longue, alors que l'apprenante du stade 3 a pu imiter la structure du stimulus à l'infinitif, la forme étant celle attendue par le contexte cette fois-ci, mais elle a également su produire la forme courte après le stimulus au présent. Ce résultat correspond également à ce qui est attendu par les deux hypothèses.

La principale différence au niveau des outputs pour les verbes *habiter* et *acheter* est le nombre de formes rappelées : 21 outputs pour les stimuli *habite* et *acheter*, 18 pour *habiter* et seulement 14 pour *achète* sur un maximum possible de 33. Ce résultat implique que la forme la plus fréquente est non seulement la forme la plus accessible en production, mais aussi en compréhension, ce qui correspond bien aux propositions de Gass & Mackey (2002 ; 2007) et de VanPatten (2002), que la production des formes sensibilise les apprenant/es à un nouvel input.

Une autre différence au niveau du résultat de ces deux verbes est la proportion de constructions de l'infinitif dans l'imitation des stimuli à l'infinitif : pour le verbe *habiter*, on a 5 outputs sur 18 (28%), alors que pour le verbe *acheter* on en obtient 9 sur 21 (43%). La production de la construction de l'infinitif semble donc plus aisée avec le verbe *acheter*, un verbe fréquent à la forme longue dans l'input, et donc dans des constructions *auxiliaire + verbe en -E*. Ce résultat confirme la proposition de Bybee au niveau de l'influence de la fréquence type : “[...] a certain degree of type frequency is needed to uncover the structure of words and phrases” (Bybee, 2008 : 221) ainsi que celle d'Ellis (2002) sur la sensibilité des apprenant/es aux régularités de l'input. Ceci laisse à penser que l'auxiliaire est plus saillant avec des verbes tels que *acheter*, fréquents à la forme longue, les formes de ces verbes étant rarement présentées sans être précédées par un auxiliaire. Une étude sur le développement de la syntaxe à partir de la fréquence des constructions pourrait faire l'objet d'une étude complémentaire à celle-ci.

Finalement, l'exemple (14-6) présente les résultats du verbe *parler*, un verbe qui était attendu à la forme longue d'après l'hypothèse 1 sur l'influence de l'aspect lexical prototypique et à une variation des formes (comparé aux verbes des exemples ci-dessus) d'après l'hypothèse 2 sur la fréquence des formes, ce verbe étant fréquent aux deux formes dans l'input.

(14-6)	Stim	14	Maintenant	Marie	veut	parler	avec son chef	
1	P31	14				parlE	en su chef	L>L
2	P19	14	Maintenant	Marie		parle	avec son chef	L>C
3	P26	14	Maintenant	Marie	va	parle	avec son chef	L>C
3	P08	14	Maintenant	Carine	va	parler	avec son chef	L>L
	Stim	13	Au travail	Carl		parle	avec une collègue	
1	P31	13				parlE	dans une copain	C>L
2	P19	13	Au travail	Carl		parle	avec une collègue	C>C
3	P26	13	Au travail	Carl		parle	au une collègue	C>C
3	P08	13	Au travail	Carl		parle	avec une collègue	C>C

On relève que l'apprenante P31 du stade 1 a uniquement produit une forme verbale et un peu de matériel supplémentaire. Bien que cette manière de produire ne se voie pas dans les exemples cités ici, il s'agit d'une production assez caractéristique des apprenant/es du stade 1. Ce genre de production confirme le fait que les apprenant/es en tout début d'acquisition centrent avant tout leur attention sur ce qui est porteur de contenu comme on l'a vu à la section 5.2.1. (Schmidt, 2001 ; VanPatten, 2002 ; 2007). Le fait que le verbe *parler* ait surtout été produit à la forme longue au stade 1 pourrait être un argument pour

une influence de l'aspect lexical. Mais, contrairement aux verbes *habiter* et *acheter*, ce verbe a été produit aux deux formes après le stimulus à la forme courte comme on l'a vu dans l'exemple (12-10), une apprenante ayant même produit le verbe aux deux formes. On pourrait donc aussi dire que ce verbe a été produit avec un peu plus de variation que les autres. L'apprenante du stade 2 a produit les deux phrases au présent et à la forme courte, ce qui n'est pas attendu par l'hypothèse sur l'influence de l'aspect lexical. On s'attendait à la forme longue au moins après le stimulus à la forme longue. Finalement, pour le stade 3, on a deux cas : soit la construction syntaxique est imitée correctement, mais le verbe est toujours à la forme courte, soit le verbe est produit comme dans les stimuli avec la même construction et la forme attendue par la construction.

Le verbe *parler* a donc été produit avec un peu plus de variation que les verbes *habiter* et *acheter*, ce qui parle plutôt en faveur de l'influence de la fréquence des formes dans l'input.

Finalement, les données du test d'imitation ont donc aussi révélé un développement dans l'acquisition du français :

- les apprenant/es du stade 1 produisent des phrases avec un verbe à une forme courte ou longue et un peu de matériel supplémentaire
- les apprenant/es du stade 2 produisent des constructions du type *sujet + verbe* où le verbe est de préférence à la forme courte sauf pour les verbes clairement fréquents à la forme longue dans l'input
- les apprenant/es du stade 3 produisent en principe correctement la construction du présent, mais produisent volontiers la forme la plus fréquente du verbe dans la construction de l'infinitif.

Ces résultats correspondent à la proposition de Bartning & Schlyter (2004) qui avait été présentée à la section 3.2. dans la mesure où le développement peut se voir dans la production de plus en correcte du contexte du présent. Mais, les données montrent aussi que le développement de la construction de l'infinitif semble plus tardif pour des apprenant/es soumis à un input massif de la forme courte du présent comme on peut le présupposer ici. Ces données confirment aussi les recherches faites à partir d'une approche fonctionnelle (section 3.1.) évoquant que les apprenant/es ont recours à des formes de base en début d'acquisition et qu'une utilisation de la morphologie grammaticale demande une certaine expérience de la langue.

Les données des apprenant/es du stade 3 semblent aussi montrer deux effets de la fréquence type : d'une part au niveau d'une certaine influence de l'aspect lexical et d'autre part au niveau de la mise en place de la construction de

l'infinitif. L'effet de la fréquence d'occurrence pourrait donc avoir perdu de l'importance à ce stade pour les items les plus fréquents.

Ce travail a donc montré combien l'aspect lexical et la fréquence des formes étaient liés. Pourtant, bien que les données aient plusieurs fois montré des différences significatives au niveau du pourcentage de formes courtes entre les verbes statiques et dynamiques, on a pu voir que les apprenant/es avaient une préférence pour la forme courte avec les verbes fréquents à la forme courte dans l'input et une préférence pour la forme longue avec les verbes clairement fréquents à la forme longue dans l'input. Bien que ces deux groupes de verbes se laissent catégoriser selon l'aspect lexical, on a aussi pu voir que le petit groupe de verbes dynamiques moins clairement fréquents à la forme longue dans l'input n'était pas préféré à la forme longue par les apprenant/es comme attendu par l'hypothèse 1 sur l'influence de l'aspect lexical. Ces résultats ont été interprétés comme étant le fruit de la fréquence des formes dans l'input, comme proposé par l'hypothèse 2 dans ce travail. On a aussi pu voir une possible influence de l'aspect lexical pour les apprenant/es de niveau intermédiaire (stade 3), un stade auquel les apprenant/es commencent aussi à produire la construction de l'infinitif. Un tel résultat pourrait être le fruit des effets de la fréquence type de cette construction plutôt que de la fréquence d'occurrence des verbes isolés.

On peut donc conclure que les analyses présentées dans ce travail montrent que la production des formes verbales par les apprenant/es débutant/es est surtout influencée par la fréquence des formes dans l'input.

Références bibliographiques

Ågren, M. (2008), *À la recherche de la morphologie silencieuse. Sur le développement du pluriel en français L2 écrit*, Thèse de doctorat, Université de Lund.

Allen, R. & Hulme, C. (2006), « Speech and language processing mechanisms in verbal serial recall », *Journal of Memory and Language*, 55, 64-88.

Andersen, R.W. (1990), « Models, processes, principles and strategies : second language acquisition inside and outside the classroom », in B. VanPatten & J.F. Lee (éds), *Second language acquisition / Foreign language learning*, Clevedon : Multilingual Matters, 45-68.

Andersen, R.W. (1991), « Developmental sequences: the emergence of aspect marking in second language acquisition », in T. Huebner & C. Ferguson (éds), *Crosscurrents in second language acquisition and linguistic theories*, Amsterdam/Philadelphia : John Benjamins, 305-324.

Andersen, R.W. (1993), « Four operating principles and input distribution as explanations for underdeveloped and mature morphological systems », in K. Hyltenstam & Å. Viberg (éds.), *Progression and regression in language. Sociocultural, neuropsychological, and linguistic perspectives*, Cambridge : Cambridge University Press, 309-339.

Andersen, R. (2002), « The dimensions of 'pastness' », in R. Salaberry & Y. Shirai (éds), *The L2 acquisition of tense-aspect morphology*, Amsterdam/Philadelphia : John Benjamins, 79-105.

Anderson, B. (2007), « Pedagogical rules and their relationship to frequency in the input : observational and empirical data from L2 French », *Applied Linguistics*, 28-2, 286-308.

Baddeley, A. (2000), « The episodic buffer : a new component of working memory ? », *Trends in cognitive sciences*, 4-11, 417-423.

Baddeley, A.D., Gathercole, S., Papagno, C. (1998), « The phonological loop as a language learning device », *Psychological Review* 105-1, 158-173.

Bardovi-Harlig, K. (1987), « Markedness and salience in second-language acquisition », *Language Learning*, 37-3, 385-407.

Bardovi-Harlig, K. (2000), « Tense and aspect in second language acquisition : form, meaning, and use », *Language Learning*, 50-s1.

Bardovi-Harlig, K. (2002), « Analyzing aspect », in R. Salaberry & Y. Shirai (éds), *The L2 acquisition of tense-aspect morphology*, Amsterdam/Philadelphia : John Benjamins, 129-154.

Bartning, I. (1998), « Procédés de grammaticalisation dans l'acquisition des prédications verbales en français parlé », *Travaux de Linguistique*, 36, 223-234.

Bartning, I. & Schlyter, S. (2004), « Itinéraires acquisitionnels et stades de développement en français L2 », *Journal of French Language Studies*, 14, 281-299.

Bergström, A. (1997), « L'influence des distinctions aspectuelles sur l'acquisition des temps en français langue étrangère », *AILE*, 9, 51-82.

Billières, M. & Spanghero-Gaillard, N. (2005), « Pour une prise en compte des faits cognitifs dans l'enseignement des langues étrangères », in *Actes des journées d'études de l'AFVT : méthodologie verbo-tonale et apprentissages*, Nantes, mars 2004, 45, 19-27.

Bley-Vroman, R. & Chaudron C. (1994), « Elicited imitation as a measure of second-language competence », in: E.E. Tarone, S.M. Gass, A.D. Cohen, *Research methodology in second-language acquisition*, Hillsdale : Lawrence Erlbaum Associates, 245-261.

Bonin, P. (2002), *Production verbale de mots. Approche cognitive*, Bruxelles : DeBoeck Université.

Bonnotte, I. (2008), « The role of semantic features in verb processing », *Journal of Psycholinguistic Research*, 37, 199-217.

Bonnotte, I. & Fayol, M. (2000), « L'imparfait : genèse et utilisation. Approche de psycholinguistique développementale », in J.-E. Tyvaert (éd), *L'imparfait. Philologie électronique et assistance à l'interprétation des textes*, Reims : Presses Universitaires de Reims, 15-34.

Bonnotte, I. & Fayol, M. (2001), « Gestion de relations sémantiques entre locutions temporelles, types de procès et temps grammaticaux. Étude développementale en production et en compréhension », in F. Cordier, J. François & B. Victorri (éds), *Syntaxe et Sémantique 2 : Sémantique du lexique verbal*, Caen : Presses Universitaires de Caen, 179-201.

Bonnotte, I., Fayol, M., Gombert, J.E. (1991), « La représentation cognitive des verbes. Approche descriptive et développementale », *Travaux de linguistique et de philologie*, 29, 209-225.

Bonnotte, I., Kaifer, A., Fayol, M., Idiazabal- Gorrochategui, I. (1993), « Rôle des types de procès et du co-texte dans l'emploi des formes verbales de la narration. Étude de productions d'adultes en français, en castillan et en basque », *Langue Française*, 97, 81-101.

Bozier, C. (2005), *La sollicitation dans l'interaction exolingue en français*, Thèse de doctorat, Université de Lund.

Bronckart, J.P. & Sinclair, H. (1973), « Time, tense and aspect », *Cognition*, 2-1, 107-130.

Bybee, J.L. (1985), *Morphology. A study of the relation between meaning and form*, Amsterdam/Philadelphia : John Benjamins.

Bybee, J. (1991), « Natural morphology: the organization of paradigms and language acquisition », in T. Huebner & C. Ferguson (éds), *Crosscurrents in second language acquisition and linguistic theories*, Amsterdam/Philadelphia : John Benjamins, 67-91.

Bybee, J. (1995), « Regular morphology and the lexicon », *Language and cognitive processes*, 10-5, 425-455.

Bybee, J. (2001), *Phonology and language use*, Cambridge : Cambridge University Press.

Bybee, J. (2006), « From usage to grammar : the mind's response to repetition », *Language*, 82-4, 711-733.

Bybee, J. (2008), « Usage-based grammar and second language acquisition », in P. Robinson & N.C. Ellis, *Handbook of cognitive linguistics and second language acquisition*, New-York : Routledge, 216-236.

Bybee, J.L. & Dahl, Ö. (1989), « The creation of tense and aspect systems in the languages of the world », *Studies in Language*, 13, 51-103.

Bybee, J. & Thompson S. (2000), « Three frequency effects in syntax », *Berkeley Linguistic Society*, 23, 65-85.

Carroll, S.E. (2001), *Input and evidence. The raw material of second language acquisition*, Amsterdam/Philadelphia : John Benjamins.

Carroll, S.E. (2004), « Segmentation : learning how to 'hear words' in the L2 speech stream », *Transactions of the Philological Society*, 102-2, 227-254.

Chomsky, N. (1995), *The minimalist program*, Cambridge (MA) : MIT Press.

Christensen, L. (à paraître), « Vendler revisited from a Swedish point of departure ».

Clahsen, H. & Felser, C. (2006), « Grammatical processing in language learners », *Applied Psycholinguistics*, 27, 3-42.

C-ORAL-ROM (2005), *Integrated reference corpora for spoken romance languages*, E. Cresti & M. Moneglia (éds), Amsterdam/Philadelphia : John Benjamins.

Corder, S.P. (1967), « The significance of learner's errors », *IRAL*, 5-4 : 161-170.

Cowan, N., Baddeley, A.D., Elliott, E., Norris, J. (2003), « List composition and the word length effect in immediate recall: a comparison of localist and globalist assumptions », *Psychonomic Bulletin & Review*, 10-1, 74-79.

David, J., Guyon, O., Brissaud, C. (2006), « Apprendre à orthographier les verbes : le cas de l'homophonie des finales en /E/ », *Langue Française*, 151, 109-126.

DeSwart, H. (1998), « Aspect shift and coercion », *Natural Language and Linguistic Theory*, 16, 347-385.

Defays, J.-M., Deltour S. & Thonard, A. (2007), « Les compétences syntaxiques et lexicales dans les notions de fluidité, de complexité et d'exactitude », in S. van Daele, A. Housen, F. Kuiken, M. Pierrard & I. Vedder (éds), *Complexity, accuracy and fluency in second language use, learning & teaching*, Bruxelles : KVAB, 43-52.

Dietrich, R., Klein, W. & Noyau, C. (éds) (1995), *The acquisition of temporality in a second language*, Amsterdam/Philadelphia : John Benjamins.

Dostie, G. (2004), *Pragmaticalisation et marqueurs discursifs. Analyse sémantique. Analyse sémantique et traitement lexicographique*, Bruxelles : DeBoeck/Duculot.

Duperron, L. (2003), *The role of focused input on the acquisition of tense and aspect by L2 learners of French*, Thèse de doctorat, University of Illinois at Urbana-Champaign.

Ellis, N.C. (2001), « Memory for language », in P. Robinson, *Cognition and second language instruction*, Cambridge : Cambridge University Press, 33-68.

Ellis, N.C. (2002), « Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition », *Studies in Second Language Acquisition*, 24, 143-188.

Ellis, N.C. (2003), « Constructions, chunking, and connectionism : the emergence of second language structure », in C.J. Doughty & M.H. Long, *The handbook of second language acquisition*, Oxford: Blackwell, 63-103.

Ellis, N.C. (2008), « Usage-based and form-focused language acquisition », in P. Robinson & N.C. Ellis, *Handbook of cognitive linguistics and second language acquisition*, New-York : Routledge, 372-405.

Ellis, R. (1994), *The study of second language acquisition*, Oxford : Oxford University Press.

Ellis, R. (2005), « Measuring implicit and explicit knowledge of a second language. A psychometric study », *Studies in Second Language Acquisition*, 27, 141-172.

Ellis, R. (2008), « Investigating grammatical difficulty in second language learning : implications for second language acquisition research and language testing », *International Journal of Applied Linguistics*, 18-1, 4-22.

Ellis, R. & Barkhuizen, G. (2005), *Analysing learner language*, Oxford : Oxford University Press.

Erlam, R. (2006), « Elicited imitation as a measure of L2 implicit knowledge : an empirical validation study », *Applied Linguistics*, 27-3, 464-491.

Farrell, S. (2006), « Mixed-list phonological similarity effects in delayed serial recall », *Journal of Memory and Language*, 55, 587-600.

Ferdinand, A. (1996), *The development of functional categories : The acquisition of the subject in French*, The Hague : Holland Academic Graphics.

Florby, G. & Fält, E. (1998), *Grammatisk ABC. Grammatiska termer och begrepp för effektiva språkstudier*, Uppsala : Scripta textproduction.

Flyman Mattsson, A. (2003), *Teaching, learning, and student output. A study of French in the classroom*, Thèse de doctorat, Université de Lund.

Fournet, N., Juphard, A., Monnier, C., Roulin, J.-L. (2003), « Phonological similarity in free and serial recall : the effet of increasing retention intervals », *International Journal of Psychology*, 38-6, 384-389.

Freed, B., Segalowitz, N. & Dewey, D.P. (2004), « Context of learning and second language fluency in French. Comparing regular classroom, study abroad, and intensive domestic immersion programs », *Studies in Second Language Acquisition*, 26, 275-301.

Gaonac'h, D. & Larigauderie, P. (2000), *Mémoire et fonctionnement cognitif. La mémoire de travail*, Paris : Armand Colin.

Gallimore, R. & Tharp, R.G. (1981), « The interpretation of elicited sentence imitation in a standardized context », *Language Learning*, 31-2, 369-392.

Gass, S.M. & Mackey, A. (2002), « Frequency effects and second language acquisition. A complex picture ?, *Studies in Second Language Acquisition* 24, 249-260.

Gass, S.M. & Mackey, A. (2007), « Input, interaction, and output in second language acquisition », in B. VanPatten & J. Williams (éds), *Theories in second language acquisition. An introduction*, Mahwah (NJ) : Lawrence Erlbaum Associates, 175-199.

Gathercole, S.E. & Baddeley, A.D. (1993), *Working memory and language*, Hove : Psychology Press.

Gathercole, S.E. (1995), « Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords », *Memory and Cognition*, 23-1, 83-94.

Gathercole, S.E., Frankish, C.R., Pickering, S.J. & Peaker, S. (1999), « Phonotactic influences on short-term memory », *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25-1, 84-95.

Giacalone-Ramat, A. (2002), « How do learners acquire the classical three categories of temporality », in R. Salaberry & Y. Shirai (éds), *The L2 acquisition of tense-aspect morphology*, Amsterdam/Philadelphia : John Benjamins, 221-247.

Gineste, M.-D., & Le Ny, J.-F. (2002), *Psychologie cognitive du langage. De la reconnaissance à la compréhension*, Paris : Dunod.

Goad, H., White, L. & Steele, J. (2003), « Missing inflection in L2 acquisition : defective syntax or L1-constrained prosodic representations ? », *Canadian Journal of Linguistics / Revue canadienne de linguistique*, 48-3/4, 243-263.

Goldschneider, J.M. & DeKeyser, R. (2001), « Explaining the “Natural order of L2 morpheme acquisition” in English : a meta-analysis of multiple determinants », *Language Learning*, 51-1, 1-50.

Granfeldt, J. (2003), *L'acquisition des catégories fonctionnelles. Étude comparative du développement du DP français chez des enfants et des apprenants adultes*, Thèse de doctorat, Université de Lund.

Granfeldt, J. (2005), « Direkt Profil et deux études sur la morphologie verbale et les stades de développement », in J. Granfeldt et S. Schlyter (éds.) *Acquisition et production de la morphologie flexionnelle. Actes du « Festival de la morphologie »*, mars 2005 à Lund, PERLES n° 20, Université de Lund, 65-85.

Granfeldt, J. (2007), « Speaking and writing in L2 French : exploring effects on fluency, complexity and accuracy », in S. van Daele, A. Housen, F. Kuiken, M. Pierrard & I. Vedder (éds), *Complexity, accuracy and fluency in second language use, learning & teaching*, Bruxelles : KVAB, 87-98.

Granfeldt, J., Nugues, P., Persson E., Persson, L., Kostadinov, F., Ågren M., Schlyter, S. (2005) « Direkt profil : un système d'évaluation de textes d'élèves de français langue étrangère fondé sur les itinéraires d'acquisition », in *Actes de la conférence Traitement Automatique des Langues Naturelles, TALN & RECITAL 2005, Tome 1 – Conférences principales*, Dourdan, France, 6-10 juin 2005, 113–122.

Granfeldt, J. & Nugues, P. (2007). « Évaluation des stades de développement en français langue étrangère », in N. Hathout and P. Muller (éds), *Actes de la 14e conférence sur le Traitement Automatique des Langues Naturelles (communications orales)*, volume 1, 357-366.

Granget, C. (2004), « L'acquisition par des apprenants germanophones de la référence au passé en français : un cas d'école », Thèse de doctorat, Université de Paris III – Sorbonne Nouvelle.

Gretsch, P. & Perdue, C. (2007), « Finiteness in first and second language acquisition », in I. Nikolaeva (éd), *Finiteness. Theoretical and empirical foundations*, Oxford : Oxford University Press, 432-484.

Gunnarsson, C. (2006), *Fluidité, complexité et morphosyntaxe dans la production écrite en FLE*, Thèse de doctorat, Université de Lund.

Heiman, G.W. (2006), *Basic statistics for the behavioral sciences*, Boston : Houghton Mifflin.

Herschensohn, J. (2000), *The second time around. Minimalism and L2 acquisition*, Amsterdam/Philadelphia : John Benjamins.

Herschensohn, J. (2001), « Missing inflection in second language French: accidental infinitives and other verbal deficits », *Second Language Research*, 17-3, 273-305.

Hilton, H. (2006), « La compétence lexicale en production orale », in *Actes du colloque international Recherches en acquisition et en didactique des langues étrangères et secondes*, publication en ligne (page consultée le 23 août 2009), <http://www.groupeelca.org/h/colloque2006/actespdf/hilton.pdf>.

Hoekstra, T. & Hyams, N. (1998), « Aspects of root infinitives », *Lingua*, 106, 81-112.

Housen, A. (1994), « Tense and aspect in second language acquisition : the Dutch interlanguage of a native speaker of English », in C. Vet & C. Vettters, *Tense and aspect in discourse*, Berlin : Mouton de Gruyter, 257-291.

Howard, M. (2004), « On the interactional effect of linguistic constraints on interlanguage variation : the case of past time marking », *IRAL*, 42, 319-334.

Hulstijn, J.H. (1997), « Second language acquisition research in the laboratory », *Studies in Second Language Acquisition*, 19, 131-143.

Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008), « Assessed levels of second language speaking proficiency : how distinct ?, *Applied Linguistics* 29, 1, 24-49.

Jaffré, J.P. & Brissaud, C. (éds) (2006), *Morphographie et homophones verbaux*, Langue Française 151, Paris : Larousse /Armand Colin.

Kaplan, M.A. (1987), « Developmental patterns of past tense acquisition among foreign language learners of French », in B. VanPatten, T.R. Dvorak, J.F. Lee (éds), *Foreign language learning. A research perspective*, Boston : Heinle & Heinle, 52-60.

Kemmer, S. & Barlow, M. (2000), « Introduction : a usage-based conception of language », in M. Barlow & S. Kemmer, *Usage-based models of language*, Stanford (CA) : CSLI Publications, vii-xxviii.

Kihlstedt, M. (1998), *La référence au passé dans le dialogue. Étude de l'acquisition de la temporalité chez des apprenants dits avancés de français*, Thèse de doctorat, Université de Stockholm.

Klein, W. (1989), *L'acquisition de langue étrangère*, Paris : Armand Colin.

Klein, W. (1995), « Frame of analysis », in R. Dietrich, W. Klein & C. Noyau (éds), *The acquisition of temporality in a second language*, Amsterdam/Philadelphia : John Benjamins, 17-29.

Klein, W., Dietrich R., & Noyau C. (1995), « Conclusions » in R. Dietrich, W. Klein & C. Noyau (éds), *The acquisition of temporality in a second language*, Amsterdam/Philadelphia : John Benjamins, 261-280.

Klein, W. & Perdue, C. (1997), The Basic Variety (or : Couldn't natural languages be much simpler ?), *Second Language Research*, 13-4, 301-347.

Labeau, E. (2005), « Beyond the Aspect Hypothesis. Tense-aspect development in advanced L2 French », in S. Foster-Cohen, M. García Mayo, J. Cenoz (éds), *EUROSLA Yearbook 5*, Amsterdam/Philadelphia : John Benjamins, 77-101.

Labelle, M. (2005), « The acquisition of grammatical categories : the state of the art », in H. Cohen & C. Lefebvre, *Handbook of categorization in cognitive science*, Amsterdam : Elsevier, 433-457.

Lardiere, D. (1998), « Dissociating syntax from morphology in a divergent L2 end-state grammar », *Second Language Research*, 14-4, 359- 375.

MacWhinney, B. (2000), *The CHILDES project : tools for analyzing talk*, Mahwah (NJ) : Lawrence Erlbaum Associates.

Markman, B.R., Spilka, I.V. & Tucker, G.R. (1975), « The use of elicited imitation in search of an interim French grammar », *Language Learning*, 25-1, 31-41.

McDade, H., Simpson, M.A. & Lamb, D.E. (1982), « The use of elicited imitation as a measure of expressive grammar. A question of validity », *Journal of Speech and Hearing Disorders*, 47, 19-24.

Mehnert, U. (1998), « The effects of different lengths of time for planning on second language performance », *Studies in Second Language Acquisition*, 20, 83-108.

Meisel, J.M. (1997), « The acquisition of the syntax of negation in French and German contrasting first and second language development », *Second Language Research*, 13-3, 227-263.

Mobayyen, F. & de Almeida, R.G. (2005), « The influence of semantic and morphological complexity of verbs on sentence recall : implications for the nature of conceptual representation and category-specific deficits », *Brain and Cognition*, 57, 168-171.

Munnich E., Flynn, S & Martohardjono, G. (1994), « Elicited imitation and grammaticality judgement tasks : what they measure and how they relate to each other », in E.E. Tarone, S.M. Gass, A.D. Cohen, *Research methodology in second-language acquisition*, Hillsdale : Lawrence Erlbaum Associates, 227-243.

Naiman, N. (1974), « The use of elicited imitation in second language acquisition research », *Working Papers on Bilingualism*, 2, 1-37.

Noyau, C. (1990), « Structure conceptuelle, mise en texte et acquisition d'une langue étrangère, *Langages*, 25, 101-114.

Noyau, C., Houdaïfa E.-T., Vasseur M.-T., Véronique, D. (1995), « The acquisition of French », in R. Dietrich, W. Klein & C. Noyau (éds), *The Acquisition of temporality in a second language*, Amsterdam/Philadelphia : John Benjamins, 145-209.

Olive, T. & Piolat, A. (2005), « Le rôle de la mémoire de travail dans la production écrite de textes », *Psychologie française*, 50, 373-390.

Öman, M. & Hedman-Ekman, M. (1990), *A propos 2*, Stockholm : Liber.

Perdue, C. (éd) (1993), *Adult language acquisition: cross-linguistic perspectives. Vol II : the results*, Cambridge : Cambridge University Press.

Perdue, C., Benazzo, S. & Guiliano, P. (2002), « When finiteness gets marked : the relation between morphosyntactic development and use of scopal items in adult language acquisition », *Linguistics*, 40-4, 849-890.

Persson, U.-B. & Lindfors, E. (2002), *Un air de Paris*, Solna : Ekelund.

Persson, U.-B. & Lindfors, E. (2002), *Un air de France*, Solna : Ekelund.

Pinker, S. (1998), « Words and rules », *Lingua* 106, 219-242.

Platzack C. (1979), *The semantic interpretation of aspect and aktionsarten. A study of internal time reference in Swedish*, Dordrecht : Foris.

Platzack, C. (1998), *Svenskan inre grammatik – det minimalistiska programmet. En introduktion till modern generativ grammatik*, Lund : Studentlitteratur.

Platzack, C. (2004), « The Verb Phrase », in G. Josefsson, C. Platzack & G. Håkansson, *The acquisition of swedish grammar*, Amsterdam/Philadelphia : John Benjamins, 135-154.

Poirier, M. & Saint-Aubin, J. (1995), « Memory for related and unrelated words : further evidence on the influence of semantic factors in immediate serial recall », *The Quarterly Journal of Experimental Psychology*, 48A-2, 384-404.

Prévost, P. (2004), « Morphological variation in early adult second language French. A cross-sectional study », in S. Foster-Cohen, M. Sharwood Smith, A. Sorace, M. Ota (éds), *EUROSLA Yearbook 4*, Amsterdam/Philadelphia : John Benjamins, 147-175.

Prévost, P. (2007), « L'influence du paradigme verbal de la langue maternelle sur la variabilité morphologique en acquisition du français L2 par de apprenants adultes », *Journal of French Language Studies*, 17, 49-80.

Prévost, P. & White, L. (2000a), « Accounting for morphological variation in second language acquisition : truncation or missing inflection ? », in M.-A. Friedemann & L. Rizzi (éds), *The acquisition of syntax. Studies in comparative developmental linguistics*, Harlow : Longman, 202-235.

Prévost, P. & White, L. (2000b), « Missing surface inflection or impairment in second language acquisition ? Evidence from tense and agreement », *Second Language Research*, 16-2, 103-133.

Rast, R. (2008), *Foreign language input : initial processing*, Clevedon : Multilingual Matters.

Riegel, M., Pellat, J.-C, Rioul, R. (1994), *Grammaire méthodique du français*, Paris: PUF.

Robison, R.E. (1995), « The Aspect Hypothesis revisited : a cross-sectional study of tense and aspect marking in interlanguage », *Applied Linguistics*, 16-3, 344-370.

Rocca, S. (2002), « Lexical aspect in child second language acquisition of temporal morphology. A bi-directional study », in R. Salaberry & Y. Shirai (éds), *The L2 acquisition of tense-aspect morphology*, Amsterdam/Philadelphia : John Benjamins, 249-284.

Saint-Aubin, J. & LeBlanc, J. (2005), « Word frequency effects in immediate serial recall of pure and mixed lists : test of the associative link hypothesis », *Canadian Journal of Experimental Psychology*, 59-4, 219-227.

Salaberry, R. & Ayoun D. (2005), « The development of L2 tense-aspect in the romance languages », in D. Ayoun & R. Salaberry, *Tense and aspect in romance languages*, Amsterdam/Philadelphia : John Benjamins, 1-33.

Sanner, M.-L. & Wennberg, L. (2002), *Génial 1, allt-i-ett-bok*, Stockholm : Natur och Kultur.

Sasse, H.-J. (2002), « Recent activity in the theory of aspect : accomplishments, achievements, or just non-progressive state ? », *Linguistic Typology*, 199-271.

Schlyter, S. (2003), *Stades de développement en français L2. Exemples d'apprenants suédophones, guidés et non guidés, du "Corpus Lund"*, publication en ligne (page consultée le 23 août 2009), <http://luur.lub.lu.se/luur?func=downloadFile&fileOID=624010>.

Schmidt, R. (2001), « Attention », in P. Robinson, *Cognition and second language instruction*, Cambridge : Cambridge University Press, 3-32.

Segui, J. & Ferrand, L. (2002), « The role of the syllable in speech perception and production », in J. Durand & B. Laks (éds), *Phonetics, phonology and cognition*, Oxford : Oxford University Press, 151-167.

Service, E. (1992), « Phonology, working memory, and foreign-language learning », *The Quarterly Journal of Experimental Psychology*, 45A-1, 21-50.

Smith, C.S. (1991), *The parameter of aspect*, Dordrecht: Kluwer.

Speciale G., Ellis N.C. & Bywater T. (2004), « Phonological sequence learning and short-term memory store capacity determine second language vocabulary acquisition », *Applied Psycholinguistics*, 25, 293-321.

Stroh-Wollin, U. (1998), *Koncentrerad nusvensk formlära och syntax*, Lund : Studentlitteratur.

Tarone, E. (1998), « Research on interlanguage variation : implications for language testing », in L.F. Bachman & A.D. Cohen, *Interfaces between second language acquisition and language testing research*, Cambridge : Cambridge University Press, 71-89.

Tarone, E. (2002), « Frequency effects, noticing, and creativity. Factors in a variationist interlanguage framework », *Studies in Second Language Acquisition*, 24, 287-296.

Thomas, A. (2003), *Stades initiaux de l'acquisition du français. Exemples d'adultes suédophones*, Mémoire de licence, Université de Lund.

Thomas, Anita (2006), *Les formes du type je parle dans l'interlangue d'apprenants du FLE et la fréquence dans l'input*, PERLES, 23, Université de Lund.

Tillman, B., Winblad, M., Martin, S., Lönnerblad, V. (1997), *Escalade B*, Stockholm : Liber.

Tillman, B., Winblad, M., Martin, S., Lönnerblad, V., Pettersson, M. & Sandberg, E. (2001), *Escalade 1*, Stockholm : Liber.

Tillman, B., Winblad, M., Martin, S., Lönnerblad, V., Pettersson, M. & Sandberg, E. (2002), *Escalade 2*, Stockholm : Liber.

Towell, R., Hawkins, R. & Bazergui, N. (1996), « The development of fluency in advanced learners of French », *Applied Linguistics*, 17-1, 84-119.

VanPatten, B. (1996), *Input processing and grammar instruction in second language acquisition. Theory and research*, Norwood : Ablex.

VanPatten, B. (2002), « Processing instruction : an update », *Language Learning*, 52, 4, 755-803.

VanPatten, B. (2007), « Input processing in adult second language acquisition », in B. VanPatten & J. Williams (éds), *Theories in second language acquisition. An introduction*, Mahwah (NJ) : Lawrence Erlbaum Associates, 115-135.

Vendler, Z. (1957), « Verbs and times », *The philosophical Review*, 66-2, 143-160.

Véronis, J. (2002), *Fréquence des mots en français parlé*, <http://sites.univ-provence.fr/veronis/data/freq-oral.txt> (page consultée le 22 août 2009).

Vet, C. (1980), *Temps, aspects et adverbes de temps en français contemporain. Essai de sémantique formelle*, Genève : Droz.

Vinther, T. (2002), « Elicited imitation : a brief overview », *International Journal of Applied Linguistics*, 12-1, 54-73.

Weinrich, H. (1989), *Grammaire textuelle du français*, Paris : Didier/Hatier.

Weist, R.M. (2002), « The first language acquisition of tense and aspect : a review », in R. Salaberry & Y. Shirai (éds), *The L2 acquisition of tense-aspect morphology*, Amsterdam/Philadelphia : John Benjamins, 21-78.

White, L. (2003), *Second language acquisition and universal grammar*, Cambridge : Cambridge University Press.

Wiberg, E. (1997), *Il riferimento temporale nel dialogo. Un confronto tra giovani bilingui italo-svedesi e giovani monolingui romani*, Thèse de doctorat, Université de Lund.

Wolfe-Quintero, K., Inagaki, S. & Kim, H.-Y. (1998), *Second language development in writing : measures of fluency, accuracy & complexity*, Honolulu : University of Hawai'i Press.

Annexes

Index des annexes :

Annexe 1 : Description des stades 1 à 3 selon Bartning & Schlyter (2004)

Annexe 2 : Formes des verbes en *-er* dans l'input

Annexe 3 : Formes des verbes irréguliers dans l'input oral

Annexe 4 : Tâche de narration : verbes en *-er* par participant/es

Annexe 5 : Test d'imitation : données par participant/es

Annexe 6 : Test d'imitation : verbes par stade

Annexe 7 : Test de rappel : données des participant/es

Annexe 8 : Test de rappel : verbes par stade

Annexe 1 : Description des stades 1 à 3 selon Bartning & Schlyter (2004)

Stade 1 – le stade initial

Au stade initial l'apprenant se sert d'une structuration en grande partie nominale mais aussi de moyens grammaticaux. Il/elle emploie des 'formes finies courtes' et des formes non-finies, ce qui montre que cette opposition n'est pas maîtrisée. Il n'y a pas d'opposition entre les personnes des formes verbales. On trouve pourtant à ce stade l'article défini et indéfini ainsi que le pronom sujet *je*, bien que très souvent accentués et non élidés (Granfeldt et Schlyter, 2004). D'autres traits caractéristiques sont l'emploi de la négation Nég X (*non grand-lit*) et les formules telles que *je(ne)sais pas* et *je m'appelle*. L'apprenant utilise quelques rares formes du passé composé mais très peu de contextes du passé sont marqués. Il y a aussi un début d'emploi de connecteurs *et*, *mais* et *puis*.

Ce stade peut être comparé au *niveau prébasique* ainsi qu'à la transition vers la *variété basique* (ESF).

Stade 2 – le stade post-initial

Ce stade est caractérisé par l'apparition de quelques phénomènes grammaticaux, mais encore très variables.

La subordination simple introduite par *quand*, *parce que* et *qui*, *que* apparaît. Il y a apparition de la négation préverbale *ne* sans *pas*, à côté de la négation postverbale. L'apprenant commence à utiliser des formes verbales modales (suivies d'infinitif) et le futur périphrastique. Son emploi du PC est devenu plus productif. Ainsi on voit que

certain morphèmes grammaticaux libres apparaissent tôt. Quelques apprenants guidés utilisent parfois l'imparfait sous forme de *était* et *avait*. Les formes verbales non-finies dans un contexte fini sont encore fréquentes mais le nombre de 'formes finies courtes' augmente. L'accord sujet-verbe est marqué par l'opposition entre la 1^{ère} et 2^{ème} personne au singulier pour les verbes non-thématiques *être*, *avoir* ainsi que par *nous V-ons* en alternance avec **nous V*, sans désinence. Les pronoms objet sont généralement postposés. L'apprenant à ce stade recourt fréquemment à *c'est* en tant que constructeur d'énoncés et de structure passe-partout (Bartning, 1997b).

Une comparaison avec le projet ESF suggère que ce niveau correspond en partie à la *variété basique*: la plupart des énoncés contiennent un verbe, mais la morphologie de ces verbes est loin d'être établie.

Stade 3 – le stade intermédiaire

A ce stade l'apprenant semble avoir développé une interlangue plus systématique et régulière, bien que très simple encore. Cette interlangue contient des surextensions et des régularisations qui font qu'elle n'est pas toujours en accord avec les normes de la langue cible. L'apprenant utilise, cependant, la négation (*ne Vfin pas*) comme dans la langue cible. De plus, elle/il emploie le passé avec la plupart des références au passé marquées ainsi que le futur (périphrastique; le futur simple, dans des cas isolés) pour les références à l'avenir. Les pronoms objet sont placés devant le verbe lexical pour les temps composés et simples – souvent incorrectement après l'auxiliaire *est/a*. Il y a aussi un enrichissement de la subordination avec des subordonnées causales, temporelles, relatives, complétives et interrogatives. Il y a même quelques premiers contextes du subjonctif.

Les formes verbales non-finies dans des contextes finis existent encore mais se raréfient. *Nous V-ons* est marqué la plupart du temps. L'opposition entre la 3^{ème} personne du singulier et du pluriel pour *ont*, *sont*, etc commence à s'établir relativement bien (la moitié des cas environ) mais l'apprenant à ce stade continue à employer en alternance *ils *prendre*, *ils *prend* et quelques cas de *ils prennent*. (Bartning & Schlyter, 2004 : 293-295)

Annexe 2 : Formes des verbes en –er dans l’input

Pour éviter de surcharger le tableau, les occurrences des formes du présent singulier et 3^e du pluriel (*parle, parles, parlent*) sont représentés par la forme *parle* de même que les occurrences des formes de l’imparfait singulier et 3^e du pluriel (*parlais, parlait, parlaient*) sont représentées par la forme *parlais*.

Le calcul de la fréquence de chaque forme a été obtenu en divisant le nombre d’occurrences de la forme en question par la somme des formes courtes et longues produites pour ce verbe (colonne *occ*).

	occ	adorer		adoré		adorais		adorez		adore	
INT-C	0	-	-	-	-	-	-	-	-	-	-
INT-A	0	-	-	-	-	-	-	-	-	-	-
LA	0	-	-	-	-	-	-	-	-	-	-
LB	0	-	-	-	-	-	-	-	-	-	-
LC	0	-	-	-	-	-	-	-	-	-	-
C-oral	11	0%	0	0%	0	18%	2	9%	1	73%	8
ESC	21	14%	3	0%	0	0%	0	0%	0	86%	18
Σ	32	9%	3	0%	0	6%	2	3%	1	81%	26
	occ	acheter		acheté		achetais		achetez		achète	
INT-C	4	50%	2	0%	0	0%	0	0%	0	50%	2
INT-A	1	100%	1	0%	0	0%	0	0%	0	0%	0
LA	2	0%	0	50%	1	0%	0	50%	1	0%	0
LB	5	40%	2	40%	2	0%	0	20%	1	0%	0
LC	12	58%	7	0%	0	0%	0	42%	5	0%	0
C-oral	72	29%	21	56%	40	3%	2	1%	1	11%	8
ESC	14	36%	5	7%	1	0%	0	14%	2	43%	6
Σ	110	35%	38	40%	44	2%	2	9%	10	15%	16
	occ	aimer		aimé		aimais		aimez		aime	
INT-C	17	0%	0	41%	7	0%	0	0%	0	59%	10
INT-A	87	0%	0	3%	3	0%	0	84%	73	13%	11
LA	8	0%	0	0%	0	0%	0	0%	0	100%	8
LB	10	10%	1	0%	0	0%	0	0%	0	90%	9
LC	8	25%	2	0%	0	0%	0	38%	3	38%	3
C-oral	142	11%	16	12%	17	14%	20	5%	7	58%	82
ESC	69	4%	3	0%	0	0%	0	3%	2	93%	64
Σ	341	6%	22	8%	27	6%	20	25%	85	55%	187
	occ	détester		détesté		détestais		détestez		déteste	
INT-C	0	-	-	-	-	-	-	-	-	-	-
INT-A	0	-	-	-	-	-	-	-	-	-	-
LA	1	0%		0%		0%		0%		100%	1
LB	1	100%	1	0%		0%		0%		0%	
LC	0	-	-	-	-	-	-	-	-	-	-
C-oral	4	0%		0%		0%		0%		100%	4
ESC	4	50%	2	0%	0	0%	0	0%	0	50%	2
Σ	10	30%	3	0%	0	0%	0	0%	0	70%	7

	occ	étudier		étudié		étudiais		étudiez		étudie	
INT-C	9	56%	5	0%	0	0%	0	33%	3	11%	1
INT-A	24	4%	1	46%	11	0%	0	46%	11	4%	1
LA	2	0%	0	50%	1	0%	0	50%	1	0%	0
LB	10	80%	8	0%	0	0%	0	0%	0	20%	2
LC	5	40%	2	40%	2	0%	0	0%	0	20%	1
C-oral	14	50%	7	36%	5	7%	1	0%	0	7%	1
ESC	4	25%	1	50%	2	0%	0	0%	0	25%	1
Σ	68	35%	24	31%	21	1%	1	22%	15	10%	7
	occ	habiter		habité		habitais		habitez		habite	
INT-C	3	0%	0	0%	0	0%	0	0%	0	100%	3
INT-A	31	6%	2	0%	0	0%	0	23%	7	71%	22
LA	0	-	-	-	-	-	-	-	-	-	-
LB	12	0%	0	0%	0	0%	0	0%	0	100%	12
LC	6	0%	0	0%	0	0%	0	67%	4	33%	2
C-oral	54	4%	2	2%	1	37%	20	4%	2	54%	29
ESC	44	5%	2	0%	0	0%	0	5%	2	91%	40
Σ	150	4%	6	1%	1	13%	20	10%	15	72%	108
	occ	jouer		joué		jouais		jouez		joue	
INT-C	1	0%	0	0%	0	0%	0	0%	0	100%	1
INT-A	32	9%	3	6%	2	0%	0	53%	17	31%	10
LA	0	-	-	-	-	-	-	-	-	-	-
LB	2	100%	2	0%	0	0%	0	0%	0	0%	0
LC	3	33%	1	33%	1	0%	0	33%	1	0%	0
C-oral	90	41%	37	14%	13	7%	6	2%	2	36%	32
ESC	43	30%	13	0%	0	0%	0	2%	1	67%	29
Σ	171	33%	56	9%	16	4%	6	12%	21	42%	72
	occ	manger		mangé		mangais		mangez		mange	
INT-C	1	0%	0	100%	1	0%	0	0%	0	0%	0
INT-A	10	70%	7	20%	2	0%	0	0%	0	10%	1
LA	1	100%	1	0%	0	0%	0	0%	0	0%	0
LB	51	16%	8	55%	28	0%	0	8%	4	22%	11
LC	23	13%	3	0%	0	0%	0	30%	7	57%	13
C-oral	62	63%	39	11%	7	5%	3	0%	0	21%	13
ESC	12	25%	3	0%	0	0%	0	0%	0	75%	9
Σ	160	38%	61	24%	38	2%	3	7%	11	29%	47
	occ	parler		parlé		parlais		parlez		parle	
INT-C	40	35%	14	10%	4	0%	0	15%	6	40%	16
INT-A	50	20%	10	8%	4	0%	0	52%	26	20%	10
LA	63	24%	15	46%	29	11%	7	3%	2	16%	10
LB	61	20%	12	13%	8	2%	1	7%	4	59%	36
LC	18	22%	4	28%	5	0%	0	6%	1	44%	8
C-oral	319	36%	115	13%	41	13%	41	2%	6	36%	116
ESC	30	13%	4	3%	1	0%	0	3%	1	80%	24
Σ	581	30%	174	16%	92	8%	49	8%	46	38%	220

	occ	penser		pensé		pensais		pensez		pense	
INT-C	15	27%	4	0%	0	13%	2	0%	0	60%	9
INT-A	14	0%	0	0%	0	0%	0	21%	3	79%	11
LA	8	38%	3	0%	0	13%	1	13%	1	38%	3
LB	7	0%	0	0%	0	0%	0	29%	2	71%	5
LC	16	13%	2	13%	2	6%	1	19%	3	50%	8
C-oral	307	6%	19	5%	14	8%	26	5%	14	76%	234
ESC	1	0%	0	100%	1	0%	0	0%	0	0%	0
Σ	368	8%	28	5%	17	8%	30	6%	23	73%	270
	occ	préférer		préfére		préfèrais		préfèrez		préfère	
INT-C	0	-	-	-	-	-	-	-	-	-	-
INT-A	6	0%	0	17%	1	0%	0	83%	5	0%	0
LA	0	-	-	-	-	-	-	-	-	-	-
LB	4	25%	1	0%	0	0%	0	0%	0	75%	3
LC	4	0%	0	0%	0	0%	0	50%	2	50%	2
C-oral	27	4%	1	11%	3	11%	3	11%	3	63%	17
ESC	4	0%	0	0%	0	0%	0	0%	0	100%	4
Σ	45	4%	2	9%	4	7%	3	22%	10	58%	26
	occ	regarder		regardé		regardais		regardez		regarde	
INT-C	7	29%	2	14%	1	0%	0	0%	0	57%	4
INT-A	32	9%	3	13%	4	0%	0	75%	24	3%	1
LA	6	50%	3	17%	1	0%	0	17%	1	17%	1
LB	18	61%	11	6%	1	0%	0	28%	5	6%	1
LC	13	23%	3	31%	4	0%	0	23%	3	23%	3
C-oral	128	23%	29	18%	23	6%	8	6%	8	47%	60
ESC	19	11%	2	5%	1	0%	0	0%	0	84%	16
Σ	223	24%	53	16%	35	4%	8	18%	41	39%	86
	occ	travailler		travaillé		travaillais		travaillez		travaille	
INT-C	11	45%	5	18%	2	0%	0	0%	0	36%	4
INT-A	30	13%	4	7%	2	0%	0	47%	14	33%	10
LA	1	0%	0	0%	0	100%	1	0%	0	0%	0
LB	21	38%	8	10%	2	0%	0	5%	1	48%	10
LC	13	31%	4	23%	3	0%	0	38%	5	8%	1
C-oral	171	40%	69	12%	21	11%	19	4%	6	33%	56
ESC	20	10%	2	0%	0	0%	0	5%	1	85%	17
Σ	267	34%	92	11%	30	7%	20	10%	27	37%	98
	occ	trouver S		trouvé S		trouvais S		trouvez S		trouve S	
INT-C	3	0%	0	0%	0	0%	0	0%	0	100%	3
INT-A	11	0%	0	0%	0	0%	0	55%	6	45%	5
LA	7	0%	0	71%	5	0%	0	0%	0	29%	2
LB	2	0%	0	0%	0	50%	1	0%	0	50%	1
LC	1	0%	0	0%	0	0%	0	0%	0	100%	1
C-oral	153	4%	6	8%	13	10%	16	3%	4	75%	114
ESC	5	0%	0	0%	0	0%	0	0%	0	100%	5
Σ	182	3%	6	10%	18	9%	17	5%	10	72%	131

	occ	trouver D		trouvé D		trouvais D		trouvez D		trouve D	
INT-C	4	50%	2	50%	2	0%	0	0%	0	0%	0
INT-A	1	0%	0	100%	1	0%	0	0%	0	0%	0
LA	12	0%	0	92%	11	0%	0	0%	0	8%	1
LB	7	57%	4	29%	2	0%	0	0%	0	14%	1
LC	22	14%	3	77%	17	0%	0	5%	1	5%	1
C-oral	133	40%	53	29%	39	3%	4	2%	2	26%	35
ESC	32	9%	3	38%	12	0%	0	3%	1	50%	16
Σ	211	31%	65	40%	84	2%	4	2%	4	26%	54
	occ	trouver Σ		trouvé Σ		trouvais Σ		trouvez Σ		trouve Σ	
INT-C	7	29%	2	29%	2	0%	0	0%	0	43%	3
INT-A	12	0%	0	8%	1	0%	0	50%	6	42%	5
LA	19	0%	0	84%	16	0%	0	0%	0	16%	3
LB	9	44%	4	22%	2	11%	1	0%	0	22%	2
LC	23	13%	3	74%	17	0%	0	4%	1	9%	2
C-oral	286	21%	59	18%	52	7%	20	2%	6	52%	149
ESC	37	8%	3	32%	12	0%	0	3%	1	57%	21
Σ	393	18%	71	26%	102	5%	21	4%	14	47%	185
	occ	visiter		visité		visitais		visitez		visite	
INT-C	2	0%	0	100%	2	0%	0	0%	0	0%	0
INT-A	6	83%	5	17%	1	0%	0	0%	0	0%	0
LA	0	-	-	-	-	-	-	-	-	-	-
LB	2	0%	0	100%	2	0%	0	0%	0	0%	0
LC	0	-	-	-	-	-	-	-	-	-	-
C-oral	13	31%	4	46%	6	0%	0	8%	1	15%	2
ESC	2	50%	1	50%	1	0%	0	0%	0	0%	0
Σ	25	40%	10	48%	12	0%	0	4%	1	8%	2
	occ	voyager		voyagé		voyageais		voyagez		voyage	
INT-C	2	0%	0	100%	2	0%	0	0%	0	0%	0
INT-A	26	38%	10	19%	5	0%	0	35%	9	8%	2
LA	0	-	-	-	-	-	-	-	-	-	-
LB	2	0%	0	100%	2	0%	0	0%	0	0%	0
LC	2	50%	1	0%	0	0%	0	50%	1	0%	0
C-oral	16	0%	0	56%	9	19%	3	0%	0	25%	4
ESC	4	75%	3	0%	0	0%	0	0%	0	25%	1
Σ	52	27%	14	35%	18	6%	3	19%	10	13%	7

Légende : INT-C : interlocutrice corpus Lund ; INT-A : interlocutrice corpus Thomas ; LA, LB, LC : enseignant/es ; C-oral : C-ORAL-ROM (2005) ; ESC : manuel Escalade 1 ; Σ : total ; occ : occurrences ; trouver S : trouver statique ; trouver D : trouver dynamique ; trouver Σ : trouver au total.

Annexe 3 : Formes des verbes irréguliers dans l'input oral

	occ	aller lex		allé lex		allais lex		allez lex		pré sg lex		vont lex	
INTC	17	71%	12	12%	2	6%	1	6%	1	6%	1	0%	0
INTA	73	27%	20	21%	15	0%	0	48%	35	3%	2	1%	1
LA	15	27%	4	27%	4	0%	0	7%	1	33%	5	7%	1
LB	49	47%	23	12%	6	0%	0	6%	3	31%	15	4%	2
LC	45	22%	10	13%	6	2%	1	27%	12	29%	13	7%	3
Coral	352	31%	108	19%	66	13%	45	3%	11	29%	101	6%	21
Σ	551	32%	177	18%	99	9%	47	11%	63	25%	137	5%	28
	occ	aller otr		allé otr		allais otr		allez otr		pré sg otr		vont otr	
INTC	65	6%	4	0%	0	0%	0	8%	5	85%	55	2%	1
INTA	68	3%	2	0%	0	0%	0	57%	39	31%	21	9%	6
LA	26	0%	0	0%	0	0%	0	0%	0	100%	26	0%	0
LB	62	2%	1	0%	0	0%	0	8%	5	84%	52	6%	4
LC	74	0%	0	0%	0	0%	0	18%	13	82%	61	0%	0
Coral	1132	5%	51	3%	29	5%	58	4%	46	78%	881	6%	67
Σ	1427	4%	58	2%	29	4%	58	8%	108	77%	1096	5%	78
	occ	comprendre		compris		comprenais		comprenez		comprend		comprennent	
INTC	11	45%	5	0%	0	9%	1	0%	0	45%	5	0%	0
INTA	9	44%	4	11%	1	0%	0	11%	1	33%	3	0%	0
LA	2	0%	0	50%	1	0%	0	50%	1	0%	0	0%	0
LB	8	25%	2	38%	3	0%	0	13%	1	25%	2	0%	0
LC	37	16%	6	22%	8	0%	0	32%	12	30%	11	0%	0
Coral	125	33%	41	23%	29	1%	1	5%	6	34%	43	4%	5
Σ	192	30%	58	22%	42	1%	2	11%	21	33%	64	3%	5
	occ	connaître		connu		connaissais		connaissiez		connais		connaissent	
INTC	16	19%	3	6%	1	0%	0	0%	0	75%	12	0%	0
INTA	21	0%	0	5%	1	0%	0	43%	9	52%	11	0%	0
LA	20	45%	9	5%	1	5%	1	45%	9	0%	0	0%	0
LB	17	0%	0	0%	0	6%	1	59%	10	29%	5	6%	1
LC	18	28%	5	0%	0	0%	0	61%	11	11%	2	0%	0
Coral	182	12%	22	20%	36	13%	24	7%	12	43%	79	5%	9
Σ	274	14%	39	14%	39	9%	26	19%	51	40%	109	4%	10
	occ	croire		cru		croyais		croyez		crois/nt		-	
INTC	22	0%	0	0%	0	5%	1	0%	0	95%	21		
INTA	2	0%	0	100%	2	0%	0	0%	0	0%	0		
LA	3	0%	0	0%	0	0%	0	0%	0	100%	3		
LB	8	0%	0	0%	0	13%	1	0%	0	88%	7		
LC	13	0%	0	0%	0	0%	0	23%	3	77%	10		
Coral	206	11%	22	3%	7	5%	11	1%	3	79%	163		
Σ	254	9%	22	4%	9	5%	13	2%	6	80%	204		

	occ	dire		dit		disais		dites		dis		disent	
INTC	49	39%	19	14%	7	4%	2	0%	0	37%	18	6%	3
INTA	18	67%	12	11%	2	0%	0	6%	1	11%	2	6%	1
LA	46	52%	24	11%	5	2%	1	13%	6	22%	10	0%	0
LB	28	75%	21	4%	1	4%	1	0%	0	14%	4	4%	1
LC	154	38%	58	22%	34	0%	0	8%	12	32%	50	0%	0
Coral	1312	34%	450	26%	340	7%	98	2%	32	28%	365	2%	27
Σ	1607	36%	584	24%	389	6%	102	3%	51	28%	449	2%	32
	occ	faire		fait ppé		faisais		faites		fais		font	
INTC	82	46%	38	12%	10	5%	4	0%	0	35%	29	1%	1
INTA	165	25%	42	8%	13	1%	2	39%	64	25%	42	1%	2
LA	14	57%	8	43%	6	0%	0	0%	0	0%	0	0%	0
LB	75	33%	25	28%	21	17%	13	4%	3	16%	12	1%	1
LC	104	22%	23	16%	17	1%	1	13%	14	46%	48	1%	1
Coral	2099	36%	756	20%	429	7%	152	2%	49	29%	615	5%	98
Σ	2539	35%	892	20%	496	7%	172	5%	130	29%	746	4%	103
	occ	lire		lu		lisais		lisez		lis		lisent	
INTC	9	67%	6	0%	0	0%	0	0%	0	0%	0	33%	3
INTA	75	29%	22	15%	11	0%	0	52%	39	3%	2	1%	1
LA	4	0%	0	100%	4	0%	0	0%	0	0%	0	0%	0
LB	12	58%	7	33%	4	0%	0	0%	0	8%	1	0%	0
LC	28	18%	5	39%	11	0%	0	21%	6	21%	6	0%	0
Coral	50	44%	22	26%	13	4%	2	0%	0	26%	13	0%	0
Σ	178	35%	62	24%	43	1%	2	25%	45	12%	22	2%	4
	occ	pouvoir		pu		pouvais		pouvez		peux		peuvent	
INTC	42	2%	1	2%	1	2%	1	2%	1	90%	38	0%	0
INTA	33	3%	1	0%	0	0%	0	64%	21	33%	11	0%	0
LA	21	0%	0	0%	0	10%	2	43%	9	48%	10	0%	0
LB	47	2%	1	4%	2	0%	0	0%	0	91%	43	2%	1
LC	95	4%	4	0%	0	0%	0	40%	38	56%	53	0%	0
Coral	962	7%	69	11%	102	7%	68	5%	49	65%	627	5%	47
Σ	1200	6%	76	9%	105	6%	71	10%	118	65%	782	4%	48
	occ	prendre		pris		prenais		prenez		prend		prennent	
INTC	12	25%	3	33%	4	17%	2	0%	0	25%	3	0%	0
INTA	9	33%	3	22%	2	0%	0	0%	0	33%	3	11%	1
LA	1	0%	0	0%	0	0%	0	0%	0	100%	1	0%	0
LB	21	57%	12	14%	3	10%	2	10%	2	10%	2	0%	0
LC	26	31%	8	15%	4	0%	0	8%	2	46%	12	0%	0
Coral	351	29%	102	33%	115	4%	15	5%	17	25%	89	4%	13
Σ	420	30%	128	30%	128	5%	19	5%	21	26%	110	3%	14
	occ	savoir		su		savais		savez		sais		savent	
INTC	31	3%	1	0%	0	3%	1	0%	0	94%	29	0%	0
INTA	22	0%	0	0%	0	0%	0	27%	6	68%	15	5%	1
LA	7	14%	1	0%	0	0%	0	43%	3	43%	3	0%	0
LB	38	0%	0	0%	0	0%	0	50%	19	47%	18	3%	1
LC	30	17%	5	0%	0	0%	0	53%	16	30%	9	0%	0
Coral	796	12%	92	2%	12	5%	42	7%	59	72%	572	2%	19
Σ	924	11%	99	1%	12	5%	43	11%	103	70%	646	2%	21

	occ	voir		vu		voyais		voyez		vois/nt		-	
INTC	21	33%	7	24%	5	0%	0	5%	1	38%	8		
INTA	24	33%	8	63%	15	0%	0	0%	0	4%	1		
LA	3	0%	0	33%	1	0%	0	33%	1	33%	1		
LB	23	30%	7	22%	5	0%	0	30%	7	17%	4		
LC	48	35%	17	19%	9	0%	0	17%	8	29%	14		
Coral	769	30%	228	17%	129	5%	41	3%	25	45%	346		
Σ	888	30%	267	18%	164	5%	41	5%	42	42%	374		
	occ	vouloir		voulu		voulais		voulez		veux		veulent	
INTC	16	0%	0	0%	0	6%	1	13%	2	69%	11	13%	2
INTA	16	0%	0	0%	0	6%	1	63%	10	25%	4	6%	1
LA	13	0%	0	0%	0	0%	0	15%	2	85%	11	0%	0
LB	6	0%	0	0%	0	0%	0	83%	5	17%	1	0%	0
LC	59	2%	1	0%	0	0%	0	24%	14	73%	43	2%	1
Coral	572	2%	9	6%	37	17%	96	10%	60	61%	347	4%	23
Σ	682	1%	10	5%	37	14%	98	14%	93	61%	417	4%	27

Légende : voir tableau précédent.

Annexe 4 : Tâche de narration : verbes en -er par participant/es

stade	part	Σ er	Cds *Lds					Lds *Cds					
			Σ PRE	%PRE	PRE	PRE	%*L	Σ INF	INF	INF	Σ C	Σ L	%C
1	P05	4	4	100%	1	3	75%	0	0	0	1	3	25%
1	P10	6	6	100%	3	3	50%	0	0	0	3	3	50%
1	P15	3	3	100%	3	0	0%	0	0	0	3	0	100%
1	P25	1	0	0%	-	-	-	1	1	0	0	1	0%
1	P27	2	2	100%	0	2	100%	0	0	0	0	2	0%
1	P28	14	12	86%	7	5	42%	2	2	0	7	7	50%
1	P30	7	7	100%	5	2	29%	0	0	0	5	2	71%
1	P31	3	3	100%	1	2	67%	0	0	0	1	2	33%
1	P33	9	9	100%	7	2	22%	0	0	0	7	2	78%
1	Σ	49	46	94%	27	19	41%	3	3	0	27	22	55%
2	P01	9	9	100%	0	9	100%	0	0	0	0	9	0%
2	P03	14	14	100%	5	9	64%	0	0	0	5	9	36%
2	P04	10	10	100%	4	6	60%	0	0	0	4	6	40%
2	P07	4	4	100%	4	0	0%	0	0	0	4	0	100%
2	P09	5	5	100%	2	3	60%	0	0	0	2	3	40%
2	P12	10	10	100%	9	1	10%	0	0	0	9	1	90%
2	P13	11	9	82%	7	2	22%	2	1	1	8	3	73%
2	P16	14	12	86%	7	5	42%	2	2	0	7	7	50%
2	P19	9	9	100%	9	0	0%	0	0	0	9	0	100%
2	P20	5	5	100%	5	0	0%	0	0	0	5	0	100%
2	P21	10	10	100%	2	8	80%	0	0	0	2	8	20%
2	P23	-	-	-	-	-	-	-	-	-	-	-	-
2	P24	12	12	100%	12	0	0%	0	0	0	12	0	100%
2	P29	3	3	100%	3	0	0%	0	0	0	3	0	100%
2	Σ	116	112	97%	69	43	38%	4	3	1	70	46	60%

stade	part	Σer	Cds *Lds					Lds *Cds					
			ΣPRE	%PRE	PRE	PRE	%*L	ΣINF	INF	INF	ΣC	ΣL	%C
3	P02	9	9	100%	6	3	33%	0	0	0	6	3	67%
3	P06	5	5	100%	2	3	60%	0	0	0	2	3	40%
3	P08	13	11	85%	5	6	55%	2	1	1	6	7	46%
3	P11	19	5	26%	2	3	60%	14	14	0	2	17	11%
3	P14	11	10	91%	5	5	50%	1	1	0	5	6	45%
3	P17	19	17	89%	15	2	12%	2	1	1	16	3	84%
3	P18	13	11	85%	3	8	73%	2	2	0	3	10	23%
3	P22	19	12	63%	7	5	42%	7	7	0	7	12	37%
3	P26	6	6	100%	5	1	17%	0	0	0	5	1	83%
3	P32	12	12	100%	11	1	8%	0	0	0	11	1	92%
3	Σ	126	98	78%	61	37	38%	28	26	2	63	63	50%
N	N01	29	14	48%				15					
N	N02	37	24	65%				13					
N	N03	23	11	48%				12					
N	N04	36	24	67%				12					
N	N05	32	20	63%				12					
N	N06	14	9	64%				5					
N	N07	30	13	43%				17					
N	N08	50	30	60%				20					
N	N09	16	9	56%				7					
N	N10	25	20	80%				5					
N	Σ	292	174	60%				118					

Légende : part : participant/es ; Σer : somme des occurrences C+L ; ΣPRE : somme contextes du présent ; %PRE : pourcentage de contexte du présent ; CdsPRE : forme courte dans contexte du présent ; *LdsPRE : forme longue dans contexte du présent ; %*L : pourcentage de formes longues dans contexte du présent ; ΣINF : somme contextes de l'infinif ; LdsINF : forme longue dans contexte de l'infinif ; *CdsINF : forme courte dans contexte de l'infinif ; ΣC : somme formes courtes ; ΣL : somme formes longues ; %C : pourcentage de formes courtes.

Annexe 5 : Test d'imitation : données par participant/es

stade	part	C>C	C>L	C>øtr	L>L	L>C	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
1	P05	4	1	8	3	2	7	10	40%	70%	6	4	60%
1	P10	0	2	11	4	1	8	7	27%	57%	1	6	14%
1	P15	3	2	8	1	5	7	11	42%	36%	8	3	73%
1	P25	7	0	6	4	2	7	13	50%	85%	9	4	69%
1	P27	0	1	12	0	0	13	1	4%	0%	0	1	0%
1	P28	5	2	6	2	2	9	11	42%	64%	7	4	64%
1	P30	5	0	8	1	2	10	8	31%	75%	7	1	88%
1	P31	4	3	6	2	1	10	10	38%	60%	5	5	50%
1	P33	1	0	11	1	3	9	5	20%	40%	4	1	80%
1	Σ	29	11	76	18	18	80	76	33%	62%	47	29	62%
2	P01	4	2	7	8	1	4	15	58%	80%	5	10	33%
2	P03	3	1	9	1	3	9	8	31%	50%	6	2	75%
2	P04	8	2	3	5	4	4	19	73%	68%	12	7	63%
2	P07	5	1	7	0	2	11	8	31%	63%	7	1	88%
2	P09	2	1	10	2	3	8	8	31%	50%	5	3	63%
2	P12	8	1	4	4	4	5	17	65%	71%	12	5	71%
2	P13	7	0	6	5	4	4	16	62%	75%	11	5	69%
2	P16	5	5	3	5	2	6	17	65%	59%	7	10	41%
2	P19	11	0	2	2	6	5	19	73%	68%	17	2	89%
2	P20	5	1	7	1	4	8	11	42%	55%	9	2	82%
2	P21	3	3	7	5	3	5	14	54%	57%	6	8	43%
2	P23	3	1	9	1	3	9	8	31%	50%	6	2	75%
2	P24	5	0	8	0	6	7	11	42%	45%	11	0	100%
2	P29	2	1	10	2	0	11	5	19%	80%	2	3	40%
2	Σ	71	19	92	41	45	96	176	48%	64%	116	60	66%
3	P02	9	2	2	4	5	4	20	77%	65%	14	6	70%
3	P06	7	1	5	5	4	4	17	65%	71%	11	6	65%
3	P08	8	3	2	6	1	6	18	69%	78%	9	9	50%
3	P11	7	5	1	10	2	1	24	92%	71%	9	15	38%
3	P14	10	0	3	2	5	6	17	65%	71%	15	2	88%
3	P17	6	2	5	2	1	10	11	42%	73%	7	4	64%
3	P18	3	1	9	2	2	9	8	31%	63%	5	3	63%
3	P22	10	1	2	9	1	3	21	81%	90%	11	10	52%
3	P26	11	0	2	3	7	3	21	81%	67%	18	3	86%
3	P32	7	1	5	3	6	4	17	65%	59%	13	4	76%
3	Σ	78	16	36	46	34	50	174	67%	71%	112	62	64%
Σ	P	178	46	204	105	97	226	426	50%	66%	275	151	65%
N	Natifs	117			117			234	100%	100%	117	117	50%

Légende : s : stade ; part : participant/es ; codage : voir tableau I2-2 ; C : forme courte ; L : forme longue ; Σoutp : somme outputs avec verbe stimulus ; %outp : pourcentage d'outputs avec verbe stimulus ; %corsp : pourcentage de correspondance stimulus-output (C>C + L>L divisé par Σoutp).

Annexe 6 : Test d'imitation : verbes par stade

s	C>C	C>L	C>øtr	L>L	L>C	L>øtr	Σoutp	%outp	%corssp	ΣC	ΣL	%C
1 adorer	1	0	8	0	0	9	1	6%	100%	1	0	100%
1 aimer	2	0	7	0	2	7	4	22%	50%	4	0	100%
1 détester	3	0	6	2	1	6	6	33%	83%	4	2	67%
1 habiter	4	0	5	0	3	5	7	41%	57%	7	0	100%
1 penser	0	1	8	0	0	9	1	6%	0%	0	1	0%
1 préférer	5	2	2	2	4	3	13	72%	54%	9	4	69%
1 trouver S	0	0	8	1	0	8	1	6%	100%	0	1	0%
1 ΣETA	15	3	44	5	10	47	33	27%	61%	25	8	76%
1 acheter	0	2	7	4	0	5	6	33%	67%	0	6	0%
1 manger	5	0	4	0	3	6	8	44%	63%	8	0	100%
1 parler	2	2	5	5	0	4	9	50%	78%	2	7	22%
1 regarder	5	2	2	2	3	4	12	67%	58%	8	4	67%
1 trouver D	0	0	9	0	1	8	1	6%	0%	1	0	100%
1 visiter	2	2	5	2	1	6	7	39%	57%	3	4	43%
1 ΣDYN	14	8	32	13	8	33	43	40%	63%	22	21	51%
2 adorer	3	0	11	1	0	13	4	14%	100%	3	1	75%
2 aimer	4	0	10	1	4	9	9	32%	56%	8	1	89%
2 détester	8	0	6	4	2	8	14	50%	86%	10	4	71%
2 habiter	8	1	5	2	7	5	18	64%	56%	15	3	83%
2 penser	2	0	12	0	2	12	4	14%	50%	4	0	100%
2 préférer	10	1	3	5	4	5	20	71%	75%	14	6	70%
2 trouver S	2	1	11	3	1	10	7	25%	71%	3	4	43%
2 ΣETA	37	3	58	16	20	62	76	39%	70%	57	19	75%
2 acheter	3	4	7	7	1	6	15	54%	67%	4	11	27%
2 manger	5	5	4	4	1	9	15	54%	60%	6	9	40%
2 parler	7	0	7	3	7	4	17	61%	59%	14	3	82%
2 regarder	10	2	2	3	7	4	22	79%	59%	17	5	77%
2 trouver D	4	0	10	1	4	9	9	32%	56%	8	1	89%
2 visiter	5	5	4	7	5	2	22	79%	55%	10	12	45%
2 ΣDYN	34	16	34	25	25	34	100	60%	59%	59	41	59%
3 adorer	4	0	6	1	1	8	6	30%	83%	5	1	83%
3 aimer	8	0	2	0	2	8	10	50%	80%	10	0	100%
3 détester	7	1	2	4	4	2	16	80%	69%	11	5	69%
3 habiter	8	0	2	3	3	4	14	70%	79%	11	3	79%
3 penser	6	1	3	1	3	6	11	55%	64%	9	2	82%
3 préférer	8	2	0	5	2	3	17	85%	76%	10	7	59%
3 trouver S	6	0	4	1	5	4	12	60%	58%	11	1	92%
3 ΣETA	47	4	19	15	20	35	86	61%	72%	67	19	78%
3 acheter	3	2	5	8	1	1	14	70%	79%	4	10	29%
3 manger	7	0	3	3	4	3	14	70%	71%	11	3	79%
3 parler	6	2	2	5	3	2	16	80%	69%	9	7	56%
3 regarder	9	1	0	4	2	4	16	80%	81%	11	5	69%
3 trouver D	2	2	6	3	3	4	10	50%	50%	5	5	50%
3 visiter	4	5	1	8	1	1	18	90%	67%	5	13	28%
3 ΣDYN	31	12	17	31	14	15	88	73%	70%	45	43	51%

s	C>C	C>L	C>øtr	L>L	L>C	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
Σ adorer	8	0	25	2	1	30	11	17%	91%	9	2	82%
Σ aimer	14	0	19	1	8	24	23	35%	65%	22	1	96%
Σ détester	18	1	14	10	7	16	36	55%	78%	25	11	69%
Σ habiter	20	1	12	5	13	14	39	60%	64%	33	6	85%
Σ penser	8	2	23	1	5	27	16	24%	56%	13	3	81%
Σ préférer	23	5	5	12	10	11	50	76%	70%	33	17	66%
Σ trouver S	8	1	23	5	6	22	20	31%	65%	14	6	70%
Σ ΣETA	99	10	121	36	50	144	195	42%	69%	149	46	76%
Σ acheter	6	8	19	19	2	12	35	53%	71%	8	27	23%
Σ manger	17	5	11	7	8	18	37	56%	65%	25	12	68%
Σ parler	15	4	14	13	10	10	42	64%	67%	25	17	60%
Σ regarder	24	5	4	9	12	12	50	76%	66%	36	14	72%
Σ trouver D	6	2	25	4	8	21	20	30%	50%	14	6	70%
Σ visiter	11	12	10	17	7	9	47	71%	60%	18	29	38%
Σ ΣDYN	79	36	83	69	47	82	231	58%	64%	126	105	55%
1 trouvé	0	0	0	0	1	8	1	11%	0%	1	0	100%
2 trouvé	0	0	0	1	3	10	4	29%	25%	3	1	75%
3 trouvé	0	0	0	2	2	6	4	40%	50%	2	2	50%
Σ trouvé	0	0	0	3	6	24	9	27%	33%	6	3	67%

Légende : voir tableau précédent ; ETA : statiques ; DYN : dynamiques.

Annexe 7 : Test de rappel : données par participant/es

stade	part	C>C	C>L	C>øtr	L>C	L>L	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
1	P05	21	2	13	2	14	20	39	54%	90%	23	16	59%
1	P10	13	2	21	0	22	14	37	51%	95%	13	24	35%
1	P15	24	1	11	5	13	18	43	60%	86%	29	14	67%
1	P25	31	0	5	1	24	11	56	78%	98%	32	24	57%
1	P27	14	4	18	1	21	14	40	56%	88%	15	25	38%
1	P28	15	11	11	0	15	20	41	57%	73%	15	26	37%
1	P30	24	6	6	2	10	24	42	58%	81%	26	16	62%
1	P31	25	2	9	1	16	19	44	61%	93%	26	18	59%
1	P33	21	0	15	6	15	15	42	58%	86%	27	15	64%
I	Σ	188	28	109	18	150	155	384	59%	88%	206	178	54%

stade	part	C>C	C>L	C>øtr	L>C	L>L	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
2	P01	19	8	9	3	17	16	47	65%	77%	22	25	47%
2	P03	17	5	14	2	17	17	41	57%	83%	19	22	46%
2	P04	28	4	4	3	26	7	61	85%	89%	31	30	51%
2	P07	16	6	14	5	17	14	44	61%	75%	21	23	48%
2	P09	19	3	14	2	13	21	37	51%	86%	21	16	57%
2	P12	17	4	15	1	15	20	37	51%	86%	18	19	49%
2	P13	20	2	14	1	14	21	37	51%	92%	21	16	57%
2	P16	20	8	8	5	10	21	43	60%	70%	25	18	58%
2	P19	19	1	16	0	12	24	32	44%	97%	19	13	59%
2	P20	20	1	15	1	11	24	33	46%	94%	21	12	64%
2	P21	8	11	17	3	14	19	36	50%	61%	11	25	31%
2	P23	20	3	14	3	18	14	44	61%	86%	23	21	52%
2	P24	8	15	13	2	24	10	49	68%	65%	10	39	20%
2	P29	17	2	17	2	12	22	33	46%	88%	19	14	58%
2	Σ	248	73	184	33	220	250	574	57%	82%	281	293	49%
3	P02	22	5	8	3	18	16	48	67%	83%	25	23	52%
3	P06	19	3	14	0	15	21	37	51%	92%	19	18	51%
3	P08	26	2	6	3	29	6	60	83%	92%	29	31	48%
3	P11	23	2	10	3	13	21	41	57%	88%	26	15	63%
3	P14	19	12	6	0	20	15	51	71%	76%	19	32	37%
3	P17	26	4	7	3	18	14	51	71%	86%	29	22	57%
3	P18	22	0	14	2	15	19	39	54%	95%	24	15	62%
3	P22	19	6	12	3	28	4	56	78%	84%	22	34	39%
3	P26	23	5	8	3	23	10	54	75%	85%	26	28	48%
3	P32	19	3	14	8	19	9	49	68%	78%	27	22	55%
3	Σ	218	42	99	28	198	135	486	68%	86%	246	240	51%
N	N01	36	0	0	0	36	0	72	100%	100%	36	36	50%
N	N02	33	2	1	1	29	6	65	90%	95%	34	31	52%
N	N03	33	0	3	2	29	5	64	89%	97%	35	29	55%
N	N04	30	1	5	1	27	8	59	82%	97%	31	28	53%
N	N05	32	1	3	2	24	10	59	82%	95%	34	25	58%
N	N06	36	0	0	0	36	0	72	100%	100%	36	36	50%
N	N07	32	2	2	2	25	9	61	85%	93%	34	27	56%
N	N08	33	2	1	4	27	5	66	92%	91%	37	29	56%
N	N09	36	0	0	0	36	0	72	100%	100%	36	36	50%
N	Σ	301	8	15	12	269	43	590	91%	97%	313	277	53%

Légende : voir annexe 5 ; codage : voir tableau I3-3 ; N : natifs

Annexe 8 : Test de rappel : verbes par stade

stade		C>C	C>L	C>øtr	L>C	L>L	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
1	adorer	14	6	7	1	14	12	35	65%	80%	15	20	43%
1	aimer	24	0	3	0	11	16	35	65%	100%	24	11	69%
1	détester	19	0	8	1	8	18	28	52%	96%	20	8	71%
1	habiter	22	1	5	1	15	10	39	72%	95%	23	16	59%
1	penser	12	0	15	0	3	24	15	28%	100%	12	3	80%
1	préférer	7	7	13	2	15	10	31	57%	71%	9	22	29%
1	ΣETA	98	14	51	5	66	90	183	56%	90%	103	80	56%
1	acheter	8	0	19	0	10	17	18	33%	100%	8	10	44%
1	manger	16	3	8	4	12	11	35	65%	80%	20	15	57%
1	parler	16	2	9	2	19	6	39	72%	90%	18	21	46%
1	regarder	19	1	7	1	17	9	38	70%	95%	20	18	53%
1	visiter	13	6	8	1	17	9	37	69%	81%	14	23	38%
1	ΣDYN	72	12	51	8	75	52	167	62%	88%	80	87	48%
2	adorer	23	4	15	0	15	27	42	50%	90%	23	19	55%
2	aimer	26	5	12	5	16	20	52	62%	81%	31	21	60%
2	détester	20	3	19	1	20	21	44	52%	91%	21	23	48%
2	habiter	21	7	14	1	22	19	51	61%	84%	22	29	43%
2	penser	21	1	20	1	10	31	33	39%	94%	22	11	67%
2	préférer	13	8	21	0	21	21	42	50%	81%	13	29	31%
2	ΣETA	124	28	101	8	104	139	264	52%	86%	132	132	50%
2	acheter	17	6	19	2	23	17	48	57%	83%	19	29	40%
2	manger	21	6	15	5	19	18	51	61%	78%	26	25	51%
2	parler	23	7	12	4	12	26	46	55%	76%	27	19	59%
2	regarder	10	18	14	0	23	19	51	61%	65%	10	41	20%
2	visiter	23	6	13	2	24	16	55	65%	85%	25	30	45%
2	ΣDYN	94	43	73	13	101	96	251	60%	78%	107	144	43%
3	adorer	21	2	7	0	21	9	44	73%	95%	21	23	48%
3	aimer	23	2	6	3	17	9	45	75%	89%	26	19	58%
3	détester	23	2	5	7	14	9	46	77%	80%	30	16	65%
3	habiter	19	2	9	1	18	11	40	67%	93%	20	20	50%
3	penser	18	2	10	2	12	16	34	57%	88%	20	14	59%
3	préférer	16	6	7	2	21	8	45	75%	82%	18	27	40%
3	ΣETA	120	16	44	15	103	62	254	71%	88%	135	119	53%
3	acheter	8	7	14	0	15	16	30	50%	77%	8	22	27%
3	manger	23	1	6	6	15	9	45	75%	84%	29	16	64%
3	parler	17	6	7	1	17	12	41	68%	83%	18	23	44%
3	regarder	15	4	11	1	20	9	40	67%	88%	16	24	40%
3	visiter	16	4	11	0	18	11	38	63%	89%	16	22	42%
3	ΣDYN	79	22	49	8	85	57	194	65%	85%	87	107	45%

stade		C>C	C>L	C>øtr	L>C	L>L	L>øtr	Σoutp	%outp	%corsp	ΣC	ΣL	%C
Σ	adorer	58	12	29	1	50	48	121	61%	89%	59	62	49%
Σ	aimer	73	7	21	8	44	45	132	67%	89%	81	51	61%
Σ	détester	62	5	32	9	42	48	118	60%	88%	71	47	60%
Σ	habiter	62	10	28	3	55	40	130	66%	90%	65	65	50%
Σ	penser	51	3	45	3	25	71	82	41%	93%	54	28	66%
Σ	préférer	36	21	41	4	57	39	118	60%	79%	40	78	34%
Σ	<i>ΣETA</i>	<i>342</i>	<i>58</i>	<i>196</i>	<i>28</i>	<i>273</i>	<i>291</i>	<i>701</i>	<i>59%</i>	<i>88%</i>	<i>370</i>	<i>331</i>	<i>53%</i>
Σ	acheter	33	13	52	2	48	50	96	48%	84%	35	61	36%
Σ	manger	60	10	29	15	46	38	131	66%	81%	75	56	57%
Σ	parler	56	15	28	7	48	44	126	64%	83%	63	63	50%
Σ	regarder	44	23	32	2	60	37	129	65%	81%	46	83	36%
Σ	visiter	52	16	32	3	59	36	130	66%	85%	55	75	42%
Σ	<i>ΣDYN</i>	<i>245</i>	<i>77</i>	<i>173</i>	<i>29</i>	<i>261</i>	<i>205</i>	<i>612</i>	<i>62%</i>	<i>83%</i>	<i>274</i>	<i>338</i>	<i>45%</i>
N	adorer	26	1	0	2	18	7	47	87%	94%	28	19	60%
N	aimer	27	0	0	0	21	6	48	89%	100%	27	21	56%
N	détester	26	0	1	1	21	5	48	89%	98%	27	21	56%
N	habiter	23	1	3	0	24	3	48	89%	98%	23	25	48%
N	penser	26	1	0	0	26	1	53	98%	98%	26	27	49%
N	préférer	27	0	0	1	24	2	52	96%	98%	28	24	54%
N	<i>ΣETA</i>	<i>155</i>	<i>3</i>	<i>4</i>	<i>4</i>	<i>134</i>	<i>24</i>	<i>296</i>	<i>91%</i>	<i>98%</i>	<i>159</i>	<i>137</i>	<i>54%</i>
N	acheter	25	2	0	0	20	7	47	87%	96%	25	22	53%
N	manger	26	0	1	3	22	2	51	94%	94%	29	22	57%
N	parler	24	0	3	3	22	2	49	91%	94%	27	22	55%
N	regarder	22	2	3	1	22	4	47	87%	94%	23	24	49%
N	visiter	24	1	2	0	26	1	51	94%	98%	24	27	47%
N	<i>ΣDYN</i>	<i>121</i>	<i>5</i>	<i>9</i>	<i>7</i>	<i>112</i>	<i>16</i>	<i>245</i>	<i>91%</i>	<i>95%</i>	<i>128</i>	<i>117</i>	<i>52%</i>
1	trouver	18	2	7	5	9	13	34	63%	79%	23	11	68%
2	trouver	30	2	10	12	15	15	59	70%	76%	42	17	71%
3	trouver	19	4	6	5	10	16	38	63%	76%	24	14	63%
Σ	trouver	67	8	23	22	34	44	131	66%	77%	89	42	68%
N	trouver	25	0	2	1	23	3	49	91%	98%	26	23	53%

Légende : voir tableau précédent ; ETA : statiques ; DYN : dynamiques.

ÉTUDES ROMANES DE LUND
SÉRIE FONDÉE PAR ALF LOMBARD

ED. ALF LOMBARD

1. MALMBERG, BERTIL, *Le roman du Comte de Poitiers, poème français du XIII^e siècle*, publié avec introduction, notes et glossaire. 1940.
2. THORDSTEIN, ARVID, *Le bestiaire d'amour rimé, poème inédit du XIII^e siècle*, publié avec introduction, notes et glossaire. 1940.
3. NILSSON-EHLE, HANS, *Les adverbes en -ment compléments d'un verbe en français moderne. Étude de classement syntaxique et sémantique*. 1941.
4. SCHLYTER, BÖRJE, *La vie de Thomas Becket par Beneit. Poème anglo-normand du XII^e siècle*, publié d'après tous les manuscrits. 1941.
5. RONSJÖ, EINAR, *La vie de saint Nicolas par Wace. Poème religieux du XII^e siècle*, publié d'après tous les manuscrits. 1942.
6. THORNÉ HAMMAR, EVA, *Le développement de sens du suffixe latin -bilis en français*. 1942.
7. MALMBERG, BERTIL, *Le système consonantique du français moderne. Études de phonétique et de phonologie*. 1944.
8. BRANDT, GUSTAF, *La concurrence entre soi et lui, eux, elle(s). Étude de syntaxe historique française*. 1944.
9. NILSSON-EHLE, HANS, *Les propositions complétives juxtaposées en italien moderne*. 1947.
10. MALMBERG, BERTIL, *Études sur la phonétique de l'espagnol parlé en Argentine*. 1950.
11. ANDERSSON, SVEN, *Études sur la syntaxe et la sémantique du mot français tout*. 1954.
12. BOSTRÖM, INGEMAR, *Les noms abstraits accompagnés d'un infinitif combinés avec avoir. Étude historique sur la syntaxe des articles et des prépositions dans ce genre de constructions françaises*. 1957.
13. NEUMANN, SVEN-GÖSTA, *Recherches sur le français des XV^e et XVI^e siècles et sur sa codification par les théoriciens de l'époque*. 1959.
14. ANDERSSON, SVEN, *Nouvelles études sur la syntaxe et la sémantique du mot français tout*. 1961.
15. BORNÄS, GÖRAN, *Trois contes français du XIII^e siècle, tirés du recueil des Vies des Pères*. 1968.
16. JACOBSSON, HARRY, *L'expression imagée dans Les Thibault de Roger Martin du Gard*. 1968.
17. NILSSON, ELSA, *Les termes relatifs et les propositions relatives en roumain moderne. Étude de syntaxe descriptive*. 1969.
18. *Mélanges de philologie offerts à Alf Lombard*. 1969.

19. BRODIN, GRETA, *Termini dimostrativi toscani. Studio storico di morfologia, sintassi e semantica*. 1970.

ED. ÖSTEN SÖDERGÅRD

20. GUNNARSON, KJELL-ÅKE, *Le complément de lieu dans le syntagme nominal*. 1972.
21. WESTRIN, MAIBRIT, *Étude sur la concurrence de davantage avec plus dans la période allant de 1200 à la Révolution. Comparaison avec l'usage actuel*. 1973.
22. SCHLYTER, KERSTIN, *Les énumérations des personnages dans la Chanson de Roland. Étude comparative*. 1974.
23. ROBACH, INGER-BRITT, *Étude socio-linguistique de la segmentation syntaxique du français parlé*. 1974.
24. BRODIN, BRITA, *Criaturas ficticias y su mundo, en « Rayuela » de Cortázar*. 1975.
25. UNDHAGEN, LYDIA, *Morale et les autres lexèmes formés sur le radical moral- étudiés dans des dictionnaires et dans des textes littéraires français de la seconde moitié du XVIII^e siècle. Étude de sémantique structurale*. 1975.
26. SANDQVIST, SVEN, *Études syntaxiques sur la Chronique des Ducs de Normandie par Benoit*. 1976.
27. SWAHN, SIGBRIT, *Proust dans la recherche littéraire. Problèmes, méthodes, approches nouvelles*. 1979.
28. LARSSON, EVA, *La dislocation en français. Étude de syntaxe générative*. 1979.
29. SWEDENBORG, EKY, *Jean Barois de Roger Martin du Gard. Étude des manuscrits et des techniques narratives*. 1979.
30. GRAUMANN, GUNNAR, *« La guerre de Troie » aura lieu. La préparation de la pièce de Giraudoux*. 1979.
31. KELLNER, SVEN, *« Le Docteur Pascal » de Zola: Rétrospective des Rougon-Macquart, Livre de Documents, Roman à Thèse*. 1980.
32. LLAVADOR, YVONNE, *La poésie algérienne de langue française et la guerre d'Algérie*. 1980.
33. BIRGANDER, PIA, *Boris Vian romancier. Étude des techniques narratives*. 1981.
34. GRELSSON, SIGVARD, *Les adverbes en -ment. Étude psycho-mécanique et psycho-systématique*. 1981.
35. JOSEFSON, EVA-KARIN, *La vision citadine et sociale dans l'œuvre d'Emile Verhaeren*. 1982.
36. WIJK, MARGARETH, *Guillaume Apollinaire et l'esprit nouveau*. 1982.
37. HEED, SVEN-ÅKE, *Le coco du dada. Victor ou les Enfants au pouvoir de Roger Vitrac : texte et représentation*. 1983.
38. ORFALI, INGRID, *Fiction érogène à partir de Klossowski*. 1983.
39. SANDQVIST, SVEN, *Notes textuelles sur le Roman de Tristan de Bérout*. 1984.

ED. LARS LINDVALL

40. BORNÄS, GÖRAN, *Ordre alphabétique et classement méthodique du lexique. Étude de quelques dictionnaires d'apprentissage français*. 1986.

41. LARSSON, BJÖRN, *La réception des Mandarins. Le roman de Simone de Beauvoir face à la critique littéraire en France*. 1988.
42. SANDQVIST, SVEN, *Le Dyalogue saint Gregore. Les Dialogues de saint Grégoire le Grand traduits en vers français à rimes léonines par un Normand anonyme du XIV^e siècle*. Édition avec introduction, notes et glossaire. 2 vol. 1989.
43. SANDQVIST, OLLE, *La Vie saint Gregore. Poème normand du XIV^e siècle*, publié avec introduction, notes et glossaire. 1989.
44. ANGELFORS, CHRISTINA, *La Double Conscience. La prise de conscience féminine chez Colette, Simone de Beauvoir et Marie Cardinal*. 1989.
45. *Actes du X^e Congrès des Romanistes Scandinaves, Lund, 10-14 août 1987*, édités par LARS LINDVALL, 1990.

ED. SUZANNE SCHLYTER

46. SWAHN, SIGBRIT, *Balzac et le merveilleux. Étude du roman balzacien 1822-1832*. 1991.
47. ELGENIUS, BERNT, *Studio sull'uso delle congiunzioni concessive nell'italiano del Novecento*. 1991.
48. SANDQVIST, SVEN, *La Vie de saint Évrout. Poème normand du XIV^e siècle*, publié avec introduction, notes et glossaire. 1992.
49. HERMERÉN, INGRID, *El uso de la forma en RA con valor no-subjunctivo en el español moderno*. 1992.
50. LARSSON, BJÖRN, *La place et le sens des adjectifs épithètes de valorisation positive*. 1994.
51. EKBLAD, SVEN, *Studi sui sottofondi strutturali nel Nome della rosa di Umberto Eco. Parte I. La Divina Commedia di Dante*. 1994.
52. ZETTERBERG, ANDERS, *Les propriétés des choses selon le Rosarius (B.N. f. fr. 12483)*. Édition revue et complétée par SVEN SANDQVIST. 1994.
53. EGERLAND, VERNER, *The Syntax of Past Participles. A Generative Study on Nonfinite Constructions in Ancient and Modern Italian*. 1996.
54. BENGTTSSON, ANDERS, *La Vie de sainte Bathilde. Quatre versions en prose des XIII^e et XV^e siècles, publiées avec introduction, notes et glossaire*. 1996.
55. SANDQVIST, SVEN, *Le Bestiaire et le Lapidaire du Rosarius (B.N. f. fr. 12483)*. 1996.
56. JÖNSSON, NILS-OLOF, *La Vie de saint Germer et la Vie de saint Josse de Pierre de Beauvais. Deux poèmes du XIII^e siècle, publiés avec introduction, notes et glossaire*. 1997.
57. LARSSON, BJÖRN, *Le bon sens commun. Remarques sur le rôle de la (re)cognition intersubjective dans l'épistémologie et l'ontologie du sens*. 1997.
58. WIBERG, EVA, *Il riferimento temporale nel dialogo. Un confronto tra giovani bilingui italo-svedesi e giovani monolingui romani*. 1997.
59. SANDBERG, VESTA, *Temps et Traduction. Étude contrastive des temps de l'indicatif du français et du suédois*. 1997.

60. DITVALL, CORALIA, *Études sur la syntaxe et la sémantique de "tot" en roumain ancien et moderne*. 1997.
61. BARDEL, CAMILLA, *La negazione nell'italiano degli svedesi. Sequenze acquisizionali e influssi translinguistici*. 2000.
62. CARIBONI KILLANDER, CARLA, *De la théorie de la description à la description chez Julien Gracq*. 2000.
63. FORNÉ, ANNA, *La piratería textual. Un estudio hipertextual de Son vacas, somos puercos y El médico de los piratas de Carmen Boullosa*. 2001.
64. LENNARTSSON, VIVI-ANNE, *L'Effet-sincérité. L'Autobiographie littéraire vue à travers la critique journalistique. L'Exemple de La Force des choses de Simone de Beauvoir*. 2001.
65. MÖRTE ALLING, ANNIKA, *Le désir selon l'Autre. Étude du Rouge et le Noir et de la Chartreuse de Parme à la lumière du « désir triangulaire » de René Girard*. 2003.
66. JARLSBO, JEANA, *Ecriture et altérité dans trois romans de J. M. G. Le Clézio : Désert, Onitsha et La quarantaine*. 2003.
67. GRANFELDT, JONAS, *L'Acquisition des catégories fonctionnelles. Étude comparative du développement du DP français chez des enfants et des apprenants adultes*. 2003.
68. WESTIN, EVA, *Le récit conversationnel en situation exolingue de français - Formes, types et fonctions*. 2003.
69. BÖRJESSON, ANNE, *La syntaxe de seul et seulement*. 2004.
70. WILHELM, JUAN - ENKVIST, INGER, *Literatura y Compromiso*. Serie de estudios hispánicos. 2004.

ED. INGER ENKVIST, BJÖRN LARSSON, SUZANNE SCHLYTER

71. BERNARDINI, PETRA, *L'italiano come prima e seconda (madre)lingua. Indagine longitudinale sullo sviluppo del DP*. 2004.
72. ÁLVAREZ SALAMANNA, MARÍA DEL PILAR, *De Sobremesa, 1887-1896. José Asunción Silva: El poeta novelista*. 2004.
73. CONWAY, ÅSA, *Le paragraphe oral en français L1, en suédois L1 et en français L2. Étude syntaxique, prosodique et discursive*. 2005.
74. JABET, MARITA, *L'omission de l'article et du pronom sujet dans le français abidjanais*. 2005.
75. BOZIER, CHRISTINE, *La sollicitation dans l'interaction exolingue en français*. 2005.
76. WIKMAN, CHRISTINE, *L'immagine pubblicitaria dell'olio d'oliva, della pasta e del caffè. Uno studio comparativo sulla ricezione*. 2005.
77. ENKVIST, INGER – IZQUIERDO, JOSÉ MARÍA (ed), *Aprender a pensar. Simposio internacional en la Universidad de Lund 2005*. 2006.
78. GUNNARSSON, CECILIA, *Fluidité, complexité et morphosyntaxe dans la production écrite en FLE*. 2006.
79. DAICIU, VIOLETA, *Enjeux idéologiques dans Le Conte du Graal de Chrétien de Troyes*. 2007.

80. JOHANSSON, INGELA, *El personaje femenino de la novela indigenista*. 2008.
81. BACQUIN, MARI, Theseus de Cologne, *édition partielle d'une chanson de geste du XIV^e siècle*. 2008.
82. LUTAS, LIVIU, *Biblique des derniers gestes de Patrick Chamoiseau: Fantastique et histoire*. 2008.
83. LEON-VEGAS, CAROLINA, *Ausencia, prohibición y carencia, Estudio de los personajes masculinos y el deseo frustrado en tres obras de García Lorca*. 2008.
84. ÅGREN, MALIN, *À la recherche de la morphologie silencieuse: sur le développement du pluriel en français L2 écrit*. 2008.
85. BERNARDINI, PETRA, EGERLAND, VERNER & GRANFELDT, JONAS, *Mélanges plurilingues offerts à Suzanne Schlyter à l'occasion de son 65^{ème} anniversaire*.
86. JONSSON, PETTER, *Tres lecturas de las novelas de Mario Vargas Llosa, Interpretación psicoanalítica de la producción novelesca de un autor*. 2009.
87. THOMAS, ANITA, *Les débutants parlent-ils à l'infinif ? Influence de l'input sur la production des verbes par des apprenants adultes du français*. 2009

