

CONTEMPORARY STUDIES IN DESCRIPTIVE LINGUISTICS

VOL. 20

PETER LANG

LE LANGAGE PRÉFABRIQUÉ

Formes, fonctions et fréquences
en français parlé L2 et L1

Fanny Forsberg

CONTEMPORARY STUDIES IN DESCRIPTIVE LINGUISTICS

VOL. 20

PETER LANG

LE LANGAGE PRÉFABRIQUÉ

Formes, fonctions et fréquences
en français parlé L2 et L1

Fanny Forsberg

1 Introduction

Comme le fait remarquer à juste titre Warren (2005, p. 35), on considère traditionnellement, notamment dans le courant chomskyen, que la maîtrise d'une langue implique la connaissance de ses règles grammaticales et de ses items lexicaux. Dans l'optique de cette théorie, la langue offre des possibilités de combinaison quasiment illimitées. Or, au cours des dernières décennies, des études sur corpus (cf. Sinclair 1991) ont montré que, bien que l'être humain dispose de moyens de création élaborés, il n'en fait pas tout le temps usage dans ses productions langagières. Tout comme notre vie abonde d'habitudes, qu'on le veuille ou non, notre pratique langagière se caractérise aussi par des routines et des répétitions, en d'autres mots, par la récurrence. Nous répétons, à un degré important, ce qui a déjà été énoncé, au lieu de recourir à des constructions créatives. Le système langagier admet donc de très grandes variations, mais l'usage est soumis à certaines contraintes et présente un grand nombre de « séquences préfabriquées ». C'est pourquoi le locuteur natif ne conçoit pas d'autre possibilité de dire *poser une question*, alors que le choix de *poser* est ici complètement arbitraire. Pourquoi ne pas utiliser *faire*, *émettre* ou *demander* ? Pour le locuteur natif, ces restrictions combinatoires ne posent, en principe, pas problème ; c'est surtout le locuteur L2, dans son usage de la langue cible, qui met en évidence l'existence de ces séquences parce qu'il se heurte à des difficultés en les acquérant. C'est à ce locuteur L2 que les possibilités de combinaison différentes sont sensibles, dans la mesure où il tentera probablement de transférer des mots et des structures des langues qu'il a déjà acquises. Le répertoire des séquences préfabriquées du locuteur L2 sera par conséquent différent de celui du locuteur L1, puisqu'il sera composé de séquences de la langue cible, de séquences de la langue cible déviantes ainsi que de séquences qui semblent récurrentes et préfabriquées, mais qui ne font pas partie du répertoire de la langue cible.

Pour mieux illustrer les différentes réalisations du terme « séquence préfabriquée », citons quelques exemples provenant d'études récentes. Schmitt & Underwood (2004) s'intéressent à l'usage de *formulaic sequences*¹ par des apprenants d'anglais L2 avancés et discutent de séquences comme *on the other hand* et *I see what you mean*, alors que Myles *et al.* (1998) examinent le lien entre *formulaic language* et règles créatives, se concentrant sur les séquences *j'aime* et *j'adore*, qui, de prime abord, n'ont rien en commun avec les séquences étudiées par Schmitt & Underwood (2004). Ces deux types de séquences sont cependant souvent discutés sous la même étiquette de *formulaic sequence / language*, bien qu'ils soient assez distincts, ce que nous commenterons plus en détail dans 1.3. Des traits communs les unissent cependant et justifient qu'on les regroupe. Leur usage et leur acquisition ne dépendent en effet pas de règles grammaticales ou lexicales, mais de l'imitation, de la mémorisation et de la préfabrication.

L'objet principal de ce travail est d'examiner comment l'usage du langage préfabriqué, dans les différents sens que nous venons de mentionner, se développe au cours de l'acquisition du français L2 et en quoi il se distingue de l'usage qu'en font les locuteurs natifs. L'analyse sera acquisitionnelle et comparative, ainsi que quantitative et qualitative. Avant de préciser nos questions de recherches et nos objectifs, il convient de présenter notre domaine de recherche et les difficultés terminologiques qui le caractérisent.

1.1 Survol historique

La *grammaire générative*, telle qu'elle a été conçue dans les années 1960, s'est développée en réaction aux théories behavioristes, représentées notamment par Skinner et sa publication *Verbal Behavior* (1957). La théorie générativiste a rejeté l'idée que l'imitation serait

1 Nous proposons de traduire en français *formulaic sequence* par *séquence préfabriquée*.

une stratégie d'apprentissage, et elle a considéré la propriété créative du système langagier comme essentielle. Il a toutefois été reconnu que certaines séquences dans la langue sont impérativement imitées et apprises par cœur, mais on les a considérées comme un phénomène marginal et comme une exception dans un système créatif. Étant donné l'impact et l'ampleur de la théorie chomskyenne dans la recherche linguistique, les expressions figées et le caractère imitatif de la langue n'ont donc pas reçu beaucoup d'attention de la part du courant générativiste. Cependant, il y a toujours eu des contre-courants qui se sont intéressés à la propriété imitative de la langue naturelle. L'un des représentants les plus engagés de ces courants est Charles Fillmore, selon qui « *An enormously large amount of natural language is formulaic, automatic and rehearsed, rather than propositional, creative or freely generated* » (Fillmore in Vihman 1982). Avec nombre d'autres chercheurs, il a développé une autre approche de la théorie linguistique, la « Grammaire constructionnelle » (*Construction Grammar*), (voir section 2.1), qui pose l'existence de nombreuses constructions plus ou moins figées. Parallèlement, le traitement informatique des données linguistiques (*corpus linguistics*) a permis la découverte d'un grand nombre de « collocations » (pour une définition, voir section 1.2). Par la suite, on a commencé à reconnaître un rôle de plus en plus important dans la langue aux combinaisons conventionnalisées.

Pour ce qui est de la linguistique française, citons des travaux de G. Gross (1996) et de Schapira (1999), qui s'intéressent aux expressions figées et à la stéréotypisation de la langue. Au Québec, l'équipe de Mel'čuk (1993–1999) a élaboré un dictionnaire de collocations, en s'appuyant sur la théorie Sens-Texte. À l'heure actuelle, une équipe de l'université de Cologne et de l'ATILF à Nancy sont en train d'élaborer un dictionnaire français de collocations (Blumenthal & Bernard 2002). Toutefois, le nombre de recherches dans ce domaine en français demeure restreint.

C'est surtout dans la sphère anglo-saxonne que l'intérêt pour le langage préfabriqué s'est développé, Wray (2002) et Schmitt *et al.* (2004) accordant une place tout à fait centrale aux séquences préfabriquées dans la production langagière. Ils vont jusqu'à dire : « *Formulaic language has become one of the major issues in applied*

linguistics in the new millennium » (2004, p. 55). Nous pouvons donc suivre un développement historique de l'intérêt de la recherche pour les séquences préfabriquées, qui, d'une place périphérique, en est venu aujourd'hui à occuper une place pivot.

1.2 Survol terminologique – du figement à la préfabrication

Définir ce qu'est le langage préfabriqué est l'une des tâches les plus difficiles de notre travail. Dans la section 7, nous entrerons plus en détail dans le problème définitoire, mais il convient d'aborder déjà dans cette introduction la question de la terminologie qui foisonne dans ce domaine.

Wray (2002) signale qu'un des problèmes de l'étude théorique des SP est que la terminologie est très riche et qu'il n'est pas toujours certain que les phénomènes visés soient les mêmes. La tâche qui consiste à trouver une terminologie pertinente devient d'autant plus difficile que nous avons affaire à au moins deux traditions différentes. D'un côté la tradition française, où les études sur le figement en linguistique générale ont dominé, et de l'autre côté la tradition anglo-saxonne où les études statistiques sur les collocations ont été prépondérantes. À cela vient s'ajouter la tradition acquisitionnelle, qui est avant tout d'origine anglo-saxonne : jusqu'ici aucune étude générale sur le langage préfabriqué n'a été effectuée en linguistique acquisitionnelle française.

En linguistique française, les termes les plus fréquemment employés sont, entre autres, *expression idiomatique* (terme qui constitue un problème en soi, cf. G. Gross 1996), *expression figée* (voir 2.1.3 pour une définition), *stéréotype*, *locution* etc.² Il faut ici se souvenir que ces termes ne correspondent pas entièrement à ce

2 Pendant les dernières décennies, on a aussi commencé à employer le terme de *collocation*. Nous y reviendrons dans un instant. Voir aussi le no 150 de la revue *Langue française* (Juin 2006) pour une mise à jour des études sur les *collocations* en français.

que nous appelons *séquence préfabriquée*. Ils ont une portée plus restreinte. Les phénomènes étudiés dans la tradition française sont inclus sous le terme de SP, mais tout ce qu’englobe le terme de SP ne correspond pas aux termes français. La raison en est qu’une séquence préfabriquée ne suit pas les critères stricts de figement (voir 2.1.3 et 2.1.4), mais qu’elle peut être considérée comme préfabriquée pour d’autres raisons, telles que sa complexité par rapport au reste de la production (chez des apprenants débutants) ou sa force pragmatique (formules de politesse ou marqueurs de discours). Comme nous concevons, à l’instar de Wray (2002), le langage préfabriqué dans une perspective vaste, le terme de séquence préfabriquée nous paraît adéquat. Naturellement, on pourrait ici se poser la question de l’utilité d’étudier des phénomènes aussi disparates. Cependant, il s’est avéré, dans nos études pilotes, que l’étude de différents types de SP est très fructueuse dans la perspective développementale et comparative de l’acquisition d’une L2.

En linguistique anglo-saxonne et acquisitionnelle, les dénominations sont entre autres *chunks*, *prefabs*, *fixed expressions*, *formulae*, *gambits*, *prefabricated routines* et *patterns*, pour n’en citer que quelques-unes (cf. Wray 2002). Bien évidemment, le terme de *collocation* est aussi très utilisé, dans des sens différents. Sinclair (1991) le définit ainsi : « *Collocation is the occurrence of two or more words within a short space of each other in a text [...]. When two words of different frequencies collocate significantly, the collocation has a different value in the description of each of the two words* ».

D’après Stubbs (2001, p. 29), la collocation peut être représentée de la façon suivante, lors de la recherche automatique : *Collocation – Node – Collocation*. Le *node* constitue le mot / lemme examiné et la *collocation* les mots / lemmes qui co-apparaissent avec le *node* dans le corpus étudié. Étant donné que la linguistique de corpus s’intéresse surtout aux fréquences, il faut également que la co-occurrence soit fréquente pour qu’elle soit qualifiée de collocation. La probabilité d’une co-occurrence dépend bien évidemment du mot au centre d’une collocation donnée. Si nous prenons l’exemple *prendre le petit déjeuner*, il est largement plus probable que *petit déjeuner* co-existe avec *prendre*, que l’inverse, étant donné que *prendre* est un verbe fort fréquent alors que *petit déjeuner* a moins de possibilités de

combinaison. Si *petit déjeuner* constitue notre *node*, on pourra sûrement dire qu'il y a une relation collocationnelle forte, et tel n'est pas le cas si notre *node* dans l'analyse est *prendre*.

Nous n'avons pas l'intention de mener une étude de cette envergure technique, vu que notre but primordial n'est pas de vérifier le statut collocationnel d'un certain nombre de séquences mais de décrire un développement acquisitionnel en le comparant avec la production des natifs. C'est pourquoi le terme de collocation ne nous convient pas. Citons en dernier lieu une définition française de la collocation, qui n'est pas aussi exhaustive que ses équivalents anglo-saxons. Riegel *et al.* (2004, p. 122–123) qualifient la collocation comme une « affinité sélective » qui apparaît entre certains mots. Cela veut dire qu'un verbe comme *hocher* s'emploie uniquement avec le complément *tête* ou que l'on dit *grièvement blessé* et non *gravement blessé*. Les possibilités de combinaison sont donc extrêmement réduites pour certains mots, combinaisons qu'on appelle « collocations ». Cette définition nous semble trop peu opérationnelle pour une étude quantitative.

Les autres termes (*expression figée* par exemple) sont trop limités ou trop vagues (*stéréotype* ou *expression idiomatique*) : c'est la raison pour laquelle nous avons finalement adopté le terme de Wray, à savoir *séquence préfabriquée* (formulaic sequence). La synthèse de Wray est la plus approfondie que nous ayons trouvée et nous pensons qu'elle nous propose une vision novatrice de l'usage langagier (voir 3.2 pour une présentation). Comme son terme s'applique à la fois à la production des natifs et des non-natifs, nous l'avons jugé approprié pour notre étude acquisitionnelle. Il convient de dire que c'est également le terme adopté par Schmitt (éd., 2004). Ce dernier argumente, comme nous, en faveur de ce terme, soulignant qu'il a une portée très vaste. Il nous arrivera aussi d'utiliser le terme de *langage préfabriqué* pour parler du phénomène de façon plus générale. Le terme de *pré-fabrication* qui est considéré comme synonyme des deux autres sera également employé.

Après ce survol général, passons à notre domaine de spécialisation, l'acquisition L2, où *formulaic* peut signifier encore autre chose que la description faite dans la présente section.

1.3 La notion de *formulaic*³ – tour d’horizon d’une problématique acquisitionnelle centrale

Comme nous l’avons déjà évoqué dans la section 1, la notion de *formulaic sequence* en acquisition est compliquée, vu qu’elle réfère à des phénomènes distincts.

1.3.1 *Des unités non analysées aux règles créatives ?*

Dans la recherche acquisitionnelle, qu’elle soit en L1 ou en L2, on s’intéresse depuis longtemps à la relation entre les unités acquises de façon holistique – nous les nommerons ici les *unités non analysées* – et l’émergence d’un système de règles créatives. Il est fréquent d’observer l’usage dit non-analysé ou *formulaic* de séquences comme *je ne sais pas* et *s’il vous plaît*, dans la production par ailleurs peu développée d’apprenants débutants. C’est aussi l’usage soupçonné d’être *formulaic* qui est souvent écarté dans les études générativistes ou par le modèle de la processabilité (cf. Pallotti 2006), car cet usage ne serait pas révélateur d’une application de règles. Faisant une étude sur les *formulaic sequences* en acquisition, nous avons donc été amenée à considérer également cette description du terme, qui est assez éloigné de celle qu’on fait des collocations et des expressions figées, ayant pourtant en commun le manque d’analyse grammaticale ou lexicale.

Wong-Fillmore (1976) a été parmi les premiers à avancer l’hypothèse que les unités non analysées servent de base de données pour l’acquisition des règles créatives. Dans son étude, elle a observé que les apprenants qui progressaient le plus rapidement possédaient au départ un vaste répertoire d’unités non analysées. Ensuite, ils commencent à décomposer le matériau non-analysé ; d’abord en se servant des unités comme « cadres » permettant l’insertion de différents

3 Nous gardons le mot anglais, car les études auxquelles nous nous référons sont de langue anglaise. Le terme « préfabriqué » est, comme nous venons de le dire, notre traduction de l’anglais.

éléments, et ensuite en construisant des énoncés librement, sans recours ni aux unités non analysées ni aux cadres. Wong-Fillmore (1976) a ainsi suggéré qu'un grand stock d'unités non analysées facilite le chemin vers un système de règles créatives.

Cette séquence acquisitionnelle, allant des unités aux constructions créatives, correspond aux suggestions faites par N. Ellis (2002a, 2005) au cours des dernières années (voir aussi section 3.1). D'après ce chercheur, l'acquisition peut être décrite comme une séquence « *formula – low-scope pattern – creative constructions* ». L'idée est exactement la même que celle proposée par Wong-Fillmore trente ans plus tôt, à ceci près que Wong-Fillmore (1976) estime que seuls certains apprenants ont recours à cette stratégie, alors que N. Ellis suggère que cette séquence acquisitionnelle est la stratégie d'apprentissage préférée et naturelle, en d'autres termes, c'est ainsi que se déroule l'acquisition d'une langue.

Myles *et al.* (1998, 1999) ont conduit des recherches dans cette même direction. Elles ont observé un processus semblable à celui observé par Wong-Fillmore. Il nous paraît difficile cependant d'établir un lien entre les unités non analysées et l'émergence d'un système de règles créatives. On peut bien évidemment observer que les apprenants produisent un grand nombre d'unités non analysées, que celles-ci sont ensuite décomposées et modifiées et qu'enfin des règles créatives semblent être employées. En revanche, il est plus délicat de se prononcer sur le rapport qu'entretient l'acquisition des unités non analysées avec l'émergence de la capacité créative. Peut-on montrer qu'on apprend les règles grâce aux unités non analysées, comme le propose par exemple Wong-Fillmore (1976) ? N'est-il pas concevable qu'il puisse s'agir de deux systèmes séparés et que les règles puissent être acquises indépendamment des unités non analysées ? L'hypothèse proposée par tous ces chercheurs nous paraît plausible, mais sa confirmation demanderait d'énormes quantités de données sur les mêmes apprenants. Tomasello (2000) a suivi de façon très intensive quelques enfants, pour tenter de prouver l'existence de cette séquence acquisitionnelle. En réalité, il faudrait accompagner l'enfant quasiment sans interruption pour voir quand et dans quel contexte les premières unités apparaissent, comment elles sont décomposées et ainsi de suite. On peut même se demander si la confirmation empirique de cette

décomposition est possible, étant donné qu'il s'agit de processus sous-jacents. Il est possible que le raisonnement ne puisse pas dépasser le stade hypothétique.

Ces constatations impliquent qu'il est impossible, avec les données dont nous disposons (voir 8.1 et 9.1), de faire une étude dans cette optique. C'est pourquoi cette perspective « décomposante » n'a pas pu être utilisée et approfondie dans cette thèse. Remarquons aussi que cette problématique se rapporte aux étapes initiales de l'acquisition, alors que nous étudions, outre ces étapes, des étapes avancées et très avancées.

1.3.2 Unités non analysées et séquences préfabriquées idiomatiques : acquisition et stockage

Dans sa synthèse des études en langage préfabriqué, Wray (2002) propose de représenter l'acquisition holistique / analytique par la courbe suivante :



Figure 1. Le traitement holistique dans le processus acquisitionnel (d'après Wray (2002), p. 133)

Selon cette courbe, l'acquisition commencerait par une phase très holistique, conformément au schéma proposé par N. Ellis (2002a, 2005, voir 1.3.1). Ensuite on se retrouverait dans une phase analytique, ce qui correspondrait à l'étape *creative constructions* chez N. Ellis. Jusqu'ici la courbe de Wray ci-dessus est conforme à la séquence acquisitionnelle proposée par N. Ellis. Mais que se passe-t-il après ? On voit que le traitement holistique augmente à nouveau, pour enfin se stabiliser à un niveau relativement élevé. Cette dernière phase est aussi présente chez N. Ellis (2002a) lorsqu'il traite de l'acquisition de

l'idiomaticité. Il propose que l'acquisition d'une langue se caractérise par un *piecemeal-learning* et qu'il faille une grande quantité d'input pour deviner la combinatoire des mots. Nous acquérons alors de plus en plus de *formulaic sequences*, mais dans cette troisième et dernière phase, le terme prend un autre sens (que dans la section 1.3.1) et c'est ce sens-là que nous traduisons ci-dessous par *séquence préfabriquée idiomatique*. Ce dernier terme correspond à peu près aux termes de *collocation* et d'*expression figée*, que nous venons de voir dans la section 1.2. Compte tenu du raisonnement de N. Ellis (2002a, 2002b), le terme *formulaic sequence* / *séquence préfabriquée* englobe donc, à notre sens, deux types principaux, à savoir les *unités non analysées* et les *séquences préfabriquées idiomatiques*, comme on le voit dans la figure 2 ci-dessous. Préfabriqué (de même que *formulaic* en anglais) peut donc, dans notre optique, signifier non-analytique et / ou idiomatique.

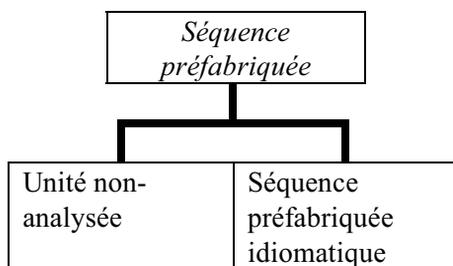


Figure 2. Le terme séquence préfabriquée : les deux réalisations principales

Dans la mesure où nous effectuons une étude couvrant quasiment toutes les étapes acquisitionnelles, des débutants aux apprenants très avancés, nous avons la possibilité d'étudier ces deux types et de démontrer la pertinence de cette distinction, qui n'a pas été précisée dans les études précédentes aux corpus plus limités.

Comme nous l'avons vu, les unités non analysées se rapportent aux étapes initiales de l'acquisition d'une langue (et présupposent une décomposition), alors que la maîtrise des séquences préfabriquées

idiomatiques survient assez tardivement (*cf.* par ex. Yorio 1989). Mais les deux types sont-ils acquis de la même manière ?⁴

Wray (2002) suggère que toute unité préfabriquée dépend du même principe, à savoir « l'analyse au-cas-où » (*needs-only-analysis*). D'après ce principe, le traitement holistique est toujours préféré au traitement analytique en langue première (L1). Le fait qu'on peut décomposer des collocations en deux mots distincts, par ex. *avoir + faim*, relève de nos compétences métalinguistiques et de la dominance de l'écrit sur l'oral. La L1 favorise de manière générale le traitement holistique, et c'est grâce à des décompositions que l'apprenant prend connaissance de l'unité mot. En L2, la situation est toute autre, vu que l'apprenant part d'une perspective analytique : il est tout à fait conscient de l'unité mot, et, il doit, petit à petit, apprendre la combinatoire des mots. Ce n'est que pendant la toute première période d'acquisition que l'apprenant L2 applique un mode holistique, mais il passe rapidement à un mode analytique. L'acquisition L2 est donc un processus qui évolue plutôt vers un mode holistique. Wray (2002) se sert des différents processus pour expliquer le manque de séquences préfabriquées chez les locuteurs L2 avancés par rapport aux locuteurs natifs (voir aussi section 3.2).

Nous avons vu ci-dessus comment Wray (2002) explique la différence en L1 et L2. Mais comment N. Ellis l'explique-t-il ? La séquence acquisitionnelle qu'il propose serait toujours valable en L2, mais elle peut l'être moins en fonction du milieu d'acquisition, de l'âge des apprenants et du système cognitif et perceptuel déjà présent chez les apprenants (2002b, p. 320). La conséquence est que plus les apprenants sont guidés, plus ils sont âgés et plus ils ont de connaissances métalinguistiques, moins ils vont suivre la séquence acquisitionnelle proposée par Ellis, vu que celle-ci se veut naturelle. Que se passe-t-il alors, en L2, avec les séquences préfabriquées idiomatiques, à savoir l'acquisition incrémentielle de combinaisons récurrentes ? N. Ellis propose que ces séquences s'acquièrent encore plus tardivement et avec plus de difficulté chez les apprenants L2,

4 Nous n'avons ni l'intention ni la possibilité pour résoudre dans ce livre cette problématique psycholinguistique. Néanmoins, il nous paraît indispensable de mentionner ce débat théorique.

vu leur fréquence dans l'input. Notons cependant que l'input en soi n'est pas suffisant, car des différences individuelles dans la mémoire phonologique entrent aussi en jeu (N. Ellis 1998, 2005). Les individus n'ont pas la même capacité de repérer puis de retenir des séquences, et cela constitue une explication de la diversité des productions L2. N. Ellis (1998, p.12) propose en fait que la capacité de mémoire phonologique soit un indice de la disposition linguistique (*language aptitude*).

N. Ellis (2002a) avance que l'exposition continue à certaines combinaisons assure leur consolidation en tant que séquences préfabriquées. Nous avons l'impression que la courbe de Wray (voir figure 1 ci-dessus) exprime plus ou moins le même raisonnement, mais Ellis précise que même le locuteur L1 a besoin d'un certain temps pour s'approprier toutes les combinaisons idiomatiques. Peut-on donc qualifier cette acquisition, dans la phase tardive, de *holis-tique*, comme le fait Wray ? Nous inspirant du raisonnement de Bybee (2003), « *words used together fuse together* », nous avons l'impression qu'il s'agit plutôt d'un processus fusionnel.

D'après N. Ellis (2002b, p. 299–303), les *formulas* peuvent être acquises de deux façons différentes. L'acquisition initiale d'unités non analysées relève d'une acquisition explicite (*episodic, one-off, explicit*),⁵ tandis que l'acquisition tardive de séquences préfabriquées idiomatiques relève d'une acquisition implicite (*incremental, implicit*).⁶ Nous aurions donc affaire à deux types d'apprentissage différents. D'abord, les unités non analysées sont acquises très tôt, grâce à une attention explicite portée à une unité forme-sens (*form-meaning pairing*). Prenons l'exemple de *je m'appelle*. L'apprenant conçoit, bien qu'il ne discerne pas la structure interne de l'unité, qu'il s'agit d'un énoncé d'autoprésentation. Une seule rencontre (*one-off*) avec cette séquence suffit pour qu'il commence à s'en servir, Or, pour

5 Pour une explication approfondie de ces termes, voir N. Ellis 2002b. *Episodic* et *one-off* signifie (de façon simplifiée) que l'on saisit la séquence à un moment particulier, en un événement unique. *Explicite* veut dire conscient mais pas nécessairement analysé.

6 Comme pour les autres termes, voir N. Ellis 2002b. Très brièvement, on peut dire que *incremental* réfère à une augmentation graduelle, en l'occurrence un ancrage qui se consolide graduellement. *Implicite* veut dire inconscient.

que la séquence soit ancrée et consolidée dans le lexique mental, une certaine quantité d'input de la même séquence est nécessaire.

Selon ce raisonnement on peut supposer que moins la séquence est fréquente, plus cela prendra de temps avant que la séquence soit bien ancrée dans le lexique mental de l'apprenant. Certaines séquences demandent davantage de temps que d'autres pour être maîtrisées. Dans le bilan de ce travail (section 10), nous suggérerons comment la distinction unité non-analysée / séquence préfabriqué idiomatique peut être liée aux résultats empiriques que nous avons obtenus. Soulignons cependant que nous touchons ici au débat de l'acquisition explicite/ implicite, qui ne pourra être que soulevé dans le cadre de cette étude. Les propos de N. Ellis à cet égard ne peuvent que pousser à la réflexion dans le présent travail.

Comme nous l'avons vu ci-dessous, Wray (2002) propose que l'acquisition des séquences préfabriquées en L2 soit rendue difficile par une approche trop analytique. Mais n'est-il pas possible que la fréquence dans l'input aide, au bout d'un certain temps, le locuteur L2 à s'apercevoir de l'existence de ces unités qui dépassent le mot simple ? Les observations qu'a pu faire Wray (2002) concernant les apprenants adultes L2 se fondent surtout sur des études faites sur des apprenants avancés guidés ou semi-guidés (voir par exemple section 5). Il est possible que des études s'intéressant à des apprenants encore plus avancés, effectuées dans d'autres contextes, permettraient de nuancer les propos de Wray (2002), pour les rapprocher davantage des hypothèses de N. Ellis. Selon N. Ellis, il ne serait pas impossible d'acquérir un stock considérable de séquences préfabriquées, pourvu que l'input soit suffisant. Mais combien de temps et combien d'input de la langue cible faudrait-il avant que les combinaisons conventionnalisées ne soient bien ancrées chez les locuteurs L2 ? Nous espérons que la présente étude contribuera à répondre à cette question.

Après avoir essayé de clarifier le concept de *formulaic* en acquisition, qui s'est avéré pour le moins polysémique, nous allons maintenant exposer notre approche et les questions qui seront examinées dans cette étude.

1.4 Objectifs et questions de recherche

Comme on a pu le voir, le terme *séquence préfabriquée* est vaste. D'après Wray (2002), les unités non analysées du début de l'acquisition ont la même nature que les séquences préfabriquées idiomatiques qui constituent un trait plus tardif dans l'acquisition. Selon N. Ellis, par contre, ces deux phénomènes seraient distincts et ils relèveraient de deux types d'acquisition différents (communication personnelle). Nous avons aussi noté que certains chercheurs estiment qu'il y ait un lien entre acquisition initiale non analysée et développement de règles créatives. Cependant, il a été constaté que les données dont nous disposons ne sont pas suffisantes pour examiner cette dernière question. En revanche, ces données vont nous permettre d'effectuer une étude acquisitionnelle et comparative de l'usage des séquences préfabriquées en français L2 (suivant la définition que nous proposons dans 7.1). Les séquences incluses sont aussi bien des séquences de la langue cible que des séquences issues / typiques de l'interlangue. Nous servant d'interviews semi-guidées d'apprenants débutants, de lycéens, d'apprenants universitaires et de locuteurs L2 très avancés en France, nous espérons pouvoir décrire le développement de la mise en œuvre de séquences préfabriquées en langue seconde. Par ailleurs, d'une part deux groupes de locuteurs natifs, des étudiants Erasmus français en Suède et d'autre part des Français vivant à Paris (corpus parallèle à celui des locuteurs L2 très avancés), seront étudiés, afin de disposer d'un modèle contrastif de l'usage en français parlé. L'étude des locuteurs natifs s'est avérée d'autant plus importante que nous ne disposons d'aucune étude de référence en français parlé pour les séquences préfabriquées. L'objectif de l'étude est donc double : d'une part décrire l'usage et le développement en L2, de l'autre l'usage en L1. Comme ce travail n'a pas de précédent en français, il a fallu effectuer une étude exploratoire pour arriver à une compréhension globale du fonctionnement du langage préfabriqué en français L2 et L1. Les résultats de cette étude permettront, comme l'indiquent les différentes pistes que nous proposerons dans les analyses et plus particulièrement dans la section Perspectives (section 10.2), de réaliser ultérieurement des études approfondies plus délimitées. Le présent

travail se veut cependant exploratoire et d'une portée relativement générale.

Les questions de recherche que nous nous posons sont les suivantes :

- Quel est le rôle principal des SP dans l'usage L1 et L2 du français parlé ?
- Quelle est la proportion de SP dans la production orale de chaque groupe d'apprenants ? Les différences sont-elles significatives ?
- Quelle est la distribution par catégories dans les différents groupes ? Les différences sont-elles significatives ?
- Quels sont les rapports types / occurrences dans les différentes catégories au sein des groupes (voir section 7 pour une présentation des catégories) ? Les différences sont-elles significatives ?
- Quelles sont les variations individuelles dans l'usage du langage préfabriqué ? Sont-elles plus importantes au début ou à la fin de l'acquisition ?
- Est-il possible d'atteindre une maîtrise native dans le domaine du langage préfabriqué ?
- Quels facteurs pourraient expliquer l'acquisition de SP ?

Les hypothèses accompagnant ces questions de recherche sont exposées dans la section 7.7, puisque nous voulons d'abord présenter au lecteur les études antérieures et la méthode, afin qu'il puisse mieux comprendre les fondements des hypothèses que nous émettons.

1.4.1 Disposition du livre

Le présent travail commencera par la description des séquences pré-fabriquées, aussi bien en linguistique générale qu'en linguistique acquisitionnelle (chapitre 2). Seront ensuite présentés les modèles psycholinguistiques qui peuvent expliquer la (non)-acquisition des SP (chapitre 3). Suivra un chapitre sur les études antérieures empiriques faites dans le cadre de l'acquisition L2, notre domaine de spécialisation (chapitre 4). Après quoi, deux chapitres traiteront des notions ayant des implications surtout pour la deuxième analyse : les différences individuelles, l'intégration socio-culturelle et la notion du locuteur natif (chapitres 5 et 6).

À la lumière des théories et des études antérieures, il s'agira par la suite de présenter les définitions, le modèle d'analyse, en d'autres termes, la méthode que nous jugeons adéquate pour répondre à nos questions de recherche (chapitre 7). À la fin de ce chapitre, on trouvera les hypothèses que nous serons alors en mesure de proposer, après avoir achevé le parcours théorique. La partie empirique contient deux analyses (chapitres 8 et 9). Compte tenu du manque d'études de référence en français, il a été nécessaire de réaliser une première étude pilote permettant d'avoir une compréhension générale du phénomène des SP en L2. Cette étude, qui constitue la première analyse, nous a ensuite permis de poser des questions de recherches complémentaires et de nous concentrer sur une étape qui manquait dans l'étude pilote : l'étape finale de l'acquisition d'une L2. Le plan en deux analyses séparées représente à la fois l'ordre dans lequel le travail s'est fait, et, nous l'espérons, un ordre logique de raisonnement pour le lecteur.