



GERFLINT

ISSN 1768-2649

ISSN en ligne 2261-2769

Les séquences préfabriquées dans la production écrite dans le cas de scripteurs finnophones de français et suédois L2

Maarit Mutta

Université de Turku, Finlande

maamut@utu.fi

<https://orcid.org/0000-0003-1227-1484>

Susanna Salminen

Université de Turku, Finlande

susanna.e.salminen@utu.fi

Reçu le 22-08-2020 / Évalué le 16-11-2020 / Accepté le 10-12-2020

Résumé

Cet article examine la manière dont les apprenants multilingues utilisent les séquences préfabriquées (SP) dans leur production écrite. Cette étude de cas analyse les productions écrites en langue étrangère, notamment en français et en suédois, de quatre apprenants finlandais au niveau universitaire (niveau Licence). Les données ont été recueillies à l'aide du logiciel *ScriptLog* permettant l'enregistrement des frappes de clavier et à l'aide des verbalisations rétrospectives avec rappel stimulé. Les résultats montrent que les apprenants produisent plus de SP en français qu'en suédois, et que les catégories des SP lexicales et des SP discursives sont les plus importantes dans les deux langues, même si leur distribution en sous-catégories varie selon les langues. Les résultats témoignent aussi des profils individuels d'utilisation des SP et soutiennent l'idée des SP automatisées facilitant les processus d'écriture.

Mots-clés : séquence préfabriquée, production écrite, apprenant multilingue, processus automatique

**Formulaic sequences in written production:
the case of Finnish learners of L2 French and Swedish**

Abstract

This article examines how multilingual learners use formulaic sequences (FS) in their written production. This case study presents the analysis of foreign language written productions, especially in French and Swedish, by four Finnish learners at university level (BA-level). The material was collected using *ScriptLog* program which records computer keystrokes and retrospective verbalisations with stimulated recall. The results show that learners produce more FS in French than in Swedish, and that the categories of lexical FS and discourse FS are the most important in both languages,

although their distribution in sub-categories varies between languages. The results also reflect individual patterns of FS use and support the idea of automaticity of using FS that facilitate the writing process.

Keywords: formulaic sequence, written production, multilingual learner, automaticity of process

Introduction

La recherche en linguistique et en psycholinguistique utilise le terme « séquence préfabriquée » pour parler de suites de mots comme à mon avis, *faire du sport*, *il y a*. Le phénomène des séquences préfabriquées (SP) intéresse les linguistes et les psycholinguistes depuis les années 60 mais sans qu'il y ait une théorie unifiée ni une terminologie commune (Myles, Cordier, 2017 ; Perez-Bettan, 2015). La recherche autour des SP a d'abord concerné leur usage en langue maternelle (L1) et, plus récemment, elle a été étendue à celui des langues étrangères (L2). Dans les études antérieures, les chercheurs ont trouvé qu'en L1 la production orale contient environ 58 % de SP et la production écrite autour de 52 % de SP (Erman, Warren, 2000 : 37). Il est à noter que ces études se sont concentrées principalement sur la langue parlée, bien que les SP existent à l'oral comme à l'écrit. L'utilisation des SP dans la production écrite en L2 est un sujet digne d'intérêt mais relativement peu étudié jusqu'ici. Ainsi, dans cet article, nous nous concentrons sur la production écrite des apprenants d'une L2, ou plus précisément de deux L2, notamment le français et le suédois.

On peut se demander pourquoi ce phénomène des SP est tellement crucial dans la production verbale. Selon Perez-Bettan (2015 : 153), les études antérieures montrent que l'utilisation des SP facilite la production verbale en L1 et en L2 et ceci aussi bien dans la réception du texte ou du discours que dans la production. Cela est lié à l'automatisation des processus de traitement des SP, qui sont gardées en mémoire à long terme comme un tout ; cela permet de les récupérer plus rapidement dans la mémoire. L'utilisation des SP automatisées exige moins d'effort, limite la charge cognitive en mémoire de travail et libère des ressources pour d'autres processus cognitifs, dont la planification et la révision lors de la production verbale (Perez-Bettan, 2015 : 25). De plus, selon Cislaru et Olive (2018 : 13), le recours aux préfabriqués linguistiques, comme les SP, aide à « fluidifier le flux d'écriture ». Comme l'écriture en L2 est encore plus exigeante qu'en L1 (entre autres, Mutta, 2020), le recours aux SP aide les apprenants à s'exprimer de manière plus aisée et leur donne plus de temps pour se concentrer sur le contenu et la structure de leur rédaction.

Traditionnellement, les productions écrites ont été analysées à partir des textes finalisés, c'est-à-dire des produits finis. Or, le développement technologique rend possible l'analyse des processus d'écriture et l'enregistrement des frappes de clavier lors de la rédaction ; le logiciel *ScriptLog* utilisé dans cette recherche fait partie de ce type de technologie. En ayant recours à ce logiciel, on peut analyser aussi bien le processus que le produit fini des textes écrits. Dans ce qui suit, nous parlerons de productions écrites dans les deux cas et précisons quand il s'agit spécifiquement du processus.

Ainsi, dans cet article, nous nous intéressons aux SP auxquelles les apprenants ont recours pendant la rédaction en français L2 et en suédois L2 au niveau Licence dans une université finlandaise. L'article présente les résultats d'un corpus ramassé pour un projet en cours sur la l'aisance rédactionnelle et la visualisation mené à l'Université de Turku et dont les premières données ont été analysées dans le mémoire de Salminen (2020)¹. Notre réflexion est guidée par les questions de recherche suivantes :

1. Quels types de SP les étudiants finnophones produisent-ils dans leur rédaction en français et en suédois L2 ? Peut-on trouver des différences dans l'utilisation des SP dans ces deux langues ?
2. Qu'est-ce que l'analyse des processus révèle de l'utilisation des SP ?

Dans la première question, la qualité réfère aux catégories et types de SP analysées (voir le Tableau 1). Dans la section suivante seront présentées les SP et leur classification, ensuite nous présenterons les données et méthodes d'analyse et, pour finir, nous discuterons les résultats de l'analyse.

1. Les séquences préfabriquées (SP)

La définition du phénomène des SP a souvent fait l'objet des recherches antérieures, entre autres, Wray (2002), Forsberg (2006), Sundman (2011), Cordier (2013), Perez-Bettan (2015), Rasmussen (2015) et Myles et Cordier (2017). En effet, la terminologie est riche dans ce domaine et les différents termes ne se réfèrent pas toujours de manière unanime aux mêmes constructions ; le même terme peut renvoyer à des constructions différentes (Myles, Cordier, 2017 : 4). De plus, les critères d'identification des SP peuvent aussi varier (Perez-Bettan, 2015 : 35-36), ce qui reflète les centres d'intérêt différents dans les recherches (Cordier, 2013). Par exemple, Perez-Bettan (2015 : 35-36) liste dans sa thèse, entre autres, les termes suivants utilisés par différents chercheurs provenant de domaines divers (didactique, linguistique, psycholinguistique et sociolinguistique) : *formulaic chunks*, *holophrases*, *gambits*, *lexical phrases*, *phrasèmes*, *preassembled speech*,

prefabricated routines, sentence stems, séquences figées et unité de construction préformée. Il est à noter que les traditions divergent dans les traditions française et anglo-saxonne : selon Forsberg (2006), la première se concentre sur le figement en linguistique de manière générale, alors que la seconde s'intéresse à la fréquence et aux collocations. Il existe des listes terminologiques exhaustives sur le sujet, telles celles réalisées par Wray (2002 : 9) et Perez-Bettan (2015 : 36) ; voir aussi le tableau récapitulatif figurant dans le mémoire de master de Salminen (2020 : 13). Il est à noter que nous avons recours au terme de SP tout au long de cet article, même si les recherches mentionnées utilisent un autre terme. Dans cet article, les SP sont définies, à l'instar de Forsberg (2006 : 44), comme suit :

Une séquence préfabriquée est composée d'au moins deux mots graphiques. Elle est préférée, c'est-à-dire plus fréquente dans un contexte donné, chez les locuteurs natifs, qu'une autre combinaison qui aurait pu être équivalente si la première n'avait pas été conventionnalisée. Or le répertoire des séquences préfabriquées chez les locuteurs L2 diverge de celui des locuteurs natifs. Par conséquent, une séquence préfabriquée peut aussi être une séquence préférée par un locuteur L2, identifiée à cause de son emploi idiosyncrasique (morphosyntaxique, sémantique, pragmatique) et récurrent dans l'interlangue d'un locuteur L2.

Wray a élaboré ses idées concernant les SP dans une étude ultérieure (2008 : 11). Ainsi, selon elle, il est important de distinguer entre, d'une part, l'utilisation des SP dans la langue en général - autrement dit, ce qui est reconnu comme SP par tous - et, d'autre part, l'utilisation qui en est faite par un groupe spécifique ou par un individu - à savoir, les SP typiques dans un discours particulier. Cette distinction de Wray est proche de la catégorisation de Cordier (2013) et de Myles et Cordier (2017), qui ont recours aux termes *speaker-external use* (utilisation centrée sur la langue même) pour caractériser l'approche linguistique et *speaker-internal use* (utilisation centrée sur le locuteur) pour résumer l'idée de l'approche psycholinguistique. Par conséquent, à l'instar de leurs études, nous étudions les SP au niveau individuel et, ainsi, nous pouvons parler des SP propres à l'apprenant et analyser la variation interne et externe des SP produites.

Comme nous l'avons déjà mentionné, les SP aident à récupérer du matériel linguistique dans la mémoire à long terme et même à diminuer le coût des traitements en mémoire de travail (Cislaru, Olive, 2018 : 64), facilitant ainsi les processus verbaux. Or, leur rôle a été principalement étudié dans la langue orale et il existe ainsi une lacune importante par rapport à l'écrit en français L2. De son côté, l'avancement technologique² permet de raffiner l'étude des SP aussi bien à l'oral qu'à l'écrit. Entre autres, les logiciels *ScriptLog* et *GGXLog* permettent d'étudier les

processus derrière les SP et aide à identifier leur forme dans la production écrite parce que, par exemple, les pauses suivent ou précèdent souvent les SP. En général, il n'y a pas de pauses à l'intérieur des séquences automatisées, ce qui peut venir en aide au moment de l'identification des SP, étant donné qu'il n'est pas toujours facile de décider si une séquence de mots forme une SP ou s'il s'agit d'une suite de mots dans un texte linéaire (par exemple <pause> en même temps <pause> vs. <pause> il parle de son chien <pause>, rétrospectivement). Cislaru et Olive (2018 : 69) nous rappellent pourtant que le moindre coût cognitif des SP pourrait induire une pause à l'intérieur d'une SP permettant le chevauchement de plusieurs tâches cognitives en même temps.

2. Données recueillies et catégorisation des SP

Le corpus de cette étude de cas a été collecté auprès de quatre étudiants multilingues de Finlande ; nous reprenons ici la notation utilisée par Salminen (2020) dans son mémoire de master : A, B, C et D. Les étudiants étaient âgés de 22 à 26 ans, leur L1 était le finnois et ils étudiaient l'anglais ou le français comme matière principale et le suédois comme matière secondaire. Il est à noter que la plupart des apprenants en Finlande choisissent l'anglais comme leur première langue étrangère à l'école, à l'âge de 10 ans (à partir du printemps 2020, à l'âge de 7 ans) et que le suédois est la deuxième langue officielle en Finlande, mais le niveau de compétences en suédois varie beaucoup d'un élève à l'autre. Tous les étudiants ont dû passer un concours d'entrée pour pouvoir commencer les études de ces langues au niveau universitaire ; le niveau des compétences linguistiques est au moins le niveau B1 selon les critères du CECR³. Dans notre corpus, trois des quatre participants ont considéré que l'anglais était la langue qu'ils maîtrisaient le mieux. De plus, les étudiants de français matière principale (A et C) considéraient que leurs compétences en français dépassaient celles en suédois, tandis que les étudiants d'anglais matière principale (B et D) pensaient avoir de meilleures compétences en suédois qu'en français. Tous les étudiants ont rempli un formulaire d'informations préalables et de consentement avant de participer à l'expérience.

Les participants ont écrit deux textes, l'un en français et l'autre en suédois, à l'aide du logiciel *ScriptLog*. Ce logiciel enregistre les touches du clavier et les mouvements de la souris et permet d'analyser les processus d'écriture de différentes manières (par exemple, la durée des pauses et les retours sur le texte ; cf. Cislaru et Olive, 2018 ; Lindgren et Sullivan, 2019 ; Mutta, 2020). Nous avons recueilli au total huit productions écrites (produits finis et processus d'écriture).

Les participants ont eu 30 minutes pour écrire le texte sans s'aider d'aucun moyen supplémentaire comme les dictionnaires. Ils ont ainsi dû avoir recours uniquement à leurs propres ressources linguistiques, c'est-à-dire à leur lexique mental. Le premier et le second texte ont été écrits lors de séances séparées, avec un intervalle d'un à quatre jours entre les deux séances ; aléatoirement choisis, deux participants ont commencé par le texte en français, les deux autres par celui en suédois. Des verbalisations rétrospectives avec rappel stimulé (*stimulated recall*) ont été collectées juste après les sessions d'écriture, mais elles ne sont pas traitées en détail dans cet article.

Comme les étudiants n'avaient pas accès aux dictionnaires pendant la rédaction, les sujets de rédaction choisis étaient familiers, liés à leur vie étudiante, pour leur permettre d'élaborer facilement leur texte dans le temps imparti. Les deux sujets proposés étaient les suivants (pour plus de précisions, voir Salminen, 2020 : 28) :

- En français : Qu'est-ce que vous pensez des études en échange ? Quels sont les avantages ? Quels sont les inconvénients ?
- En suédois : Vad tycker du om att jobba vid sidan av studier ? Vilka fördelar eller nackdelar finns det? (Trad. Qu'est-ce que vous pensez du fait de travailler à côté des études ? Quels sont les avantages ou les inconvénients ?)

Sur la base des études antérieures (Forsberg, 2006 ; Sundman, 2011 ; Rasmussen, 2015 ; Jaworska *et al.*, 2015) et en nous appuyant sur notre analyse des processus d'écriture, nous avons catégorisé les SP trouvées en 7 catégories principales avec des sous-catégories, dont nous donnons des exemples dans le Tableau 1 - tableau tiré du travail de Salminen avec de légères modifications formelles (2020 : 23). Parmi ces catégories, il y en a qui ne sont pas définies préalablement, mais qui émergent de l'analyse des processus : c'est notamment le cas des *SP individuelles* et des *SP modifiées* (voir aussi Cordier, 2013 et Cordier, Myles, 2017).

Catégorie	Sous-catégorie/ type	Abréviation	Exemples (fr, sué)
SP Lexicales		L	
	Syntagmes nominaux	SN	études <i>en échange</i> , <i>tid för att</i> 'temps pour'
	Syntagmes adjectivaux	SAdj	<i>väldigt bra</i> 'très bien'
	Syntagmes adverbiaux	SAdv	à <i>l'université</i> , <i>per vecka</i> 'par semaine'
	Syntagmes verbaux	SV	<i>faire des études</i> , <i>ha tillräckligt med</i> 'avoir assez de'
	Syntagmes prépositionnels	SPré	à <i>cause de</i>

Catégorie	Sous-catégorie/ type	Abréviation	Exemples (fr, sué)
SP Discursives		D	
	Marqueurs textuels	MT	<i>il y a, till exempel</i> 'par exemple'
	Marqueurs autorégulateurs et expressions de positionnement	MAEP	<i>je pense que, enligt min åsikt</i> 'à mon avis'
	Marqueurs interactifs	MI	<i>bien sûr</i>
SP Grammaticales		G	
	Expressions de quantité et de degré	QD	<i>beaucoup de, för mycket</i> 'trop beaucoup'
	Pronoms indéfinis	PI	<i>pour quelqu'un, någon annan</i> 'quelqu'un autre'
SP Idiosyncrasiques intérieures		SPIDI	
	Lexicales	L	* <i>tous les choses, på sommarna</i> 'en été'
	Grammaticales	G	* <i>tellement beaucoup</i>
	Discursives	D	* <i>il est aussi le possibilité que, a min åsikt</i> 'à mon avis'
SP Thématiques		T	
	Syntagmes adverbiaux	SAdv	<i>en France, i Finland</i> 'en Finlande'
	Syntagmes nominaux	SN	<i>en français</i>
SP Individuelles		I	
	Toutes les catégories sont possibles		<i>ce pays, extra jobbet</i> 'extra travail'
SP Modifiées		M	
	Toutes les catégories sont possibles		<i>dans le cas</i> → <i>dans le meilleur cas, avsikter om</i> → <i>åsikter om</i> 'avis'

Tableau 1. Catégorisation des SP avec des exemples

Il n'y a pas de façon unanime de classer les SP, et ainsi certains cas auraient pu être catégorisés d'une autre manière. Pour clarifier les catégories, notons que les *SP thématiques* sont liées aux sujets traités et sont ainsi plus fréquentes dans ce corpus qu'elles ne le seraient dans un autre corpus sur un sujet différent. Il en va de même en ce qui concerne les *SP individuelles*, qui représentent des SP chez un certain individu. En d'autres mots, ces groupes de mots sont relativement fréquents et utilisés de manière automatisée chez cet individu. Les *SP modifiées* sont des SP que le participant a produites lors des processus d'écriture mais qui ont été ensuite modifiées pour la version finale, tandis que les *SPIDI* sont des formes déviantes de

la norme. Les *SP individuelles*, les *SP thématiques*, *modifiées* et les *SP idiosyncrasiques intérimaires* décrivent un usage individuel des SP chez les participants.

3. L'utilisation des SP par des apprenants multilingues

3.1. La catégorisation des SP dans la rédaction en français et en suédois L2

Pour répondre à la première question de recherche, nous avons calculé les SP et analysé les différents types de SP. Il ressort de l'analyse que les participants ont écrit un peu plus de mots en suédois qu'en français, tandis que le pourcentage des SP dans les textes rédigés en français était plus élevé que dans les textes en suédois, 30 % et 24 %, respectivement. Les résultats chiffrés au niveau de chaque groupe sont présentés dans le Tableau 2 (voir aussi Salminen, 2020 : 35, 37).

	Nombre de mots dans le produit final	Mots faisant partie des SP dans le produit final	Pourcentage des SP dans le produit final
En français	1315	393	30 %
En suédois	1350	321	24 %

Tableau 2. Nombre de mots contenant des SP en français et en suédois

Une raison pour les différences visibles dans le Tableau 2 se trouve dans la façon de calculer le nombre de mots faisant partie des SP, liée en partie à des différences morphosyntaxiques des langues en question. Pour donner un exemple, en suédois, les formes indéfinies et définies ont une morphologie différente : on écrit *en flicka* 'une fille' (deux mots) vs. *flickan* 'la fille' (un seul mot). Ceci a un impact sur les chiffres obtenus. Il se peut aussi que les participants n'aient pas eu autant de SP automatisées à leur disposition en suédois qu'en français. Au niveau individuel, le nombre des SP varie entre 24 % et 36 % en français et entre 22 % et 29 % en suédois ; les valeurs sont présentées dans le Tableau 3 (voir aussi Salminen, 2020 : 35, 38).

Étudiant	En français			En suédois		
	Nombre de mots dans le produit final	Mots faisant partie des SP dans le produit final	Pourcentage des SP dans le produit final	Nombre de mots dans le produit final	Mots faisant partie des SP dans le produit final	Pourcentage des SP dans le produit final
A	452	128	28 %	484	112	23 %
B	421	150	36 %	434	97	22 %
C	248	60	24 %	206	47	23 %
D	194	55	28 %	226	65	29 %
Moyenne	329	98	30 %	338	80	24 %

Tableau 3. Nombre de mots contenant des SP au niveau individuel

Il ressort du tableau que le nombre des mots produits et le nombre des mots faisant partie des SP varient d'un individu à l'autre. De plus, il existe des différences d'utilisation dépendant de la langue en question. Nous avons également analysé s'il y avait des différences entre les deux langues par rapport à la qualité des SP utilisées en examinant les types de SP. Notre analyse a montré que les textes en français contenaient 149 occurrences (*tokens*) des SP, dont 97 étaient de types (*types*) différents, alors qu'en suédois les nombres étaient de 128 et 82, respectivement. Les sept types différents des SP par langues sont présentés dans le Tableau 4 : les SP lexicales (L), discursives (D), grammaticales (G), thématiques (T), individuelles (I), modifiées (M) et idiosyncrasiques intérimaires (les SPIDI) (voir aussi Salminen, 2020 : 46, 50).

SP	L	D	G	T	I	M	SPIDI	Au total
Français	58	53	10	9	2	7	10	149
Suédois	60	36	14	2	5	3	8	128

Tableau 4. La distribution des SP par catégories

En français et en suédois, les catégories dominantes comprennent les SP lexicales (respectivement, 39 % et 47 %) et les SP discursives (respectivement, 35% et 28 %). Ces résultats corroborent les résultats des études antérieures de Forsberg (2006) et de Rasmussen (2015), mais dans un ordre inversé, comme ces chercheuses ont étudié la langue parlée, où le recours aux formes discursives est plus fréquent. De plus, leurs pourcentages étaient plus élevés, mais il est à noter que nous avons plus de catégories qu'elles. Pour donner un exemple concret des SP utilisées, nous listons, dans les Tableaux 5 et 6 suivants (d'après Salminen, 2020 : 45, 49), les SP ayant au moins 3 occurrences dans le corpus.

SP	Occurrence	Catégorie	Type
En échange	8	L	SAdv
À l'université	7	L	SAdv
Par exemple	6	D	MT
En Finlande	6	T	SAdv
Faire des/les études	6	L	SV
Études en échange	6	L	SN
Il y a	5	D	MT
À mon avis	3	D	MAEP
Pour moi	3	D	MAEP
Parce que	3	D	MT

SP	Occurrence	Catégorie	Type
Types : SAdv= syntagmes adverbiaux, SN= syntagmes nominaux, SV= syntagmes verbaux, MT= marqueurs textuels, MAEP= marqueurs autorégulateurs et expressions de positionnement			

Tableau 5. Les SP le plus fréquentes dans les productions en français

SP	Occurrence	Catégorie	Type
Vid sidan av studier/na, 'parallèlement à des/aux études'	13	L	SAdv
Det är, 'il est/c'est'	5	D	MT
Till exempel, 'par exemple'	5	D	MT
För mig, 'pour moi'	5	D	MAEP
Så mycket, 'tellement'	5	G	QD
Hur mycket, 'combien'	3	G	QD
Types : SAdv= syntagmes adverbiaux, MT= marqueurs textuels, MAEP= marqueurs autorégulateurs et expressions de positionnement, QD= expressions de quantité et de degré			

Tableau 6. Les SP les plus fréquentes dans les productions en suédois

Il ressort des tableaux que les types ou sous-catégories de SP varient légèrement d'une langue à l'autre. Le sujet de rédaction semble avoir influé sur l'utilisation de certaines SP (comme mentionné *supra*), et tout particulièrement des SP lexicales. Même si les consignes de la rédaction étaient relativement courtes, il est évident que plusieurs SP faisaient partie du sujet de rédaction. Nous n'avons pas voulu exclure ces SP de notre corpus, puisque nous étions également intéressées par les processus qui menaient au produit fini. En effet, l'analyse des processus a révélé que les étudiants ont eu recours à des séquences copiées de manières diverses (cf. aussi *infra*). Les résultats montrent aussi que la matière principale n'a pas influencé directement l'utilisation de SP et les compétences perçues par les étudiants non plus (cf. section 3 *supra*). Cela va dans le même sens que les résultats de Jaworska *et al.* (2015 : 519) qui ont trouvé dans leur étude sur les scripteurs L1 et L2 qu'il n'y avait pas de lien direct entre le niveau de compétences langagières et l'utilisation des SP copiées. Les deux catégories de scripteurs avaient eu recours à ce type de SP dans leurs productions. Dans notre cas, il est à noter aussi que nous n'avons pas évalué l'utilisation des SP par rapport aux compétences linguistiques de manière

systématique ; d'ailleurs, le nombre d'occurrences dans le corpus entier n'est pas très élevé.

Les catégories et les types variaient selon les participants, ce qui était prévu, puisque les scripteurs ont des profils d'écriture individuels. Par exemple, le participant A avait en français le plus de SP discursives (21/47, 45 %, par exemple à mon avis, *il y a*), tandis qu'en suédois il avait recours le plus souvent au SP lexicales (22/44, 50 %, par exemple *vida sidan av studier* 'parallèlement à des/aux études'). Quant au participant B, il avait le plus d'occurrence de SP lexicales en français (23/55, 42 %, par exemple à l'université, *en échange*) et en suédois (17/40, 42 %). Les participants C et D avaient des textes plus courts, et par conséquent, le nombre d'occurrences des SP dans leurs textes était moindre par rapport aux participants A et B. Le participant C avait le plus recours aux SP lexicales en français et en suédois (respectivement, 14/24, 58 % et 9/18, 50 %). De la même manière, le participant D avait plus de SP lexicales (respectivement, 8/23, 35 % et 12/26, 46 %) mais par rapport aux autres participants, il était le seul à avoir des occurrences dans chacune des sept catégories.

En somme, les catégories des SP utilisées dans les productions en français et en suédois variaient selon les langues mais il y avait beaucoup de convergences aussi. Pour le participant A, les SP discursives (du type MAEP et MT) et les SP lexicales (du type SAdv et SV) semblaient être automatisées dans ses productions aussi bien en français qu'en suédois. De son côté, le participant B a eu beaucoup recours à des SP similaires, dont les SP lexicales, dans ses productions dans les deux langues. Le participant C, quant à lui, a fait beaucoup de pauses pendant la rédaction et ses pauses ont été relativement longues. Cela nous a aidées à identifier les SP plus facilement (cf. *infra*), mais en même temps, comme le temps de rédaction était limité, il n'a pas produit autant de texte ni de SP que les participants A et B ; les SP lexicales étaient les plus utilisées dans ses productions. De la même manière, le participant D a rédigé un texte plutôt court, mais, contrairement aux autres, il a utilisé des SP de chaque catégorie dans ses productions en français et en suédois. Mais, d'un autre côté, les pourcentages des SP dans les deux langues étaient relativement élevés, ce qui introduit la question de la répétitivité du texte final - nous n'avons pas évalué ici la qualité du texte, ce qui reste à faire ultérieurement, puisque l'utilisation abondante des mêmes SP peut témoigner des compétences langagières réduites. Ainsi, il existe beaucoup de variations individuelles dans notre corpus et on ne peut pas tirer de conclusions, vu la petite taille du corpus. Il reste à examiner ce que révèlent les processus d'écriture quant à l'utilisation des SP.

3.2. Les SP dans les processus d'écriture

Nous avons également examiné les processus d'écriture de chaque participant, *primo*, pour identifier les SP utilisées, et *secundo*, pour étudier l'automatisation des processus de traitement des SP, à savoir si ces SP sont utilisées de manière automatique sans pauses (cf. *supra*). Lors des processus d'écriture, les SP ont été identifiées à l'aide des pauses et plus précisément des *jets textuels* qui segmentent les processus. Il s'agit des séquences textuelles que les scripteurs produisent entre deux pauses de deux secondes ; ces jets correspondent à l'écriture fluente (Cislaru, Olive, 2018 : 44-46). Nous illustrons cela avec l'exemple suivant ; les exemples sont tirés de Salminen (2020) et ils sont tous présentés sous leur forme originale, avec des erreurs éventuelles :

Exemple 1. [...] <2.251> <BACKSPACE1>, *je pense qu'on* <2.255> doit le faire. <12.048> [...] (participant D)

Puisqu'il n'y a pas de pause à l'intérieure de l'unité *je pense qu'on*, nous pouvons la considérer comme une SP (ici, une SP discursive, de type MAEP, marqueurs autorégulateurs et expressions de positionnement). Le participant semble avoir produit cette SP de manière automatisée, c'est-à-dire qu'il a récupéré ces mots comme un tout.

Les jets textuels ont aidé à identifier la catégorie des SP individuelles qui était une catégorie émergente lors de l'analyse (cf. *supra*). Tous les étudiants n'y ont pas recours, mais illustrons ce phénomène avec l'exemple 2 suivant trouvé chez le participant D.

Exemple 2. [...] <3.243> <BACKSPACE1> et aussi le culu<BACKSPACE1>ture <3.002> de <2.145> <18.289> *ce pays*. <4.418> [...] seulement le culture de pays <BACKSPACE5>*ce pays* qu'on peut <6.780> recevoir. [...] (participant D)

Dans l'exemple 2, nous considérons la suite de mots *ce pays* comme étant une SP individuelle, car le participant les utilise de manière particulière : une longue pause de réflexion précède l'expression et elle est entourée d'autres pauses aussi. Le mot *culture* est lié à cette SP plus loin. En général, ce type de suites de mots ne constitue pas une SP ; nous discuterons cette interprétation dans des recherches ultérieures.

En principe, tous les participants ont utilisé plus de SP lors des processus d'écriture que dans le produit terminé : en français le nombre était de 167 (vs. 149 dans le produit final) et en suédois de 137 (vs. 128) (Salminen, 2020 : 46, 50, 52). Cela signifie que certaines SP ont disparu pendant la construction du texte dû aux modifications et choix lexicaux et/ou grammaticaux et d'autres SP ont disparu dû

à l'hésitation dans leur emploi (cf. *infra*). Ce dernier témoigne de l'instabilité de l'utilisation des SP et, par conséquent, des formes non automatisées. L'exemple 3 illustre ce genre d'hésitation :

Exemple 3. [...] <6.359> à <BACKSPACE2> À <2.535> vivre <4.080> da<BACKSPACE2> <4.124> dans un pays étrangere<BACKSPACE1> <4.860> <LEFT18> <2.553> <BACKSPACE4> en <4.961> <RIGHT18> <9.762> (participant D)

Dans cet exemple, le participant D a produit la suite de mots *dans un pays étranger* - en supprimant la lettre *e* tout de suite après l'avoir écrite - en tant qu'unité, ici une SP modifiée représentant en même temps une SPIDI (lexicale, de type SAdv). L'unité est entourée de pauses mais aucune interruption ne se trouve à l'intérieure de l'expression. Or, le participant a hésité quelle préposition devrait être utilisée et il a fini par la remplacer par la préposition *en*. Dans le produit final, nous trouvons la phrase *À vivre en un pays étranger* sans traces d'une hésitation quelconque, ce qui montre combien il est important d'étudier aussi les processus pour comprendre le fonctionnement et l'apprentissage des SP.

L'exemple 4 montre comment l'écriture du participant A était influencée par la langue dominante. Chez cet étudiant, la catégorie des SPIDI est plus marquante en suédois qu'en français. Les trois SPIDI en suédois sont restées les mêmes dans le processus et dans le produit final, dont une représente la SP discursive (de type MAEP).

Exemple 4. [...] <6.399> Ja<BACKSPACE2> A min asivkt är <2.917> [...] /[...] <6.399> [Je]<BACKSPACE2> A mon avis est' <2.917> [...] (participant A)

Cet exemple illustre bien le fait que l'expression française à mon avis a eu un impact sur l'expression correspondante en suédois *enligt min åsikt* 'selon mon avis'. Il existe une expression en suédois dans le même sens qui est encore plus similaire, notamment *å min åsikt* 'à mon avis', mais elle est beaucoup plus rare dans l'usage. Ce qui est intéressant ici aussi est que l'étudiant a commencé à écrire *À min* en mélangeant les deux langues, mais en supprimant ce texte sur-le-champ. Lors de la verbalisation après la rédaction (voir *supra*), cet étudiant a commenté qu'il avait « en effet pensé à comment dire 'à mon avis' en suédois et il a[vait] constaté aussi qu'il s'était souvenu de plus de mots en français qu'en suédois » (Salminen, 2020 : 68). L'étudiant a, de plus, constaté dans le questionnaire que ses compétences langagières étaient meilleures en français qu'en suédois, le français étant sa matière principale à l'université.

Conclusion

Notre analyse illustre bien que l'utilisation des SP dans les productions des apprenants de langues étrangères varie d'une langue à l'autre, mais qu'il y a aussi des convergences. Pour répondre à notre première question de recherche, nous pouvons conclure que les participants ont utilisé plus de SP en français qu'en suédois et que, dans ces langues, les catégories dominantes étaient les SP lexicales et SP discursives. Le nombre d'occurrences variait selon les langues. Dans les recherches antérieures sur le français oral L2 (Forsberg, 2006 ; Rasmussen, 2015), la catégorie des SP discursives était la plus utilisée, suivie par les SP lexicales. Par rapport à leurs études, les pourcentages des SP dans les productions étaient plus élevés, ce qui peut être dû au mode de production verbale. Toutefois, la taille limitée de notre corpus ne permet pas de tirer de conclusions définitives à ce sujet.

Ayant eu recours aux nouveaux moyens technologiques, nous avons pu ouvrir de nouvelles perspectives dans l'étude de l'interlangue des apprenants. Nous avons pu distinguer des manières individuelles et diversifiées dans le recours aux SP, c'est-à-dire des profils d'utilisateurs individuels des SP. Les participants ont varié leur utilisation de différents types de SP entre les deux langues, même si les catégories les plus utilisées avaient été les mêmes dans leurs deux productions en français et en suédois.

De plus, l'analyse des processus nous a permis d'identifier les SP à l'aide des jets textuels entourés de pauses et elle a révélé aussi que le produit final ne donnait qu'un point de vue partiel sur l'utilisation des SP. Par l'intermédiaire de l'étude des processus d'écriture, nous avons pu déterminer que les séquences produites étaient en effet des SP. Le processus nous a montré, entre autres, que certaines séquences étaient produites plusieurs fois comme unités lors du processus, même si dans le produit final il n'y avait qu'une seule séquence ; cela témoigne, à notre avis, du fait que ces séquences doivent être considérées comme des SP si elles remplissent d'autres critères, notamment la location des pauses. Donnons un exemple du processus du participant A : lors de la rédaction en suédois, il a écrit trois fois la SP *det är* 'il est', mais dans le produit fini, cette expression n'apparaît qu'une seule fois. Par conséquent, cette expression représente une SP dans le lexique mental du participant. Cette approche permet ainsi d'étudier le phénomène des SP de manière dynamique et de relever des aspects non observables auparavant.

Pour conclure, nous voulons souligner que l'analyse des processus d'écriture par l'intermédiaire d'un logiciel comme *ScriptLog* ou *GGXLog* est très pertinente. Il est même nécessaire, en étudiant les SP dans les productions écrites en L2, d'examiner non seulement les produits finis et terminés, mais aussi les processus. Les processus

ont révélé des informations importantes, par exemple des SP automatisées chez les participants qui semblaient faciliter la production écrite. Les résultats de cette étude contribuent au savoir sur l'utilisation des SP en L2. Les résultats ont aussi un intérêt pédagogique auprès des apprenants eux-mêmes mais aussi chez les enseignants quand ils présentent des moyens stratégiques ; dans l'enseignement, on pourrait, par exemple donner des listes de différents types d'expressions figées telles que *grièvement blessé* mais *gravement malade*, *faire attention*, *prendre une décision* - en finnois on dit **faire une décision* - ou l'on pourrait enseigner l'utilisation des marqueurs textuels à l'écrit, par exemple *de plus*, *en outre*, *je crois que*, etc. Tout cela aide les apprenants à surmonter des problèmes éventuels lors des activités rédactionnelles en langues étrangères.

Bibliographie

Cislaru, G., Olive, T. 2018. *Le processus de textualisation : Analyse des unités linguistiques de performance écrite*. Louvain-la-Neuve : De Boeck Supérieur.

Cordier, C. 2013. *The presence, nature and role of formulaic sequences in English advanced learners of French: a longitudinal study*. Thèse de doctorat. Université de Newcastle. [En ligne] : <http://theses.ncl.ac.uk/jspui/handle/10443/2265> [consulté le 20 août 2020].

Erman, B., Warren, B. 2000. « The idiom principle and the open choice principle ». *Text*, n° 20(1), p. 29-62.

Forsberg, F. 2006. *Le langage préfabriqué en français parlé L2 : étude acquisitionnelle et comparative*. Thèse de doctorat. Université de Stockholm. [En ligne] : https://www.researchgate.net/publication/267978735_Le_langage_prefabrique_en_francais_parle_L2_Etude_acquisitionnelle_et_comparative [consulté le 20 août 2020].

Jaworska, S., Krummes, C., Ensslin, A. 2015. « Formulaic sequences in native and non-native argumentative writing in German ». *International Journal of Corpus Linguistics*, n° 20(4), p. 500-525. [En ligne] : <https://doi.org/10.1075/ijcl.20.4.04jaw> [consulté le 20 août 2020].

Lindgren, E., Sullivan, K. P. H. (éds.) 2019. *Observing writing: insights from keystroke logging and handwriting*. Leiden: Brill.

Mutta, M. 2020. L2 fluency and writer profiles. In : P. Lintunen, M. Mutta, P. Peltonen (éds.), *Fluency in L2 learning and use*. Bristol: Multilingual Matters, p. 63-80.

Myles, F., Cordier, C. 2017. « Formulaic sequence (FS) cannot be an umbrella term in SLA: Focusing on psycholinguistic FSs and their identification ». *Studies in Second Language Acquisition*, n° 39(1), p. 3-28. [En ligne] : <https://doi.org/10.1017/S027226311600036X> [consulté le 20 août 2020].

Perez-Bettan, A. 2015. *Apprentissage et utilisation du langage préfabriqué chez des apprenants de français langue étrangère*. Thèse de doctorat. Université Paris 8 - Vincennes - Saint-Denis. [En ligne] : <https://tel.archives-ouvertes.fr/tel-01714531/document> [consulté le 20 août 2020].

Rasmussen, L. 2015. *L'usage des séquences préfabriquées dans le langage parlé d'étudiants finnophones du français L2 au niveau universitaire : Une étude de cas*. Mémoire de master. Université de Turku.

Salminen, S. 2020. *Les séquences préfabriquées dans les productions écrites des étudiants du français L2 et du suédois L2*. Mémoire de master. Université de Turku. [En ligne] : <https://www.utupub.fi/handle/10024/149713> [consulté le 20 août 2020].

Sundman, M. 2011. « Helsekvenser i finska grundskoleelevers skrivande ». *Svenskans beskrivning*, n° 31, p. 327-336.

Vassiliadou, H. 2020. « Quels liens entre la pause, la syntaxe et les marqueurs discursifs ? ». In : F. Hirsch, I. Didirková, C. Dodane (éds.), *Manuel de pausologie. Recueil de recherches sur les pauses présentes dans la parole et le discours*. Paris: L'Harmattan, p. 51-79.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. 2008. *Formulaic sequences: pushing boundaries*. Oxford: Oxford University Press.

Notes

1. Le projet de recherche sur l'aisance rédactionnelle et la visualisation a été initié par Maarit Mutta ; le mémoire de master de Susanna Salminen a été réalisé dans le cadre de ce projet et a permis de tester la méthodologie du projet.

2. Entre autres, le logiciel PRAAT® peut être utilisé pour étudier le lien entre les pauses, la prosodie et les marqueurs discursifs (Vassiliadou, 2020), tandis que le logiciel *GenoGraphiX-Log* (GGXLog) permet en plus la visualisation des processus rédactionnels. Ce logiciel est disponible à partir du site <https://ggxlog.net/index.php>.

3. *Cadre européen commun de référence pour les langues* (CECR). Disponible sur : <https://www.coe.int/fr/web/common-european-framework-reference-languages>.