

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
FILOZOFICKÁ FAKULTA
ÚSTAV ROMANISTIKY

DIPLOMOVÁ PRÁCE

MÉMOIRE DE MASTER

Constructions verbonominales françaises du domaine de la santé
et leur comparaison avec le tchèque

Vedoucí práce / Directeur du mémoire : prof. PhDr. Jan Radimský, Ph. D.

Autor práce / Auteur du mémoire : Petr Březina

Studijní obor / Discipline étudiée : Francouzský jazyk

Ročník / Année : 2

2021

Je déclare que je suis l'auteur de ce mémoire et que je ne l'ai préparé qu'en utilisant les sources et la littérature mentionnées dans la liste des sources utilisées.

Le 30 juillet 2021

Petr Březina

Je tiens à remercier M. le professeur Radimský de m'avoir initié à la linguistique de corpus et d'avoir dirigé mon mémoire de master. J'exprime mes remerciements également à tous mes autres enseignants de l'Université de Bohême du Sud pour avoir élargi l'horizon de mes connaissances et à mes collègues étudiants qui ont contribué à une ambiance amicale et stimulante pendant mon séjour à České Budějovice.

Petr Březina

Anotace

Předkládaná diplomová práce je zaměřena na francouzské verbonominální kolokace a na jejich české překlady. Obsahuje dvě praktické úlohy: první úloha spočívá ve vytvoření srovnatelného francouzsko-českého korpusu textů z oblasti zdravotnictví; následně se korpus stává primárním zdrojem pro druhou úlohu, v níž je studována množina verbonominálních kolokací nalezených ve francouzském modulu korpusu, v českém modulu jsou potom dohledány jejich možné překlady a je provedena translatologická analýza. Praktické části předchází čtyři teoretické kapitoly, z nichž první se zaměřuje na vymezení pojmu kolokace, druhá se věnuje některým aspektům klasifikace verbonominálních kolokací, ve třetí se přechází k teorii překladu a čtvrtá se zabývá problematikou týkající se vytvoření korpusu, zejména srovnatelného korpusu. K práci je připojen francouzsko-český slovníček.

Klíčová slova

francouzština; čeština; zdravotnický jazyk; korpus; kolokace; translatologie

Annotatio

In hoc libello ad magistri gradum consequendum scripto de Francogallicis collocationibus uerbonominalibus eorumque translationibus Bohemicis disseritur. Libello duo opera practica continentur, quorum unum spectat ad conficiendum corpus comparabile textuum Francogallicorum et Bohemicorum ad curam ualetudinis pertinentium, quod deinde fit fons primarius in altero opere efficiendo, auctor enim certum numerum collocationum uerbonominalium in corporis parte Francogallica repertarum perscrutatur atque in corporis parte Bohemica quaerit uocabula quibus transferri possint et analysim translaticam facit. Partem practicam praecedunt quattuor capita theoretica : primum collocationis notioni definiendae opera datur, deinde classificatio collocationum uerbonominalium attingitur, tum ad theoriam translationis transitur, postremo ea tractantur quae ad corpus conficiendum, praesertim comparabile, pertinent. Libello additum est uocabularium Francogallico-Bohemicum.

Vocabula quae clauae uocantur

lingua Francogallica; lingua Bohemica; lingua in cura ualetudinis usitata; corpus; collocatio; translaticologia

Annotation

This master's thesis deals with French verbonominal collocations and their Czech translations. It contains two practical tasks: the first task consists in creating a comparable French-Czech corpus of healthcare texts; subsequently, the corpus becomes the primary source for the second task, in which several verbonominal collocations found in the French module of the corpus are studied, their possible translations are looked up in the Czech module of the corpus, and a translational analysis is made. Practical part is preceded by four theoretical chapters focused on: the definition of collocation, some aspects of the classification of verbonominal collocations, the theory of translation, and issues related to the creation of a corpus, especially a comparable corpus. A French-Czech vocabulary is attached to the thesis.

Keywords

French language; Czech language; healthcare language; corpus; collocation; translational

Table des matières

1.	Introduction	9
2.	Collocations	11
2.1.	Collocations dans la perspective d'E. JEŽEK	11
2.2.	Collocations dans la perspective d'I. MEL'ČUK	16
2.3.	Collocations dans la perspective de F. ČERMÁK	19
2.4.	Conclusions	21
3.	Classification des collocations verbonominales	22
3.1.	Classification selon les critères syntaxiques	22
3.2.	Fonctions lexicales	24
3.3.	Conclusions	27
4.	Approche traductologique	28
4.1.	Équivalence fonctionnelle	28
4.1.1.	Aspect stylistique	31
4.1.2.	Aspect pragmatique	35
4.2.	Différences typologiques entre le français et le tchèque	37
4.3.	Procédés de traduction	41
4.3.1.	Stylistique comparée	41
4.3.2.	Emprunt	42
4.3.3.	Calque	43
4.3.4.	Traduction littérale	44
4.3.5.	Transposition	45
4.3.6.	Modulation	47
4.3.7.	Équivalence et adaptation	47
4.4.	Conclusions	47
5.	Corpus	49
5.1.	Corpus comparable	50
5.2.	Choix de textes pour un corpus	53
5.2.1.	Représentativité du corpus	53

5.2.2.	Classification de textes	55
5.2.3.	Étendue temporelle	57
5.2.4.	Taille du corpus	58
5.3.	Préparation de textes pour un corpus	59
5.4.	Exploitation du corpus	63
5.5.	Conclusions	64
6.	Partie pratique	67
6.1.	Création d'un corpus comparable	67
6.2.	Objectifs et méthode de la recherche	68
6.3.	Recherche	69
6.3.1.	Collocations verbonominales françaises	69
6.3.2.	Collocations françaises et leurs traductions tchèques	71
6.4.	Conclusions	76
7.	Liste des sources utilisées	78
8.	Résumé	87
	ANNEXE : Vocabulaire français-tchèque	90

1. Introduction

Le présent mémoire est consacré aux collocations verbonominales françaises et à leurs traductions tchèques. Il comporte deux tâches pratiques : la première tâche consiste à créer un corpus comparable français-tchèque de textes du domaine de la santé ; par la suite, le corpus devient notre source primaire pour la seconde tâche, dans laquelle nous étudions un ensemble de collocations verbonominales trouvées dans le module français du corpus, nous essayons d'identifier, dans le module tchèque du corpus, leurs traductions possibles et nous faisons une analyse traductologique. La partie pratique est précédée de quatre chapitres théoriques : tout d'abord, nous nous concentrons sur la délimitation de la notion de la collocation ; ensuite, nous traitons différents aspects de la classification des collocations verbonominales en prenant en considération les rapports syntaxiques dans le champ valentiel du verbe et les fonctions lexicales ; puis, nous passons à la théorie de traduction ; enfin, nous nous occupons de la problématique liée à la création d'un corpus, en particulier d'un corpus comparable. Un résumé tchèque et un vocabulaire français-tchèque sont attachés à ce mémoire.

Pour les citations des sources, nous utilisons le système numérique, appelé aussi de Vancouver. La liste des sources est classée selon l'ordre dans lequel les sources sont citées pour la première fois dans le texte. Les sources non citées, typiquement des dictionnaires, se trouvent dans un ordre logique à la fin de la liste. Nous tenons pour nécessaire d'inclure dans la liste des sources notre appareil cognitif formé au cours de notre vie, sans lequel la liste ne pourrait pas être complète. Dans la liste, un numéro univoque est associé à chaque source ; dans le texte, nous renvoyons à la source en indiquant son numéro entre crochets. S'il en est besoin, la localisation précise (d'ordinaire la pagination) se trouve entre les crochets séparée du numéro de la source par une virgule. Plusieurs sources, séparées par des points-virgules, peuvent être incluses entre les mêmes crochets.

Quant aux citations latines, nous adoptons l'orthographe utilisée par les latinistes français, c.-à-d. nous ne distinguons pas entre l'*i* voyelle et l'*i* consonne ni entre l'*u* voyelle et l'*u* consonne ; cela concerne également l'abréviation *s. u.* = *sub uoce* qui renvoie à

l'article d'un dictionnaire. Les autres abréviations sont soit habituelles (*cf.*, *etc.*) soit intuitives (*fr.* = *français*, *tch.* = *tchèque*, *all.* = *allemand*).

Les règles typographiques françaises [1, p. 79-80 et 104-5] veulent que les mots étrangers soient composés en italique dans un texte en romain ; c'est notamment le cas des locutions latines, telles que *ibidem*, *infra*, *passim* ; cependant, les mots qui sont entrés dans le langage courant tels que *week-end*, *match*, *leitmotiv*, les abréviations *cf.* et *etc.* ainsi que les noms d'organismes étrangers se composent en romain. En général, nous suivons ces règles qui nous poussent à mettre en italique notamment les abréviations *et al.* et *s. u.* tandis que les abréviations *cf.* et *etc.* restent en caractères romains, mais elles ne nous empêchent pas de mettre en italique dans cet alinéa les deux dernières abréviations parce que l'emploi de l'italique est ici justifié par d'autres raisons. Nous jugeons approprié de remarquer à cet endroit que, en ce qui concerne la ponctuation après les mots en italique, nous avons opté, dans le présent mémoire, pour l'usage ancien qui se diffère de celui qui est prôné par le *Lexique des règles typographiques en usage à l'Imprimerie nationale* [1, p. 148] ; nous mettons donc dans ce cas la ponctuation systématiquement en italique.

2. Collocations

La maîtrise active d'une langue comporte la habileté de combiner les mots du point de vue lexical. Pour produire un texte bien formé, il faut savoir quels mots peuvent aller ensemble ; sinon, le texte peut rester compréhensible, mais le locuteur natif le percevra comme peu naturel, bizarre ou incorrect. À cet égard, Igor MEL'ČUK [2, p. 3-4] note que le lexique et la grammaire ne suffisent pas pour parler bien une langue et conclut que « le natif parle en phrasèmes ».

Un type particulier des combinaisons de mots prend le nom de collocations. Le statut de ces dernières parmi les combinaisons de mots varie selon les auteurs. Dans les sections qui suivent, nous nous occuperons de trois vues différentes des collocations.

2.1. Collocations dans la perspective d'E. JEŽEK

Pour déterminer quels mots sont susceptibles de se combiner, Elisabetta JEŽEK [3, p. 190-4] distingue trois types de contraintes des combinaisons de mots : 1^o restrictions conceptuelles (ou ontologiques), 2^o restrictions lexicales basées sur une solidarité sémantique et 3^o restrictions lexicales basées sur une solidarité institutionnalisée par l'usage. En fait, en cadre des deux premières contraintes, la capacité de se combiner est définie plutôt de manière négative par le conflit qui se produit par le non-respect de ces contraintes.

Les restrictions conceptuelles sont fondées sur la connaissance de la réalité extralinguistique et sur les relations logiques. La violation de ces restrictions conduit à un conflit conceptuel ; un tel conflit se trouve dans la fameuse phrase que Noam CHOMSKY [4, p. 116] donne comme exemple d'une phrase grammaticalement correcte mais dont aucun couple des mots n'a vraisemblablement jamais apparu ensemble auparavant, évidemment à cause de leurs incompatibilités sémantiques : *Colorless green ideas sleep furiously* « D'incolores idées vertes dorment furieusement ». František ČERMÁK [5, p. 10] en déduit la priorité des règles sémantiques qui régissent les combinaisons de mots par rapport aux règles grammaticales. Cependant, nous sommes d'avis que le conflit conceptuel n'empêche pas absolument la combinaison : c'est la figure rhétorique

d'oxymore qui est fondée précisément sur ce conflit ; ainsi, selon SUÉTONE [6, *Aug.* 25], l'empereur Auguste disait souvent le proverbe grec « Hâte-toi lentement », où la contradiction entre le verbe et l'adverbe retient l'attention de l'auditeur, le fait réfléchir, et finalement c'est l'adverbe qui prévaut en prenant la fonction d'une négation.

Eugenio COSERIU [7] développe, dans la tradition structuraliste, le concept de la solidarité lexicale. Elle repose sur le fait qu'un lexème est inclus dans le contenu définitoire d'un autre lexème et fonctionne comme un de ses traits distinctifs, par exemple *chien* est contenu dans *aboyer* qui signifie « pousser son cri, en parlant du chien » [8, *s. u.*]. Un lexème (*chien*) en détermine un autre (*aboyer*) ; il s'agit donc d'une relation orientée. L'élément déterminant de cette relation admet une hiérarchisation de sorte qu'un lexème peut être déterminé par un groupe ou une classe de lexèmes ; ainsi, le lexème *pasteuriser* est déterminé par le lexème *lait* mais aussi par les lexèmes *bière*, *vin*, *jus de fruits*, etc., qui se regroupent tous sous l'archilexème *liquide fermentescible*.

Dans différentes langues, la lexicalisation est effectuée différemment ; par exemple, pour exprimer « se marier », le tchèque dispose de deux verbes : *oženit se* s'utilise en parlant d'un homme et *vdát se*, en parlant d'une femme. Le verbe français *aller* correspond, entre autres, aux verbes tchèques *jít* et *jet* ; le verbe *jít* entretient la solidarité lexicale avec le lexème *pěšky* « à pied », le verbe *jet*, avec d'autres modes de locomotion, tels que sur un animal (cheval, âne, chameau, etc.), en voiture, en bus, en train, à vélo ou à skis. L'allemand fait une différenciation encore plus détaillée : l'équivalent allemand du tchèque *jít* est *gehen*, mais le sens du verbe tchèque *jet* se divise entre *reiten* « aller sur un animal » et *fahren* « aller mais autrement qu'à pied ou sur un animal ».

E. JEŽEK emprunte le concept de la solidarité lexicale pour définir le conflit lexical. Il se produit lorsque l'on dit, par exemple : *Le cheval a aboyé*, ou lorsqu'un étranger trouve dans son dictionnaire le verbe *vdát se* comme un des équivalents tchèques du verbe *se marier* et qu'il formule, en tchèque, la phrase suivante : *Pavel se vdal*. À la différence du conflit conceptuel, le conflit lexical est réversible ; on peut y remédier soit par l'emploi du mot approprié : *Le cheval a henni* ; *Pavel se oženil* « Paul s'est marié », soit, selon E. JEŽEK, par l'emploi d'un hyperonyme : *Le cheval a poussé un cri* ; *Pavel uzavřel sňatek* « Paul a contracté mariage ».

D'après E. COSERIU, lorsque l'on transmet les propriétés du lexème déterminant à un autre lexème, une métaphore voit le jour ; par exemple, il y a une solidarité lexicale entre *mordre* (déterminé) et *dent* (déterminant) ; quand on dit : *Le froid mord*, on représente le froid comme un être pourvu de dents.

Quant à la solidarité institutionnalisée par l'usage, il existe, dans la langue donnée, des combinaisons de mots préférées aux autres combinaisons possibles du même sens sans que le locuteur natif soit souvent capable d'expliquer clairement cette préférence. Par exemple, en tchèque, les noms de pays se combinent d'ordinaire avec la préposition *v* « dans », mais certains noms de pays se combinent avec la préposition *na* « sur » : *v Rakousku* « en Autriche » × *na Slovensku* « en Slovaquie ».

Abstraction faite de cas spéciaux tels que l'oxymore, chaque combinaison respecte, au moins, des restrictions conceptuelles. Les deux autres classes de restrictions entrent dans la classification des combinaisons de mots. En effet, E. JEŽEK [3, 195-210] classe les combinaisons de mots sur la base de trois critères : 1^o les contraintes lexicales, 2^o la prévisibilité du sens de la combinaison et 3^o la variabilité lexicale et syntaxique de la combinaison.

Le sens d'une combinaison de mots est prévisible lorsqu'il est composé du sens des mots qui la constituent. La variabilité lexicale consiste dans le fait qu'un mot de la combinaison peut être remplacé par d'autres mots, par exemple : *écrire / lire / publier / acheter / trouver / voir / perdre un livre*. La variabilité syntaxique permet diverses transformations telles que la relativisation ou la passivation, par exemple : *Paul a écrit un livre — le livre que Paul a écrit — le livre a été écrit par Paul*.

La classification d'E. JEŽEK comporte les types de combinaisons de mots suivants : combinaisons libres, combinaisons restreintes, collocations, constructions à verbe support, expressions idiomatiques.

Les combinaisons libres appliquent des restrictions conceptuelles, leur sens est compositionnel et prévisible, elles admettent la variabilité lexicale et syntaxique.

Les combinaisons restreintes sont fondées sur une solidarité lexicale, leur sens est compositionnel, elles admettent la variabilité syntaxique, mais leur variabilité lexicale est restreinte à différents degrés. Par exemple, dans la combinaison *nez aquilin*, le mot

nez ne peut être remplacé par un autre ; dans la combinaison *pasteuriser le lait*, on peut remplacer *le lait* par un nombre restreint de mots (*la bière, le vin, le jus de fruits*) ; d'autres verbes peuvent se combiner avec des groupes de mots plus nombreux. Plus le groupe est hétérogène, moins la combinaison est restreinte. Il nous semble donc que l'on peut imaginer une échelle de combinaisons de mots qui exprimerait la diversité (ou inversement la spécificité) de mots qui peuvent s'associer avec le mot concerné ; l'un bout de cette échelle serait occupé par des combinaisons telles que *nez aquilin* et *le chien aboie* où se manifeste la solidarité lexicale avec un seul mot et l'autre bout, par des combinaisons du genre *vendre* + la classe de toutes les entités dont la possession peut être transmise à quelqu'un d'autre (objets matériels aussi bien qu'immatériels, voire des personnes, cf. l'esclavage). Cette échelle montrerait la transition graduelle entre les combinaisons restreintes et les combinaisons libres.

Une collocation est une combinaison de mots restreinte pour laquelle la sélection d'un mot particulier (collocatif) pour exprimer un sens donné est influencé par un autre mot (base) auquel ce sens s'applique. Le rôle d'un collocatif est souvent joué par une expression d'intensification ; quoique les adjectifs *haut, profond* et *battant* puissent exprimer une grandeur ou une grande intensité, on ne peut les combiner arbitrairement avec les substantifs *qualité, silence* et *pluie*, on dit : *une haute qualité, un silence profond* et *une pluie battante* ; c'est l'usage qui a institutionnalisé ces combinaisons.

E. JEŽEK ne prend pas en considération la variabilité de l'étendue de groupes des mots avec lesquels différents mots peuvent se combiner, et fait une distinction binaire : alors que dans une combinaison restreinte, l'adjectif ou le verbe implique le substantif⁽¹⁾ (*aquilin* impliquerait *nez* ; *garer* impliquerait *véhicule*), dans une collocation, une telle implication n'est pas présente (*haut* n'impliquerait pas *qualité* ; *prêter* n'impliquerait pas *attention*). Par conséquent, elle formule le test suivant : Est-ce qu'un locuteur étranger qui connaît le sens des mots individuels mais non celui de combinaisons de mots d'une langue peut deviner le nom à partir de l'adjectif ou du verbe ? Peut-il deviner *nez* à partir de *aquilin* ? Peut-il deviner *qualité* à partir de *haut(e)* ?

1. En fait, il faudrait parler d'un groupe restreint de substantifs plutôt que d'un seul substantif.

La direction de la restriction dans les combinaisons restreintes et dans les collocations est, selon E. JEŽEK, opposée : dans les combinaisons restreintes, c'est l'adjectif ou le verbe qui impose la restriction sur le choix du nom, dans les collocations c'est le nom qui sélectionne l'adjectif ou le verbe.

Le sens des collocations est prévisible. Elles admettent la variabilité syntaxique, le choix du collocatif est restreint par l'usage.

Les constructions à verbe support représentent un type de collocations. Elles se composent d'un verbe qui contribue très peu au sens de la collocation et d'un nom d'événement sur lequel repose la charge sémantique principale ; celui-ci joue le rôle d'un prédicat nominal auquel le verbe support prête des catégories grammaticales telles que le temps et le mode. Le verbe support peut aussi exprimer le mode d'action.

À la différence des autres collocations qui admettent la variabilité syntaxique, les constructions à verbe support se divisent en deux groupes : les unes sont syntaxiquement flexibles, par exemple l'anglais *make a call*, les autres sont syntaxiquement figées ou à demi figées, par exemple l'anglais *pay attention*. E. JEŽEK explique la différence entre ces deux groupes par la référentialité du nom : les constructions dans lesquelles le nom est non-référentiel ne sont pas syntaxiquement flexibles.

Étant donné que les constructions à verbe support correspondent souvent à des verbes synthétiques, par exemple *faire une promenade — se promener*, E. JEŽEK leur attribue le statut d'unités lexicales et de prédicats analytiques.

Les expressions idiomatiques, quant à elles, ne sont pas transparentes, leur sens n'est pas compositionnel, elles sont plutôt créées en bloc par recours à l'usage figuratif de la langue, par exemple *vider son sac*, signifie « dire le fond de sa pensée ; avouer une chose que l'on tenait cachée » [8, s. u. « sac »] ; cette expression repose sur l'image de choses qui sont cachées dans un sac ; lorsque l'on les fait sortir du sac, elles se présentent à la vue. Du point de vue sémantique et syntaxique, l'expression idiomatique se comporte comme un seul mot.

2.2. Collocations dans la perspective d'I. MEL'ČUK

I. MEL'ČUK [2] part de la distinction entre les syntagmes libres et non libres. Ils se distinguent selon qu'ils sont construits de façon régulière et de façon non contrainte. Si l'on s'appuie uniquement sur un dictionnaire de lexèmes et les règles de grammaire, on produit un syntagme libre. L'opposé est le syntagme non libre, appelé aussi phrasème. Les phrasèmes se divisent en phrasèmes pragmatiques (ou pragmatèmes) et phrasèmes sémantiques. Les phrasèmes sémantiques se divisent à leur tour en trois sous-classes : phrasèmes complets, semi-phrasèmes et quasi-phrasèmes.

Les pragmatèmes sont conditionnés par la situation où l'on utilise des expressions conventionnelles pour transmettre un contenu informationnel, par exemple, au téléphone, on dit : *Ne quittez pas !* On ne peut pas le remplacer par un autre syntagme qui exprime la même idée, par exemple : *N'interrompez pas cet appel !*

Les phrasèmes complets correspondent, en principe, aux expressions idiomatiques dans la classification d'E. JEŽEK. I. MEL'ČUK en distingue deux types. Le premier type, beaucoup plus fréquent, comporte les phrasèmes dont le sens n'inclut pas le sens d'aucun de leurs constituants, par exemple *le coq du village* « l'homme le plus admiré par les femmes ». L'autre type inclut le sens d'un des constituants mais non dans la position communicativement dominante, par exemple le tchèque *soukromé očko* « détective privé », litt. « petit œil privé », calqué sans doute sur l'anglais *private eye*, la position dominante étant occupée par *očko* « petit œil » au sens figuré et non par *soukromé* « privé ».

Les sémi-phrasèmes sont un autre nom pour les collocations. La base d'une collocation est choisie de façon régulière et non contrainte, son sens occupe la position dominante dans le sens de la collocation. Le collocatif, dont le sens peut être ou ne pas être inclus dans le sens de la collocation, est choisi en fonction de la base et donc de façon irrégulière et / ou contrainte.

À la différence d'E. JEŽEK, I. MEL'ČUK n'opère pas, dans son concept, avec la solidarité lexicale et n'a pas donc besoin de distinguer, sur la base d'un test, entre les combinaisons dites restreintes et les collocations (*nez aquilin* × *haute qualité*). Si, pour E. JEŽEK c'est l'adjectif *aquilin* qui impose la restriction sur le choix du nom dans la

combinaison de mots *nez aquilin*, I. MEL'ČUK voit une combinaison analogue *année bissextile* à l'inverse : l'adjectif *bissextile* « qui a 366 jours » ne se combine qu'avec *année* et c'est cet adjectif qui est sélectionné de façon contrainte ; cette direction nous semble plus logique : on choisit librement la base *nez* et pour y ajouter le sens « fin et recourbé en bec d'aigle » on a recours à l'adjectif *aquilin* qui est conditionné par la base. Nous pouvons corroborer ce point de vue par les principes utilisés dans la distinction des catégories de mots de base (substantifs, adjectifs, verbes et adverbes) [9, p. 16] ; ces principes constituent, en effet, une hiérarchie qui se développe à partir du substantif : les substantifs désignent des phénomènes considérés comme des faits autonomes de la réalité extralinguistique ; les faits désignés par les substantifs reçoivent diverses caractéristiques, soit à l'aide des verbes, des caractéristiques dynamiques qui se déroulent dans le temps soit à l'aide des adjectifs, des caractéristiques qui ne se déroulent pas dans le temps ; les deux types de caractéristiques sont à leur tour caractérisées à l'aide des adverbes. L'orientation du développement a de l'impact sur l'organisation d'ouvrages lexicographiques ; par exemple, dans le *Dictionnaire des combinaisons de mots* [10], les entrées sont uniquement des substantifs, parce que les auteurs considèrent le substantif comme le porteur du concept, et les adjectifs et les verbes qui se combinent avec les substantifs en leur ajoutant des caractéristiques spécifiques sont à chercher dans les articles consacrés aux substantifs respectifs ; ainsi, on trouve l'adjectif *aquilin* dans l'article consacré à *nez* et le verbe *garer* dans les articles consacrés à *véhicule*, à *voiture* et peut-être encore à d'autres substantifs. Pour un type spécifique des collocations, celui des constructions à verbe support, J. RADIMSKÝ [11, p. 55] justifie, sur la base de liens sémantiques, l'évaluation à partir du nom vers le verbe dit support.

Quant au quasi-phrasème, son sens inclut les sens de ses constituants mais ils ne sont pas communicativement dominants, par exemple *donner le sein* « nourrir en mettant le sein à la portée de la bouche du bébé » où le sens dominant est « nourrir », le sens de *donner* et celui de *sein* étant inclus mais pas de manière dominante.

Alors que la classification d'I. MEL'ČUK permet de distinguer aisément les trois sous-classes de phrasèmes sémantiques selon le rôle que les sens des constituants jouent dans le sens du phrasème, la distinction entre les collocations et les syntagmes libres

semble problématique : dans ce cas, la considération du sens n'apportant pas beaucoup d'aide, on doit s'appuyer sur le critère relativement vague qui postule que le collocatif soit choisi de façon irrégulière et / ou contrainte. Il existe des mots qui se combinent couramment avec d'autres mots, par exemple l'adverbe *très* qui sert à intensifier le sens des adjectifs ; toutefois, I. MEL'ČUK signale que cet adverbe ne se combine pas avec certains adjectifs : on dit, par exemple, *très fatigué* mais *bien reposé* et non **très reposé* ; afin de former la combinaison d'un adjectif avec *très*, il faut *a priori* savoir si l'adjectif en question peut se combiner avec cet intensificateur ; en reconnaissant, pour cette raison, le caractère contraint de la sélection de l'adverbe *très*, I. MEL'ČUK classe les combinaisons d'adjectifs avec l'adverbe *très* parmi les collocations. Sa notion de la collocation est donc extensif, et on pourrait aller jusqu'à mettre en question l'existence des syntagmes libres ce qui justifierait la citation mentionnée plus haut que le natif parle en phrasèmes. Quant à la contrainte, elle doit consister, avant tout, dans l'usage qui constitue, pour un sens donné, un ensemble des combinaisons de mots préférés par les locuteurs natifs aux autres combinaisons possibles du même sens ; lorsque l'étranger construit un syntagme qui se trouve hors cet ensemble, le syntagme est considéré comme mal formé ou incorrect quoique son sens puisse être bien compréhensible. Cependant, la frontière de cet ensemble n'est pas nette ; au centre se trouvent des syntagmes préférés par la plupart des locuteurs natifs et vers la périphérie se manifestent des préférences individuelles. Il nous semble donc problématique de conditionner la collocation par la non-universalité du collocatif pour le sens donné, c.-à-d. que, pour un sens donné, toutes les bases ne doivent pas partager le même collocatif ; on ne peut être sûr que la périphérie n'offre pas des exemples positifs qui brisent la non-universalité d'un collocatif et mettent en doute toute une classe de collocations. À l'inverse, il est difficile de prouver que, pour exprimer le sens donné, un même mot (collocatif) peut se combiner avec n'importe quel autre mot (base), et donc prétendre que, pour ce motif, il ne s'agit pas de collocations.

Une autre source de la contrainte est l'analogie : lorsque, en 2019, une nouvelle maladie, appelée *covid-19*, apparut et provoqua, l'année suivante, une pandémie qui persiste jusqu'aujourd'hui, on n'hésita pas de combiner son nom avec les mêmes verbes que l'on combinait avec les noms des autres maladies infectieuses : de même que l'on

dit *contracter la grippe* on peut aussi dire *contracter le covid-19*. I. MEL'ČUK considère ce phénomène plutôt du point de vue de l'hyponymie et l'hyperonymie et parle de la généralisation ; nous pouvons exemplifier ce procédé comme suit : si *on contracte une maladie*, alors *on contracte le covid-19* aussi. Néanmoins, il faut traiter l'analogie ou la généralisation avec précaution ; nous avons vu plus haut que l'adverbe *très* convient à l'adjectif *fatigué* mais non à l'adjectif *reposé*, quoique les deux adjectifs semblent appartenir à la même classe sémantique.

2.3. Collocations dans la perspective de F. ČERMÁK

Pour F. ČERMÁK [5] la collocation est une combinaison de mots douée du sens ; par là, elle se distingue, d'une part, des n-grammes qui sont des combinaisons aléatoires de mots retrouvées dans le texte, et, de l'autre part, des combinaisons de mots sémantiquement inadmissibles, telles que *dormir furieusement* de la fameuse phrase de N. CHOMSKY citée plus haut.

En s'appuyant sur la distinction entre la langue et la parole, il distingue entre les combinaisons systémiques et les combinaisons textuelles, les premières étant figées, les secondes, non figées. Les combinaisons systémiques incluent des termes à plusieurs mots, tels que *acide sulfurique*, des noms propres à plusieurs mots, tels que *Grande Bretagne*, et des expressions idiomatiques ; tous ces sous-groupes sont classés par F. ČERMÁK parmi les collocations. Le figement semble donc être, en termes de la logique, une condition suffisante de la collocation mais non nécessaire ; car certains des combinaisons textuelles, et donc non figées, sont aussi des collocations. Les combinaisons textuelles comprennent des collocations courantes, telles que *réponse facile*, des formes composées, telles que *étant venu*, des métaphores d'auteur (et donc individuelles), des combinaisons aléatoires de mots voisins et d'autres combinaisons, telles que le bafouillage. Parmi ces sous-groupes, seules les collocations courantes et les métaphores d'auteur sont considérées par F. ČERMÁK comme collocations ; elles ont en commun la compatibilité sémantique de leurs constituants. Cette propriété est d'ailleurs présumée dans les combinaisons figées, et nous pouvons donc la considérer comme

une condition nécessaire de la collocation ; elle nous fait aussi penser à l'application des restrictions ontologiques et sémantiques dans la perspective d'E. JEŽEK.

F. ČERMÁK reconnaît que les frontières ne sont pas nettes et admet un type transitoire de combinaisons textuelles-systémiques qui comporte des « collocations usuelles courantes ». C'est déjà la dénomination qui pose des problèmes étant donné que les mots *usuel* (*uzuální* en tchèque) et *courant* (*běžný* en tchèque) sont des synonymes, et il est difficile de concevoir la différence entre les collocations *snadná odpověď* « réponse facile » et *umýt si ruce* « se laver les mains », dont l'une est donnée comme un exemple des collocations courantes et l'autre, comme un exemple des collocations usuelles courantes ; c'est, en effet, l'usage qui rend des expressions courantes.

En connexion avec les collocations, on parle souvent de la fréquence, mais celle-ci sert plutôt à distinguer les collocations du cœur de la langue de celles de la périphérie et son application peut être utile, par exemple, pour des fins lexicographiques. L'inclusion des métaphores d'auteur, dont les occurrences sont rares voire uniques, montre que la fréquence n'est pas un élément constitutif de la notion de la collocation.

F. ČERMÁK mentionne que les constituants d'une collocation peuvent ne pas se suivre immédiatement dans le texte, et il présente l'opinion de J. SINCLAIR selon laquelle la distance entre les constituants d'une collocation peut s'étendre jusqu'à 5 mots. Ce point a surtout un intérêt pratique pour la recherche automatisée de collocations dans le texte. Toutefois, la limite indiquée n'est pas universelle ; nous pouvons, en effet, bien imaginer des cas où elle soit dépassée : en latin, l'ordre des mots est, avec prédilection, marqué par l'hyperbate, qui consiste dans la séparation des éléments de phrase qui sont naturellement liés, par exemple : *Magna dis immortalibus habenda est atque huic ipsi Ioui Statori, antiquissimo custodi huius urbis, gratia [...]* « Vouons aux dieux immortels une reconnaissance infinie, et, en particulier, à ce Jupiter Stator le plus ancien protecteur de la cité [...] » [12, CIC., *Cat.* 1, 11] ; en allemand, les formes impersonnelles du verbe (l'infinitif et le participe passé) et, dans les subordonnées, même les formes personnelles sont de règle rejetées à la fin de la phrase, ce qui peut bien éloigner, l'un de l'autre, les constituants d'une collocation verbonominale ; et nous pourrions encore continuer cette énumération.

Le concept de la collocation proposé par F. ČERMÁK est d'une étendue vaste et d'une nature hétérogène, regroupant des types divers de combinaisons de mots : d'un côté, des syntagmes sémantiquement transparents et de l'autre côté, des expressions idiomatiques. Ainsi, il recouvre, au moins, tout ce que I. MEL'ČUK comprend sous l'appellation de phrasème sémantique avec ses trois sous-classes, à savoir phrasème complet, semi-phrasème et quasi-phrasème, tandis qu'I. MEL'ČUK ne tient pour collocations que les sémi-phrasèmes.

2.4. Conclusions

Dans ce chapitre, nous avons présenté trois différentes perspectives des collocations. La perspective de F. ČERMÁK est la plus large ; il postule comme condition nécessaire et, semble-t-il, aussi suffisante de la collocation la compatibilité sémantique des mots qui se combinent. Les deux autres perspectives, celle d'E. JEŽEK et celle d'I. MEL'ČUK, appliquent aux collocations de mots une différenciation plus détaillée ; il s'agit surtout de la discrimination entre les combinaisons sémantiquement plus ou moins transparentes et les expressions idiomatiques. Toutefois, E. JEŽEK justifie, à notre avis, insuffisamment la distinction entre ce qu'elle appelle combinaisons restreintes et les collocations. Dans le présent mémoire, nous adoptons donc la perspective d'I. MEL'ČUK sauf que nous n'insistons pas sur la non-universalité des collocatifs. En règle générale, nous prenons, conformément à l'opinion d'I. MEL'ČUK, le nom d'une collocation verbonominale pour base et le verbe pour collocatif.

3. Classification des collocations verbo-nominales

La base et le collocatif peuvent être constitué de différentes espèces des mots. Nous nous restreignons aux collocations verbonominales, et donc formées d'un verbe et d'un nom ; cependant, le nom est le plus fréquent accompagné d'un article, peut s'attacher au verbe par l'intermédiaire d'une préposition et peut éventuellement être développé par une épithète.

3.1. Classification selon les critères syntaxiques

Pour les relations syntaxiques, nous nous appuyons sur la tradition tchèque de la valence verbale, détaillée notamment dans la grammaire dite académique [13] et dans l'ouvrage de M. GREPL et P. KARLÍK [14]. Nous essayons de l'appliquer sur la réalité linguistique française.

À travers sa capacité prédicative, le verbe sert d'un pivot qui a le potentiel d'organiser la structure syntaxique de la proposition. Le verbe crée autour de lui des positions valentielles, obligatoires ou potentielles, qui sont occupées par différentes fonctions syntaxiques. Le nom remplit, dans le champ valentiel du verbe, sa fonction primaire comme le sujet ou le complément d'objet [9, p. 17-18]. La première classe de collocations que nous détachons est donc formée par le sujet et le verbe, par exemple *la nuit tombe* ou *le vent tombe*, les noms *nuit* et *vent* étant des bases qui conditionnent la sélection du collocatif *tombe(r)*, dans la première collocation pour exprimer le sens « arriver », et dans la seconde collocation pour exprimer le sens « perdre de sa force ».

Il est notoirement connu que, dans la langue française, les compléments d'objet se divisent en directs et indirects selon qu'ils sont accompagnés d'une préposition ou non, tandis que, dans la langue tchèque qui dispose d'une flexion nominale, seules les compléments d'objet exprimés par l'accusatif sans préposition sont considérés comme directs, les autres, avec ou sans préposition, comme indirects. Les deux types de compléments

d'objet, direct et indirect, avec le verbe constituent deux autres classes de collocations, par exemple : *écrire une lettre* et *profiter de l'occasion*.

Il y a évidemment une différence entre la collocation *assister au concert* et la collocation *aller au concert*. Dans le premier cas, c'est le verbe qui impose la préposition à ; dans le second cas, *au concert* a plutôt le caractère d'une locution adverbiale, le verbe ne déterminant pas une préposition précise mais une classe de locutions que l'on peut utiliser, cf. *aller en ville*, *aller chez coiffeur*, et aussi les questions *à quoi a-t-il assisté ?* et *où est-il allé ?* Tout en restant valentiel, le complément de ce type peut être appelé complément adverbial.

Prenons la collocation *élire un président*, où le nom *président* joue le rôle de complément d'objet direct. On peut la transformer en *élire [qqn] président* ; cette fois-ci, c'est *qqn* qui est le complément d'objet direct, et le nom *président* en tant que complément qualificatif⁽²⁾ dénote le caractère du complément d'objet. Le complément qualificatif peut aussi être prépositionnel, par exemple : *prendre [qqn] pour époux*. La transformation passive résulte en *être élu président*, ce qui est très proche des collocations *devenir président* et *être président*, le nom *président* étant appelé selon la terminologie française « attribut du sujet » ; toutefois, dans la tradition tchèque, on prend le nom pour le vrai prédicat et le verbe *être* (*být* en tchèque) est considéré comme copule, l'ensemble constituant un prédicat verbonominal.

La grammaire académique [13, p. 60] traite les expressions qui expriment le moyen d'une action comme transitives entre les éléments valentiels et les éléments non valentiels : elles ont, d'une part, la nature d'objets et présentent peu de variabilité, et de l'autre part, elles sont applicables à un grand nombre de verbes. Elles comprennent l'instrument, le matériel, l'intermédiaire et le moyen au sens étroit. Exemples : *écrire à la machine*, *aller en voiture*, *couper avec des ciseaux*, *saupoudrer avec du sucre*, *envoyer par avion*. C'est notamment l'expression de l'instrument qui entretient une forte solidarité lexicale avec le verbe, cf. les différents instruments utilisés pour couper, qui, en tchèque, se lient à des verbes spéciaux : *couper avec un couteau* — *krájet (nožem)*, *couper avec une hache* — *sekat (sekerou)*, *couper avec des ciseaux* — *stříhat (nůžkami)*,

2. Selon une autre terminologie, on l'appellerait « attribut du complément d'objet », cf. [15, p. 239-41].

couper avec une scie — *řezat (pilou)*. Parfois, les expressions du moyen s'assimilent aux compléments locales, cf. la collocation *aller à bicyclette* avec la collocation critiquée [8, s. u. « bicyclette »] mais utilisée *aller en bicyclette*, où la critique est, sans doute, dirigée vers le fait que, en utilisant la bicyclette, on ne se trouve pas dedans.

Il y a des locutions non valentielles, telles que *avec intérêt*, *avec zèle* ou *avec assiduité*, qui sont susceptibles de qualifier ou quantifier l'action exprimée par le verbe. Nous sommes enclin de considérer, dans la collocation formée d'un verbe et d'une de ces locutions, contrairement à la règle générale exprimée dans le chapitre précédent, le verbe comme base et la locution comme collocatif. D'ailleurs, dans ces cas, le nom ne remplit pas sa fonction primaire mais une fonction secondaire d'adverbe, et la logique, dont est issue la règle générale, reste donc maintenue.

3.2. Fonctions lexicales

I. MEL'ČUK [2, p. 10 et suiv.] propose un moyen utile pour décrire et classer les collocations. Le classement se fait en premier lieu selon des critères sémantiques ; secondairement, on peut tenir compte de relations syntaxiques, ce qui est le cas notamment pour les collocations à verbe support.

Soit un sens donné ; la fonction lexicale qui représente ce sens, prend pour son argument un mot qui doit servir de base d'une collocation, et lui associe tous les collocatifs possibles ; on obtient ainsi un ensemble de collocations prêtes à être utilisées dans le texte. Exemple : la fonction lexicale qui exprime le sens « émettre le son typique » fait correspondre à l'argument *chien* les verbes *aboyer* et *japper*, et à l'argument *moteur*, les verbes *ronronner* et *vrombir*, d'où les collocations *le chien aboie*, *le chien jappe*, *le moteur ronronne*, *le moteur vrombit*. On peut, par convention, noter cette fonction lexicale « Son » et écrire schématiquement : $\text{Son}(\text{chien}) = \text{aboyer, japper}$.

Quoique I. MEL'ČUK le prétende, la fonction lexicale n'est pas une fonction au sens mathématique parce que, dans les mathématiques, la fonction établit une correspondance telle que, à une valeur de l'argument corresponde une seule valeur de la fonction [16, p. 112], tandis que la fonction lexicale peut produire pour une valeur de l'argument (une base) plusieurs résultats (plusieurs collocatifs).

Du point de vue pratique, les fonctions lexicales sont intéressantes lorsqu'elles sont susceptibles de prendre pour l'argument un grand nombre de bases et qu'elles produisent des collocatifs variés ; le sens qu'elles expriment doit être suffisamment général. I. MEL'ČUK les appelle « fonctions lexicales standard » et en présente une liste (non exhaustive).

Différentes classes des collocations verbonominales sont décrites notamment par les fonctions lexicales standard simples suivantes : ceux qui associent à la base nominale un verbe support dont la base peut être sujet, complément d'objet direct ou complément d'objet indirect, ceux qui expriment la réalisation des objectifs inhérents de la base nominale (*exécuter un ordre*), ceux qui expriment des effets de la base nominale (*la lumière se répand [qq. part]*, *le vent agite [qqch.]*), une manifestation de la base nominale (*la joie éclate*), une dégradation de la base nominale (*le lait tourne*), une émission du son typique (*le chien aboie*), un fonctionnement difficile (*le souffle manque [à qqn]*), un arrêt du fonctionnement (*perdre le souffle*), un fonctionnement excessif (*le cœur palpite*). La fonction lexicale qui exprime l'intensification peut prendre comme argument des verbes et produire des collocatifs de nature adverbiale dont certains sont constitués de noms, cf. *remercier de tout cœur* à côté de *remercier chaleureusement*.

Les fonctions lexicales simples peuvent se combiner pour former des fonctions lexicales complexes. Par exemple, la fonction qui produit un verbe support peut se combiner avec les fonctions standard simples qui expriment les trois phases de l'action (*acquérir de l'influence*, *garder son influence*, *perdre son influence*) ou la fonction lexicale standard causative avec celle qui exprime une manifestation de la base (*exprimer une opinion*).

Il faut encore mentionner la fonction lexicale « Pred » qui représente la copule *être*, et certains de ses combinaisons (dans la notation schématique, pour l'argument *joie*) : $\text{IncepPredPlus}(joie) = \textit{grandir}$, $\text{IncepPredMinus}(joie) = \textit{faiblir}$, $\text{CausPredPejor}(joie) = \textit{altérer, gâcher [~]}$.

Les fonctions lexicales standard et leurs combinaisons ne couvrent pas toutes les collocations possibles. Pour les sens spécifiques, on est libre de créer de nouvelles

fonctions lexicales appelées « non standard ». Dans ce cas, il faut donner une brève définition de la nouvelle fonction.

Regardons encore un aspect des fonctions lexicales : leur utilité pour la comparaison des langues et en particulier pour la traduction est évidente : lorsque l'on a découvert, dans l'analyse de l'original, une fonction lexicale, on peut la réutiliser dans la synthèse de la traduction. Or, le passage d'une construction verbonominale à l'expression monolexicale (verbale) n'est pas directement accessible ; par exemple, on peut facilement comparer les valeurs de la fonction lexicale $Oper_1$ du substantif qui signifie « résistance » en tchèque, en allemand ou en français :

$Oper_1(odpor)$ = *klást* [$N_{dat} \sim acc$]

$Oper_1(Widerstand)$ = [$N_{dat} ART \sim acc$] *leisten*

$Oper_1(résistance)$ = *opposer* [ART ~ à N]

et par conséquent on sera capable de traduire la collocation tchèque *klást odpor* par la collocation française *opposer de la résistance*, mais on n'arrivera pas à obtenir le verbe *résister*, ce qui veut dire que la possibilité de traduire ainsi la collocation tchèque restera cachée au traducteur à moins qu'il n'ait recours à la fonction paradigmaticque⁽³⁾ (dérivationale) V_0 :

$V_0(résistance)$ = *résister*

La fonction lexicale V_0 peut se combiner avec les autres fonctions lexicales pour former des fonctions lexicales complexes, par exemple :

$CausOper_1(mort)$ = *provoquer* [ART ~ de N]

$V_0CausOper_1(mort)$ = *tuer* [N]

Les fonctions lexicales assurent, théoriquement, la correspondance sémantique complète des expressions de différentes langues, pourvu que l'on les applique à des mots qui ont la même signification dans ces langues.

3. Outre les fonctions lexicales syntagmatiques qui sont bien convenables à la description et à la classification des collocations, I. MEL'ČUK propose encore des fonctions lexicales paradigmaticques qui n'établissent pas de relations syntagmatiques entre leur argument et leurs valeurs mais remplacent leur argument par leurs valeurs.

3.3. Conclusions

La théorie de la valence verbale permet de différencier les différentes positions syntaxiques que le nom occupe dans les collocations verbonominales. Il est souvent dans la position valentiel du sujet ou du complément d'objet. Cependant, l'appartenance du nom au champ valentiel du verbe n'est pas toujours univoque, notamment dans le cas des compléments qui expriment l'instrument et gardent une forte solidarité lexicale avec le verbe. Dans d'autres collocations verbonominales, le nom n'est pas en position valentielle. Pour une description et un classement basés sur les critères sémantiques, I. MEL'ČUK propose les fonctions lexicales. Elle sont utiles pour la description d'une seule langue et pour la lexicographie monolinguale tandis que leur utilité pour la comparaison de langues et pour la traduction est limitée par les différences structurelles de langues.

4. Approche traductologique

4.1. Équivalence fonctionnelle

Le processus de traduction se concentre sur la recherche de l'équivalence. De nos jours, on parle plutôt d'une équivalence fonctionnelle, ce qui veut dire que, indépendamment des moyens d'expression, la traduction doit remplir les mêmes fonctions que l'original [17, p. 7]⁽⁴⁾. Le principe fonctionnel conduit à une traduction que nous identifions, dans la classification proposée par Bronislava GRÝGOVÁ [17, p. 16-7], avec la traduction communicative. Celle-ci donne au récepteur l'impression qu'il a sous les yeux un texte original et non une traduction ; elle se distingue, d'un côté, de la traduction littérale et de sa forme extrême, à savoir la traduction interlinéaire à laquelle nous pouvons assimiler les traductions juxtalinéaires qui furent publiées à des fins pédagogiques notamment au XIX^e et au début du XX^e siècle⁽⁵⁾ et, de l'autre côté, de la traduction libre qui tend à altérer le sens et d'autres aspects de l'original et qui nous fait penser, par exemple, à des traductions des XVII^e et XVIII^e siècles connues sous l'appellation de « belles infidèles ».

Le fait que l'original et la traduction font partie de deux actes de communication nous donne lieu de considérer l'action de traduire comme médiation, cf. J. ŠABRŠULA [19, p. 7], dans laquelle le traducteur sert d'intermédiaire entre deux milieux linguistiques jouant à la fois le rôle de récepteur de l'original et celui d'émetteur de la traduction⁽⁶⁾.

En améliorant le modèle de Karl BÜHLER qui ne reconnaît que trois fonctions du langage, Robert JAKOBSON [21, p. 353-7] attribue à chacun des six facteurs de l'acte de communication une fonction du langage. Le schéma de l'acte de communication est le suivant, les différents facteurs étant mis en italique : l'*émetteur* (ou destinataire)

4. Cf. le concept de la traduction équifonctionnelle expliqué par Ch. NORD [18, p. 48].

5. La collection des « auteurs latins expliqués d'après une méthode nouvelle par deux traductions françaises l'une littérale et juxtalinéaire présentant le mot à mot français en regard des mots latins correspondants l'autre correcte et précédée du texte latin » et une collection analogue des auteurs grecs ; les deux collections parurent chez Hachette.

6. Cf. le schéma communicationnel de la traduction proposé par J. SVOBODOVÁ [20, p. 292-3].

envoie un *message* au *récepteur* (ou destinataire) ; le message se réfère à un *contexte* ; un *contact* est établi entre l'émetteur et le récepteur et ceux-ci partagent *un code* qui sert à encoder et à décoder le message. Orientée vers le contexte, la fonction référentielle (ou dénotative) est d'une grande importance dans la communication, permettant de parler de la réalité extralinguistique. La fonction émotive (ou expressive) donne à l'émetteur la possibilité d'exprimer son attitude envers ce dont il parle. La fonction conative⁽⁷⁾ est utilisée pour exercer de l'influence sur le récepteur. R. JAKOBSON a les trois fonctions ci-dessus en commun avec K. BÜHLER et y ajoute encore les trois qui suivent. La fonction phatique sert à établir, maintenir ou interrompre le contact. La fonction métalinguistique permet de parler du code lui-même en posant, par exemple, des questions comme celle du petit prince⁽⁸⁾ : *Qu'est-ce que signifie « apprivoiser » ?* Enfin, la fonction poétique est liée à la forme du message.

Dans sa théorie de traduction, Christiane NORD [18] s'appuie sur les trois fonctions de K. BÜHLER auxquelles elle juge nécessaire d'ajouter une quatrième, empruntée à R. JAKOBSON, à savoir la fonction phatique. La fonction métalinguistique ne représente pour elle qu'une sous-fonction de la fonction référentielle, le référent étant, dans ce cas, non la réalité extralinguistique mais le langage. Pareillement, la fonction poétique semble intégrée dans la fonction conative, vu que la forme du message est susceptible de produire un effet voulu sur le récepteur.

Pour illustrer les fonctions du langage, prenons comme exemple une formule de politesse qui termine la lettre : *Je vous prie d'agréer, Monsieur, l'expression de mes salutations distinguées.* Elle a avant tout la fonction phatique ; de plus, l'auteur de la lettre exprime son respect envers le destinataire (fonction expressive) et cherche à s'assurer la faveur de celui-ci (fonction conative) d'autant plus qu'il attend une réponse ou des contacts successifs. La formule peut être traduite en tchèque par un simple syntagme

7. Nous ne partageons pas l'opinion de R. JAKOBSON lorsqu'il compte le vocatif parmi les expressions grammaticales les plus pures de la fonction conative [21, p. 355] ; il nous semble, en effet, que le vocatif est plutôt une expression de la fonction phatique : de même que l'on peut établir le contact à l'aide d'interjections telles que *ohé!* ou *allô!*, de même on peut le faire en utilisant le vocatif, notamment au début d'une lettre. Curieusement, M. ARRIVÉ *et al.* assignent à l'apostrophe tantôt la fonction conative [22, p. 364], tantôt la fonction phatique [22, p. 67].

8. A. de SAINT-EXUPÉRY.

S pozdravem litt. « Avec salutation » qui remplit les mêmes fonctions dans les lettres tchèques. Si le traducteur avait essayé de traduire la formule littéralement, la traduction serait fautive dans la mesure où le récepteur considérerait la fin de la lettre comme quelque chose d'étrange, d'inapproprié et d'exagéré ce qui n'est pas le cas de l'original.

Les auteurs de *La grammaire d'aujourd'hui* [22, p. 365] étudient les fonctions du langage dans le slogan électoral *Mitterrand, président*⁽⁹⁾! Nous pouvons le comparer avec le slogan tchèque *Havel na Hrad!* litt. « Havel au Château! » apparu en Tchécoslovaquie lors de la révolution de 1989. Le château de Prague, étant la résidence du président de la république, fait, par métonymie, référence à la présidence; on peut y sentir une connotation militante, cf. le cri des révolutionnaires français *À la Bastille!* Le slogan tchèque ainsi que le français renvoient, bien sûr, aussi aux candidats respectifs. L'émetteur du slogan exprime son attachement au candidat et cherche à gagner les récepteurs à sa cause. Quant à la fonction poétique, les slogans se composent de deux termes symétriques: le premier terme du slogan tchèque (*Havel*) est dissyllabique et a l'accent sur la première syllabe de même que le second (*na Hrad*), tandis que les termes du slogan français sont trisyllabiques avec l'accent sur la syllabe finale; à la différence du slogan français, les deux termes du slogan tchèque ne sont pas liés par une rime, mais nous repérons dans le slogan tchèque une sorte d'allitération, la lettre *h* étant prononcée. Si l'on voulait traduire ces deux slogans dans la perspective fonctionnelle, non littéralement, c'est surtout la fonction poétique qui poserait des problèmes; le slogan tchèque suppose, de plus, une connaissance suffisante de la réalité tchèque (ou tchécoslovaque) pour comprendre la métonymie et déceler sa valeur dénotative.

La fonction de la traduction peut ne pas coïncider avec celle de l'original: la traduction du slogan électoral se trouvera, par exemple, dans un article de journal pour documenter la communication de l'émetteur de l'original avec les récepteurs de l'original; sa fonction sera donc métatextuelle⁽¹⁰⁾. Par contre, lorsqu'une entreprise

9. Quant à la fonction métalinguistique, nous ne la repérons pas dans ce slogan contrairement aux auteurs de *La grammaire d'aujourd'hui*; l'explication par laquelle ils cherchent à justifier sa présence nous semble peu convaincante.

10. Dans sa typologie des traductions, Ch. NORD [18, p. 45-48] appelle les traductions de ce type « documentaires »; elles tendent à la littéralité.

lancera un produit commercial sur le marché étranger, la traduction du slogan publicitaire gardera les fonctions de l'original tout en tenant compte de la culture d'arrivée.

En transposant un texte de la langue de départ dans la langue d'arrivée, le traducteur est supposé de prendre en considération les aspects sémantique, stylistique et pragmatique ; la plus grande importance est d'ordinaire attachée à l'aspect sémantique alors que les deux autres aspects sont, dans la pratique, observés avec moins de rigueur quoique le manque de souci pour ceux-ci soit perçu négativement par les récepteurs de la traduction, cf. [17, p. 13 et 15].

4.1.1. Aspect stylistique

Le styl consiste dans le choix entre plusieurs moyens d'expression et dans leur emploi et leur organisation dans le discours [23, p. 441] ; il est donc conditionné par l'existence de la synonymie [24, p. 131] : au même signifié correspondent plusieurs signifiants. Les moyens d'expression se regroupent selon leur appartenance à différentes variétés de la langue. Jan ŠABRŠULA [19, p. 116 ; 25, p. 10-3], appelant les variétés sous-codes, propose un modèle tridimensionnel de leur classification : l'axe verticale représente ce que l'on appelle communément registres ou niveaux de langue ; sur l'axe horizontal sont réparties les variétés relevant de domaines spécialisés tel que la médecine, diverses disciplines scientifiques ou techniques, etc. ; le troisième axe est destiné aux variétés régionales.

L'axe vertical reflète en principe l'appréciation sociale de différentes manières de parler. Cette appréciation peut découler soit de la situation de communication soit des compétences linguistiques du locuteur : le même locuteur recourt, dans des situations banales de la vie quotidienne, à une langue moins soignée que dans des situations qui ont plus d'importance du point de vue social telles qu'un discours public ; d'autre part, les compétences linguistiques du locuteur dépendent de son instruction qui est en rapport avec son statut social ; ainsi, la langue des milieux populaires est censée très basse ; par contre, un locuteur tchèqu connaissant les langues classiques sera capable d'apprécier le génitif *Polluka* du nom de l'un des Dioscures, bâti sur la déclinaison

latine, tandis que le génitif *Polluxe*, créé systématiquement sur le nominatif, offenserait ses oreilles.

La division verticale du français varie selon les auteurs. M. GREVISSE et A. GOOSSE [26, 23-4] jugent utile de distinguer entre les registres, qui renvoient à la situation, et les niveaux, qui concernent les compétences linguistiques ; les registres sont comme suit : très familier, familier, soigné ou soutenu et très soutenu ; les niveaux sont les suivants : populaire, moyen et intellectuel. Pareillement, J. CHMELOVÁ-SVOBODOVÁ⁽¹¹⁾ distingue deux gammes de variétés du français sur l'axe verticale : 1^o le français vulgaire, populaire, familier et tenu ; 2^o le français relâché, neutre et soigné. M. ARRIVÉ *et al.* [22, p. 598] disent qu'il serait difficile d'isoler plus de trois ou quatre registres, à savoir populaire, familier, courant ou moyen, et soigné (soutenu, éventuellement littéraire). *La nouvelle grammaire du français* [27] se contente de trois niveaux : familier, courant et soutenu.

La langue tchèque se divise en tchèque standard (*spisovná čeština*)⁽¹²⁾ et tchèque commun (*obecná čeština*). Seul le tchèque standard est considéré comme forme correcte de la langue tchèque étant utilisé dans la vie publique, dans l'administration, dans les journaux, dans les publications scientifiques, à l'école, etc. À l'intérieur du tchèque standard on peut distinguer trois variétés de l'axe vertical : une variété neutre, c.-à-d. non marquée, et deux variétés marquées qui s'opposent, à savoir le tchèque familier (*hovorová čeština*) et le tchèque littéraire ou soutenu (*knižní čeština*). Le tchèque familier est défini comme langue parlée dans la vie quotidienne par les locuteurs qui maîtrisent le tchèque standard de manière active et pour qui le tchèque standard est un moyen de communication courant et conscient [28, p. 101]. Le tchèque commun⁽¹³⁾, quant à lui, est utilisé dans la vie quotidienne, surtout à l'oral, il est inadmissible dans les devoirs scolaires ou dans les actes officiels. Or, la frontière entre le tchèque familier et le tchèque commun est très floue ; B. HAVRÁNEK et A. JEDLIČKA reconnaissent que le tchèque commun se rapproche progressivement du tchèque familier [23, p. 4] et que ces deux variétés du tchèque partagent certains phénomènes morphologiques [23, p. 483-4]

11. Citée par J. ŠABRŠULA [25, p. 11-2].

12. Pour la problématique terminologique, cf. [28, p. 49-51 ; 24, p. 62].

13. Pour la caractéristique du tchèque commun, voir [23, p. 481-4].

et beaucoup d'expressions lexicales [23, p. 448]. Par conséquent, il est souvent difficile aux locuteurs de distinguer si un discours relève encore du tchèque familial (et est donc considéré comme correct) ou déjà du tchèque commun (et donc non correct) ; certains locuteurs confondent même la notion du tchèque familial et du tchèque commun. À cela accède le fait que la division du tchèque en standard et commun est parfois mise en question, cf. [24, p. 63].

Il se pose la question de savoir comment les différents niveaux du français et du tchèque se correspondent. Otomar RADINA [29, p. 129] explique que le mot français *prof* ne peut être traduit par le tchèque *prófa* quoique les deux mots signifient « professeur » tout en relevant d'un niveau plus bas que la langue courante : le professeur français peut dire de lui-même : *Moi, je suis prof*, tandis que le professeur tchèque n'utiliserait pas le mot *prófa* pour se désigner lui-même⁽¹⁴⁾. Pareillement, le mot tchèque *matura* n'a pas la même valeur stylistique que le français *bac*. J. CHMELOVÁ-SVOBODOVÁ⁽¹⁵⁾ dit que le français familial correspond au tchèque familial, tandis que le tchèque commun est situé à un niveau plus bas que le français familial et plus élevé que le français populaire.

Le discours est souvent constitué de moyens d'expression pris à différents endroits de l'axe vertical. Celui-ci représente un spectre continu plutôt que discret ; dans ce sens, M. ARRIVÉ *et al.* [22, p. 598-9] expliquent que, sur le plan phonétique, la langue passe du niveau populaire au niveau soigné en fonction de l'augmentation de différents traits, tels que nombre de liaisons, nombre des *e* muets ou maintien des groupes consonantiques. La distinction des variétés de langue se fait avec subjectivité qui se manifeste, sur le plan lexical, dans le désaccord des dictionnaires sur l'indication de niveau : les substantifs *piaule*, *bagnole* et *boulot*, par exemple, sont pour *Le Petit Robert* [8] encore familiaux, mais le *Dictionnaire de l'Académie française* [30, 9^e éd.] les considère déjà comme populaires, tandis que d'autres mots, tels que *pantouflard*, reçoivent la même étiquette de style familial dans les deux dictionnaires. Cf. aussi ci-dessus le statut incertain du tchèque familial par rapport au tchèque commun. Ces faits pris en considération, il nous semble utile d'adopter pour le français aussi bien que pour le tchèque une division

14. Précisons que la différence entre le français *prof* et le tchèque *prófa* repose plutôt sur l'axe horizontal, ce dernier mot relevant de l'argot des élèves qui lui attribue une connotation péjorative.

15. Citée par J. ŠABRŠULA [25, p. 11].

tripartite : niveau neutre, niveau familier regroupant tous les moyens d'expression qui sont décalés vers le bas, et niveau soutenu regroupant tous les moyens d'expression qui sont décalés vers le haut.

Les variétés de l'axe horizontal sont des compétences supplémentaires exigées par le sujet du discours. Elles se manifestent surtout sur le plan lexical et elles subissent, à leur tour, une division verticale : une sous-variété, disons standard, est basée sur la terminologie et s'utilise dans la communication officielle, dans les articles des revues de spécialité, aux conférences de spécialistes, dans les manuels, etc., tandis qu'une sous-variété plus basse, familière, trouve sa place dans les entretiens de la pratique quotidienne de spécialistes. Toutefois, dans un discours réel, les deux sous-variétés peuvent se combiner.

La sous-variété familière se distingue de la sous-variété standard par l'effort de brièveté (l'économie de langue), cf. [31, p. 20], qui se manifeste par la tendance à la monolexicalité, par exemple, le terme standard *fyzilogický roztok* « sérum physiologique » devient *fýzák*, et *emitní miska* « haricot (récipient) », *emitka*; et à l'abréviation des mots, par exemple, *nebulizace* « nébulisation » devient *nebulka*, et *glykémie* « glycémie », *glycka*; les suffixes les plus productifs dans ces procédés sont *-ák* pour les masculins et *-ka* pour les féminins, les variantes familières ayant tendance à garder le genre de leurs contreparties standard. Les abréviations tchèques prennent souvent un suffixe, ce qui n'est pas le cas pour les abréviations françaises, par exemple, *perf* pour *perfusion* par une simple apocope. Des expressions monolexicales peuvent être obtenues aussi par métonymie : *heimlich* de *Heimlichův manévr* « manœuvre de Heimlich », ou bien par métaphore : *bažant* litt. « faisan » pour *močová láhev* « urinal », *gramofon* litt. « phonographe à disques » pour *toaletní křeslo* « chaise percée ». Quoique on n'ait isolé que deux sous-variétés, l'axe vertical garde son caractère de spectre continu, ce qui devient évident lorsqu'on compare des synonymes : par exemple, le mot *gramofon* au sens de « chaise percée » et son abréviation *gramec* sont tous les deux familiers mais celle-ci est plus familière que celui-là.

4.1.2. Aspect pragmatique

Le milieu linguistique se distingue non seulement par sa langue mais aussi il est étroitement lié à une réalité culturelle, sociale et environnementale de sorte que les locuteurs d'un même milieu linguistique ont en commun une expérience linguistique et extralinguistique qui encadre leur communication. L'auteur d'un texte suppose cette expérience aux récepteurs et bâtit son texte sur elle. Or, l'original et la traduction relevant de deux milieux linguistiques différents, le récepteur de la traduction ne partage pas la même expérience linguistique et extralinguistique avec l'auteur de l'original. La médiation qu'effectue le traducteur consiste à composer, dans des conditions différentes de celles de l'original, un nouveau texte qui garde le sens de l'original, cf. [19, p. 8].

La même situation, lorsqu'elle se présente dans deux milieux linguistiques, peut entraîner le recours à des moyens linguistiques différents qui sont dictés par la convention du milieu respectif, cf. [17, p. 13]. Dans les consignes d'exercices, par exemple, on utilise l'infinitif en français et la 2^e personne du pluriel de l'impératif en tchèque. L'usage de vouvoiement et tutoiement diffère dans les deux langues : dans le contexte liturgique, par exemple, le tutoiement est de règle pour le locuteur tchèque (sur le modèle latin), tandis qu'en français on observe une hésitation ; certes, depuis le second concile du Vatican, le vouvoiement est en régression, mais on vouvoie le prêtre pendant la messe en lui disant : *Avec votre esprit*⁽¹⁶⁾, et la version officielle de la prière mariale la plus connue utilise toujours le vouvoiement : *Je vous salue, Marie*⁽¹⁷⁾. Il existe des syntagmes entiers qui sont sélectionnés par le locuteur dans la situation exigeant leur sens mais qui ne peuvent pas être librement remplacés par d'autres syntagmes ayant la même valeur dénotative et grammaticalement corrects ; I. MEL'ČUK [2, p. 4 ; 33, p. 11-2] les appelle « pragmatèmes ». Plus haut, nous avons mentionné la formule finale d'une lettre, elle compte parmi les pragmatèmes ; d'autres exemples : *Craint la chaleur et l'humidité* ou *Craint le gel* sur l'emballage d'un produit et *Ne quittez pas !* au téléphone. Le pragmatème d'une langue doit être traduit par le pragmatème qu'utilisent

16. En tchèque le tutoiement : *I s tebou*.

17. L'actuelle période de la transition du vouvoiement au tutoiement dans la prière catholique est bien illustrée par l'*Appendice du Compendium du Catéchisme de l'Église catholique* [32] qui contient un mélange de prières dont les unes utilisent le tutoiement, les autres le vouvoiement.

les locuteurs de la langue d'arrivée dans la même situation ; ainsi les trois derniers pragmatèmes ci-dessus se traduiront en tchèque comme suit : *Uchovávejte v chladu a suchu* litt. « Conservez au frais et au sec », *Chraňte před mrazem* litt. « Protégez contre le gel » et *Nezavěšujte!* litt. « Ne raccrochez pas ! ».

Lorsque la même situation n'est pas connue dans le milieu d'arrivée, le traducteur lui substitue une situation analogue⁽¹⁸⁾. Par exemple, pour souhaiter bonne chance à quelqu'un, les locuteurs tchèques peuvent faire un geste appelé *držet palce* « tenir les pouces » qui consiste à serrer les pouces dans les mains fermées ; dans le milieu français, un geste analogue est le croisement des doigts. Dans le milieu français, la notation scolaire se fait habituellement de 0 à 20, la moyenne étant la limite de la réussite mais le locuteur tchèque ne connaît pas cette notation, il est familier avec une échelle de cinq notes, le un étant la meilleure note et le cinq la pire note qui signifie l'échec ; la situation décrite dans le milieu français par l'expression *avoir une note inférieure à la moyenne* correspond à la situation décrite dans le milieu tchèque par *dostat pětku* litt. « recevoir un cinq », et *avoir un zéro* à *dostat čistou pětku* litt. « recevoir un cinq pur ».

La différence de la réalité extralinguistique peut mener au remplacement d'une expression plus spécifique par une plus générale ; on utilise un hyperonyme ou substitue un appellatif au nom propre [17, p. 59 et 93].

O. RADINA [29, p. 213-4] touche la problématique de la traduction de textes concernant les personnages légendaires et historiques du milieu linguistique en question. Selon lui, le message perd dans la traduction sa dimension affective. Il est déjà difficile de trouver une désignation appropriée de ces personnages ; O. RADINA se demande comment traduire, par exemple, *praotec Čech* en français : *le premier père Čech ?* ou *l'ancêtre légendaire des Tchèques ?*

Parfois, il faut ajouter de l'information. Les éléments de la vie quotidienne sont si familiers aux locuteurs qu'il les considèrent comme quelque chose de naturel et ne se rendent pas compte que, dans d'autres milieux, ils peuvent être différents. Telles est la manière de faire le lit : activité quotidienne non seulement à domicile mais aussi à l'hôpital. Lorsqu'une infirmière française, voulant refaire le lit d'un patient et changer en

18. Cf. [17, p. 93-5] ; ce procédé de traduction s'appelle « adaptation ».

même temps le linge de lit, demande à sa collègue : *Apporte-moi du linge propre !* cette dernière sait qu'il faut apporter, entre autres, deux draps, tandis qu'une infirmière tchèque irait chercher un seul drap et une housse de couette. Dans un roman ou dans un film présentant le milieu hospitalier, on n'hésitera pas à traduire la demande littéralement : *Přines mi čisté prádlo !* Or, supposons qu'une étudiante tchèque vient en France pour faire un stage infirmier, qu'elle ne connaît pas le français et qu'elle a par chance un interprète à ses côtés ; dans ce cas-là, ce dernier explicitera dans sa traduction, qu'il s'agit d'apporter, entre autres, un drap de dessous et un drap de dessus ; sinon la traduction ne fera pas sur le récepteur le même effet que l'original.

4.2. Différences typologiques entre le français et le tchèque

Différentes langues disposent de différents moyens d'expressions selon leurs propriétés structurelles. C'est pourquoi Jitka SVOBODOVÁ [20, p. 274-5 et 286] compte parmi les fonctions langagières que le traducteur devrait déceler dans l'original et appliquer dans la traduction la « fonction de conformité aux modèles systémiques ». Cependant, un écart du système peut être intentionnel pour produire des effets spéciaux tels que l'imitation de la parole d'un étranger ; dans ce cas-là, le traducteur découvrirait dans l'original la fonction de non-conformité aux modèles du système, dont la matérialisation dans la traduction pourrait être guidée par des règles conventionnelles ; en tchèque, par exemple, afin de signaler que c'est un étranger qui parle, on est habitué de substituer des infinitifs aux formes fléchies des verbes ; les règles dans tels cas particuliers dépendent quand même du système.

Suivant les caractéristiques de structure, Vladimír SKALIČKA [34] distingue cinq types de langues : isolantes, agglutinantes, flexionnelles, introflexionnelles et polysynthétiques. En gros, ces types peuvent être décrits comme suit : le type isolant exprime les catégories grammaticales par des mots auxiliaires et par l'ordre des mots ; le type agglutinant accumule des affixes dont chacun n'a qu'une seule fonction ; le type flexionnel utilise des affixes qui sont susceptibles d'exprimer plusieurs fonctions à la fois ; dans le

type introflexionnel, la racine des mots subit des changements ; le type polysynthétique est caractérisé par la composition.

Chaque langue pouvant avoir des caractéristiques de plusieurs types, l'assignation d'une langue à un type se fait selon les caractéristiques de la langue en question qui prévalent. Ainsi, le tchèque est classé parmi les langues flexionnelles, quoique V. SKALIČKA [34, p. 12] soit capable de donner des exemples de tous les cinq types empruntés au tchèque :

— isolant : *budu nosit* « je porterai » (futur analytique imperfectif formé de l'auxiliaire « être » au présent et de l'infinitif) ;

— agglutinant : *orám — vyorám — dovyorám — nedovyorám* « je laboure — je creuserai (ou arracherai) en labourant — j'achèverai de creuser (ou d'arracher) en labourant — je n'achèverai pas de creuser (ou d'arracher) en labourant » ;

— flexionnel : *dobrý* « bon » (-ý ayant trois sens, soit nominatif, singulier et masculin) ;

— introflexionnel : *vojáci — vojáky* « soldats (nominatif et vocatif) — soldats (accusatif et instrumental) », *přítel — přátel* « ami (nominatif singulier) — amis (génitif pluriel) », *střelím — střílím* « je tirerai (une seule fois) — je tire » ;

— polysynthétique : *lidojed, maloměsto, modrobílý* « anthropophage, petite ville, bleu et blanc ».

De même, le français est une langue isolante et pourtant nous pouvons y trouver des exemples de tous les cinq types :

— isolant : *elle a chanté* ;

— agglutinant : *croyable — incroyable — incroyablement* ;

— flexionnel : *chantâmes* (1^{re} personne, pluriel, passé simple) ;

— introflexionnel : *buvons — boirons* ;

— polysynthétique : *garde-robe*.

Regardons de près les tendances d'ordre structurel par lesquelles le français diffère du tchèque. O. RADINA a consacré à une analyse détaillée des différences systémiques du français et du tchèque un livre entier [29], qui, outre l'ouvrage mentionné de V. SKALIČKA [34], nous a fourni à ce sujet des observations précieuses.

Au cours de son évolution depuis le latin vulgaire vers le français moderne, le français a perdu la flexion du nom. Le nom français est donc incapable de distinguer par sa forme les fonctions de différents cas. L'opposition des fonctions syntaxiques du nominatif et de l'accusatif, est liée, en français, à l'ordre des mots : le sujet précède, de règle, le verbe tandis que le complément d'objet direct le suit, par exemple : *Paul aime Marie* ; *Marie n'aime pas Paul*. Les fonctions des autres cas sont exprimées par diverses prépositions, la fonction du datif, par exemple, le plus souvent par la préposition *à* : *Paul donne un cadeau à Marie*. Le tchèque, en revanche, a gardé sept de huit cas de la déclinaison indo-européenne : nominatif, génitif, datif, accusatif, vocatif, locatif et instrumental, de sorte que la position du sujet et celle du complément d'objet direct ne sont pas fixes dans la phrase et que les deux phrases *Pavel miluje Marii* et *Marii miluje Pavel* ont le même sens « Paul aime Marie ». Toutefois, le pronom français a partiellement conservé la flexion : *je × me* ; *tu × te* ; *il × le × lui* ; *qui × que* ; la variabilité du pronom relatif permet au locuteur de choisir de placer le sujet avant ou après le verbe dans des phrases relatives telles que : *[l'homme] que Marie hait* et *[l'homme] que hait Marie* où *Marie* est le sujet à la différence de la phrase relative *[l'homme] qui hait Marie* où *Marie* joue le rôle syntaxique du complément d'objet direct.

L'élimination de la flexion a conduit en français à la réduction des genres : les neutres latins sont passés soit aux masculins, par exemple : lat. *uīnum* > fr. *le vin*, soit aux féminins, par exemple lat. *folium* > fr. *la feuille* ; dans le cas des noms utilisés souvent au pluriel, le passage au féminin a été favorisé par la finale *-a* qui marque le pluriel des neutres en latin. Le tchèque a conservé les trois genres.

Quant au pluriel, des formes particulières se restreignent, en français, à certains types de noms : sg. *travail* ~ pl. *travaux* ; sg. *animal* ~ pl. *animaux*. Or, la marque habituelle du pluriel *-s* n'est qu'un signe graphique qui ne se manifeste pas à l'oral. C'est, en effet, l'article qui permet de distinguer le pluriel du singulier, par exemple *le crayon × les crayons*⁽¹⁹⁾ ; *un crayon × des crayons*. Le tchèque en tant que langue flexionnelle a en principe des formes différentes pour le pluriel, par exemple nom. sg. *žena* « femme » × nom. pl. *ženy*. Néanmoins, certaines formes sont homonymes, par

19. Alors que le *-s* ajouté aux noms ne se manifeste pas à l'oral, la prononciation du pluriel de l'article *les* est bien différente de celle du singulier *le*.

exemple *ženy* est non seulement le nominatif, l'accusatif et le vocatif du pluriel mais aussi le génitif du singulier ; pareillement, *moře* « mer » est le nominatif, l'accusatif et le vocatif aussi bien du singulier que du pluriel.

Le nom français est d'ordinaire accompagné d'un déterminant. Nous avons mentionné le rôle du déterminant dans la distinction du singulier et du pluriel des noms qui sont invariables à l'oral. L'article est le déterminant par excellence. Le français dispose de trois sortes de l'article : l'article défini qui indique soit une entité déterminée d'une classe, par exemple *l'homme que j'ai vu*, soit toute une classe, par exemple *l'homme est mortel* ; l'article indéfini qui marque une entité indéterminée d'une classe, par exemple *un homme est venu* ; et l'article partitif qui indique une quantité indéterminée d'une entité non dénombrable. Si le français exprime la différence de fonctions syntaxiques, à savoir l'opposition du sujet et du complément d'objet direct, par l'ordre des mots, le tchèque a recours au même moyen d'expression pour exprimer le défini et l'indéfini, par exemple : *Na stole jsou květiny* « Il y a des fleurs sur la table » et *Květiny jsou na stole* « Les fleurs sont sur la table ». Ce rôle de l'ordre des mots est directement lié à la perspective fonctionnelle de la phrase et à la répartition des places du thème et du rhème, un sujet majeur étudié au sein de l'école de Prague dès le temps de V. MATHESIUS et qui a été intégré, entre autres, dans la grammaire dite académique [13]. Cependant, les pronoms démonstratifs (*ten*) et indéfinis (*nějaký*) ainsi que le mot de nombre *jeden* « un » peuvent être employés en tchèque de manière semblable à l'article français, mais leur emploi n'est pas le plus souvent nécessaire et notamment le pronom démonstratif *ten* ajoute à l'énonciation le caractère de la langue familière, à comparer *Zavři dveře!* et *Zavři ty dveře!* « Ferme la porte ! ».

Le rôle du déterminant peut être joué par l'adjectif possessif ; par conséquent, l'emploi des adjectifs possessifs est plus fréquent en français qu'en tchèque. En plus, le caractère obligatoire du déterminant en français prédispose les adjectifs possessifs à remplir la même fonction que le datif d'intérêt en tchèque, par exemple : fr. *ôter son chapeau* × tch. *sundat si klobouk*.

On souligne souvent le caractère nominal du français, cf. [35, p. 69-71]. Une des causes de ce caractère est le fait que la dérivation des adjectifs relationnels n'est pas

productif en français. Ainsi, ce qui s'exprime en tchèque par un adjectif épithète, entraîne, en français, l'attachement d'un nom à un autre à l'aide d'une préposition, par exemple : fr. *arrêt d'autobus* × tch. *autobusová zastávka* ; fr. *bague en or* × tch. *zlatý prsten*. Étant donné l'absence des adjectifs relationnels, l'économie de la langue conduit à l'utilisation de noms apposés dont le rôle s'assimile à celui de l'épithète : on dit, par exemple, *assurance maladie*, expression qui résulte d'une simplification de la tournure *assurance contre la maladie* ; ce phénomène, qui concerne systématiquement les couleurs (*orange, marron, etc.*), contribue à l'effacement des frontières entre différentes parties du discours ; pareillement, la frontière entre l'adjectif et l'adverbe n'est pas en français toujours tellement nette qu'en tchèque, à comparer fr. *voir clair* et tch. *vidět jasně*. Plus généralement, le caractère nominal du français est en relation avec son caractère analytique : le sens que le tchèque condense en un seul mot est exprimé, en français, par l'enchaînement de plusieurs mots, par exemple : tch. *železnice* ~ fr. *chemin de fer* ; tch. *čekárna* ~ fr. *salle d'attente* ; mais aussi tch. *pračka* ~ fr. *machine à laver* ; tch. *květák* ~ fr. *chou-fleur* ; tch. *křížovka* ~ fr. *mots croisés*. Le verbe simple tchèque *zneškodnit* correspond à l'enchaînement de six mots en français *mettre hors d'état de nuire*.

4.3. Procédés de traduction

4.3.1. Stylistique comparée

L'idée d'étudier les moyens d'expression par la confrontation de deux langues plutôt que par l'analyse interne d'une seule langue remonte à Charles BALLY qui, dans un de ses ouvrages, paru en 1932, mit des exemples allemands en face des exemples français. Son approche ainsi que son choix de langues furent suivis en 1944 par Alfred MALBLANC [36, p. 15]. La lignée a été complétée par Jean-Paul VINAY et Jean DARBELNET qui publièrent en 1958 une *Stylistique comparée du français et de l'anglais* conçue comme manuel de traduction et qui connut plusieurs rééditions [37]. Alfred MALBLANC rédigea à son tour en 1961 un ouvrage analogue intitulé *Stylistique comparée du français et de l'allemand*

et réédité, lui aussi, plusieurs fois [36]. Le mérite de cette école franco-canadienne est d'avoir élaboré une théorie traductologique centrée sur une série de sept procédés de traduction : emprunt, calque, traduction littérale, transposition, modulation, équivalence et adaptation.

Quant aux théoriciens tchèques, ils ont généralement adopté, quoique avec certaines modifications, les bases méthodologiques franco-canadiennes et plus particulièrement leurs procédés de traduction [17, p. 19]. Ceux-ci sont systématiquement appliqués à la comparaison du français et du tchèque notamment par Jan ŠABRŠULA dans son ouvrage traductologique [19] et par Eva JANOVCOVÁ dans un manuel de français pour les avancés [38] et leur influence est bien sensible dans un livre d'Otomar RADINA consacré à la comparaison du tchèque et du français [29].

4.3.2. Emprunt

Lorsque la langue d'arrivée ne dispose pas de mot approprié pour exprimer une réalité qui est exprimé dans le texte de départ, on peut emprunter le mot étranger ; le tchèque a ainsi emprunté au français les mots *drén*, *drenáž*, *drénovat* ; le français, à son tour, les avait empruntés à l'anglais, cf. [39, s. u.]. L'orthographe des mots empruntés peut être accommodée à la langue d'arrivée ; il peuvent prendre des désinences nécessaires pour s'intégrer dans le système grammatical de la langue d'arrivée, cf. fr. *drainer* et tch. *drénovat*. Ce n'est pas uniquement l'absence d'équivalent approprié qui conditionne l'usage d'un emprunt ; on l'utilise aussi pour des raisons stylistiques ; dans le cas d'une langue de spécialité, on peut être mené par la volonté de garder les liens avec la communauté étrangère ou internationale ; ainsi, pour traduire le mot français *bandage* qui a l'équivalent tchèque *obvaz* ou *obvázání*, on utilise de préférence l'emprunt *bandáž*. Ce mot montre encore une propriété des emprunts : ils permettent une dérivation : la collocation française *appliquer un bandage* se traduit en tchèque par le verbe dérivé *bandážovat*.

D'autre part, il faut éviter les « faux-amis », il peuvent être étymologiquement apparentés comme fr. *infusion* et tch. *infúze* dont le premier se traduit par tch. *(bylinkový) čaj* et le second, par fr. *perfusion* ; pareillement fr. *gymnase* se traduit par tch. *tělocvična*,

et tch. *gymnázium* par fr. *lycée* ou *collège*. La ressemblance d'autres mots est fortuite : par exemple, fr. *liste* et tch. *list* « feuille ».

4.3.3. Calque

Le calque est l'imitation de la langue de départ en utilisant des éléments de la langue d'arrivée. Pour appeler un élément chimique, A. LAVOISIER et ses collègues [40, p. 33] ont créé à partir des éléments grecs le mot *hydrogène* en voulant exprimer le sens « engendrant l'eau »⁽²⁰⁾; le mot a été par la suite latinisé en *hydrogenium* et de nombreuses langues l'ont emprunté comme un internationalisme. Cependant, sur cette dénomination, la terminologie allemande a calqué *Wasserstoff* et la terminologie tchèque, fondée par Jan Svatopluk PRESL, *vodík*. Parfois, l'emprunt et le calque se trouvent en concurrence : pour traduire en tchèque *adverbe*, on peut emprunter l'internationalisme *adverbium*⁽²¹⁾ ou bien employer le calque *příslovce*.

Le calque concerne non seulement la création des mots mais aussi la structure des expressions et les constructions syntaxiques. Pour exprimer le sens « mettre en doute » en tchèque, on utilise d'ordinaire le verbe simple *zpochybnit* mais, pour traduire la collocation française *mettre en doute*, le traducteur pourrait être séduit à employer la collocation structurellement semblable *uvést v pochybnost* qui existe en tchèque et est correcte ; toutefois, si son emploi n'est pas justifié par des raisons stylistiques et qu'il résulte simplement d'une imitation de la langue de départ, le texte traduit sera peu naturel. D'autres constructions s'écartent tellement de la nature de la langue d'arrivée que leur imitation est communément considérée comme fautive ; c'est le cas, par exemple, du présentatif *c'est . . . qui / que* utilisé en français pour mettre en relief un membre de

20. Du point de vue linguistique, cette appellation a été créée incorrectement, parce que le second terme des composés grecs *-γενής* avec le vocalisme *ε* a régulièrement le sens passif « qui est né », par exemple *θεογενής* « né d'un dieu », tandis que les composés avec le second terme *-γονος* avec le vocalisme *ο* ont soit le sens passif, par exemple *θεόγονος* « né d'un dieu » soit le sens actif « faire naître », « engendrer », par exemple *δακρυγόνοτος* « qui fait naître les larmes », cf. [41, s. u. « γίγνομαι »], pour l'accent des composés en *-γονος* v. [42, p. 433]. Il faudrait donc dire *hydrogone*, cf. l'appellation néogrecque de cet élément chimique : *υδρογόνο*. La dénomination *hydrogène* souleva des critiques à cause de sa formation fautive déjà dans le passé, par exemple [43, p. 166].

21. Calqué par les grammairiens romains sur le grec *ἐπίρρημα*.

la phrase ; en tchèque, on cherche d'obtenir le même effet le plus souvent par l'ordre des mots.

4.3.4. Traduction littérale

Si la langue de départ et celle d'arrivée étaient typologiquement semblables et que leurs locuteurs appartenaient au même milieu socioculturel, la traduction littérale serait le procédé le plus naturel de la traduction. Néanmoins, souvent ce n'est pas le cas, par exemple, le français et le tchèque diffèrent dans les deux aspects. Pour des raisons historiques, le tchèque est socioculturellement plus proche de l'allemand que du français, ce qui se manifeste par des tournures que le tchèque partage avec l'allemand et qui peuvent donc être traduites d'une langue dans l'autre littéralement tandis que, en français, on exprime les mêmes choses différemment_ nous avons plus haut déjà mentionné la locution tchèque *držet palce* qui correspond à la locution française *croiser les doigts* ; elle peut être traduite en allemand littéralement *die Daumen halten* ; cela concerne aussi le choix de collocations, par exemple tch. *dostat horečku* ~ all. *Fieber bekommen* × fr. *prendre la fièvre*. Quant aux différences structurelles, la seule présence des déterminants en français et leur absence en tchèque mène J. ŠABRŠULA [19, p. 22] à mettre en question le caractère littéral de la traduction tchèque *Muži mají radši blondýnky* de la phrase française *Les hommes préfèrent les blondes*. la traduction littérale se restreint, selon le même auteur, à des cas élémentaires tels que : *Marie est malade* ~ *Marie je nemocná*. Là où on ne peut pas se servir de la traduction littérale, parce qu'elle conduirait à une construction agrammaticale ou qu'elle ne serait pas un équivalent fonctionnel de l'original, il faut avoir recours à d'autres procédés de traduction ; à la différence de l'emprunt, du calque et de la traduction littérale qui sont directs, les autres procédés de traduction nécessitent « un bouleversement plus ou moins grand de l'agencement ou même du lexique » et sont considérés comme indirects [37, p. 46-47].

4.3.5. Transposition

Les transpositions reflètent les différences typologiques des langues. Elles consistent dans le remplacement d'une partie du discours qui exprime un sens par une autre [37, p. 50]. Cependant, J. ŠABRŠULA [19, p. 27] élargie cette définition en prenant en compte également les morphèmes (préfixes et suffixes) et souligne que les transpositions incluent aussi le zéro linguistique.

Le caractère nominal prédispose le français à l'usage des constructions verbonominales là où le tchèque se contente d'un verbe simple. La structure syntaxique de ces constructions est variée : verbe avec sujet, avec complément d'objet direct, avec complément d'objet indirect ou adverbial (avec préposition), par exemple : *la nuit tombe ~ stmívá se ; garder le silence ~ mlčet ; mettre en colère ~ rozzlobit*. On peut, bien entendu, trouver des cas contraires, où le tchèque préfère une construction verbonominale tandis que le français utilise un verbe simple, par exemple *se suicider ~ spáchat sebevraždu*.

Le nom, ou plus précisément sa catégorie grammaticale du nombre, permet d'exprimer la quantité de l'action : le singulier sert à désigner une action sémelfactive tandis que le pluriel donne à l'action le caractère itératif ou distributif. Le tchèque est capable d'exprimer cette différence à l'aide de l'aspect verbal, par exemple : *pousser un cri ~ vykřiknout ; pousser des cris ~ vykřikovat*.

Le choix du verbe, en revanche, peut nuancer l'action d'une manière semblable à l'usage des préfixes en tchèque, par exemple : *avoir un sourire ~ usmát se ; esquisser un sourire ~ pousmát se*.

Les constructions avec le verbe *échanger* expriment la réciprocité, qui peut être rendu en tchèque par le pronom réflexif, éventuellement renforcé par l'adverbe *vzájemně*, par exemple : *ils échangèrent un murmure ~ zamručeli na sebe*.

Une construction répandue, constituée du verbe *avoir* et d'un nom abstrait correspond en tchèque à la copule *být* « être » + adjectif, par exemple : *avoir de l'importance ~ být důležitý*. Cette construction est concurrencée par celle formée du verbe *être* + la préposition *de* + le nom et aussi par celle formée du verbe *être* + adjectif : *être d'importance ; être important*. La construction *être* + *de* + nom s'utilise notamment lorsque le nom est qualifié : *être d'une grande importance*.

L'emploi d'un nom dans les constructions françaises en face d'un verbe ou d'un adjectif en tchèque entraîne une transposition secondaire : l'adjectif qui qualifie le nom est remplacé par un adverbe, par exemple : *pousser un cri terrible* ~ *strašlivě vykřiknout* ; *avoir une grave maladie* ~ *být vážně nemocen*.

Lorsqu'une construction verbonominale française a en tchèque un équivalent également verbonominal, l'adjectif qualificatif français peut être transposé par la dérivation d'un diminutif, par exemple : *allumer un petit feu* ~ *zažehnout ohýnek*. Dans ce cas, E. JANOVCOVÁ [38, p. 296-7] parle d'une concentration, de même que dans le cas des verbes simples qui correspondent aux constructions verbonominales : le contenu sémantique est exprimé dans la langue de départ par plusieurs mots que dans la langue d'arrivée. Le contraire est la dilution qui consiste dans le fait que la langue d'arrivée emploie plusieurs mots que celle de départ, par exemple la collocation française *subir un interview* se traduit en tchèque par *absolvovat osobní pohovor*, où le nom *interview* est exprimé par deux mots *osobní pohovor*.

La différence entre le complément d'objet direct et le complément d'objet indirect ou adverbial est, dans la perspective de J. ŠABRŠULA, considéré comme une transposition de la préposition zéro, par exemple : *demander pardon* ~ *žádat o prominutí*. De même, l'article français se transpose, en tchèque, dans l'article zéro.

J. ŠABRŠULA [19, p. 60-4 ; 35, p. 104] prête attention au déterminant possessif. En français, l'emploi d'un déterminant est plus ou moins obligatoire ; le déterminant possessif s'utilise parfois aux endroits où sa valeur découle du contexte ; sa traduction en tchèque par l'adjectif possessif serait superflue : *j'écris mon devoir* ~ *píšu úkol* ; *il passe son examen* ~ *skládá zkoušku*. Il peut être en concurrence avec l'article défini : *fumer sa pipe* ~ *fumer la pipe* ~ *kouřit dýmku*. Dans d'autres cas, le déterminant possessif est important pour le sens de la locution et on le rend en tchèque par le datif du pronom personnel ou réflexif, par exemple : *il est venu à mon secours* ~ *přišel mi na pomoc* ; *se ôter ses gants* ~ *sundat si rukavice* ; à comparer avec la locution *se laver les mains*. Parfois, le déterminant possessif signale une manifestation actuelle d'un état qui se répète habituellement ; on peut le transposer, en tchèque, au pronom démonstratif *ten* et

la combinaison d'adverbes *už zase*, par exemple : *elle a sa migraine*, J. ŠABRŠULA traduit *přišla na ni ta její migréna*, nous proposons de traduire *už zase má tu svoji migrénu*.

4.3.6. Modulation

La modulation consiste dans le changement du point de vue, par exemple : *être pris d'une fièvre* × *dostat horečku* litt. « recevoir une fièvre ». Elle peut être fondée sur différents types d'oppositions : concret / abstrait, action / son résultat, une partie / le tout, deux parties différentes telles que « les lèvres » / « les dents » (*siffler entre ses lèvres* ~ *syknout mezi zuby*), verbes corrélatifs tels que « donner » / « recevoir », etc.

4.3.7. Équivalence et adaptation

L'équivalence associe à l'image de la langue de départ une image différente de la langue d'arrivée qui produit pourtant le même effet informationnel ; elle concerne surtout les idiomes, les proverbes, les comparaisons, etc. L'adaptation est le remplacement d'une situation qui n'existe pas dans le milieu de la langue d'arrivée par une situation semblable ; c'est donc une sorte d'équivalence. J. ŠABRŠULA [19, p. 36] en donne l'exemple suivant : si l'on voulait traduire la fable du loup et de l'agneau pour les enfants d'une communauté qui ne connaît pas le loup, on replacerait le loup par une autre bête féroce.

4.4. Conclusions

Dans le processus de la traduction, il s'agit de créer un texte qui remplisse, dans le milieu de la langue d'arrivée, les mêmes fonctions communicationnelles que l'original remplit chez ses récepteurs. Outre la sémantique, il faut respecter aussi le style du texte. Pour les variétés stylistiques, on peut employer un modèle tridimensionnel : l'axe vertical contient différents niveau de langue, sur l'axe horizontal sont réparties différentes variétés relevant de domaines spécialisés, et le troisième axe est destiné au variétés régionales. Les locuteurs de la langue d'arrivée n'ont pas la même expérience

linguistique et extralinguistique que les locuteurs de la langue de départ, ce qui se manifeste par la nécessité de recourir à des moyens linguistiques différents pour décrire la même situation, voire de remplacer la situation par une situation analogue. L'école franco-canadienne a élaboré un ensemble de sept procédés de traduction qui aident à surmonter des différences structurelles aussi bien que pragmatiques entre la langue de départ et la langue d'arrivée.

5. Corpus

La notion du *corpus* comme regroupement de textes est connue depuis des siècles ; on connaît, par exemple, le *Corpus iuris civilis* qui regroupe des textes juridiques romains ou encore le *Corpus Hippocraticum* contenant des textes anciens de médecine attribués traditionnellement à HIPPOCRATE mais provenant de différents auteurs et de différentes époques quoique la majorité d'entre eux correspondent à la durée de la vie active de ce médecin fameux [44, p. 60-61]. Or, pour le linguiste, un corpus comme ensemble de textes choisis selon certains critères constitue le matériel de base pour ses recherches. F. ČERMÁK [45, p. 119] souligne deux avantages du corpus : les données se trouvent dans leur contexte naturel et des usages marginaux de la langue ne prévalent pas sur les usages typiques et de base ; ainsi, on peut étudier les phénomènes linguistiques objectivement et sous tous les aspects, mais l'induction de conclusions ne s'appuie pas sur des analogies insuffisantes. Ces avantages doivent être considérés par rapport à d'autres sources, telles l'introspection et l'élicitation, où se manifeste l'invention plutôt que la spontanéité d'expression.

Un problème crucial lié au corpus est celui de la recherche d'informations dans le corpus. La recherche séquentielle, c.-à-d. la relecture de tous les textes du corpus, est peu efficace, sa complexité en temps étant d'ordre $O(n)$ et donc dépendant de la taille du corpus ; c'est pourquoi on recourt à une seule lecture du corpus pendant laquelle toutes les informations souhaitées sont extraites du corpus et stockées dans des structures qui permettent un accès beaucoup plus rapide, de sorte que, dans le cas d'une table de hachage, l'accès est d'ordre $O(1)$ et donc presque immédiat. Autrefois, le linguiste, ou bien ses assistants, devait lire tous les textes du corpus et noter sur des fiches les informations d'intérêt, tous les mots avec leur contexte par exemple ; les fiches étaient ensuite triées, le plus souvent alphabétiquement, dans un fichier ; ainsi, on obtint une concordance. Si le corpus était à certains égards remarquable, le résultat pouvait être donné à la portée des autres chercheurs sous forme d'un livre imprimé, telle est, par exemple, la concordance des livres du Nouveau Testament éditée par W. F. MOULTON et A. S. GEDEN [46]. Dans la seconde moitié du XX^e siècle, la tâche a été

facilitée par des ordinateurs qui permettent d'effectuer des opérations algorithmisables de manière automatique, sans fatigue et plus vite que les humains. Ainsi, autour des corpus composés de textes numériques s'est développée une branche de la linguistique, dite linguistique de corpus, qui se spécialise aux méthodes de construction de corpus et de leur exploitation.

Notre objectif est d'employer les fruits de la linguistique de corpus pour créer un corpus comparable français-tchèque de textes du domaine de la santé en vue de l'étude de constructions verbonominales françaises et de leur équivalents tchèques.

5.1. Corpus comparable

Pour comparer deux langues, on peut avoir recours à un corpus parallèle. Celui-ci est composé de textes en deux versions (si possible, l'original et sa traduction) qui sont alignées de manière que leurs portions correspondantes soient présentées vis-à-vis. Toutefois, on est limité par la disponibilité de textes en deux versions : si, par exemple, le *Lexique illustré d'anatomie Feneis* est disponible en version française [47] aussi bien qu'en version tchèque [48], ce n'est pas le cas pour le manuel d'anatomie [49] écrit par R. ČIHÁK et utilisé comme ouvrage standard dans les facultés de médecine en République tchèque. Un deuxième inconvénient des corpus parallèles, mentionné par F. MANIEZ [50, p. 552], réside dans le fait que la traduction est influencée par la langue de départ et ne reflète pas parfaitement l'expression spontanée de la langue d'arrivée ; ainsi une traduction tchèque souffrirait de gallicismes, et à l'inverse, une traduction française, de bohémismes. De plus, la relation entre les deux versions peut être plus compliquée : par exemple, ni la version française ni la version tchèque du *Lexique illustré d'anatomie Feneis* n'est l'original, l'ouvrage original étant rédigé en allemand. Pour remédier à ces deux inconvénients des corpus parallèles, on peut utiliser des corpus comparables, qui sont, selon F. MANIEZ [50, p. 552], « constitués de textes originaux partageant une similitude de style et de contenu », tandis que F. ČERMÁK [51, p. 76 et 267] ne restreint pas expressément la définition du corpus comparable à des textes originaux⁽²²⁾.

22. Si F. MANIEZ [50] considère les corpus comparables comme un complément avantageux des corpus parallèles, F. ČERMÁK [51, p. 76], au contraire, insinue l'infériorité des corpus comparables par

Au lieu de la dichotomie corpus parallèle / corpus comparable, S. SHAROFF *et al.* [52, p. 2-3] proposent une classification graduée des corpus bilingues selon qu'ils sont constitués de :

— textes parallèles, comportant des traductions, éventuellement avec de légères modifications telles que le remplacement de *New York* par *Pékin* dans un exemple de la version chinoise du guide de l'utilisateur d'un logiciel de traitement de texte ;

— textes fortement comparables, comportant des traductions considérablement remaniées ou des textes indépendants rapportant le même événement ou décrivant le même sujet, par exemple, les articles de Wikipédia dans des langues différentes ;

— textes faiblement comparables, comportant des textes du même domaine et genre plus étroits mais décrivant des faits différents, par exemple, débats parlementaires sur les soins de santé dans le Bundestag, dans la Chambre des communes et dans la Douma⁽²³⁾, ou bien des textes du même domaine et genre plus vastes mais variables en sous-domaines et en genres spécifiques, par exemple, dans les forums de discussion consacrés à l'informatique, des discussions techniques sur l'administration d'un serveur Linux en anglais et des discussions d'utilisateurs sur AutoCAD en français ;

— textes sans rapport.

Apparemment, ce sont les deux premiers types de corpus de cette classification graduée qui trouvent le plus de faveur parmi les chercheurs, cf. [52, p. 4].

Si, dans un corpus parallèle, les équivalents sont aisément accessibles grâce à l'alignement de textes, c'est un défi de trouver des fragments parallèles dans un corpus comparable. Pour cette tâche, diverses approches ont été élaborées, dont SHAROFF *et al.* [52] donnent un aperçu.

F. MANIEZ [50] se concentre sur les collocations verbonominales avec la base *diagnosis* en anglais et *diagnostic* en français dans des corpus du domaine médical. D'abord, il trouve des collocations dans un corpus parallèle constitué d'articles en anglais et de leur traduction en français et établit une liste de collocatifs et de leurs équivalents français. Ensuite, il extrait des collocatifs d'un corpus comparable, en pro-

rapport aux corpus parallèles en disant que, à cause de l'absence de corpus parallèles, on utilise parfois *au moins* des corpus comparables.

23. Dans ce cas-là, on obtiendrait un corpus trilingue.

duisant deux listes de verbes, l'une pour la partie anglaise du corpus et l'autre pour la partie française ; il regroupe les verbes anglais selon leurs liens sémantiques et essaie d'attribuer à chaque groupe de synonymes les verbes français. Il croit avoir ainsi enrichi le répertoire d'équivalents de traduction mis à la disposition du traducteur, ce qui se justifie par la constatation, résultant de l'observation du corpus parallèle, que la langue d'une traduction est du point de vue lexical moins riche que celle de l'original. Or, un défaut de cette approche consiste dans le fait qu'elle s'appuie sur le sens du collocatif bien que celui-ci puisse ne pas être inclus dans le sens de la collocation, ce qui est de règle dans le cas des constructions à verbe support ; par exemple, les collocations tchèques *přijmout rozhodnutí* et *vydat rozhodnutí* sont synonymes quoique le sens de leurs collocatifs soit opposé, cf. [11, p. 50]. De surcroît, cette approche ne tient pas compte des transpositions grammaticales en laissant de côté les cas où le sens d'une collocation du type verbe + nom d'une langue n'est pas exprimé, dans l'autre langue, par une collocation du même type.

Une approche centrée sur les transpositions est proposée par L. DELÉGER et P. ZWEIFGENBAUM [53] qui font des recherches sur des corpus comparables constitués de documents français d'une part de spécialité, de l'autre part laïques traitant de thèmes médicaux. Selon une de leurs hypothèses, la langue de spécialité contient plus de constructions nominales que la langue laïque. Afin de trouver des paraphrases laïques de constructions nominales, tout d'abord, il extraient, à l'aide d'un dictionnaire, des paires de noms déverbatifs et de verbes correspondants avec leurs contextes ; ensuite, en se basant sur ces paires, ils déterminent des patrons grammaticaux de paraphrases ; enfin, ils recherchent des mots correspondants dans les contextes, en prenant en considération la synonymie et le changement des parties du discours. Cette approche a permis de découvrir des paraphrases telles que *consommation régulière* ~ *consommer de façon régulière*, *gêne à la lecture* ~ *empêche de lire* et *évolution de l'affection* ~ *la maladie évolue*. Toutefois, la synonymie s'est montrée trompeuse dans le cas d'une fausse paraphrase *prise en charge* ~ *prendre du poids*.

5.2. Choix de textes pour un corpus

La langue est accessible au chercheur indirectement à travers ses réalisations concrètes, c.-à-d. textes [54, p. 7-8]. Alors que la quantité des textes possibles est illimitée, la quantité des textes réellement produits dans un temps donné par un locuteur ou un groupe de locuteurs donné est finie. Dans certains cas, il est possible d'inclure dans le corpus la totalité ou quasi-totalité des textes d'un champs bien déterminé ; par exemple, lors de l'étude de l'idiolecte d'un romancier, on inclurait dans le corpus tous les romans écrits par celui-ci. Mais ordinairement, le corpus ne contient qu'un échantillon de textes.

5.2.1. Représentativité du corpus

Un échantillon représentatif d'une population donne une image fidèle de celle-ci. Or, il n'y a pas d'accord commun sur la nature de la population du corpus. J. CHROMÝ [55, p. 188] se demande si l'unité de base de la population du corpus est le texte ou le locuteur ; alors que la population de tous les textes tchèques serait mal connue, on aurait, grâce au recensement et à d'autres enquêtes sociologiques, une connaissance assez ample pour faire un échantillon représentatif des locuteurs tchèques ; on partirait donc de l'assomption que différents groupes démographiques et sociologiques ont différentes habitudes linguistiques.

Tous les textes n'ont pas la même portée et le même potentiel d'influencer la langue, par exemple un quotidien national et un quotidien régional. C'est pourquoi F. ČERMÁK [51, p. 19-20] distingue entre la réception et la production de textes. La réception concerne les habitudes des lecteurs et des auditeurs. La production, quant à elle, peut être divisée entre les producteurs, c.-à-d. auteurs de textes, et les textes produits. Ainsi, on obtient trois possibilités de concevoir le corpus, proposées par D. BIBER [56, p. 111] : dans les deux premiers cas, l'unité de base est la personne et l'on inclut dans le corpus les textes dont chaque unité choisie est le récepteur ou le producteur, dans le troisième cas, l'unité de base est le texte même.

Pour V. CVRČEK *et al.* [57] la représentativité du corpus est guidée par l'effort de saisir la diversité de la réalité linguistique sans s'occuper des proportions des textes qui

sont porteurs de cette diversité ; si dans l'échantillon les proportions sont les mêmes que dans la population, ils parlent de l'équilibrage du corpus. Ce concept donne lieu à constituer des corpus représentatifs mais non équilibrés, tel SYN2015 ; F. ČERMÁK [51, p. 21], au contraire, fonde la représentativité juste sur des proportions des types de texte du corpus.

Du point de vue statistique, un échantillon représentatif est choisi de manière aléatoire d'une population d'individus dans le but de caractériser la population sur la base de l'échantillon ; si la population présente plusieurs sous-ensembles homogènes, le choix aléatoire peut être effectué dans chaque sous-ensemble (échantillonnage stratifié). Le choix aléatoire assure que chaque individu a la même chance d'être choisi, ce qui permet d'appliquer la statistique inductive qui est fondée sur des modèles probabilistes. Grâce au dépôt légal, tous les documents publiés sont répertoriés dans les bibliographies nationales et accessibles dans les bibliothèques nationales. Il est donc possible, notamment pour des langues limitées à une nation telles que le tchèque, d'établir la liste de tous les documents publiés dans la période étudiée, de les numéroter et d'effectuer un échantillonnage à l'aide d'un générateur de nombres aléatoires (ou pseudo-aléatoires), l'unité de base étant le texte produit. Si l'on s'intéressait aux producteurs, on établirait une liste d'auteurs. Pour la réception, la tâche serait plus compliquée ; les enregistrements de lecteurs dans les bibliothèques ou les indications de tirage ne fourniraient que des informations partielles sur la circulation d'écrits publiés ; un sondage auprès de la population serait nécessaire. Pareillement, obtenir un échantillon représentatif des textes écrits non publiés ou des textes parlés est problématique.

La représentativité du corpus est-elle une condition nécessaire de la recherche linguistique ? À cet égard, J. CHROMÝ [55, p. 189] distingue trois fonctions possibles du corpus. 1^o Il peut servir de réservoir de données pour trouver des exemples de l'usage de mots dans le contexte et des exemples de structures grammaticales ; le remplissement de cette fonction résulte de la taille du corpus plutôt que de sa représentativité. 2^o Il peut rapporter la distribution de moyens et permettre des analyses statistiques. 3^o Il peut guider la recherche. Ces deux dernières fonctions supposent la représentativité du corpus afin que l'on puisse généraliser les résultats ; autrement, les résultats se

rappellent uniquement aux textes du corpus. Selon R. ČECH [58], les données statistiques ne sont qu'un supplément illustratif de la description à moins d'être mises en relation avec d'autres caractéristiques ; néanmoins, il propose de quitter l'approche descriptive en faveur d'une approche basée sur la vérification d'hypothèses pour laquelle la représentativité ne serait pas nécessaire, parce que les mécanismes qui guident le comportement linguistique des utilisateurs de la langue, ayant la nature d'une loi, se manifesteraient dans des textes individuels ; le mélange de différentes influences dans un échantillon, au contraire, dégraderait la vue du fonctionnement de ces mécanismes.

5.2.2. Classification de textes

Si certains [51, p. 21-23 ; 59, p. 37] considèrent la classification de textes comme une condition nécessaire à la représentativité, elle peut aussi servir à établir la comparabilité [57, p. 97].

Tout d'abord, on distingue entre les textes écrits et les textes oraux. Les deux formes de texte exigent des approches distinctes : le texte oral doit être enregistré, ensuite transcrit, la transcription peut être alignée sur la piste sonore dans le corpus. De plus, les textes oraux, notamment ceux produits de manière spontanée, sont moins facilement accessibles aux linguistes que les textes écrits, quoique la langue parlée est plus utilisée au quotidien que la langue écrite, comme le remarque F. ČERMÁK [51, p. 22], ce qui est une entrave à inclure dans un corpus des textes écrits et des textes oraux en juste proportion. Par conséquent, on fait d'ordinaire séparément des corpus écrits et des corpus oraux. Quant aux corpus écrits, ils sont le plus souvent constitués de textes publiés.

V. CVRČEK *et al.* [57, p. 95] divisent les textes écrits de corpus en trois groupes : fictionnels, non fictionnels et journalistiques. C'est le deuxième groupe qui nous intéresse : outre des textes administratifs et des « mémoires »⁽²⁴⁾, il comprend des textes scientifiques, des textes professionnels et des textes de vulgarisation, qui sont, à leur tour divisés selon leur appartenance aux sciences humaines, sociales, naturelles, et for-

24. Quant à la catégorie « mémoires », « il s'agit surtout de mémoires, autobiographies, de livres d'interviews de personnes célèbres, de récits de voyage, etc., dont l'ambition primaire n'est pas de créer un texte artistique » [57, p. 96].

melles et techniques, les branches finales de cette classification arborescente représentant différentes disciplines ; il y est encore ajouté des textes interdisciplinaires.

Les trois types des textes non fictionnels nous font penser à la différenciation du style fonctionnel des textes de spécialité dans la tradition linguistique tchèque. M. ČECHOVÁ *et al.* [24, p. 223-224] proposent la différenciation suivante :

— style de textes scientifiques ou théoriques, qui s'adressent aux spécialistes et transmettent des idées et des connaissances nouvelles ;

— style de textes de spécialité pratiques, dont le lecteur connaît la problématique de base de la discipline en question et qui, en s'orientant vers une activité pratique, peuvent ne pas apporter des connaissances nouvelles ;

— style de textes didactiques, qui sont supposés non seulement de transmettre des connaissances mais aussi d'éveiller l'intérêt pour elles et de contrôler le progrès de leur acquisition ;

— style de textes de vulgarisation, destinés aux non-spécialistes et qui se rapprochent de textes journalistiques ou littéraires ;

— style essayistique, qui mélange des éléments du style de spécialité et du style artistique.

Il est évident que les frontières entre ces cinq types de texte ne sont pas infranchissables, et l'on peut parfois hésiter à classer un texte : alors qu'un manuel pour les lycéens représenterait bien les textes didactiques, un photocopié destiné aux étudiants contiendrait moins d'éléments didactiques et ressemblerait à un texte de spécialité pratique. Toutefois, nous retenons cette typologie de textes pour un corpus spécialisé ; en outre, il faut déterminer la discipline ou le domaine au sein duquel on se trouve, par exemple la santé ou bien les sciences médicales et paramédicales, et préciser les thèmes de textes. Nous arrivons ainsi à un classement bidimensionnel à valeurs discrètes où chaque texte est caractérisé par un couple type — thème. Le corpus peut contenir un ou plusieurs types et un ou plusieurs thèmes. Pour constituer un corpus comparable bilingue, on recueillera des textes en deux langues caractérisés par les mêmes couples type — thème.

5.2.3. Étendue temporelle

Le corpus permet d'étudier soit la synchronie, c.-à-d. un seul état de langue, soit la diachronie, c.-à-d. la transition entre deux ou plusieurs états de langue. « En pratique, un état de langue n'est pas un point, mais un espace de temps plus ou moins long pendant lequel la somme des modifications survenues est minimale » [60, p. 142].

F. ČERMÁK [51, p. 76-77] considère comme le critère d'un corpus synchronique le caractère vivant de la langue qui coïncide avec la coexistence de plusieurs générations et était autrefois estimé à 30 ans. Selon le même auteur, de grandes transformations de la langue sont indiquées par des facteurs sociaux et des événements historiques, tels que guerres (les deux guerres mondiales) et révolutions (celle de 1989 en Tchécoslovaquie). L'an 1989 était, en fait, appliqué comme le *terminus post quem* pour les textes non fictionnels dans les corpus tchèques [57, p. 84]. Cet an annonçant la fin de la guerre froide signale également une rupture à l'échelle internationale et convient donc aussi aux textes français et cela d'autant plus que ce sont à peu près les 30 ans mentionnés par F. ČERMÁK qui le séparent du moment présent.

Cependant, nous rencontrons à nouveau une tension entre la production et la réception de textes : tout en reconnaissant la réception éphémère de textes journalistiques, V. CVRČEK *et al.* [57, p. 85] incluent dans le corpus SYN2015 des textes journalistiques produits (publiés) dans un intervalle de cinq derniers ans, tandis que parmi les textes fictionnels et non fictionnels sont choisis seulement ceux qui sont censés être lus en 2015, ce qui est déterminé selon la date de parution. Il faut noter que la condition de la première édition dans la période de 25 derniers ans imposée aux textes non fictionnels peut exclure plusieurs textes non fictionnels qui sont plus anciens mais toujours en circulation ; pareillement pour les textes fictionnels dont la première édition ne doit pas être de plus de 75 ans et à la fois une édition quelconque, de plus de 25 ans. Le critère déclaré de réception est donc discutable.

5.2.4. Taille du corpus

La taille des corpus numériques suit le progrès technique et la disponibilité de textes en format approprié. Si les corpus de première génération *Brown* et *LOB* contenaient environ un million de mots [59, p. 30-31], aujourd'hui des corpus écrits dépassant un milliard de mots ne font pas exception. Étant donné des difficultés liées au recueil et à la transcription de textes oraux, la taille des corpus oraux s'attarde derrière celle des corpus écrits : alors que le corpus écrit SYN2015 compte 100 millions de mots [57], le corpus oral ORAL2013 en a moins de 3 millions [61, p. 48]⁽²⁵⁾. Les corpus spécialisés sont souvent encore plus modestes, ce qui est compréhensible parce qu'ils ne couvrent qu'une variété ou un segment de la langue.

F. MANIEZ a utilisé un corpus comparable de langue de spécialité dans le domaine médical, constitué d'un module anglais et d'un module français ; chacun d'entre eux contenait environ 300 mille mots [50, p. 558]. L. DELÉGER et P. ZWEIGENBAUM ont créé trois paires de corpus (chaquefois l'un de langue laïque, l'autre de langue de spécialité) consacrés à trois thèmes : la dépendance à la nicotine, le diabète et le cancer ; les corpus comptaient entre 228 742 et 641 584 mots [53, p. 7].

Une taille relativement petite permet d'étudier des phénomènes typiques ; plus les phénomènes étudiés se trouvent à la périphérie, plus le corpus doit être grand. À titre d'exemple, *American Heritage Intermediate Corpus* qui compte 5 millions de mots a été employé pour créer un dictionnaire de fréquences, et on a estimé que pour déterminer de manière fiable les fréquences des mots qui y sont rares, on aurait besoin d'un corpus 100 fois plus grand [59, p. 33].

La statistique donne des consignes pour déterminer la taille de l'échantillon afin de refléter des phénomènes de la population avec une exactitude souhaitée. En nous appuyant sur des calculs statistiques utilisés dans l'épidémiologie [63, p. 117-119], nous pouvons distinguer entre la fréquence d'un phénomène dans l'ensemble de textes

25. Dans les années 1950, pour élaborer le « français élémentaire » (puis rebaptisé « français fondamental »), l'équipe de G. GOUGENHEIM a enregistré au magnétophone, transcrit et dépouillé 163 conversations comptant 312 000 mots [62]. Ce travail peut être considéré comme précurseur des corpus oraux ; néanmoins, il ne s'agissait pas d'un corpus numérique.

du corpus, la taille étant exprimée par le nombre de mots ou le nombre de positions⁽²⁶⁾, et la fréquence moyenne d'un phénomène dans différents textes du corpus, la taille étant exprimée par le nombre de textes. D. BIBER [56, p. 116-117, 128-133] montre l'application de cette dernière possibilité dans la construction du corpus et discute des difficultés qui y sont liées. Le calcul de la taille exige des informations préalables ce qui résulte dans une construction cyclique en commençant par un corpus-pilote et, le cas échéant, en introduisant des corpus intermédiaires qui fourniront les informations exigées pour déterminer la taille du corpus définitif.

5.3. Préparation de textes pour un corpus

Une fois choisis, les textes doivent subir divers traitements pour pouvoir être inclus dans le corpus, cf. [51, p. 29 et suiv.]. Tout d'abord on a besoin d'obtenir les documents en format de texte brut (*txt*). Dans le cas de documents imprimés, on peut recourir aux procédés de la reconnaissance optique de caractères; dans le cas de documents électroniques, on peut utiliser des convertisseurs tels que *pdftotext* qui fait partie du projet *Xpdf* [64]. Au besoin, les fichiers textes peuvent être nettoyés à l'aide d'outils standard d'UNIX / Linux, par exemple *sed*.

Dans une première approche, on peut considérer les textes d'un corpus comme une chaîne de tokens. Un token correspond à un mot, à un signe de ponctuation ou à un symbole tel que chiffres. La segmentation de texte en tokens ou tokénisation doit prendre en considération les particularités de la langue en question, dont la conjonction tchèque *-li* qui s'attache directement au mot précédant et l'apostrophe en français qui peut séparer deux tokens, par exemple *j'avais*, ou apparaître à l'intérieur dans un seul token, par exemple *aujourd'hui*. Pareillement, le trait d'union remplit différents rôles dans le texte : typographique, lexical et syntaxique [15, p. 79-80].

H. SCHMID [65] s'occupe de la désambiguïsation du point lors de la tokénisation : le point peut signaler la fin d'une phrase, terminer une abréviation ou servir de marque d'adjectif numéral ordinal après des chiffres dans certaines langues comme en tchèque ou en allemand; ses fonctions peuvent se combiner (une abréviation à la fin d'une

26. Pour l'explication de la notion de la « position », voir p. 60-61.

phrase). Le tokéniseur reconnaîtrait le point comme token à part entière seulement s'il signale la fin de phrase ; autrement, le point resterait attaché au chiffre ou à l'abréviation pour former avec celui-ci ou celle-ci un seul token ; dans les cas d'une combinaison de fonctions, le point serait doublé. Il nous semble discutable si les avantages de la désambiguïsation du point dépassent ses inconvénients ; elle risque, en effet, de faire entrer dans le corpus des erreurs de tokénisation, son exactitude rapportée ne dépasse pas 99,80 % [65, p. 15] ; de surcroît, l'utilisateur du corpus pourrait être embarrassé par les différents traitements du point.

À notre avis, la tokénisation devrait être simple et intuitive parce que l'utilisateur du corpus fera ses recherches sur les tokens et devra donc être toujours conscient de la manière dont les textes ont été segmentés en tokens, des inventions sophistiqués⁽²⁷⁾ de l'auteur du tokéniseur pouvant compliquer son travail.

Un texte tokénisé permet d'effectuer des requêtes sur les tokens. À l'aide d'expressions régulières, on peut atteindre un certain degré d'abstraction dans les requêtes. On s'intéresse souvent à l'usage d'un mot indépendamment de sa forme concrète. Or, dans le cas d'une flexion riche, il serait trop encombrant d'énumérer, dans une expression régulière, tous les formes possibles du mot en question ; *Bescherelle* [66] enregistre, par exemple, 41 formes différentes (les formes composées non comprises) du verbe *aller* qui sont bâties sur trois racines différentes. On a donc besoin d'une forme canonique du mot, appelée « lemme », qui représente n'importe quelle forme du mot. D'un autre côté, il serait utile de pouvoir inclure dans la requête des propriétés grammaticales ; on pourrait ainsi rechercher une sous-chaîne de tokens qui contienne, par exemple, un adjectif ou un verbe au subjontif. Dans une seconde approche, on peut donc considérer les textes d'un corpus comme une chaîne de positions, cf. [67, p. 33]. Plusieurs attributs peuvent être associés à chaque position. Un attribut est obligatoire et représente le token

27. Cela concerne aussi les expressions à plusieurs mots, telles que *fer à repasser* ou *pomme de terre*, qui peuvent être considérées comme des unités lexicales à part entière. Le statut de telles expressions en anglais est discuté par E. JEŽEK [3, p. 23-28], qui propose trois tests pour vérifier si elles ont le statut d'un seul mot. Cependant, l'évaluation de ces tests n'est pas toujours univoque : alors qu'E. JEŽEK confirme pour *evening dress* le statut d'un seul mot par l'impossibilité d'introduire entre les deux éléments le mot *black*, il nous semble bien concevable de dire *an evening or cocktail dress*. Pour les fins de la tokénisation, il nous semble le plus sûr de suivre l'orthographe et de tenir l'espace pour séparateur de tokens.

correspondant à la position en question. Les autres attributs sont facultatifs et peuvent représenter, par exemple, le lemme et la partie du discours. Un exemple de la chaîne de positions est montré dans le tableau 1, où chaque position est représentée par une ligne⁽²⁸⁾. Ainsi, on peut rechercher chaque position selon n’importe quel attribut ou selon une combinaison d’attributs, par exemple un token *est* qui n’est pas un *nom* (en utilisant les opérateurs logiques « et » et « non »).

position	attributs		
	token	lemme	partie du discours
1	<i>La</i>	<i>le</i>	<i>déterminant</i>
2	<i>douleur</i>	<i>douleur</i>	<i>nom</i>
3	<i>est</i>	<i>être</i>	<i>verbe</i>
4	<i>intense</i>	<i>intense</i>	<i>adjectif</i>
5	.	<i>s/o</i>	<i>s/o</i>

TAB. 1. — *Un exemple de la chaîne de positions.*

La partie du discours n’est qu’un exemple de l’information grammaticale qui peut être ajoutée au corpus. Un attribut peut prendre la forme d’une étiquette qui encode plusieurs informations grammaticales, outre la partie du discours, par exemple le cas, le nombre, le genre, etc.

Quant aux noms d’attributs, notre expérience personnelle avec le Corpus national tchèque [68] a montré que leur taille affecte la facilité d’utilisation du corpus vu que l’utilisateur doit les écrire dans les requêtes en toutes lettres. Il est donc préférable de les réduire à une seule lettre.

La tokénisation (segmentation), la lemmatisation et l’étiquetage se font automatiquement à l’aide d’outils informatiques spécialisés, parmi lesquels *TreeTagger* et *MorphoDiTa*. Les deux peuvent être adaptés à différentes langues par l’entraînement sur des données appropriées.

28. Au lieu de *s/o*, l’attribut « lemme » de la position 5 pourrait prendre la valeur . et l’attribut « partie du discours », la valeur *punctuation*, si des raisons terminologiques n’en dissuadaient pas. Dans ce cas-là, on appellerait mieux les attributs « forme canonique » et « catégorie ».

TreeTagger est un étiqueteur probabiliste qui fait usage d'un arbre de décision [69 ; 70]. Son exactitude atteint 96,81 % pour l'anglais et 97,53 % pour l'allemand [70, p. 7-8] ; le chiffre pour le français ne nous est pas connu. Quant à la lemmatisation, *TreeTagger* est capable de rendre les lemmes qu'il a appris lors de l'entraînement ou qui lui sont fournis dans un lexique supplémentaire. Cet outil est disponible sous forme compilée pour divers systèmes d'exploitation et avec des modules déjà entraînés pour plusieurs langues, dont le français [71].

MorphoDiTa a été créé spécialement pour le tchèque et d'autres langues avec une flexion riche [72]. À la différence de *TreeTagger*, il a la capacité de deviner le lemme du mot qui ne fait pas partie du dictionnaire d'entraînement. Pour le tchèque, l'exactitude de l'étiquetage atteint 95,75 % et l'exactitude de la lemmatisation, 97,80 % ; pour l'anglais, 97,27 % [72, p. 16]. *MorphoDiTa* est disponible non seulement sous forme compilée pour plusieurs systèmes d'exploitation mais aussi sous forme de code source [73].

F. ČERMÁK [51, p. 32] souligne que l'étiquetage reflète la théorie linguistique sous-jacente qui n'est pas sans doute la seule possible et détermine l'interprétation du texte et de ses unités. L'utilisateur devient donc dépendant de la décision subjective de l'auteur de l'étiqueteur qui a opté pour une théorie particulière.

Le module français de *TreeTagger* dispose de 33 étiquettes [74]. Celles-ci permettent de diviser les mots en parties du discours. Les verbes sont, à leur tour, sous-divisés selon les modes et les temps. Les pronoms ont les sous-catégories suivantes : démonstratifs, indéfinis, personnels, possessifs et relatifs. Les numéraux sont considérés comme une partie du discours à part entière. Il y a deux étiquettes distinctes pour les noms communs et les noms propres. Seuls les articles et les adjectifs possessifs sont classés sous les déterminants. Une sous-catégorie des prépositions est dédiée aux articles contractés, y compris l'article partitif masculin. Certaines étiquettes sont associées aux signes de ponctuation, aux symboles et aux abréviations.

Pour le tchèque, il a été élaboré un système d'étiquettes à 15 caractères où chaque caractère, identifié par sa position dans l'étiquette, encode une information [75, p. 32-88]. Les significations des positions dans l'étiquette sont les suivantes : 1^o partie du discours ; 2^o précision de la partie du discours ; 3^o genre ; 4^o nombre ; 5^o cas ; 6^o genre du

possesseur ; 7^o nombre du possesseur ; 7^o personne ; 9^o temps ; 10^o degré de comparaison ; 11^o négation ; 12^o voix ; 13^o réserve pour un usage à venir ; 14^o réserve pour un usage à venir ; 15^o variante, style, registre. 3 922 étiquettes distinctes sont plausibles [72, p. 15]. La position qui est sans objet pour le mot en question est désignée par un trait d'union (ASCII 45). Par exemple, le mot *injekcemi* aura l'étiquette NNFP7-----A--- où 1^o-2^o NN = nom quelconque ; 3^o F = féminin ; 4^o P = pluriel ; 5^o 7 = instrumental ; 11^o A = affirmatif.

Le jeu d'étiquettes pour le tchèque que nous venons de décrire a été créé à Prague et est employé dans *MorphoDiTa*. Cependant, il existe un jeu d'étiquettes concurrentiel développé à Brno [76]. Les étiquettes de Brno sont formées d'une chaîne de couples xY où Y est la valeur de l'attribut x ; la longueur de la chaîne n'est pas fixe ; un ordre canonique est recommandé pour les couples. Par exemple, le mot *injekcemi* aurait l'étiquette k1gFnPc7 où k1 = partie du discours nom, gF = genre féminin, nP = nombre pluriel, c7 = cas septième (instrumental).

Dans la chaîne de positions qui constitue le corpus, chaque document peut être délimité par des balises structurelles indiquant le début et la fin du document et ajoutant des informations sur le document, telles que l'auteur, le titre, l'année de parution, le genre, etc. Ces informations serviront ensuite à limiter la requête à un groupe défini de documents. Pareillement, on peut délimiter par des balises structurelles les chapitres, les paragraphes, les phrases, etc.

5.4. Exploitation du corpus

L'exploitation du corpus nécessite une infrastructure appropriée. Dans sa thèse de doctorat [67], P. RYCHLÝ a analysé les exigences requises pour une telle infrastructure et proposé des solutions. Sur ces fondements, il a ensuite bâti un gestionnaire de corpus, *Manatee*, avec une interface utilisateur graphique, *Bonito* [77]. Cet ensemble de logiciels, dont le code source est librement disponible, est connu sous le nom de *NoSketch Engine* [78] pour être distingué de sa contrepartie commerciale, appelée *Sketch Engine* [79] et qui peut faire, en plus, des *word sketches* (c.-à-d. résumés du comportement grammatical et collocationnel de mots) [80 ; 81].

Le travail avec un corpus part du lancement de requêtes. *NoSketch Engine* offre un langage de requête très puissant, CQL; ses caractéristiques de base sont les suivantes : Une position est délimitée par des crochets. L'attribut que l'on recherche est indiqué par son nom, sa valeur sous forme d'une expression régulière est mise entre guillemets (ASCII 34) et reliée à l'attribut par l'opérateur relationnel « égal à » qui peut, au besoin, être nié par un signe d'exclamation : `[attribut="valeur"]` ou `[attribut!="valeur"]`. Plusieurs attributs peuvent être reliés par des opérateurs logiques & « et » et | « ou » dont l'ordre de l'évaluation peut être indiqué par des parenthèses : `[attribut_1="valeur_1" & (attribut_2="valeur_2" | attribut_2="valeur_3")]` ⁽²⁹⁾. Plusieurs positions peuvent être enchaînées dans la requête. Des crochets vides désignent une position quelleconque : `[]`. Les accolades mises après une position indiquent la répétition de celle-ci : `[...]{min,max}`. Il est aussi possible d'avoir recours à des balises structurales dans la requête ; la commade `within <balise/>` à la fin de la requête assurera que toute la chaîne de positions recherchée soit contenue à l'intérieur d'une seule structure délimitée par un couple de balises `<balise>` et `</balise>`.

Les résultats de la requête sont présentés sous forme de concordance. À partir des résultats de la requête, on peut faire différentes opérations : classer les résultats alphabétiquement en appliquant le tri alphabétique à une position choisie par rapport au mot recherché ; créer un échantillon aléatoire des résultats ; calculer les fréquences des mots autour du mot recherché ; affiner les résultats par des filtres positifs et négatifs ; détecter des candidats aux collocatifs à l'aide de méthodes statistiques, telles que T-score, MI-score, logDice, etc. ; visualiser la distribution des résultats dans le corpus.

5.5. Conclusions

Pour notre corpus comparable français-tchèque du domaine de la santé, nous avons proposé une classification de textes bidimensionnelle qui tient compte à la fois de la diversité de types de texte et de celle de thèmes. Dans ce cadre, nous optons pour

29. On pourrait ne pas répéter l'`attribut_2` et utiliser l'opérateur | à l'intérieur des guillemets comme suit : `[attribut_1="valeur_1" & attribut_2="valeur_2|valeur_3"]`.

un corpus monotypique et monothématique, constitué de textes didactiques relatifs aux soins infirmiers ; par conséquent, nous évitons le problème de la proportion de différentes catégories de textes. Le corpus ne servira pas à induire des caractéristiques statistiques de la population hors le corpus mais constituera plutôt une population en soi et pourra aussi remplir la fonction de réservoir de données ; nous n’insistons donc pas sur un choix strictement aléatoire de textes ; d’ailleurs, examiner la production des manuels de soins infirmiers ou leur usage dans la formation de futurs infirmiers dépasserait les limites du présent mémoire et nous gardons cette tâche pour une éventuelle étude ultérieure, possiblement dans le cadre d’une thèse de doctorat. Notre choix de textes sera influencé par des aspects matériels et économiques. Pour la date de publication⁽³⁰⁾ de textes inclus dans le corpus, nous fixons le *terminus post quem* à l’an 1989 qui à la fois signale une rupture historique et délimite un espace temporel partagé par plusieurs générations. Quant à l’étendue spaciale, nous supposons que la langue de spécialité est plus standardisée, aux niveaux national ainsi qu’international, que la langue commune ; d’éventuelles différences géographiques ne sont pas notre intérêt principal, et donc, lors du choix de textes, nous ne tiendrons pas compte de leur origine géographique. Nous considérons une taille de plusieurs centaines de milliers de positions comme suffisante pour chacun de deux modules de notre corpus comparable. Chaque module sera traité comme un corpus à part entière. Pour faciliter dans le corpus la recherche de structures à un certain degré d’abstraction, nous adoptons le modèle de chaîne de positions qui ont trois attributs : le token, la forme canonique (lemme) et la catégorie (étiquette grammaticale). Nous tokéniserons, lemmatiserons et étiqueterons le module français à l’aide de *TreeTagger* et le module tchèque, à l’aide de *MorphoDiTa*. Nous intégrerons les deux modules dans le milieu de travail *NoSketch Engine* que nous ferons tourner localement sur un serveur *Apache*. La préparation de textes pour le corpus comportera des standardisations d’usages typographiques : pour une recherche aisée dans le corpus, il sera avantageux de remplacer l’apostrophe typographique par l’apostrophe

30. Vu que l’*Anatomie* de R. ČIHÁK dont la première édition (du premier volume) date de 1987 et qui a connu une 3^e édition en 2011-2016 [49] sert toujours chaque année à des centaines d’étudiants de médecine, nous n’imposons, à la différence de V. CVRČEK *et al.* [57, p. 85], aucune limite à la première édition mais limitons seulement l’édition incluse dans le corpus.

simple (ASCII 39); de même, le digramme *oe* est plus accessible à l'utilisateur que la ligature *œ*; il faudra prêter une attention particulière aux ligatures *ff*, *fi*, *fl*, *ffi* et *ffl* qui doivent impérativement être converties en digrammes et trigrammes correspondants.

6. Partie pratique

6.1. Création d'un corpus comparable

En appliquant les principes méthodologiques exposés dans la partie théorique du présent mémoire nous avons construit un corpus comparable, constitué de deux modules : *inf-fr* (français) et *inf-cs* (tchèque). Nous y avons inclus des ouvrages relatifs aux soins infirmiers utilisés dans la formation d'infirmiers et d'autres professionnels de santé. Le module français se compose de 5 ouvrages [82 ; 83 ; 84 ; 85 ; 86] et le module tchèque, de 8 ouvrages [87 ; 88 ; 89 ; 90 ; 91 ; 92 ; 93 ; 94]. Les tailles des deux modules exprimées en nombre de tokens diffèrent : alors que le module français compte 1 797 938 tokens, le module tchèque n'en comporte que 482 894. Cette différence ne menace pas le caractère comparable du corpus, la comparabilité étant déterminée par le type et le thème de textes. Notons que L. DELÉGER et P. ZWEIGENBAUM [53] ont construit, entre autres, un corpus comparable de textes relatifs au cancer dont le module laïque comptait 228 742 mots et le module de spécialité, 641 584 mots. Si l'on voulait s'occuper de la proportion des tailles, il faudrait prendre en considération le fait que différentes langues sont « bavardes » à différents degrés ; à titre d'exemple, après avoir compté les tokens sur une page d'un texte français [60, p. 105]⁽³¹⁾ et dans la portion correspondante de sa traduction tchèque [95], nous avons trouvé la proportion suivante 392 : 324⁽³²⁾.

Pour faciliter la saisie de requêtes, nous avons voulu choisir les noms d'attributs de positions les plus courts possibles : `t` comme « token », `l` comme « lemme » et `e` comme « étiquette ». Cependant, il s'est avéré que *NoSketch Engine* exige indispensablement un attribut appelé `word` qui comporte le token. Au lieu de doubler cet attribut sous les noms `t` et `word`, nous avons profité de la possibilité offerte par *NoSketch Engine* de créer un attribut virtuel `t` qui comporte le token en bas de casse. Les attributs de chaque position sont donc au nombre de quatre : `word`, `t`, `l` et `e`.

31. Nous avons inclus dans le compte encore deux premières lignes de la page suivante par lesquelles se termine un alinéa qui débute à la page en question.

32. Nous pouvons préciser que le texte français comporte plus de tokens de mots (359 : 285) mais moins de tokens de signes de ponctuation (33 : 39) que le texte tchèque.

Quant à la version d'outils informatiques, nous avons employé : *TreeTagger* ver. 3.2.3, binaire précompilé pour Linux 64 bits ; *MorphoDiTa* ver. 1.10.1, binaire précompilé pour Linux 64 bits avec le modèle tchèque *czech-morfflex-pdt-161115* ; *Manatee* ver. 2.167.8, code source ; *Bonito* ver. 3.116.13, code source. Les deux derniers, nous les avons compilés nous-même.

6.2. Objectifs et méthode de la recherche

Un des phénomènes par lesquels le français se diffère du tchèque est l'utilisation plus ou moins obligatoire des déterminants. Comme ce sujet éveille notre curiosité, nous voulons nous focaliser sur le rôle que les déterminants, plus particulièrement les différents types de l'article et les déterminants possessifs, jouent dans les collocations verbonominales françaises. Ensuite, nous voulons comparer les collocations verbonominales françaises avec leurs équivalents tchèques ; cette comparaison sera, bien entendu, asymétrique parce que du côté français seront disponibles uniquement des collocations verbonominales et du côté tchèque, leurs traductions possibles ayant des structures variées. En nous appuyant sur la partie théorique du présent mémoire, nous voulons vérifier si le datif d'intérêt du pronom tchèque d'une part et le déterminant possessif français de l'autre part sont équivalents.

Pour notre recherche nous utiliserons notre corpus comparable constitué des modules *inf-fr* et *inf-cs*. Nous choisirons une dizaine de noms, concrets aussi bien qu'abstraites, que nous supposons être présents dans le corpus *inf-fr* comme bases de collocations verbonominales. En utilisant le moteur de recherche de *NoSketch Engine*, nous essaierons de trouver, à l'aide d'expressions régulières, des combinaisons de ces noms avec le verbe qui pourra précéder ou suivre le nom. Si, pour la base nominale donnée, la requête résulte à plusieurs collocatifs verbaux différents, nous ne les retiendrons pas tous : en premier lieu, nous tiendrons compte de la fréquence ; deuxièmement, nous retiendrons aussi des collocatifs moins fréquents pourvu qu'ils soient synonymes des collocatifs retenus pour une haute fréquence ; troisièmement, nous nous réservons la possibilité de retenir des collocatifs remarquables pour une autre raison. Une fois les collocations trouvées, nous évaluerons les données.

La structure des équivalents que nous chercherons dans le corpus *inf-cs* n'est pas au préalable déterminée. Alors, pour formuler des requêtes en termes d'expressions régulières nous utiliserons des éléments radicaux des mots tchèques qui correspondent sémantiquement aux bases des collocations françaises; pour déterminer les éléments radicaux, nous aurons recours à notre connaissance de la langue tchèque et éventuellement aux dictionnaires. Ensuite, en parcourant les résultats fournis par le moteur de recherche, nous identifierons les traductions possibles des collocations françaises. Finalement, nous ferons une comparaison traductologique en nous concentrant notamment sur les transpositions et les modulations.

6.3. Recherche

6.3.1. Collocations verbonominales françaises

Pour notre recherche nous avons choisi les noms suivants : *accord, bandage, dent, diagnostic, gant, obstacle, pansement, perfusion, prélèvement, question*. Pour ces bases nominales, nous avons retenu à peu près une trentaine de collocatifs verbaux; certains d'entre eux sont partagés par plusieurs bases.

Lorsque nous parcourions les collocations en vue d'étudier l'usage du déterminant, notre attention a été attirée par des séries de collocations que l'on peut projeter sur l'axe du temps où les collocations marquent : 1^o l'instauration d'une situation, 2^o son éventuelle évolution et 3^o son aboutissement⁽³³⁾. Du point de vue syntaxique, les noms y jouent le rôle du complément d'objet direct et sont accompagnés d'un déterminant (DÉT) :

- a) *poser* DÉT₁ *questions*
répondre à DÉT₃ *questions*
- b) *mettre / enfiler* DÉT₁ *gants*
retirer / enlever / ôter DÉT₃ *gants*

33. À comparer avec les triades que F. ČERMÁK [96, p. 289-90] a décelées parmi les collocations tchèques à nom abstrait; elles désignent les phases inchoative, durative et terminative.

- c) *appliquer / faire / mettre / réaliser / poser / effectuer* DÉT₁ *pansement*
changer / refaire / renouveler DÉT₂ *pansement*
retirer / enlever DÉT₃ *pansement*
- d) *réaliser / appliquer / effectuer* DÉT₁ *bandage*
enlever DÉT₃ *bandage*
- e) *établir / poser* DÉT₁ *diagnostic*
affiner DÉT₂ *diagnostic*
confirmer / affirmer DÉT₃ *diagnostic*

Nous avons trouvé que le déterminant DÉT₁ a tendance à être représenté par l'article indéfini tandis que la position des déterminants DÉT₂ et DÉT₃ est occupée plus ou moins régulièrement par l'article défini. Cette distribution découle des fonctions habituelles des articles : d'abord on met en scène une nouvelle entité, qui est présentée comme individualisée mais non déterminée parmi les entités de la même classe, d'où l'article indéfini ; après qu'elle a été mise en scène, explicitement ou implicitement, on peut se référer à cette entité, elle est déterminée : en effet, on répond aux questions qui ont été posées, on refait le pansement qui a été fait, on confirme le diagnostic qui a été établi, on retire les gants que l'on a mis, etc.

Regardons les divergences de la tendance découverte pour le DÉT₁. Tout en restant individualisée, l'entité peut être déterminée de manière autonymique et pour cette raison elle prend l'article défini : *poser la question* « *Est-ce que [...] ?* » ; elle peut aussi être déterminée contextuellement, on parle, par exemple, d'un pansement dont le patient a besoin et dans ce cas on dit *appliquer le pansement*. Cependant, l'article défini peut attribuer à l'entité en question la valeur générique : *apprendre à réaliser le pansement ; réaliser le pansement tous les quatre jours*. Alors que l'article indéfini singulier, qui partage la forme avec le déterminant numéral, attribue à l'action exprimée par la collocation le caractère sémelfactif, l'article défini en emploi générique n'implique ni la répétition ni l'action unique.

Les noms *bandage* et *pansement* sont dérivés de verbes. On hésite sur le sens que ces noms prennent en fonction du collocatif verbal : avec le verbe *appliquer* ils désignent

plutôt le matériel (les bandes) utilisé pour cet action, avec d'autres verbes (*faire, effectuer, réaliser*), l'action même ou son résultat.

Le nom *diagnostic* combiné avec les collocatifs *établir* et *poser* prend de préférence l'article indéfini (26 occurrences), les occurrences avec l'article défini sont moins fréquentes (12). Or, avec les collocatifs *faire* et *porter* nous avons trouvé uniquement l'article défini (14 occurrences); il est difficile de décider si c'est un pur hasard lié à notre corpus.

Le nom *gants* au pluriel est d'ordinaire pris collectivement et constitue une seule entité; le caractère collectif est accentué dans des exemples comme *mettre une paire de gants* et *mettre les deux gants* où la totalité est explicitée, cf. en tchèque les mots *dva* « deux » et *oba* « les deux ». La collocation *mettre des gants* désigne donc une action unique à la différence des collocations *réaliser des prélèvements* et *appliquer des pansements* qui ont une valeur distributive grâce à l'article indéfini pluriel.

La collocation *mettre les gants* a une variante avec le déterminant possessif *mettre ses gants*, qui souligne la participation de la personne à l'action. Néanmoins, pour la base *gants*, ce type de variantes est très rare dans notre corpus : en combinaison avec le collocatif *mettre* une occurrence du déterminant possessif contre 16 occurrences de l'article défini, en combinaison avec le collocatif *enfiler* 4 occurrences du déterminant possessif contre 18 occurrences de l'article défini, et en combinaison avec le collocatif *ôter* une occurrence du déterminant possessif contre une occurrence de l'article défini.

Parmi nos collocations se trouve une dans laquelle la base nominale en position du complément d'objet direct est sans déterminant : *faire obstacle [à qqch.]*. Ce type de constructions remonte à l'ancien français [35, 107]. Elle est concurrencée par deux constructions analogues mais avec l'article indéfini : *créer / provoquer un obstacle [à qqch.]*.

6.3.2. Collocations françaises et leurs traductions tchèques

Pour traduire les collocations françaises *être d'accord* et *être en accord*, le corpus nous a fourni le verbe *souhlasit*; il s'agit de la concentration : le nom se transpose en verbe, et le verbe *être* ainsi que la préposition disparaissent. Les collocations *dire son*

accord et *signer son accord* ont des équivalents *vyslovit souhlas* et *podepsat souhlas*, où le déterminant possessif se transpose en déterminant zéro. La traduction *dát písemný souhlas* montre un changement complexe à la fois au niveau de la modulation et au niveau de la transposition : l'action d'apposer la signature est remplacée par l'indication de la forme écrite, le syntagme s'augmente d'un adjectif, et le déterminant disparaît. Dans la traduction *vyjádřit souhlas* un verbe générique est substitué aux verbes spécifiques *dire* ou *signer*. La relation sémantique entre les collocations *obtenir l'accord* et *zajistit souhlas* correspond à peu près à la différence qui existe entre les verbes français *obtenir* et *assurer*.

Collocations *réaliser / appliquer / effectuer DÉT bandage*. Comme nous avons déjà mentionné dans la partie pratique, le mot tchèque *bandáž* a été emprunté au français. Par la suite, le tchèque en a tiré le verbe imperfectif *bandážovat*. Si l'on a besoin d'exprimer l'aspect perfectif, il faut recourir à des constructions bien analogues : *přiložit bandáž* et *vytvořit bandáž*.

Les collocations *se brosser les dents* et *se laver les dents* peuvent se traduire en tchèque *čistit si zuby*. La structure est bien symétrique : les deux langues utilisent même le pronom réflexif ayant la fonction du datif d'intérêt, qui en cas de la substantivation *čištění zubů* disparaît. La seule différence structurelle entre les constructions consiste dans l'absence d'article en tchèque (l'article zéro). Cependant, la traduction présente une modulation. Le verbe *brosser* implique l'instrument que l'on utilise pour cet action (brosse à dents) et le verbe *laver* inclut la notion de l'eau. Le verbe tchèque *čistit* est plus générique, c'est, en fait, un hyperonyme des deux verbes français.

Le nom *diagnostic*, d'origine grecque, a en tchèque deux équivalents étymologiquement apparentés *diagnóza* et *diagnostika* ; le premier désigne le résultat, la conclusion sur l'état de la santé, le second, le processus qui conduit à cette conclusion. Chacun d'entre eux se combine avec un autre collocatif ; peut-être que la sélection du déterminant pour ce nom dont nous avons parlé plus haut, dans la section précédente, est liée à la différence du sens : *stanovit diagnózu* ~ *établir / poser DÉT diagnostic* ; *provést / provádět diagnostiku* ~ *faire le diagnostic*. Avec une modulation, nous avons encore trouvé la collocation *nalézt diagnózu*. Lorsque la maladie ou le trouble de la santé est

spécifié, par exemple : *poser le diagnostic de sarcoïdose*, on utilise en tchèque le verbe transitif *diagnostikovat*.

Les collocations *mettre DÉT gants* et *enfiler DÉT gants* correspondent respectivement aux collocations tchèques *obléknout si rukavice* et *navléknout si rukavice*. Le tchèque enrichi ces syntagmes par le datif du pronom réflexif ; comme nous avons mentionné dans la section précédente, le déterminant possessif, qui est d'ailleurs dans nos exemples très rare, pourrait jouer en français un rôle analogue.

Pour la multitude des collocations synonymes françaises *faire obstacle / être un obstacle / créer un obstacle / provoquer un obstacle / il existe un obstacle [à qqch.]*, nous avons trouvé, dans le corpus, un seul équivalent tchèque : le verbe *bránit*, construit avec le datif. Nous avons donc ici la correspondance habituelle entre le datif et la préposition *à*.

La base nominale *pansement* se combine avec des collocatifs de manière semblable que la base *bandage*. Toutefois, nous avons retenu un répertoire de collocatifs plus riche.

Aux collocations *appliquer / mettre / poser DÉT pansement*, où les collocatifs sont plus ou moins synonymes, le tchèque répond par les collocations *přiložit obvaz* et *aplikovat obvaz*. Le verbe tchèque *aplikovat* issu, de même que le verbe français *appliquer*, du verbe latin *applicāre*, a communément le sens « employer, utiliser » ; cependant, le dictionnaire [39, s. u.] reconnaît que, dans le langage médical, il est synonyme du verbe *přiložit*, qui signifie « appliquer, mettre sur ». Pour les collocations *faire / réaliser / effectuer DÉT pansement* le tchèque fournit une collocation analogue : *provést obvaz*.

Les collocations *recouvrir d'un pansement* et *recouvrir avec un pansement* peuvent être transformées au passif : *être recouvert par un pansement* ; cette collocation désigne dans le corpus l'action plutôt que l'état tandis que le participe passé sans verbe *être* désigne l'état : *recouvert d'un pansement*. À noter la variabilité de la préposition. Dans le corpus tchèque, nous avons trouvé deux tournures passives, l'une au passif réflexif *krýt se obvazem*, elle désigne l'action ; l'autre au passif périphrastique *být zakryt obvazem*, elle désigne l'état. Les différentes prépositions sont en tchèque remplacées par la flexion, *obvazem* est l'instrumental.

Les collocations *changer / refaire / renouveler le pansement* peuvent se traduire par *provést převaz*; le sens du verbe *changer* ou du préfixe *re-* se transforme en préfixe *pře-* qui n'est pas attaché au verbe mais au nom; c'est parce que le nom *převaz* est apparenté au verbe simple *převázat* que l'on peut utiliser aussi mais le verbe est transitif et exige un complément d'objet direct à la différence de la construction verbonominale.

Pour les collocations avec la base *perfusion* nous n'avons pas trouvé des équivalents tchèques.

Lorsque l'on transpose le nom français *prélèvement* en verbe tchèque *odebrat / odebírat*, on a tendance à y ajouter le nom *vzorek* « échantillon », ce qui permet de traduire les collocations françaises *faire / réaliser / effectuer DÉT prélèvement* par également par une collocation verbonominale *odebrat vzorek*. Il s'agit, en fait, de la dilution parce que le nom *prélèvement* inclut la notion de « prélever » aussi bien que la notion de « échantillon ». S'il faut ajouter le matériel que l'on prélève, on peut soit garder le nom *vzorek* et y ajouter le matériel comme complément du nom : *odebrat vzorek krve* soit remplacer le nom *vzorek* par le matériel : *odebrat krev*. Le tchèque a en plus la possibilité d'imiter la collocation française : *provést odběr*, dans ce cas l'ajout du nom *vzorek* serait superflu et l'on attache le matériel comme complément du nom *odběr* : *provést odběr krve*. Il est intéressant que le français utilise ici, contrairement à ce que nous avons dit dans la partie théorique sur le caractère nominal du français, l'adjectif *sanguin* : *faire un prélèvement sanguin*; le nom *sang* est employé lorsqu'il est qualifié : *faire un prélèvement de sang veineux*.

La collocation *poser des questions* peut se traduire par le verbe simple *ptát se*. Il est possible d'ajouter le sujet des questions à l'aide de la préposition *sur* ou *concernant*⁽³⁴⁾; de même, on emploie en tchèque la préposition *na* avec le verbe *ptát se* ou bien avec le verbe *dotazovat se*. La combinaison du pluriel avec l'article indéfini implique la traduction par un verbe imperfectif; le singulier, au contraire, *poser la question* se traduit par un verbe perfectif comme *zeptat se*. La catégorie grammaticale du nombre se transpose donc à l'aspect verbal. Le tchèque connaît aussi la collocation analogue *klást otázky*; les verbes *poser* et *klást*, qui, pris isolément, ont le même sens, servent ici de

34. En fait, il s'agit du participe présent du verbe *concerner* qui s'est lexicalisé et *Le Petit Robert* [8, s. u.] le désigne comme une préposition.

verbes supports, et donc leur sens ne contribue pas au sémantisme de ces collocations ; cela peut se manifester par la difficulté de trouver un verbe approprié et, par conséquent, mener à ce que l'on appelle, en stylistique latine, *figura etymologica* : *tázat se na [dříve uvedeně] dotazy* au sens « poser de nouveau les mêmes questions ». Et encore un exemple semblable mais avec la substantivation du verbe : *zeptání se na [klíčovou screeningovou] otázku*.

La collocation *répondre aux questions* a un équivalent analogue *odpovědět na otázky* avec complément d'objet indirect attaché au verbe par une préposition. On dit aussi *zodpovědět otázky* où la préposition se transpose à la préposition zéro parce que le complément d'objet indirect devient un complément d'objet direct.

Revenons à notre hypothèse sur l'équivalence du datif d'intérêt en tchèque et le déterminant possessif en français. Dans la section précédente, les collocations avec la base *gants* nous ont montré que le français ne ressent pas le besoin d'accentuer la personne intéressée par l'action partout où le tchèque est habitué d'utiliser le datif. Le déterminant sert, dans ces cas, plutôt d'une variante stylistique.

Dans les collocations françaises *se brosser / laver les dents* le complément d'intérêt est exprimé par un pronom personnel (réflexif) de même que dans la collocation tchèque *čistit si zuby*. Le français exclut, en effet, l'emploi du déterminant possessif lorsqu'il s'agit d'une partie de la personne ou de la chose impliquée dans l'action de la phrase [15, p. 155] ; les dents sont une partie du corps de la personne qui se les brosse / lave.

Dans la collocation *poser des questions*, les questions peuvent être adressés à quiconque. Pour les adresser à soi-même, il faut utiliser le pronom exactement comme en tchèque : *se poser des questions*.

De l'autre côté, le tchèque n'est pas consistant dans l'usage du datif d'intérêt de noms. Nous trouvons, dans le corpus *inf-cs*, un exemple avec le datif d'intérêt *odpovídá pacientovi a rodinným příslušníkům na otázky* et un exemple analogue avec le génitif possessif *odpovídá na otázky pacientů a rodin*. Le corpus *inf-fr* fournit les exemples suivants, l'un avec le complément possessif et l'autre avec le déterminant possessif : *répondant aux questions du patient ; le médecin a répondu à ses questions*.

L'équivalence entre le datif d'intérêt en tchèque et le possessif en français n'est pas parfaite. On ne peut pas traduire mécaniquement le datif par le possessif et à l'inverse.

6.4. Conclusions

Nous avons créé un corpus comparable qui est devenu la source primaire pour notre recherche des collocations verbonominales françaises et leurs traductions possibles tchèques.

D'abord, nous avons cherché des collocatifs pour une dizaine de bases nominales en tenant compte notamment de la fréquence. Parmi les collocations trouvées nous avons découvert des séries qui, projetée sur l'axe du temps présentent certaines tendances de la distribution du déterminant : alors que le premier membre la série montre une hésitation entre l'article indéfini et l'article défini, les membres suivants sont presque régulièrement accompagnés de l'article défini. Nous avons expliqué ces tendances par les fonctions habituelles des différents types de l'article. Ensuite, nous nous sommes concentré sur le premier membre où se rencontrent, au singulier, trois tendances : l'individualisation par l'article indéfini qui attribue à l'action le caractère sémelfactif, la généralisation par l'article défini, et la détermination contextuelle ou autonymique qui s'exprime également par l'article défini. Quant à l'article indéfini pluriel, il prête à la collocation la valeur d'une action multipliée et distributive ; cependant cette valeur se neutralise lorsqu'il s'agit d'entités qui forme un tout.

Ensuite, nous avons recensé les traductions tchèques que nous avons trouvées pour les collocations françaises. Les procédés de traduction appliqués comportaient notamment la transposition et la modulation. Il était souvent possible de traduire la collocation verbonominale française soit par une construction analogue soit par transposition en verbe simple. Cette transposition a montré l'avantage des constructions verbonominales par rapport aux verbes simple : le verbe simple exige un complément mais, en cas d'une construction verbonominale le complément est facultatif. Nous avons aussi comparé le rôle de l'aspect verbal en tchèque avec le rôle de la catégorie du nombre en français.

Nous avons vu que le déterminant possessif est parfois une variante stylistique de l'article défini. Dans certains cas, l'emploi du pronom personnel français en fonction du datif d'intérêt est obligatoire et ne peut pas être remplacé par le déterminant possessif.

7. Liste des sources utilisées

1. *Lexique des règles typographiques en usage à l'Imprimerie nationale*. 3^e édition. Paris : Imprimerie nationale, 1990.
2. Igor MEL'ČUK. « Collocations dans le dictionnaire ». Dans Th. SZENDE (dir.). *Les écarts culturels dans les Dictionnaires bilingues*. Paris : Honoré Champion, 2003, p. 19-64. Disponible en ligne sur <http://olst.ling.umontreal.ca/pdf/Collocations-Szende.pdf> [Nous citons selon la version en ligne.]
3. Elisabetta JEŽEK. *The Lexicon. An Introduction*. New York : Oxford university press, 2016.
4. Noam CHOMSKY. « Three models for the description of language ». Dans *IRE Transactions on Information Theory*, September 1956, p. 113-124.
5. František ČERMÁK. « Kolokace v lingvistice ». Dans František ČERMÁK, Michal ŠULC (dir.). *Kolokace*. Praha : NLN, 2006.
6. *C. Suetoni Tranquilli quae supersunt omnia*. Recensuit Carolus Ludovicus ROTH. Editio stereotypa. Lipsiae : Teubner, 1904.
7. Eugenio COSERIU. « Lexikalische Solidaritäten ». Dans *Poetica*, Juli 1967, Bd. 1, Heft 3, p. 293-303.
8. *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*. Josette REY-DEBOVE, Alain REY (dir.). Paris : Dictionnaires Le Robert, DL 2005.
9. Jan PETR *et al.* *Mluvnice češtiny (2). Tvarosloví*. Praha : Academia, 1986.
10. Dominique LE FLUR (dir.). *Dictionnaire des combinaisons de mots*. Paris : Dictionnaires Le Robert, 2007.

11. Jan RADIMSKÝ. *Verbo-nominální predikát s kategoriálním slovesem*. České Budějovice : Jihočeská univerzita, 2010.
12. CICÉRON. *Discours. Tome X. Catilinaires*. Text établi par Henri BORNECQUE et traduit par Edouard BAILLY. Paris : Les Belles Lettres, 1926.
13. František DANEŠ, Zdeněk HLAUSA, Miroslav GREPL *et al.* *Mluvnice češtiny (3). Skladba*. Praha : Academia, 1987.
14. Miroslav GREPL, Petr KARLÍK. *Skladba spisovné češtiny*. Praha : SPN, 1986.
15. Martin RIEGEL, Jean-Christophe PELLAT, René RIOUL. *Grammaire méthodique du français*. Paris : PUF, 2001.
16. Josef POLÁK. *Přehled středoškolské matematiky*. Praha : Prometheus, 1995.
17. Dagmar KNITTLOVÁ *et al.* *Překlad a překládání*. Olomouc : Univerzita Palackého, 2010.
18. Christiane NORD. *Translating as a purposeful activity : Functionalist approaches explained*. 2nd edition. London, New York : Routledge, 2018.
19. Jan ŠABRŠULA. *Teorie a praxe překladu*. Ostrava : Ostravská univerzita, 2007.
20. Jitka SVOBODOVÁ. « La traduction, médiation de fonctions langagières ». Dans Jan ŠABRŠULA. *Problèmes de la stylistique comparée française-tchèque et tchèque-française*. Praha : SPN 1986, p. 266-293.
21. Robert JAKOBSON. « Closing statement : Linguistics and poetics ». Dans *Style in language*, edited by Thomas A. SEBEOK. Second printing. Cambridge, Massachusetts : MIT, 1964.
22. Michel ARRIVÉ, Françoise GADET, Michel GALMICHE. *La grammaire d'aujourd'hui : guide alphabétique de linguistique française*. Paris : Flammarion, 1986.
23. Bohuslav HAVRÁNEK, Alois JEDLIČKA. *Česká mluvnice*. 5. vyd. Praha : SPN, 1986.

24. Marie ČECHOVÁ, Marie KRČMOVÁ, Eva MINÁŘOVÁ. *Současná stylistika*. Praha : Lidové noviny, 2008.
25. Jan ŠABRŠULA. *Základy francouzské stylistiky*. Ostrava : Ostravská univerzita, 2008.
26. Maurice GREVISSE, André GOOSSE. *Le bon usage*. 14^e éd. Bruxelles : De Boeck, 2018.
27. Y. DELATOUR *et al.* *Nouvelle grammaire du français*. Paris : Hachette, 2004.
28. Alois JEDLIČKA. *Spisovný jazyk v současné komunikaci*. 2. vyd. Praha : Univerzita Karlova, 1978.
29. Otomar RADINA. *Francouzština a čestina : Systémové srovnání dvou jazyků*. Praha : SPN, 1981.
30. *Dictionnaire de l'Académie française*. 9^e édition (et toutes les éditions précédentes). Disponible en ligne sur : <https://www.dictionnaire-academie.fr/>
31. Jaroslav HUBÁČEK. *Malý slovník českých slangů*. Ostrava : Profil, 1988.
32. *Compendium du Catéchisme de l'Église catholique*. Città del Vaticano : Libreria Editrice Vaticana, 2005. Version en ligne disponible sur http://www.vatican.va/archive/compendium_ccc/documents/archive_2005_compendium-ccc_fr.html
33. Igor MEL'ČUK. « Tout ce que nous voulions savoir sur phrasèmes, mais... ». Dans *Cahiers de lexicologie*, 2013, N^o 1 (N^o 102), p. 129-149. Disponible en ligne sur <http://olst.ling.umontreal.ca/pdf/melcukphrasemes2011.pdf> [Nous citons selon la version en ligne.]
34. Vladimír SKALIČKA. *Typ češtiny*. Praha : Slovanské nakladatelství, 1951.
35. Jan ŠABRŠULA. *Nominálně verbální konstrukce a povaha děje ve francouzštině*. Praha : SP-n, 1962.

36. Alfred MALBLANC. *Stylistique comparée du français et de l'allemand*. 4^e éd. Paris : Didier, 1968.
37. Jean-Paul VINAY, Jean DARBELNET. *Stylistique comparée du français et de l'anglais*. Nouvelle édition revue et corrigée. Paris : Didier, 1972.
38. Alena TIONOVÁ *et al.* *Francoušština pro pokročilé*. Praha : SPN, 1992.
39. Bohuslav HAVRÁNEK *et al.* *Slovník spisovného jazyka českého*. 2. vyd. Praha : Academia, 1989.
40. Louis-Bernard GUYTON DE MORVEAU, Antoine LAVOISIER, Claude-Louis BERTHOLLET, Antoine-François FOURCROY. *Méthode de nomenclature chimique*. Paris : Cuchet, 1787.
41. Pierre CHANTRAINE. *Dictionnaire étymologique de la langue grecque*. Paris : Klincksieck, 1968-1980.
42. Antoine MEILLET, Joseph VENDRYES. *Traité de grammaire comparée des langues classiques*. 4^e édition. Paris : Champion, 1968.
43. Bernard JULLIEN. *Cours supérieur de grammaire. Première partie. Grammaire proprement dite*. Paris : Hachette, 1848.
44. Vivian NUTTON. *Ancient Medicine*. London and New York : Routledge, 2004.
45. František ČERMÁK. « Jazykový korpus : Prostředek a zdroj poznání ». Dans *Slovo a slovesnost*. 1995, roč. 56, č. 2, p. 119-140.
46. William F. MOULTON, Alfred S. GEDEN. *A Concordance to the Greek New Testament*. 2nd edition. Edinburgh : T. & T. Clark, 1913.
47. Wolfgang DAUBER. *Lexique illustré d'anatomie Feneis*. Paris : Flammarion, 2007.
48. Wolfgang DAUBER. *Feneisův obrazový slovník anatomie*. 3. vyd. Praha : Grada, 2007.

49. Radomír ČIHÁK. *Anatomie*. 1.-3. díl. 3. vyd. Praha : Grada, 2011-2016.
50. François MANIEZ. « Extraction d'une phraséologie bilingue en langue de spécialité : corpus parallèles et corpus comparables ». Dans *Meta*. 2001, vol. 46, n° 3, p. 552-563.
51. František ČERMÁK. *Korpus a korpusová lingvistika*. Praha : Karolinum, 2017.
52. Serge SHAROFF, Reinhard RAPP, Pierre ZWEIGENBAUM. « Overviewing Important Aspects of the Last Twenty Years of Research in Comparable Corpora ». Dans Serge SHAROFF *et al.* (dir.). *Building and Using Comparable Corpora*. Heidelberg : Springer, 2013, p. 1-17.
53. Louise DELÉGER, Pierre ZWEIGENBAUM. « Extracting Lay Paraphrases of Specialized Expressions from Monolingual Comparable Medical Corpora ». Dans *Proceedings of the 2nd Workshop on Building and Using Comparable Corpora, ACL-IJCNLP 2009*. ACL and AFNLP, 2009, p. 2-10. Disponible en ligne sur : <https://www.aclweb.org/anthology/W09-3102.pdf>
54. Adolf ERHART. *Základy jazykovědy*. Praha : SPN, 1984.
55. Jan CHROMÝ. « Korpus a reprezentativnost ». Dans *Naše řeč*. 2014, roč. 97, č. 4-5, p. 185-193.
56. Douglas BIBER. « Reprezentativnost v projektu korpusu ». Dans *Studie z korpusové lingvistiky*. Praha : Karolinum, 2000, p. 107-136.
57. Václav CVRČEK, Anna ČERMÁKOVÁ, Michal KŘEN. « Nová koncepce synchronních korpusů psané češtiny ». Dans *Slovo a slovesnost*. 2016, roč. 77, č. 2, p. 83-101.
58. Radek ČECH. « Jen popis s čísly ? Perspektivy korpusové lingvistiky ». Dans *Naše řeč*. 2014, roč. 97, č. 4-5, p. 171-184.
59. Michal ŠULC. *Korpusová lingvistika. První vstup*. Praha : Karolinum, 1999.
60. Ferdinand de SAUSSURE. *Cours de linguistique générale*. Paris : Payot, 1971.

61. Lucie BENEŠOVÁ, Michal KŘEN, Martina WACLAWIČOVÁ. « Korpus spontánní mluvené češtiny ORAL2013 ». Dans *Časopis pro moderní filologii*. 2015, roč. 97, č. 1, p. 42-50.
62. Georges GOUGENHEIM. « Le français élémentaire. Étude sur une langue de base ». Dans *International Review of Education*. 1955, vol. 1, No. 4, p. 401-412
63. Adolf ŽÁČEK. *Metody studia zdraví a nemocí v populaci*. Praha : Avicenum, 1974.
64. *Xpdf*. Site du projet : <http://www.xpdfreader.com/>
65. Helmut SCHMID. *Unsupervised Learning of Period Disambiguation for Tokenisation*. Internal report. Stuttgart : IMS, University of Stuttgart, 2000. Disponible en ligne sur : <https://www.cis.uni-muenchen.de/~schmid/papers/tokeniser.pdf>
66. *Bescherelle Poche*. Paris : Hatier, 1999.
67. Pavel RYCHLÝ. *Korpusové manažery a jejich efektivní implementace*. Disertační práce. Brno : Masarykova univerzita, Fakulta informatiky, 2000.
68. *Český národní korpus*. Site du projet : <https://www.korpus.cz/>
69. Helmut SCHMID. « Probabilistic Part-of-Speech Tagging Using Decision Trees ». Dans *Proceedings of International Conference on New Methods in Language Processing*. Manchester, 1994. Version revue disponible en ligne sur : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf> [Nous citons selon la version en ligne.]
70. Helmut SCHMID. « Improvements in Part-of-Speech Tagging with an Application to German ». Dans *Proceedings of the ACL SIGDAT-Workshop*. Dublin, 1995. Version revue disponible en ligne sur : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf> [Nous citons selon la version en ligne.]
71. *TreeTagger*. Site du projet : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

72. Jana STRAKOVÁ, Milan STRAKA, Jan HAJIČ. « Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition ». Dans *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics : System Demonstrations*. ACL, 2014, p. 13-18. Disponible en ligne sur : <https://www.aclweb.org/anthology/P14-5003.pdf>
73. *MorphoDiTa*. Site du projet : <https://ufal.mff.cuni.cz/morphodita>
74. Achim STEIN. *French TreeTagger Part-of-Speech Tags*. 2003. Disponible en ligne sur : <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/french-tagset.html>
75. Jan HAJIČ. *Disambiguation of Rich Inflection. Computational Morphology of Czech*. Praha : Karolinum, 2004.
76. Miloš JAKUBÍČEK, Vojtěch KOVÁŘ, Pavel ŠMERK. « Czech morphological tagset revisited ». Dans Aleš HORÁK, Pavel RYCHLÝ (dir.). *RASLAN 2011. Recent Advances in Slavonic Natural Language Processing*. Brno : Tribun EU, 2011, p. 29-42.
77. Pavel RYCHLÝ. « Manatee/Bonito — A Modular Corpus Manager ». Dans Petr SOJKA, Aleš HORÁK (dir.). *RASLAN 2007. Recent Advances in Slavonic Natural Language Processing*. Brno : Masaryk University, 2007, p. 65-70.
78. *NoSketch Engine*. Site du projet : <https://nlp.fi.muni.cz/trac/noske>
79. *Sketch Engine*. Site du projet : <https://www.sketchengine.eu/>
80. Adam KILGARRIFF *et al.* « The Sketch Engine ». Dans *Proceedings of the Eleventh EURALEX International Congress*. Lorient : Université de Bretagne Sud, 2004, p. 105-115.
81. Adam KILGARRIFF *et al.* « The Sketch Engine : ten years on ». Dans *Lexicography*. 2014, vol. 1, n. 1, p. 7-36.

82. Pascal HALLOUËT. *Méga Mémo IFSI. Tout le programme semestre par semestre de l'étudiant infirmier*. 2^e éd. Issy-les-Moulineaux : Elsevier Masson, 2016.
83. Pascal HALLOUËT, Jérôme EGGERS, Evelyne MALAQUIN-PAVAN. *Fiches de soins infirmiers*. 5^e éd. Issy-les-Moulineaux : Elsevier Masson, 2015.
84. Léon PERLEMUTER, Gabriel PERLEMUTER. *Guide pratique de l'infirmière*. 3^e éd. Issy-les-Moulineaux : Elsevier Masson, 2011.
85. Carole LEMIRE, Vitalie PERREAULT. *Soins infirmiers. Fondements généraux. Méthodes de soins 1*. 2^e éd. Montréal : Chenelière Éducation, 2016.
86. Carole LEMIRE, Vitalie PERREAULT. *Soins infirmiers. Médecine Chirurgie. Méthodes de soins 2*. 2^e éd. Montréal : Chenelière Éducation, 2016.
87. Daniel JIRKOVSKÝ *et al.* *Ošetrovatelské postupy a intervence*. Praha : Fakultní nemocnice v Motole, 2012.
88. Zuzana ČERVENKOVÁ *et al.* *Ošetrovatelské postupy*. Pardubice : Univerzita Pardubice, 2018.
89. Natália BEHARKOVÁ, Dana SOLDÁNOVÁ. *Základy ošetrovatelských postupů a intervencí*. Brno : Masarykova univerzita, 2016.
90. Lenka MACHÁLKOVÁ. *Kapitoly z ošetrovatelské péče v hematologii*. Olomouc : Univerzita Palackého, 2016.
91. Lenka MACHÁLKOVÁ. *Kapitoly z ošetrovatelské péče v pneumologii*. Olomouc : Univerzita Palackého, 2019.
92. Lenka ŠÁTEKOVÁ. *Vybrané kapitoly z ošetrovatelství v onkologii*. Olomouc : Univerzita Palackého, 2019.
93. Blažena ŠEVČÍKOVÁ. *Kapitoly z ošetrovatelské péče v chirurgii a traumatologii*. Olomouc : Univerzita Palackého, 2019.

94. Eva JANÍKOVÁ, Renáta ZELENÍKOVÁ. *Ošetrovatelská péče v chirurgických oborech*. Ostrava : Ostravská univerzita, 2011.
95. Ferdinand de SAUSSURE. *Kurs obecné lingvistiky*. Přel. František ČERMÁK. 2. vyd. Praha : Academia, 1996.
96. František ČERMÁK. « Víceslovná pojmenování typu verbum — substantivum v češtině ». Dans *Slovo a slovesnost*. 1974, roč. 35, p. 287-306.
97. *Trésor de la langue française informatisé*. Disponible en ligne sur : <https://www.cnrtl.fr/definition/>
98. Josef NEUMANN, Vladimír HOŘEJŠÍ *et al.* *Velký francouzsko český slovník*. 2. vyd. Praha : Academia, 1992.
99. Anatole BAILLY. *Dictionnaire grec-français*. 26^e édition. Paris : Hachette, 1963.
100. *Λεξικό της κοινής νεοελληνικής*. Disponible en ligne sur : https://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/index.html
101. Gerhard WAHRIG. *Deutsches Wörterbuch*. 6. Aufl. Gütersloh : Bertelsmann, 1997.
102. *Duden. Wörterbuch*. Disponible en ligne sur : <https://www.duden.de/woerterbuch>
103. Notre appareil cognitif formé au cours de notre vie.

8. Résumé

Předkládaná diplomová práce je zaměřena na francouzské verbonominální kolokace a na jejich české překlady. Obsahuje dvě praktické úlohy: první úloha spočívá ve vytvoření srovnatelného francouzsko-českého korpusu textů z oblasti zdravotnictví; následně se korpus stává naším primárním zdrojem pro druhou úlohu, v níž studujeme množinu verbonominálních kolokací nalezených ve francouzském modulu korpusu, v českém modulu potom hledáme jejich možné překlady a podrobujeme translatologické analýze. Praktické části předchází čtyři teoretické kapitoly: nejdříve se zaměřujeme na vymezení pojmu kolokace; dále se věnujeme některým aspektům klasifikace verbonominálních kolokací, potom přecházíme k teorii překladu; nakonec se zabýváme problematikou týkající se vytvoření korpusu, zejména srovnatelného korpusu.

Věnujeme se třem různým pohledům na kolokace. Pohled F. Čermáka má nejširší záběr; jako nutnou a zřejmě i dostačující podmínku kolokace postuluje sémantickou kompatibilitu slov, která se kombinují. Dva zbylé pohledy patří E. Ježkové a I. Mel'čukovi; v nich je na kolokace aplikována podrobnější diferenciacce; jedná se hlavně o rozlišení mezi sémanticky více či méně transparentními kombinacemi a idiomatickými výrazy. E. Ježková podle našeho názoru nedostatečně zdůvodňuje rozlišování mezi tím, co nazývá omezenými kombinacemi, a kolokacemi. V této práci se kloníme k pohledu I. Mel'čuka, ovšem s tou výhradou, že netrváme na neuniverzalitě kolokátů. Obecně pokládáme v souladu s I. Mel'čukem jméno za bázi a sloveso za kolokát.

Valenční teorie umožňuje rozlišit syntaktické pozice, které může jméno ve verbonominálních kolokacích zaujímat. Často se jedná o valenční pozici podmětu či předmětu. I. Mel'čuk nabízí zajímavé roztrídění kolokací, založené na sémantických kritériích, pomocí lexikálních funkcí; tento prostředek se hodí zejména pro popis jednoho jazyka a jednojazyčnou lexikografii, avšak jeho využití pro porovnávání jazyků a pro překladatelství je omezeno strukturálními rozdíly mezi jazyky.

V překladatelském procesu jde o vytvoření textu, který bude v prostředí cílového jazyka plnit stejné komunikační funkce, jaké plní originál u svých příjemců. Vedle sémantiky je nutné respektovat také styl textu. Pro stylistické variety jazyka je možno použít

trojrozměrný model: vertikální osa obsahuje různé úrovně jazyka, na horizontální ose jsou umístěny variety patřící různým specializovaným oborům a třetí osa je určena regionálními varietám. Mluvčí cílového jazyka nemívají stejnou jazykovou a mimojazykovou zkušenost jako mluvčí zdrojového jazyka, což se projevuje nutností použít jiné jazykové prostředky pro popsání téže situace, anebo dokonce nahradit popisovanou situaci situací obdobnou. Francouzsko-kanadská škola vytvořila soubor sedmi překladatelských postupů, které pomáhají překonat strukturní i pragmatické rozdíly mezi zdrojovým a cílovým jazykem.

Pro náš francouzsko-český korpus jsme navrhli bidimenzionální klasifikaci textů, která bere v úvahu jak rozmanitost textů, tak rozmanitost témat. V této perspektivě je náš korpus monotypický a monotematický; je složen z didaktických textů z oblasti ošetrovatelství vydaných po roce 1989. Texty jsme tokenizovali, lematizovali a otagovali pomocí nástrojů *TreeTagger* a *MorphoDiTa* a zařadili do pracovního prostředí *NoSketch Engine*.

Cílem výzkumu je nejdříve prozkoumat roli determinantů ve francouzských verbonominálních kolokacích a potom srovnání s českými ekvivalenty. Jedná se samozřejmě o nesymetrické srovnání, protože na jedné straně máme výhradně verbonominální kolokace a na straně druhé různé možnosti jejich překladu. Chceme také ověřit, zda český dativ prospěchu a francouzský přivlastňovací determinant jsou ekvivalentní. V korpusu vyhledáváme pomocí regulárních výrazů.

Nejdříve jsme vyhledali kolokáty pro deset nominálních bází, zejména s ohledem na frekvenci. Mezi nalezenými kolokacemi jsme objevili série, které při promítnutí na časovou osu vykazují určité tendence v distribuci determinantů: u prvního prvku série bývá jakási fluktuace mezi členem neurčitým a určitým, zatímco ostatní prvky série jsou téměř pravidelně doprovázeny členem určitým. Tuto tendenci vysvětlujeme obvyklými funkcemi různých druhů členu. Poté jsme se soustředili na první prvek, kde se v jednotném čísle setkávají tři tendence: individualizace neurčitým členem, který dodává ději semelfaktivní charakter, dále generalizace členem určitým, a determinace vyjádřená rovněž členem určitým. Pokud jde o člen neurčitý v množném čísle, dodává kolokaci hodnotu

opakovaného a distributivního děje; nicméně tato hodnota se neutralizuje, jedná-li se o entity které tvoří jediný celek.

Poté jsme zhodnotili české překlady, které jsme našli pro francouzské kolokace. Identifikované překladatelské postupy zahrnují zejména transpozici a modulaci. Často bylo možno přeložit francouzskou verbonominální kolokaci buď pomocí analogické konstrukce, nebo pomocí jednoduchého slovesa. Tato transpozice ukázala výhodu verbonominálních konstrukcí ve srovnání s jednoduchým slovesem: jednoduché sloveso vyžaduje doplnění, které se v případě verbonominální konstrukce stává fakultativním. Rovněž jsme srovnali roli slovesného vidu v češtině s rolí gramatického čísla ve francouzštině.

Viděli jsme, že posesivní determinant je ve francouzštině někdy jen stylistickou variantou členu určitého. V některých případech je francouzské osobní zájmeno ve funkci dativu prospěchu povinné a nemůže být nahrazeno posesivním determinanem.

ANNEXE : Vocabulaire français-tchèque

ACCORD

dire son accord : vyslovit souhlas ; vyjádřit souhlas

être d'accord : souhlasit

être en accord : souhlasit

obtenir l'accord [de qqn] : zajist souhlas [od někoho]

signer son accord : podepsat / podepisovat souhlas ; dát písemný souhlas ; (vyjádřit souhlas)

se trouver en accord : souhlasit

BANDAGE

appliquer un / le bandage : přiložit bandáž ; bandážovat

effectuer un bandage : vytvořit bandage ; bandážovat

réaliser un bandage : vytvořit bandáž ; bandážovat

DENT

se brosser les dents : čistit si zuby

se laver les dents : čistit si zuby

DIAGNOSTIC

affirmer le diagnostic : potvrdit diagnózu

confirmer le diagnostic : potvrdit diagnózu

établir un / le diagnostic : stanovit diagnózu ; nalézt diagnózu

établir le diagnostic de qqch. : diagnostikovat + A

faire le diagnostic : provést / provádět diagnostiku

faire le diagnostic de qqch. : diagnostikovat + A

porter le diagnostic de qqch. : diagnostikovat + A

poser un / le diagnostic : stanovit diagnózu

poser le diagnostic de qqch. : diagnostikovat + A

GANT

enfiler les / ses / des gants : navléknout si rukavice

mettre des / les / ses gants : obléknout si rukavice

OBSTACLE

créer un obstacle à qqch. : bránit + D

être un obstacle à qqch. : bránit + D

il existe un obstacle à qqch. : bránit + D

faire obstacle à qqch. : bránit + D

provoquer un obstacle à qqch. : bránit + D

PANSEMENT

appliquer un / le pansement : přiložit obvaz ; aplikovat obvaz

changer le pansement : provést převaz

effectuer un pansement : provést obvaz

enlever le pansement : odstranit obvaz

faire un / le pansement : provést obvaz

mettre un pansement : přiložit obvaz ; aplikovat obvaz

réaliser un / le pansement : provést obvaz

recouvrir d'un pansement / avec un pansement : krýt se obvazem

être recouvert par un pansement : krýt se obvazem

refaire le pansement : provést převaz

renouveler le pansement : provést převaz

retirer le pansement : odstranit obvaz

PRÉLÈVEMENT

effectuer un / le prélèvement (sanguin) : odebrat vzorek (krve) ; odebrat krev ; provést odběr (krve)

effectuer des prélèvements : odebírat vzorky ; provádět odběry

faire un / le prélèvement (sanguin) : odebrat vzorek (krve) ; odebrat krev ; provést odběr (krve)

faire des prélèvements : odebírat vzorky ; provádět odběry

réaliser un / le prélèvement : odebrat vzorek ; provést odběr

réaliser des prélèvements : odebírat vzorky ; provádět odběry

QUESTION

poser des questions (sur / concernant qqch.) : ptát se (na něco); dotazovat se na něco;
klást otázky

poser la question : zeptat se

répondre aux questions : odpovědět na otázky; zodpovědět otázky