

OLGA NÁDVORNÍKOVÁ

## LES CORPUS PARALLÈLES : L'ESPACE POUR L'ANALYSE CONTRASTIVE\*

### 1. Introduction

L'objectif de cet article est de présenter un nouvel outil de recherche linguistique – les corpus parallèles (section 1), et à partir d'une analyse bien ciblée (les verbes de mouvement en construction gérondive) d'esquisser les possibilités que ce type de corpus offre aux linguistes (sections 2 et 3).

### 2. Les corpus parallèles

Il est possible de définir le corpus parallèle comme un ensemble de versions électroniques de traductions et de leurs originaux respectifs, alignés au niveau des paragraphes, phrases ou même des mots<sup>1</sup>. Ainsi, chaque segment de la version originale est relié au segment correspondant de la traduction.

Les premiers corpus parallèles ont été constitués dans les années 1990, dans le but de faciliter la traduction automatique et l'extraction de terminologie bilingue. Parmi les plus larges (et les plus anciens), on compte par exemple le corpus canadien *Hansard* (anglais-français), basé sur les débats du Parlement canadien. Le français est impliqué également dans les corpus CRATER (français-anglais) ou MLCC (*Multilingual Corpus for Cooperation*, français-anglais-italien-néerlandais-espagnol-allemand). Des corpus parallèles (et oraux) ont également été

---

\* Le présent article a été réalisé grâce au soutien financier du projet de recherche *Langue – structure, évolution, communication* du Ministère de l'Éducation nationale de République tchèque (VZ MSM 112100003).

<sup>1</sup> Il existe également des corpus *comparables*, composés uniquement de textes originaux ; la *comparabilité* d'un tel corpus est assurée par le fait que les textes proviennent du même domaine thématique (par exemple la radiologie). Ce type de corpus sert principalement de base de données pour l'extraction automatique de terminologie bilingue et il est moins fréquent que le corpus parallèle basé sur les traductions.

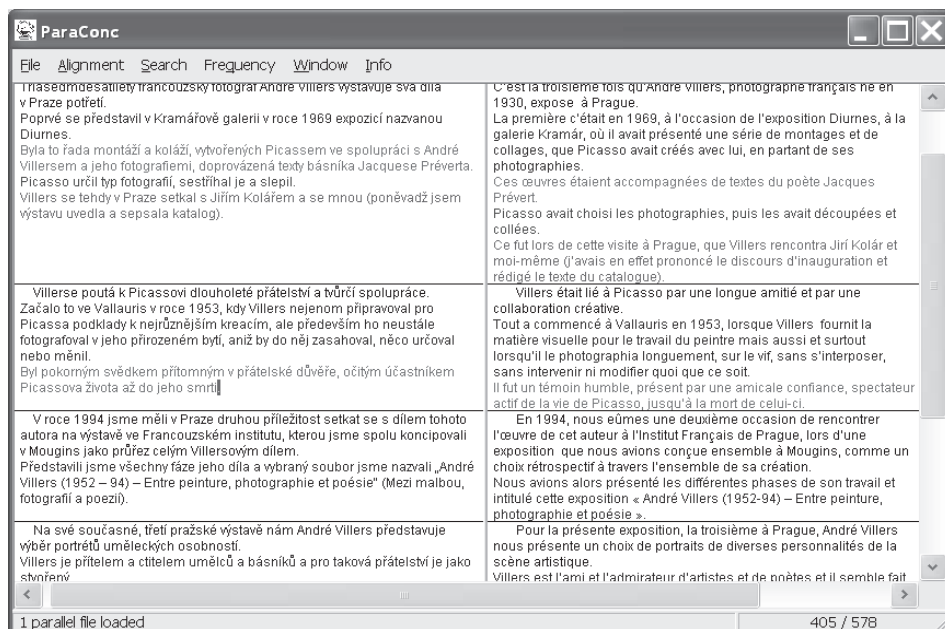


Tableau 1: Exemple de texte aligné au niveau de paragraphes

constitués par les équipes de Jean Véronis, expert en linguistique de corpus rattaché à l'Université de Provence<sup>2</sup>.

Parmi les corpus parallèles impliquant les langues slaves, citons au moins le corpus *Orwell*, composé de 11 traductions du roman *1984* (en bulgare, tchèque, slovenské, serbo-croate, russe, mais aussi en estonien, hongrois, roumain, letton et lituanien) et créé dans le cadre du projet MULTTEXT-East (<http://nl.ijs.si/ME/CD/docs/1984.html>)<sup>3</sup>; les textes sont alignés au niveau des phrases et dotés de l'annotation morphologique.

Il existe également des corpus parallèles en libre accès sur Internet, par exemple OPUS (<http://urd.let.rug.nl/tiedeman/OPUS/>), contenant entre autres le texte de la Constitution européenne en 21 langues et permettant de formuler la requête à l'aide d'expressions régulières. Certains logiciels d'exploitation de corpus parallèles sont capables d'identifier (grâce aux méthodes statistiques) l'équivalent

<sup>2</sup> Fondateur de l'équipe DELIC (à partir du 1<sup>er</sup> janvier 2008 l'équipe DELIC a rejoint l'équipe TALEP - Traitement Automatique du Langage Ecrit et Parlé - dans le cadre du *Laboratoire d'Informatique Fondamentale de Marseille*) et directeur de CILSH (*Centre Informatique pour les Lettres et Sciences Humaines*). Il mène aussi le blog *Technologies du Langage* (<http://aixtal.blogspot.com/>) consacré aux problèmes actuels du traitement automatique de la langue, mais également au fonctionnement général du langage.

<sup>3</sup> D'autres corpus de ce type sont basés par exemple sur le texte de la *Bible* ou de la *République* de Platon.

le plus probable de l'expression recherchée. Voyez ci-dessous un exemple de recherche simple (le mot *amour*) par le logiciel ParaConc (*Parallel Concordancer*):

Line	French Text	Czech Text
1	... L'Amour, roman Il est du véritable amour comme de l'apparition des esprits : tou ...	... Láška, román S opravdovou láskou je to jako se zjevením duchů: všichni o ...
2	... era ou que c'est déjà fait, on chante l'amour pour oublier qu'on en doute ou qu'on en ...	... časem, ne že ne, všechno pomine). Mluvíme o lásce, abychom zapomněli, že čas nad ní zvítězí anebo ...
3	... temps, va, tout s'en va). On parle d'amour pour oublier que le temps en triomphera ...	... časem, ne že ne, všechno pomine). Mluvíme o lásce, abychom zapomněli, že čas nad ní zvítězí anebo ...
4	... t sur 360° en quête de l'objet perdu (l'amour), ou encore si je grimpe à l'arbre géné ...	... o 360 stupňů a pátrám po ztracené věci (lásce) anebo když třeba šplhám na genealogický strom a ...
5	... tait vrai, ces histoires, le paradis, l'amour de Dieu, l'au-delà, elle qui approchait ...	... jsou pravdivé tyhle příběhy, to o ráji, o lásce Boží, o onom světě, co ona, která se ...
6	... La Rochefoucauld a beaucoup parlé de l'amour, et les amateurs de vérité cherchent da ...	... zas osamělá tak osamělá. La Rochefoucauld mluvil o lásce hodně a milovníci pravdy hledají v maximách definici, ...
7	... é - "la beauté, qui a tant de part à l'amour", écrit-il. Mais c'est une réflexion ...	... vpůli. Říkalo se, že jeho nejtrpčí myšlenky o lásce, její krátkosti či neexistenci, zapříčinilo právě tohle zranění, ...
8	... t que ses pensées les plus amères sur l'amour, sa brièveté ou son inexistence, lui fu ...	... vpůli. Říkalo se, že jeho nejtrpčí myšlenky o lásce, její krátkosti či neexistenci, zapříčinilo právě tohle zranění, ...
9	... e où se partagent le ciel et l'eau - en amour, les femmes sont plutôt de religion jui ...	... liníí doteku oblohy a vodní hladiny - v lásce jsou ženy spíš židovského vyznání: jejich bůh ještě ...
10	... siècles. - Vous parlez beaucoup de l'amour des hommes, m'a-t-elle dit. Mais que ...	... Lenclos před třemi stoletími. - "Mluvíte hodně o lásce lidí," řekla mi. "Ale co pro vás znamená ...
11	... eux s'ils n'avaient entendu parler de l'amour. " L'amour, c'est des mots, et le mot ...	... by nikdy nebyli zamilováni, kdyby neslyšeli mluvit o lásce." Láška, to jsou slova, a slovo objevuje věc; ...
12	... ère, dès qu'elle parle de lui et de son amour, finit invariablement par dire : " Mon ...	... někoho vhodného. Když o něm a o jeho lásce matka mluví, pokládá končí slovy: "Otec vždycky pro ...
13	... à parler l'amour, à parler couramment l'amour - pour moi seule. Plus tard, sa Bérén ...	... první, to kvůli mně se naučil mluvit o lásce, mluvit plynně láskyplně - kvůli mně jedině. Později ...
14	... ait pour moi qu'il apprenait à parler l'amour, à parler couramment l'amour - pour moi ...	... první, to kvůli mně se naučil mluvit o lásce, mluvit plynně láskyplně - kvůli mně jedině. Později ...
15	... qu'ils n'avaient eu ensemble, malgré l'amour, qu'une fille morte : Dieu est un enfant ...	... nemiloval, živého syna, zatímco oni dva vzdor své lásce měli spolu jen jednu mrtvou dceru: Bůh je ...
16	... ce que le hasard enlève du mystère à l'amour ou bien est-ce qu'il lui en confère, au ...	... le důležité. avy svn bvli šťastný." Ubírá náhoda lásce na tajemnosti, nebo naopak přidává? Anebo se párv ...
17	... C'était après une émission sur l'amour à la Maison de la Radio. Il s'ava	

Tableau 2: Exemple de recherche dans le corpus parallèle Intercorp

### 2.1. Le projet de corpus parallèles Intercorp

Le projet de corpus parallèles *Intercorp* (<http://www.korpus.cz/intercorp/>) a été lancé par l'Institut du Corpus national tchèque (<http://ucnk.ff.cuni.cz/>) en 2005. Il représente le prolongement naturel du large corpus unilingue, le *Corpus national tchèque*, contenant dans sa partie principale (synchronique) 200 millions de mots et doté de l'annotation morphologique et d'un logiciel d'exploitation intelligent (Bonito). La partie synchronique du *Corpus national tchèque* (SYN2000 et SYN2005) est complétée par le corpus diachronique, contenant des textes tchèques à partir du XIII<sup>e</sup> siècle, des corpus oraux (de Prague et de Brno) et d'autres corpus spéciaux, dont un corpus de lettres. Tout le projet (y compris *Intercorp*) est subventionné par le Ministère de l'Éducation nationale de République Tchèque, et l'accès aux corpus est gratuit.

*Intercorp* implique 19 langues (avec la langue tchèque au centre), à savoir l'allemand, l'anglais, le bulgare, le croate, l'espagnol, le finnois, le français, le hongrois, l'italien, le letton et le lituanien, le néerlandais, le polonais, le portugais, le russe, le serbe, le slovène et le suédois. Bien que la langue tchèque soit la langue pivot de tout le corpus, le logiciel d'exploitation permettra aux utilisateurs non-tchécoslovaques d'effectuer des recherches directement sur les langues choisies (jusqu'à quatre en même temps), sans inclure le tchèque.

Les corpus serviront à des fins scientifiques (recherche linguistique contrastive, élaboration de livres de grammaire contrastive, rédaction de mémoires et de thèses) et pédagogiques (enseignement de langues, traductologie). Les corpus parallèles permettront également de créer des outils linguistiques (dictionnaires) plus sophistiqués ainsi que des outils logiciels (visant, par exemple, la promotion de CAT, traduction assistée par ordinateur).

En ce qui concerne la composition du corpus, les textes littéraires sont les plus nombreux (compte tenu de leur grande disponibilité en versions parallèles); cependant, les coordinateurs tâchent d'inclure dans le corpus également les textes des autres genres, principalement issus des domaines scientifique et journalistique. Ainsi, dans la partie française du corpus, nous trouvons des textes parallèles de Sébastien Japrisot, Albert Camus, Amélie Nothomb, ceux de Milan Kundera, Bohumil Hrabal, Ladislav Fuks, etc., mais aussi des textes de Georges DUBY ou des programmes culturels bilingues, publiés et fournis par l'Institut français de Prague. Les textes sont obtenus dans la plupart des cas par OCR, mais un grand nombre de textes en tchèque nous ont été directement fournis par des maisons d'édition ou des traducteurs. L'obtention des textes français est compliquée, étant donné la date limite de la «synchronie» fixée à 1950 et le manque de volonté de coopération de la part des maisons d'édition françaises.

Pour le moment, la partie française du corpus contient à peu près 2 millions de mots (20 textes approx.), mais sa taille devrait bientôt doubler (grâce au nombre de textes préparés pour l'alignement). Comme le montre le tableau ci-dessous, parmi les langues les mieux représentées dans le corpus se trouvent également d'autres langues romanes – l'espagnol, l'italien et le portugais.

Langue	Nombre de mots		Total
	en tchèque	en langue étrangère	
espagnol	3 935 000	4 500 000	8 435 000
polonais	2 123 000	2 066 000	4 189 000
italien	1 815 000	2 088 000	3 903 000
anglais	1 802 000	2 099 000	3 901 000
suédois	1 729 000	1 993 000	3 722 000
portugais	1 517 000	1 744 000	3 261 000
croate	1 240 000	1 306 000	2 546 000
hongrois	1 196 000	1 190 000	2 386 000
serbe	1 049 000	1 122 000	2 171 000
russe	1 053 000	1 051 000	2 104 000

Langue	Nombre de mots		Total
	en tchèque	en langue étrangère	
FRANÇAIS	804 000	971 000	1 775 000
bulgare	867 000	868 000	1 735 000
néerlandais	795 000	919 000	1 714 000
allemand	496 000	612 000	1 108 000
letton			approx. 850 000
slovène	393 000	430 000	823 000
lithuanien	307 000	292 000	599 000
finnois	216 000	190 000	406 000

Tableau 3: Tailles des différentes sections linguistiques du corpus parallèle *Intercorp*

A présent, les coordinateurs des différentes sections linguistiques du corpus se concentrent aussi sur la constitution du « noyau commun » des textes présents dans *Intercorp*, à savoir sur le choix des textes susceptibles d'être traduits en plusieurs langues, et sur l'ajout des nouvelles variantes linguistiques des textes déjà présents en d'autres langues. Ainsi, nous avons déjà sept variantes linguistiques du *Seigneur des anneaux* de J.R.R.Tolkien, huit du *Nom de la Rose* d'Umberto Eco, neuf du *Brave Soldat Chveik* de Jaroslav Hašek, et même quinze de *L'Insoutenable légèreté de l'être* de Milan Kundera, etc.

Le corpus sera accessible sur Internet à partir de 2011 (la première version sera lancée en septembre 2008), et les premières recherches basées sur ce corpus seront présentées en septembre 2009 à Prague, dans le cadre d'une conférence internationale, consacrée entièrement à ce projet<sup>4</sup>. Le présent article ne constitue donc qu'un avant-goût des possibilités offertes par l'ensemble du corpus.

## 2.2. Le corpus spécial *Colomba*

Ce corpus, contenant la nouvelle *Colomba* de Prosper Mérimée et ses neuf traductions en tchèque<sup>5</sup>, a été créé dans le but d'analyser le répertoire des équivalents tchèques des gérondifs et participes présents français. L'intérêt d'une telle recherche consiste entre autres dans le fait que le transgressif tchèque, l'équivalent naturel de la majorité des occurrences du gérondif français, a presque disparu de l'usage contemporain. Par conséquent, comme nous le voyons dans les exemples suivants, les traducteurs tchèques doivent trouver d'autres solutions (et dans la plupart des cas expliciter le sens implicite du gérondif) – soit un verbe fini dans une proposition subordonnée ou indépendante, soit un syntagme nominal ou prépositionnel. Malgré sa petite taille, le corpus *Colomba* présente l'avantage

<sup>4</sup> D'ailleurs, en 2007, une thèse basée sur la partie française du corpus *Intercorp* a été soutenue avec succès (Martin Svášek, *Návrh, vytvoření a využití francouzsko-českého a česko-francouzského paralelního korpusu*).

<sup>5</sup> Les traductions ont paru entre 1875 et 1975 ; concrètement en 1875, 1920 (deux traductions), 1921, 1928, 1951, 1959, 1975 (deux traductions).

incontestable d'offrir neuf variantes tchèques pour chaque phrase française, en voici un extrait :

[1fr] <s lang="fr" id="30">Au thé, le capitaine charma de nouveau miss Lydia par une histoire de vendetta <i> transversale </i>\*, encore plus bizarre que la première, et il acheva de l'enthousiasmer pour la Corse **en lui décrivant** l'aspect étrange, sauvage du pays, le caractère original de ses habitants, leur hospitalité et leurs mœurs primitives. </s> [COLOMBA]

[1cz-a] <s lang="cz" transl="BAR" id="30">Při čaji roznítíl setník znovu zájem miss Lydie příběhem o krevní mstě, ještě bizarnějším nežli byl prvý a docela ji nadchl pro Korsiku **popisem** podivuhodného vzeření divoké krajiny, svérázné povahy obyvatelů, jejich pohostinství a primitivních mravů. </s> [SN – *par la description*, équivalent utilisé dans deux traductions]

[1cz-b] <s lang="cz" transl="CER" id="30">Při čaji kapitán miss Lydie znovu okouzlit příběhem o <i> dokonalé </i> krevní mstě, ještě podivnějším, než byl ten první, a nakonec ji nadchl pro Korsiku **tím, že jí popsál** podivný, divoký ráz té země, svéráz jejich obyvatel, jejich pohostinnost i jejich primitivní mravy.</s> [proposition circonstancielle de moyen]

[1cz-c] <s lang="cz" transl="CIB" id="30"> Při čaji kapitán znovu okouzlit miss Lydie příběhem o transversální krevní mstě\*) a tato historka byla ještě daleko divočejší než jeho první vyprávění; nakonec ji uvedl nad Korsikou do vytržení, **když jí popisoval** zvláštní, divoký ráz ostrova, originální charakter korsického lidu, jeho pohostinnost a jeho primitivní mravy. </s> [proposition circonstancielle de temps, équivalent utilisé dans deux traductions]

### 3. Sémantisme du gérondif (Halmøy 1982)

L'exemple précité illustre bien un des deux types principaux du sémantisme du gérondif identifiés et décrits par Odile Halmøy dans la monographie *Le Gérondif. Éléments pour une description syntaxique et sémantique* (Halmøy 1982), à savoir l'*antériorité logique* (Type A).

#### 3.1. Type A

En effet, Halmøy a observé que les sens de «moyen», «cause», «condition» et «temps» traditionnellement associés au gérondif ont en commun la notion d'*antériorité (chrono)logique* par rapport à l'action principale. Pour illustrer ce phénomène, Halmøy donne les exemples suivants :

«*Il réussira en travaillant - C'est en forgeant qu'on devient forgeron - Il a rougi en la voyant - En postant la lettre ce soir, elle arrivera demain*» (Halmøy 1982: 258)<sup>6</sup>

Les équivalents typiques des gérondifs du Type B sont principalement les propositions subordonnées circonstancielle (surtout de moyen et de temps, cf. exemple [1]).

<sup>6</sup> Halmøy note également que c'est justement le sens d'*antériorité (chrono)logique* qui exclut du sémantisme du gérondif les sens de *but* et de *conséquence* (Halmøy 1982: 389).

### 3.2. Type B

Par contre, si l'action exprimée par le gérondif n'entretient aucun rapport (chrono)logique avec l'action exprimée par le verbe régissant, il s'agit, selon Halmøy, de pure *circonstance concomitante* (*circonstance fortuite*, Type B), par exemple « *il se rase en chantant* » ou « *dit-il en riant* » (Halmøy 1982: 259). Les deux actions sont ainsi liées uniquement par le fait d'être exercées par un même agent en même temps.

Dans les textes littéraires, ce type de gérondif s'emploie le plus fréquemment dans les propositions incises (type *dit-il en riant*). Le transgressif mis à part, les équivalents tchèques les plus fréquents de ce gérondif du Type B sont d'une part le syntagme prépositionnel (dans le cas du gérondif sans expansions), principalement la préposition *avec*, exprimant clairement le sens de *circonstance concomitante*, d'autre part la proposition indépendante coordonnée avec la proposition matrice.

Tous ces équivalents possibles sont très bien illustrés par l'exemple de la construction gérondive *en souriant* (dans le corpus *Colomba*, l'expansion gérondive la plus fréquente du verbe *dire*):

[2fr] <s lang="fr" id="2189">- Oui, j'ai des nouvelles, dit Colomba **en souriant**.</s>  
s> [COLOMBA]

[2cz-a] <s lang="cz" transl="ZDA" id="2189">- Ano, mám od něho zprávy, řekla Colomba **usmívajíc se**.</s> TRANSFRESSIF

[2cz-b] <s lang="cz" transl="BAR" id="2189">"Ano, mám od něho zprávy," řekla Colomba **s úsměvem**.</s><sup>7</sup> SYNTAGME PRÉPOSITIONNEL (*avec sourire*)<sup>8</sup>

[2cz-c] <s lang="cz" transl="CER" id="2189">"Ano, mám o něm zprávy," pravila Colomba **a usmála se**.</s> PROPOSITION INDÉPENDANTE COORDONNÉE (*et elle a souri*)

<sup>7</sup> Les autres traductions confirment la tendance à traduire ce type de construction gérondive par un SP:

<s lang="cz" transl="CIB" id="2189">"Ano, mám o něm zprávy," pravila Colomba s úsměvem.</s>

<s lang="cz" transl="HOL" id="2189">"Ano, mám," řekla Colomba s úsměvem.</s>

<s lang="cz" transl="KAS" id="2189">- Ano, mám od něho zprávy, odpověděla Colomba s úsměvem.</s>

<s lang="cz" transl="LIN" id="2189">"Ano," odpověděla Colomba s úsměvem,</s>

<s lang="cz" transl="POC" id="2189">"Ano, mám od něho zprávy," řekla Colomba s úsměvem. </s>

<s lang="cz" transl="SMI" id="2189">"Ano, mám o něm zprávy," řekla Colomba s úsměvem.</s>

<sup>8</sup> D'autres constructions de ce type: *en soupirant* (*s povzdechem*), *en riant* (*se smíchem*), etc.



3.2.1. *Sous-type B' (Manière)*

Dans le cadre du Type B – *circonstance concomitante*, nous pouvons observer un sous-type spécial (d'après Halmøy, sous-type B' ; Halmøy 2003 : 104–106), par exemple :

[3] *Ludi vint à leur rencontre en galopant* – [4] *Dit-il en hurlant*

Halmøy observe que dans les deux cas, nous pouvons remplacer le verbe régissant par celui du gérondif sans changement considérable de sens; ainsi, selon Halmøy, on pourrait également dire :

[3'] *Ludi galopa à leur rencontre* – [4'] *Hurla-t-il*

Halmøy explique ce phénomène par le fait que les deux verbes (le verbe régissant et celui du gérondif) partagent un (ou plusieurs) sèmes et que le verbe du gérondif est ainsi en relation d'hyponymie avec le verbe régissant (*hurler* – *dire*, *galoper* – *venir*). Ensemble, les deux verbes désignent donc un seul procès (Halmøy 1982 : 389). (Cette analyse est confirmée aussi par les équivalents tchèques, qui suppriment en général le verbe régissant, tout en gardant le sème du gérondif – *dit-il en hurlant* – *zařval, il vint en galopant* – *přicválal*).

Les exemples de ce phénomène abondent, citons-en au moins deux du corpus de référence FRANTEXT :

*pour les verbes de dire :*

[5] LEROUX Gaston, *Le Mystère de la chambre jaune*, 1907, p. 24

[...] il descendit immédiatement de sa voiture dans un désordre d'esprit inexprimable. –allons! Allons! *Dit-il en balbutiant*.

*pour les verbes de mouvement :*

[6] BAZIN René, *Le Blé qui lève*, 1907, p. 243

[...] un abbé \*Roubiaux qui ne ressemblait plus entièrement à l'ancien. Il semblait avoir encore maigri ; sa soutane était blanche de poussière ; il *marchait en boitant* et appuyé sur un bâton : mais la petite tête noire, inattentive à la route, épanouie, dans le rêve, écoutait sûrement le cantique de la vie nouvelle.

Dans les deux exemples, il serait possible de faire économie du verbe régissant (*dire*, *marcher*), en disant seulement *balbutia-t-il* ou *il boitait*<sup>9</sup>. Par contre, selon Halmøy, il est impossible de décomposer la construction de Type B' en deux verbes finis coordonnés, ce qui confirme de nouveau le fait que les deux verbes désignent ensemble un seul procès :

<sup>9</sup> Néanmoins, dans le deuxième exemple, cette transformation serait possible uniquement si nous prenions le gérondif séparément du contexte ; en réalité, la réduction de la construction serait bloquée par la présence de l'attribut du sujet *appuyé sur un bâton*, coordonné avec le gérondif *en boitant* (*\*Il boitait et appuyé sur un bâton* est impossible, contrairement à *Il boitait appuyé sur un bâton*). Par ailleurs, la coordination du gérondif avec un attribut du sujet atteste le caractère double du gérondif: tout en étant complément circonstanciel, il garde son rapport au sujet.



*Verbes de mouvement :*

[3''] Ludi *vint* à leur rencontre **en galopant**. ≠ \*Ludi *vint* à leur rencontre **et il galopa**.

*Verbes de dire :*

[7] ...il *expliqua en criant*. ≠ \*...il *expliqua et il cria*.  
(Halmøy 1982 : 235)

En comparant ce type de sémantisme du gérondif avec celui de la *circonstance concomitante*, Halmøy postule que seul au sous-type B' peut s'appliquer l'étiquette traditionnelle de *manière*; en effet, tandis que les gérondifs dans les exemples [3] – [7] expriment vraiment la manière dont s'effectue le procès du verbe régissant (*en hurlant, en galopant, en balbutiant, en boitant, en criant*), le gérondif de *circonstance concomitante* (par exemple *Il a fermé la fenêtre en haussant les épaules*, Halmøy 2003 : 105) n'en dit rien. Néanmoins, ce type restreint de *manière* semble être limité aux verbes de dire et de mouvement.

Dans le cas prototypique, le gérondif exprimant la manière est postposé au verbe régissant (ce qui est la position typique du gérondif en général), par conséquent, il est rhématique; il est généralement lié (pas détaché) et co-référentiel (vu que dans l'absence de rapport logique entre les deux actions, l'agent commun est la seule liaison entre elles).

Cependant, Halmøy signale que la réduction de la construction à un seul verbe (type *marcher en boitant*) n'est pas toujours possible; c'est dans le cas où les sèmes du verbe régissant ne sont pas entièrement inclus dans le verbe du gérondif, par exemple *Le petit les rejoignit en courant* ou *Il se mit à danser en tournant sur lui-même* (Halmøy 2003 : 105). Nous pouvons ajouter d'autres exemples de ce type, tirés de FRANTEXT :

[8] MOINOT Pierre, *Le Sable vif*, 1963, p. 51

Lortier se retourna vers Jérôme qui désigna d'un bref regard le haut de la paroi. - On peut même *venir en nageant*, dit Françoise, penchée vers le chenal.<sup>10</sup>

ou d'INTERCORP:

[9] Alice *est arrivée en courant*, elle nous a mis la main dans la main, mariés, a-t-elle dit, il a cherché mes yeux. → Aube **přiběhla**, spojila nám ruce, "manželé", řekla, Yves se mi podíval do očí.

C'est ici que Halmøy touche (sans le développer davantage) un sujet intéressant où le point de vue contrastif peut apporter un peu plus de lumière. En effet, pour un locuteur tchèque (et même slave en général), les constructions des verbes de mouvement telles que *Alice et arrivée en courant* (VR<sub>vmt</sub> VG<sub>vmt</sub> non-réductibles à un seul verbe, contrairement à *marcher en boitant*)

<sup>10</sup> En réalité, même l'exemple cité par Halmøy (*Ludi vint à leur rencontre en galopant*) ne se prêterait pas à la réduction si le sème de *venir* n'était pas contenu également dans l'expression *à leur rencontre*, qui est gardée dans la construction réduite (*Ludi galopa à leur rencontre*).

représentent un groupe spécifique, dont l'équivalent tchèque est un seul verbe préfixé (*Alice přiběhla*)<sup>11</sup>.

Dans la dernière partie de cet article, nous allons donc nous concentrer sur la délimitation et le fonctionnement de ce type spécifique de gérondif.

#### 4. Analyses des verbes de mouvement en construction gérondive

[VR<sub>Vmt</sub> VG<sub>Vmt</sub> (VR  $\not\subset$  VG)]

##### 4.1. La classe des verbes de mouvement (Vmt)

La classe des verbes de mouvement (Vmt) semble être bien délimitée non seulement du point de vue syntaxique (voir Table 2 des *Méthodes en syntaxe* de M. Gross; Gross 1975: 237–242), mais aussi conceptuel (cf. par exemple l'article de Kekenbosch et al. *Sémantique cognitive de l'action: 2. étude expérimentale de la catégorisation des verbes*, paru dans la revue *Langages* en 1998)<sup>12</sup>. Dans l'expérience présentée dans leur article, Kekenbosch et al. ont soumis une liste de 172 verbes d'action les plus fréquents<sup>13</sup> au jugement de sept personnes (universitaires et étudiants de Paris VIII); la tâche était de classer les verbes en une vingtaine de cases (non-nommées préalablement), en fonction de leur sens. Dans cette expérience, les Vmt (et ensuite les verbes de possession) ont montré le degré de cohésion le plus élevé<sup>14</sup>; les verbes les plus représentatifs de la catégorie étant *traverser*, *venir*, *parcourir*, *avancer* et *aller*.

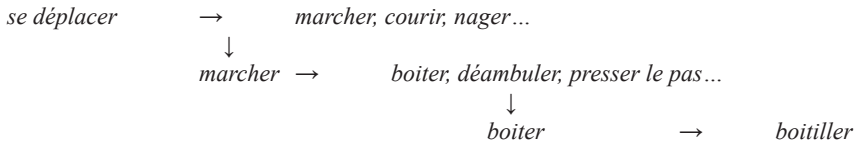
Il est possible de structurer le groupe des Vmt à l'aide des relations hiérarchiques (hyponymie et hyperonymie). Denis Le Pesant par exemple, dans son article *Introduction aux classes d'objets* (*Langages*, n°131, 1998, pp. 6–34) esquisse la structure suivante :

<sup>11</sup> Cf. par exemple *il marchait en boitant* (réductible) et un autre exemple tiré de FRANTEXT: SCHREIBER Boris, *Un silence d'environ une demi-heure*, 1996, p. 772, L'occupant de la cabine voisine sortait aussi, jeune blond en grossier uniforme kaki de l'O..T.. qui *vint* vers nous **en boitant** : - Quelle belle voix ! (non-réductible).

<sup>12</sup> L'article est disponible également en ligne, sur le site [www.persee.fr](http://www.persee.fr) créé par le Ministère de l'Éducation nationale.

<sup>13</sup> «Nous avons retenu seulement les verbes exprimant une action dont le résultat est un effet physique et observable et écarté ceux qui présentaient une polysémie manifeste.» (Kekenbosch et al. 1998: 50).

<sup>14</sup> Ainsi, à la fin de l'article, Kekenbosch et al. (1998: 65) définissent-ils le mouvement (MOU-VT) comme une des *primitives* de la Grammaire Appliquée et Cognitive (GA&C), avec APOSS (possession – sa conservation ou son transfert), FAIRE (surtout appliquée sur un objet), etc.



Le Pesant ajoute que «Les verbes hyponymiques représentent ici des modalités de leurs hyperonymes : *marcher*, c'est se déplacer «d'une certaine manière»; on peut se mouvoir plus rapidement (*courir*), dans plusieurs directions (*avancer*, *reculer*, *monter*, *descendre*), dans les airs ou dans l'eau (*voler*, *nager*), etc.; la marche elle-même peut s'effectuer de plusieurs façons, sans compter les nuances aspectuelles : *marcher/se mettre en marche*, *boiter/boitiller*... » (Le Pesant 1998 : 19–20).

#### 4.2. Organisation sémantique de $[VR_{Vmt} VG_{Vmt} (VR \not\subset VG)]$

L'analyse des rapports internes dans le cadre de la classe des Vmt est confirmée par l'organisation sémantique de la construction  $[VR_{Vmt} VG_{Vmt}]$  (et du Type B'-manière en général); en effet, le VR y est toujours hyperonyme du VG (cf. les exemples suivants):

- [5'] *dit-il en balbutiant* (dire → balbutier)  
 [6'] *[il] marchait en boitant* (marcher → boiter)  
 [9'] *[Alice] est arrivée en courant* (arriver → courir)

Dans les trois exemples, les rapports d'hyponymie-hyperonymie existant entre le VR et le VG causent l'irréversibilité des deux verbes, contrairement au Type B (*circonstance concomitante*), où le VR pourrait parfois être remplacé par le VG et vice versa, cf. :

- [10] Il était un peu plus de cinq heures du matin, et **nous buvions** des capuccinos sous l'auvent de bois d'une échoppe d'artisan, **en regardant** la neige tomber devant nous dans la ruelle.<sup>15</sup>  
 [INTERCORP] (Il serait également possible de dire *nous regardions... en buvant...*)<sup>16</sup>

Néanmoins, comme nous l'avons déjà signalé (section 2.2.1), du point de vue sémantique (et par conséquent aussi contrastif), l'exemple [9'] présente certaines spécificités: les sèmes du VR (*arriver*) n'étant pas entièrement inclus dans l'ensemble des sèmes du gérondif (*courir*), l'omission du VR (*balbutia-t-il, il boitait*) est dans ce cas impossible, parce qu'elle entraînerait un changement important

<sup>15</sup> Traduction tchèque: «Bylo něco po páté hodině ráno a **my jsme pili** kapucino pod dřevěnou stříškou stánku nějakého řemeslníka, **dívali se**, jak před námi v uličce padá sníh.» [INTERCORP]

<sup>16</sup> La substitution VR ↔ VG dans le Type B' n'est pourtant pas complètement exclue; seulement elle changerait le sens de *manière* en *circonstance concomitante* et/ou *repère temporel*, cf. par exemple *Balbutia-t-il en disant* ou *Il boitait en marchant*.

du sens de la construction (*\*elle a couru*)<sup>17</sup>. (La non-inclusion des sèmes du VR dans le SG peut être représentée par le schéma VR  $\not\subset$  SG.) En conséquence, les équivalents tchèques omettant le VR français sont aussi exclus (cf. *Zamumlal, Kulhal* vs. *\*Běžel*).

En outre, étant donné que les deux verbes du Type B' (Manière) désignent ensemble une seule action, les équivalents décomposant la construction française en deux verbes tchèques sont dans la plupart des cas également impossibles *\*Přišel a běhal přitom* (*\*il est arrivé et il a couru*); curieusement, le tchèque dans ce cas n'admet pas non plus le transgressif (*\*Přišel běhaje*), bien qu'il s'agisse dans d'autres cas de l'équivalent structurel du gérondif. Finalement, le seul équivalent tchèque admissible pour la construction [VR<sub>vmt</sub> VG<sub>vmt</sub> (VR  $\not\subset$  VG)] *il est arrivé en courant* est le verbe préfixé *Přiběhl*)<sup>18</sup>.

En ce qui concerne l'analyse sémique de cet équivalent, le préfixe tchèque (le préverbe) correspond par sa signification au verbe régissant français, tandis que la base du verbe fini tchèque transmet le sens du gérondif:

<i>il est arrivé en courant</i> →	<i>Při</i>	–	<i>běhl.</i>	
(přijít)	(běhat)	arriver		en courant

Le verbe régissant français exprime ainsi la finalité du mouvement – *arriver, entrer, rentrer, s'approcher, venir, retourner, sortir, partir, s'éloigner, monter, descendre, traverser*, mais aussi par exemple *s'en aller*, etc., tandis que le gérondif précise la manière de sa réalisation – par exemple la posture de l'agent (*ramper, boiter, galoper, chevaucher, glisser, tituber, trotter*, etc.) ou le milieu du déroulement du mouvement – l'eau (*nager*), l'air (*voler*) ou le sol (*courir, marcher*, etc.).

Le verbe régissant (VR) est donc moins riche en information, moins spécifique que le verbe du gérondif. C'est grâce à ce sémantisme restreint que le verbe régissant peut être traduit en tchèque par un préfixe, qui est dans la plupart des cas capable de transmettre seulement un sème, par exemple celui du mouvement vers l'extérieur (VY): *il est sorti en courant* – *vyběhl*)<sup>19</sup>.

17 Naturellement, pour réduire la construction [VR<sub>vmt</sub> VG<sub>vmt</sub>], le français peut dans ce cas particulier employer également un verbe préfixé (*accourir*), mais en tchèque ce procédé est plus fréquent, et parfois même obligatoire. En ce qui concerne le gérondif, il peut être parfois remplacé par un SP (*à la marche, à la nage...*).

18 Dans d'autres langues, en anglais notamment, nous pouvons trouver des procédés de simplification analogues, par exemple *Il est sorti en courant* – *He ran out*.

19 Dans ce contexte, rappelons une remarque pertinente faite par le romaniste tchèque Jan Šabršula dans sa *Vědecká mluvnice francouzštiny* (*Grammaire scientifique du français*, Šabršula 1986: 244): en comparant le français et le tchèque, Šabršula signale que le français peut profiter du caractère abstrait, général du Vmt régissant (par exemple *entrer*) et laisser la manière de la réalisation du mouvement implicite. Ainsi, en français, il est possible de dire que *l'homme, le serpent, le train, le paquebot, le hanneton est entré* (*dans la pièce, en gare, etc.*), tandis que le tchèque doit obligatoirement préciser la manière de réalisation du mouvement, puisque le préfixe (dans l'occurrence *V-/VE-*) ne peut pas exister indépendamment du verbe: *člověk/muž vešel* (= *l'homme est entré en marchant*), *had se vplazil* (= *le serpent est*

Le tableau ci-dessous dresse une liste (non-exhaustive) des sèmes de *télicité* de mouvement traduisibles par les préverbes tchèques (sur la base du verbe *courir*):

<u>VR<sub>vmt</sub> + VG<sub>vmt</sub></u>	<u>préverbe</u>
arriver <i>en courant</i>	PŘI- ( <i>přiběhl</i> )
partir	OD- ( <i>odběhl</i> )
sortir	VY- ( <i>vyběhl</i> )
entrer	V/VE- ( <i>vběhl</i> )
traverser	PŘE- ( <i>přeběhl</i> )
monter	VY- ( <i>vyběhl</i> )
descendre	S/SE- ( <i>seběhl</i> )
...	...

Néanmoins, la préfixation comme équivalent tchèque de la construction [VR<sub>vmt</sub> VG<sub>vmt</sub> (VR  $\not\subset$  VG)] peut être exclue, par exemple dans les cas où soit les verbes sont tirés des lignées synonymiques complètement différentes, soit le VR n'exprime pas la finalité (la *télicité*) du mouvement, et n'est donc pas traduisible en tchèque par un préverbe<sup>20</sup> (par exemple *danser en tournant sur lui-même*)<sup>21</sup>.

### 4.3. Comportement syntaxique de [VR<sub>vmt</sub> VG<sub>vmt</sub> (VR $\not\subset$ VG)]

En ce qui concerne les propriétés syntaxiques des SG entrant dans les [VR<sub>vmt</sub> VG<sub>vmt</sub> (VR  $\not\subset$  VG)], mentionnons au moins leur position vis-à-vis du VR et le rôle des expansions éventuelles.

Comme nous l'avons déjà signalé (2.2.1), la position typique du SG exprimant la manière est après le VR (*il est arrivé en courant*); en antéposition (position extra-prédicative), le SG<sub>vmt</sub> perdrait le sens de *manière* et désignerait plutôt un *repère temporel* (signification gérondivale typiquement thématique). Observons par exemple la phrase suivante :

[11] **En repartant**, je titubais, résolument vide. → **Cestou zpět** jsem se potácel a cítíl se dokonale vyprázdněný. [INTERCORP]

En outre, l'ancrage du SG dans le contexte précédent mène parfois à la non-coréférence des agents du SG antéposé et du VR (phénomène exclu dans le cas du SG-manière, où les deux verbes désignent ensemble une seule action) :

[12] **En repartant**<sub>[i]</sub>, la question qui le<sub>[i]</sub> taraudait depuis toujours était balayée par une autre qui s'imposait sans mal : qui surveillera les surveillants? → **Když odcházel**, smetl tuhle otázku, kte-

*entré en rampant*), vlak *vjel* (= le train est entré *en allant*<sub>[moyen de transport]</sub>), parník *vplul* (= le paquebot est entré *en flottant*), chroust *vlétl* (= le hanneton est entré *en volant*), etc.

<sup>20</sup> Évidemment, la première condition est que les deux verbes (VR et VG) soient des verbes de mouvement; cf. par contre « Et si c'était elle que je cherchais **en courant** après toutes ces filles? » → « A co když to byla ona, kterou jsem hledal, když jsem běhal za všemi těmi holkami? » [INTERCORP]

<sup>21</sup> Exemple repris de Halmøy (2003 : 105).

rá v něm odjakživa hlodala, jiný bezprostředně se nabízející problém: kdo bude hlídat hlídače?  
[INTERCORP]<sup>22</sup>

Le rapport étroit entre le VR et le SG-manière conditionne un autre trait typique de la construction [VR<sub>Vmt</sub> VG<sub>Vmt</sub> (VR  $\not\subset$  VG)], à savoir la construction liée (non-détachée), par exemple :

[13] R437/ DROIT.M / LE RETOUR / 1964 / page 60 / III  
Une nageuse arrivait **en crawlant**. [FRANTEXT]

Le SG dans la construction [VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR  $\not\subset$  VG)] est donc en prédominance intraprédicatif, par conséquent non-déplaçable et lié, et le répertoire des verbes entrant dans cette construction est limité, ce qui pourrait indiquer qu'il s'agit d'une forme figée (type *arriver en courant*). Cette interprétation pourrait être soutenue premièrement par le fait que la plupart des SG des [VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR  $\not\subset$  VG)] ne contiennent pas d'expansions, deuxièmement par le fait que les deux verbes désignent ensemble une seule action et peuvent ainsi être conceptualisables comme une seule unité (ce à quoi correspondraient également les équivalents tchèques, formés par des lexèmes uniques autonomes – par exemple *přiběhnout*).

Pour illustrer cette hypothèse, prenons le cas prototypique *arriver en courant* : dans le corpus FRANTEXT (version catégorisée, corpus de travail de 77 406 645 de mots, 1245 textes de tous les genres parus après 1900), parmi les 127 occurrences du verbe *arriver* suivi de gérondif<sup>23</sup>, 100 appartiennent au Type B ou B'<sup>24</sup> : 47 occurrences de *circonstance concomitante* (Type B) et 53 occurrences du Type B' (*manière*)<sup>25</sup>, c'est-à-dire de la construction [VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR  $\not\subset$  VG)]. Parmi

<sup>22</sup> Des cas limites peuvent apparaître également dans les constructions postposées, cf. par exemple le SG suivant faisant partie d'une proposition infinitive : J'avais refermé la porte derrière moi, et je regardais Marie<sub>[i]</sub> *avancer*<sub>[i]</sub> dans la chambre **en titubant**<sub>[i]</sub> de fatigue, son manteau de cuir noir et son pull sur un bras, son chemisier blanc qui sortait de son pantalon [...] → Zavřel jsem za sebou a díval se na Marii, jak *jde* pokojem, **potácí se** únavou, černý kožený kabát a svetr přes ruku, halenku vylézající z kalhot [...] [INTERCORP].

<sup>23</sup> Le nombre total des constructions *arriver en courant* présentes dans le FRANTEXT est néanmoins plus élevé, vu que pour cette recherche spécifique, nous avons seulement pris en compte les occurrences où le gérondif *en courant* suit immédiatement le verbe *arriver*.

<sup>24</sup> Dans les SG ayant le sens de repère temporel et/ou d'antériorité logique, le verbe *arriver* n'est pas dans la plupart des cas un Vmt, mais a une autre signification, principalement *réussir à, parvenir à*, par exemple P422/MORAND.PIERRE / CONFINS VIE BIOLOGIE DES VIRUS / 1955 / page 33 / «On y *arrive en remplaçant* le champ de gravité terrestre par un autre à qui l'on peut donner des valeurs beaucoup plus élevées : un champ de gravitation.» Parmi les SG exprimant la *circonstance concomitante* (Type B), citons par exemple K627/ GIDE.A / JOURNAL 1889–1939 / 1939 / page 1238 / 1935 «J'*arrive en lançant* mes souliers.»

<sup>25</sup> Par exemple S361/ RAMBAUD.P / LA BATAILLE / 1997 / page 196 / *CHAPITRE V, Seconde journée* Le maréchal *arriva en galopant* et vint trotter à côté de Saint-Hilaire ;

ces 53 constructions, nous avons recensé 36 (soit 68 %) dans lesquelles le SG est exprimé par le verbe *courir*, par exemple :

[14] S330/ CLAVEL.B / CELUI QUI VOULAIT VOIR LA MER / 1963 / page 234  
Elle allait parler lorsque M. Piolat *arriva en courant*.

Néanmoins, malgré toutes les spécificités sémantiques et syntaxiques et malgré certaines tendances statistiques, ni la construction *arriver en courant*, ni la [VR<sub>vmt</sub> SG<sub>vmt</sub> (VR  $\not\subset$  VG)] en général ne peuvent être considérées comme entièrement figées; premièrement, parce que les deux verbes (VR ainsi que le VG) peuvent le cas échéant prendre des expansions, citons par exemple :

[15] S039/ DURAS.M / MODERATO CANTABILE / 1958 / page 20  
Des gens arrivaient **en courant** *de toutes les rues avoisinantes* et s'aggloméraient à lui.<sup>26</sup>

d'autre part parce que le SG peut être coordonné avec un autre élément, par exemple :

[16] L487/ MAURIAC.F / ASMODEE / 1938 page 54 / *ACTE I SCÈNE 7*  
Ils arrivent **en courant** *et en criant* : à table !<sup>27</sup>

De plus, bien que très rares, l'antéposition et le détachement du SG<sub>vmt</sub> sont néanmoins possibles (tout en gardant le sens de *manière*), par exemple :

[17] S329 SCHREIBER Boris, *Un silence d'environ une demi-heure*, 1996, p. 332  
Après un court silence, la main de la femme glissa à nouveau vers la ceinture [...], lorsque Philippe *entra*, **en courant**, en sueur : - J'ai bien joué, maman ! [détachement]

et

[18] R971 FORLANI Remo, *Gouttière*, 1989, p. 168  
Un saxophoniste américain. Il me voulait, je le voulais. **En courant**, *il a monté les marches de l'escalier*. Mais... quand il vu les crapauds dans le bidet... Et mon architecte, il m'adorait celui-là aussi... [antéposition]

La position extrapredicative du SG (antéposition et/ou détachement) entraîne naturellement certains glissements de sens (mise en relief, nuance de circonstance concomitante causée par la décomposition de la [VR<sub>vmt</sub> SG<sub>vmt</sub> (VR  $\not\subset$  VG)])

<sup>26</sup> Parmi les occurrences analysées, nous n'en avons trouvé aucune où les expansions se rattacheraient au SG *en courant*; cependant, dans d'autres constructions du même type, ce type d'expansion est attesté, par exemple : CAMUS Albert, *L'Étranger*, 1942, p. 1160-1161, *PREMIÈRE PARTIE CHAP. VI* Alors j'ai laissé \*Marie et je suis rentré **en nageant régulièrement** *et en respirant bien*.

<sup>27</sup> D'autant plus que le répertoire des préfixes et/ou des bases verbales présente parfois des lacunes, empêchant la traduction de la [VR<sub>vmt</sub> SG<sub>vmt</sub> (VR  $\not\subset$  VG)] par un seul verbe tchèque.



en deux actions, etc.), mais le sens de *manière* (Type B') est toutefois préservé dans les deux exemples précités.

Enfin, même les équivalents tchèques parfois indiquent que la décomposition sémique de la construction  $[VR_{vmt} SG_{vmt} (VR \not\subset VG)]$  est possible, surtout quand le SG contient des compléments (v. 4.4).

De plus, comme le montrent d'autres exemples tirés des corpus parallèles, la réalité des équivalents tchèques est souvent plus compliquée que celle indiquée par le schéma

<i>il est arrivé en courant</i> → (přijít) (běhat)	Při –	<i>běhl.</i> en courant
	arriver	

#### 4.4. *Equivalents non-prototypiques de $[VR_{vmt} VG_{vmt} (VR \not\subset VG)]$*

Ce sont principalement les compléments (du VR ainsi que du GV) qui entrent en jeu, par exemple :

[19fr] <s lang="fr" id="1908">Obligé par la roideur de la pente à mettre pied à terre, Orso, qui avait laissé la bride sur le cou de son cheval, **descendait**<sub>1</sub> rapidement<sub>2</sub> **en glissant**<sub>3</sub> sur la cendre<sub>4</sub>; </s> [COLOMBA]

En premier lieu, contrairement à *arriver en courant*, la **décomposition** de  $[VR_{vmt} SG_{vmt} (VR \not\subset VG)]$  en **VR et le transgressif** est acceptable, et nous pouvons également observer le **dédoublé de l'indication de la direction** – *sestupoval* – [*descendait*] + *dolů* – [*en bas*], et le détachement du transgressif *klouzaje* du VR *sestupoval* :

[19cz-a] <s lang="cz" transl="CER" id="1908">Příkrý svah přinutil Orsa, aby seskočil s koně; dal koni volnost a sám rychle<sub>2</sub> **sestupoval**<sub>1</sub> **dolů**<sub>1</sub>, **klouzaje**<sub>3</sub> po popelu<sub>4</sub>. </s>

Ce type d'équivalent de la  $[VR_{vmt} SG_{vmt} (VR \not\subset VG)]$ , à savoir  $[VR, Transgr]$  est attesté dans trois autres traductions de la même phrase du corpus COLOMBA.

De plus, les deux actions (du VR et du SG) peuvent être mises au même niveau par la coordination :

[19cz-b] <s lang="cz" transl="HOL" id="1908">Svah byl tak příkrý, že Orso byl nucen seskočit na zem; nechal uzdu na šiji koně, **klouzal**<sub>3</sub> v popelu<sub>4</sub> a rychle<sub>2</sub> **sestupoval**<sub>1</sub>. </s>

Cependant, dans d'autres cas, où le VR est plus abstrait, la décomposition de la  $[VR_{vmt} SG_{vmt} (VR \not\subset VG)]$  en deux verbes en tchèque ne représente pas une traduction réussie, puisque les deux verbes expriment de façon redondante non la *télicité* du mouvement, portée en général seulement par le préverbe, mais la manière de sa réalisation, exprimée par la base du verbe fini en tchèque, cf. l'exemple suivant :

[20fr] <s lang="fr" id="1114">On eût dit qu'ils **s'étaient avancés**<sub>1</sub> **en rampant**<sub>2</sub>, comme des couleuvres au milieu du fourré de cistes et de myrtes qui couvrait le terrain. </s> [COLOMBA]

[20cz-a] <s lang="cz" transl="BAR" id="1114">Mohlo se říci, že **se připlížili**<sub>1</sub> **plazice**<sub>2</sub> se [*ramper*<sub>[hum]</sub> en *rampant*<sub>[animal]</sub>] jako užovky uprostřed houštin cistových růží a myrt, které zarůstaly krajinu.</s>

Étant donné le sens extrêmement abstrait du verbe *avancer*, un seul verbe préfixé (*připlazit se*) semble être la meilleure solution (comme dans le cas prototypique *arriver en courant – přiběhnout*).

[20cz-b] <s lang="cz" transl="CER" id="1114">Vypadalo to, že **se připlazili** jako užovky houštím cistů a myrt, kterých tam bylo všude plno.</s>

Pareillement, dans l'exemple suivant, tiré d'INTERCORP, la décomposition de [VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR ∄ VG)] en deux verbes tchèques n'est pas possible :

[21fr] Le dîner avait lieu à la Rosa Nautica, un restaurant sur pilotis, à Miraflores - on *entre*<sub>1</sub> dans la mer<sub>2</sub> **en marchant**<sub>3</sub> sur des pontons de bois<sub>4</sub>, on pourrait être à Brighton ou à Biarritz dans les années vingt, [...]. [INTERCORP]

[21cz] Večeře se konala v Rosa Nautica, restauraci zbudované na pilotech, v Miraflores - **vcházíte**<sub>1+3</sub> do moře<sub>2</sub> po dřevěném můstku<sub>4</sub>, mohli byste být v Brightonu nebo v Biaritzu dvacátých let, [...].

Le dédoublement du sème de la *télicité* du mouvement (*sestupoval dolů – il descendait en bas*), attesté dans la phrase [19cz-a], est assez fréquent dans les traductions tchèques, cf. l'exemple suivant du corpus INTERCORP :

[22fr] Nous pressâmes le pas pour descendre les quelques mètres du chemin privé de l'hôtel, et **traversâmes**<sub>1</sub> la rue<sub>2</sub> **en courant**<sub>3</sub> la main dans la main<sub>4</sub>, [...]. [INTERCORP]

[22cz] Přidali jsme do kroku, abychom překonali těch několik metrů vyhrazené cesty k hotelu, chytili se za ruce<sub>4</sub> a **přeběhli**<sub>1+3</sub> přes<sub>1</sub> ulici<sub>2</sub>, [...].

La construction [VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR ∄ VG)] *traverser en courant* trouve ici son équivalent prototypique (à savoir le verbe préfixé *přeběhnout*), mais le complément d'objet direct du verbe *traverser* (*la rue*) devient en tchèque un complément circonstanciel (*přes ulici – à travers la rue*), le sème de direction étant ainsi rendu deux fois, comme dans l'exemple [19cz-a].

Néanmoins, dans la plupart des cas, les deux variantes sont possibles – sans le dédoublement [23cz-a] ou avec lui [23cz-b] :

[23fr] <s lang="fr" id="563">La femme Pietri, ayant laissé son fardeau, **monta**<sub>1</sub> le sentier<sub>2</sub> **en courant**<sub>3</sub>, et trouva le colonel della Rebbia baigné dans son sang, percé de deux coups de feu, mais respirant encore.</s>[COLOMBA]

[23cz-a] <s lang="cz" transl="CER" id="563">Pietriová položila své břemeno, **vyběhla**<sub>1+3</sub> **po stezce**<sub>2</sub> a našla plukovníka della Rebbia zbroceného krví; měl dvě rány v těle, ale ještě dýchal.</s> [sans dédoublement]

[23cz-b] <s lang="cz" transl="HOL" id="563">Když Pietriová zanechala své břemeno na místě a **vyběhla**<sub>1+3</sub> **po stezce**<sub>2</sub> **nahoru**<sub>1</sub>, našla plukovníka della Rebbiu zbroceného krví, s dvěma průstřely, avšak ještě dýchal.</s> [avec dédoublement]

À part le dédoublement (*vyběhla nahoru* – *elle est montée en haut*), nous pouvons de nouveau observer également la transposition du COD français en complément circonstanciel tchèque (*le sentier* – *po stezce*).

Enfin, le VG français peut aussi devenir un verbe fini en tchèque; le VR français étant rendu par un simple adverbe (**VG > VF; VR > Adv**):

[19cz-c] <s lang="cz" transl="CIB" id="1908">Sráz byl tak příkrý, že Orso musil seskočit s koně; nechal koni uzdu na krku a **klouzal**<sub>3</sub> rychle **dolů**<sub>1</sub> po popelů.</s>  
[COLOMBA]<sup>28</sup>

Ce dernier procédé ressemble à l'omission du VR du type *il marchait en boitant* – *il boitait*, dont nous avons parlé dans la section 2.2.1: les sèmes de mouvement sont portés par le verbe plus spécifique – *boitait*, *glissait*, seulement le sème de *télicité* du mouvement, non-inclu dans le verbe *glisser*, doit être préservé par d'autres moyens (dans l'occurrence par l'adverbe *dolů* – *en bas*)<sup>29</sup>.

## 5. Conclusion

Les corpus parallèles (section 1) représentent une riche source de données pour la recherche (non seulement) contrastive; ils nous ont permis d'esquisser le début d'analyse d'un groupe spécifique de constructions gérondives exprimant la manière (type B' selon O. Halmøy, type *dit-il en balbutiant*, section 2.2.1), du type *il est arrivé en courant* ([VR<sub>Vmt</sub> SG<sub>Vmt</sub> (VR  $\not\subset$  VG)], section 3). Les recherches sur corpus (bilingues ainsi que unilingues) montrent que ce type spécifique de construction gérondive partage un grand nombre de traits du type B' identifié par Halmøy: il est en majorité intra-prédicatif et par conséquent postposé au VR, non détaché, non-déplaçable, irréversible et coréférentiel, et ensemble avec le VR, il désigne une seule action (sections 3.2.3). Néanmoins, contrairement à la plupart des constructions du type B', cette construction n'est pas réductible à un seul verbe (cf. *balbutia-t-il* vs. *il a couru*), étant donné que certains sèmes du VR ne sont pas inclus dans le SG, à savoir les sèmes de *télicité* du mouvement. Cette analyse est confirmée par les équivalents tchèques (*il est arrivé en courant* → *přiběhl*), où le préverbe exprime la direction (*télicité*) du mouvement (*arriver* – *při-*) et la base du verbe transmet le sens du VG français, c'est-à-dire la *manière* du mouvement (*courir* – *-běh-*).

<sup>28</sup> Le même procédé (VG > VF; VR > Adv) est observable également dans les constructions qui ne contiennent pas uniquement les Vmt, cf. par exemple: [...] je **montais**<sub>1</sub> les marches<sub>2</sub> **en traînant**<sub>3</sub> ma valise<sub>4</sub>, passais devant la porte de notre ancien appartement, maintenant loué [...] - [...] **vláčela**<sub>3</sub> **jsem** kufř<sub>4</sub> po schodech<sub>2</sub> **nahoru**<sub>1</sub>, prošla kolem dveří našeho bývalého bytu, nyní pronajatého [...] [INTERCORP].

<sup>29</sup> Dans d'autres traductions, moins fidèles, le sème porté par le SG peut se perdre complètement, cf. par exemple Des pompiers *arrivés*<sub>1[à pied]</sub> **en courant**<sub>2</sub> interrompirent notre échange muet. - Náš němý rozhovor přerušil teprve **příjezd**<sub>1[moyen de transport]</sub> hasičů.

Bien qu'il ne soit pas possible d'appliquer schématiquement les résultats de l'analyse des équivalents tchèques sur le système français (comme nous le montrent les cas non-prototypiques de la section 3.4), cette recherche contrastive basée sur les corpus parallèles ouvre plusieurs pistes de recherche intéressantes: premièrement, il faut dresser la liste des Vmt entrant dans la construction  $[VR_{Vmt} SG_{Vmt} (VR \not\subset VG)]$  et identifier leurs rapports sémantiques et combinatoires (hyponymie/hyponymie, cf. 3.1); deuxièmement, élargir la recherche sur d'autres langues, par exemple l'anglais (*the dog trotted away – le chien est parti en trot-tant*); et troisièmement, à partir de cette base de données, il serait intéressant d'analyser si les différences d'organisation des sèmes de mouvement entraînent des différences de conceptualisation de l'action chez les locuteurs des différentes langues.

### Liste des abréviations

SG	<i>syntagme gérondif (gérondif + ses expansions)</i>
VF	<i>verbe fini</i>
VG	<i>verbe en forme de gérondif</i>
Vmt	<i>verbe de mouvement</i>
VR	<i>verbe régissant</i>

### Bibliographie

- ČERMÁK, František, *et al. Jak využívat Český národní korpus*. Praha: NLN, 2005.
- KEKENBOSCH, Christiane; MEUNIER, Jean-Marc; RICHARD, Jean-François; DESCLÈS, Jean-Pierre; FLAGEUL, Valérie. Sémantique cognitive de l'action: 2. étude expérimentale de la catégorisation des verbes d'action. *Langages*, 1998, n° 132, pp. 48–68.
- GROSS, Maurice. *Méthodes en syntaxe*. Paris: Hermann, 1975.
- HALMØY, Jane-Odile. *Le Gérondif. Éléments pour une description syntaxique et sémantique*. Tapir: Université de Trondheim, 1982.
- HALMØY, Jane-Odile. *Le Gérondif*. Paris: Ophrys, 2003.
- KOCEK, Jan; KOPŘIVOVÁ, Marie; KUČERA, Karel (eds.). *Český národní korpus – Úvod a příručka uživatele*. Praha: FF UK, ÚČNK, 2000.
- LE PESANT, Denis. Introduction aux classes d'objets. *Langages*, 1998, n°131, pp. 6–34.
- SVÁŠEK, Martin. Définition, élaboration et exploitation d'un corpus parallèle bidirectionnel français – tchèque tchèque-français. Thèse de doctorat. Prague-Paris: FF UK ÚTKL, INALCO (LALIC – CERTAL), 2007.
- ŠABRŠULA, Jan. *Vědecká mluvnice francouzštiny*. Praha: Academia, 1986.
- VÉRONIS, Jean (ed.). *Parallel text processing, alignment und use of translation corpora*. Dordrecht: Kluwer, 2000.

## Les corpus utilisés

FRANTEXT [www.atilf.fr](http://www.atilf.fr)

INTERCORP <http://www.korpus.cz/intercorp/>

COLOMBA

### Abstract and key words

This article presents a new instrument of the linguistic research (parallel corpora, section 1, particularly the large project Intercorp) and an example of their application (the French gerund constructions of the verbs of movement). After a short summary of the gerund semantics, based on the works of O. Halmøy (Type A – (chrono)logical anteriority, 2.1 vs. Type B – *circonstance concomitante*, 2.2), the article focuses on the sub-type B' of the gerund, expressing the way (*la manière*) of the realization of the main process (type *dit-il en balbutiant*) and including all the semes of the main verb (*dire – balbutier*). Due to the contrastive approach, a special type of B' is identified ([VR-<sub>vmt</sub> SG<sub>vmt</sub> (VR  $\not\subset$  VG)], *il est arrivé en courant*, section 3), where the main verb and the gerund are both verbs of movement, but some semes of the main verb, expressing the *telicity* of the movement, are not included in the gerund and are usually translated in Czech by a prefix (e.g. *Přiběhl*).

Corpus linguistics; contrastive linguistics; parallel corpora; French; Czech; gerunds; verbs of movement