

LES ANGLICISMES EN FRANÇAIS ET EN TCHÈQUE CONTEMPORAINS : LE CAS DES FORMES EN *-ING*

EMMANUEL CARTIER, JAN LAZAR

ABSTRACT

Anglicisms in contemporary French and Czech: a case study on *-ing* lexical formations

The aim of this study is to analyze the neological borrowings from English that occur in contemporary French and Czech. Special emphasis is placed on neological borrowings with the *-ing* suffix. Owing to the status of Anglo-American as a *lingua franca*, Anglicisms affect all languages. This study investigates more specifically the processes involved in these *-ing* constructions. After a quantitative analysis of the respective diffusion and productivity in both languages, we focus on a few specific words (*phishing*, *géocaching*, *sharing*, *washing*), attested in both languages, and compare their integration from several perspectives (phonological, orthographical, morphological, semantic and local variants).

Keywords : Anglicism, neology, French, Czech, *-ing* construction

Mots-clés : Anglicisme, néologie, français, tchèque, formes en *-ing*

Introduction

La présente étude s'intéresse aux néologismes formés en *-ing* en français et en tchèque contemporains. En français, ce formant d'emploi suffixal est emprunté à l'anglais depuis plus d'un siècle, pour exprimer « une action, son résultat ou le lieu où se déroule l'action » (Dubois 1962 : 14) mais semble, depuis les années 2000 (généralisation d'internet et apparition des réseaux sociaux) connaître un regain d'emploi, à la fois en français (Mudrochová 2017, Cartier 2018a) et en tchèque. L'anglais, ou plutôt le *Global English* (Crystal 2007), est depuis la fin de la seconde guerre mondiale, la langue de communication internationale, la *lingua franca*, mais également et de plus en plus, en tant que langue seconde la plus parlée au monde, qui trouve sur internet son champ principal d'expression, une langue d'influence qui impacte les pratiques linguistiques individuelles et sociales.

La présente étude vise à effectuer une étude outillée et une analyse des emprunts contemporains à suffixe *-ing* (et sa variante *-ink* en tchèque) à partir de corpus dynamiques et de relevés manuels, puis à effectuer une comparaison des similitudes et des dif-

férences entre les deux langues, de plusieurs points de vue : lexies concernées, diffusions respectives, existence ou non de termes équivalents « autochtones ».

Nous utiliserons comme corpus les JSI Timestamped web Corpus 2014–2020 (Trampus *et al.* 2004), disponibles en français et en tchèque, d'une taille comparable et similaire du point de vue des sources d'informations, des sites de presse en ligne. Dans ces corpus, en dehors d'une analyse morphosyntaxique automatique, chaque contexte est doté de méta-informations que nous pourrions exploiter (date, nom du journal, pays notamment). Ces corpus sont disponibles au travers de la plateforme SketchEngine (Kilgariff *et al.*, 2014) qui fournit également un certain nombre d'outils pour l'exploitation des données linguistiques. A partir de ces corpus, nous souhaitons :

- collecter un nombre conséquent de créations lexicales en *-ing* (et *-ink* pour le tchèque) apparues récemment ou encore en phase d'émergence, dans la presse quotidienne généraliste des deux pays ;
- en effectuer une description linguistique selon une grille d'analyse qui sera explicitée ;
- en suivre la diffusion à l'aide d'outils fournis par le projet Néoveille (cf. Cartier, 2019), (et d'autres outils, le cas échéant) ;
- suivre également les termes équivalents « autochtones » éventuels, les mots dérivés, composés et transformés construits à partir des emprunts et leur implantation respective;
- effectuer des études comparatives pour un certain nombre d'emprunts pris dans des domaines spécifiques.

Le travail comprendra deux grandes parties : la première détaillera la méthodologie suivie, la seconde les résultats obtenus et leur analyse.

1. Définitions préliminaires

Dans le cadre de ce travail, nous parlerons d'emprunt lorsque le lexème identifié est directement importé dans sa totalité, dans sa forme comme dans son sens, d'une autre langue, en l'occurrence l'anglais international¹. Précisons que l'anglais international désigne la langue anglaise dans sa variante globale pratiquée par de nombreux locuteurs, non nécessairement anglophones. Nous parlerons de faux-emprunt lorsque le lexème est non-attesté dans la langue d'origine dans le même sens, mais concomitamment lorsque des caractéristiques linguistiques sont elles-mêmes empruntées. Par exemple, *parking* est un faux-emprunt puisque cette lexie n'est pas attestée en anglais dans le sens français. Néanmoins, on peut s'apercevoir de la construction *-ing*, ainsi que de la base *park-*. De même, le lexème *féminisme-washing* est un faux emprunt, puisqu'il n'est attesté qu'en français, mais *washing* est emprunté, et surtout le patron productif *Nom ou Adj + washing* (dans le sens : opération effectuée par une organisation généralement commerciale qui cherche à se donner une bonne image en effectuant des communications, opérations commerciales en lien avec une cause éthique).

Nous parlerons de néologisme emprunté ou de néologisme en général, lorsque le lexème n'a aucune attestation (dans le sens repéré) avant 2010, en basant ce jugement

¹ Nous reprenons à notre compte la notion de *Global English* (Crystal, 2012).

sur la base d'une vérification sur des corpus représentatifs et éventuellement des dictionnaires de référence. Ce critère est partiellement arbitraire, car il est souvent difficile, voire impossible d'identifier la première attestation d'un néologisme. D'autres critères complémentaires permettent également de qualifier un lexème de néologisme : sa fréquence constatée, l'existence d'emplois en mention, la présence de gloses à proximité (voir Cartier 2018b pour plus de détail). La qualification de néologisme s'éteint lorsque le lexème est adopté par la communauté linguistique, ce qui pose également des difficultés dans l'identification du moment d'adoption.

Les emprunts méritent par ailleurs qu'on détaille certaines caractéristiques de leur cycle de vie, étant donné qu'ils proviennent d'un autre système linguistique. Notamment, la phase initiale d'entrée dans le système linguistique est marquée par une adaptation éventuelle au système phonologique, orthographique et morphologique de la langue d'arrivée (Dubéda 2020 : 328). Pour ce qui concerne le tchèque, la prononciation d'un mot va de pair avec sa forme orthographique. L'orthographe de la langue tchèque est transparente et les graphèmes correspondent directement aux phonèmes. Par contre, en anglais on trouve rarement des mots dont l'écriture reproduit de façon univoque la prononciation. Certes, il y a des exceptions comme *film* ou *fit*, mais il s'agit de cas plutôt rares. Il en résulte que les locuteurs tchèques essaient d'harmoniser la forme graphique du mot avec sa prononciation. A titre d'exemple, *trénink* (*training*) s'est parfaitement intégré dans la langue tchèque et l'usage de la terminaison *-ing* pourrait plutôt surprendre un locuteur natif tchèque. Les facteurs décisifs pour la tchéquisation semblent la fréquence d'usage ainsi que la durée de la circulation parmi les locuteurs natifs (Rejzek 1993). Beaucoup de mots sont déjà parfaitement adaptés, par exemple *mítink* (*meeting*), *dispečink* (*dispatching*) ou *pudink* (*pudding*). Néanmoins, on observe aussi des emprunts sous deux formes concurrentes, p.ex. *engineering* – partiellement tchéquisé *inženýring* ou totalement tchéquisé – *inženýrink*. *Briefing* est également attesté dans la presse tchèque sous deux formes : *brífing* et *brífink*. En revanche, la terminaison *-ing* est souvent conservée notamment pour certains termes spécifiques (p.ex. *phishing*, *mobbing*, *grooming*) et des expressions populaires liées à la culture anglo-américaine (p.ex. *jumping*, *dumping*) (Svobodová, en ligne). Ajoutons que les mots tchèques se déclinent, ce qui provoque la transformation de la finale *-k* en *-g*, pour des raisons phonétiques (p.ex. *mítink* – *o mítingu*, *brífik* – *o brífingu*).

En français, la modification orthographique d'un anglicisme se terminant en *-ing* est rare et les différentes commissions essaient de trouver plutôt un équivalent français qui pourrait remplacer son homologue anglais dans l'usage quotidien (p.ex. *briefing* – *réunion préparatoire*, *doping* – *dopage*, *streaming* – *diffusion en afflux*). Ce qui semble évoluer dans le temps, c'est la prononciation des mots en *-ing*. On observe que la terminaison *-ing* prononcée [ɛ̃] est en recul en faveur de la réalisation vélaire [ŋ], ce qui rapproche le mot de la prononciation originale. Comme le confirme l'analyse de Mudrochová (2017 : 11) la prononciation [ɛ̃] est attestée dans Le *Petit Robert* seulement pour *shampo(o)ing* et *pouding*.

2. Méthodologie

Dans cette section, nous détaillons la méthodologie utilisée dans cet article pour identifier, valider puis décrire les innovations lexicales en *-ing/-ink*.

2.1 Collecte et validation des emprunts à base *-ing*

La collecte des lexies s'est déroulée en trois étapes :

Étape 1 : Récupération automatique, dans le corpus JSI Timestamped web corpus 2014–2020, dans chacune des deux langues, de l'ensemble des lexies se terminant par le suffixe *-ing/-ink* ;

Étape 2 : Validation manuelle des lexies détectées automatiquement et choix des lexies à étudier plus précisément. Pour ce qui concerne la première opération, deux cas d'extraction erronée peuvent se présenter : forme lexicale à l'évidence non liée au formant *-ing* (exemple : *ring*, *seing*, etc.), erreurs typographiques dans le texte source. Pour la seconde opération, la sélection des lexies à étudier prioritairement a été effectuée selon plusieurs critères (notamment forme commune aux deux langues, sous-formant productif, apparition récente ou moins récente, etc.).

Étape 3 : Import des contextes/phrases (et méta-données disponibles dans JSI) d'un échantillon des lexies validées et exploration/description dans le cadre des outils proposés par Néoveille et d'autres plateformes.

2.2 Description linguistique des lexies

La description linguistique en synchronie comprend cinq aspects (Cartier, 2018) :

- **description phonologique et orthographique (profil formel)** : cet aspect couvre les caractéristiques phonologiques et orthographiques de la lexie empruntée. De ce point de vue, plusieurs phénomènes peuvent se produire, généralement liés à l'adaptation phonologique et/ou orthographique à la langue d'arrivée (adaptation au système phonologique : par exemple disparition des voyelles longues ou courtes en français, variantes orthographiques : terminaison en *-ink* en tchèque) ;
- **description morphologique (profil morphologique et productif)** : on peut décomposer un lexème en formants ou morphèmes, lexicaux (base ou radical) et/ou grammaticaux (affixes et morphèmes grammaticaux proprement dits). Par exemple, *aquaplaning* se décompose en un radical (*aquaplan-*) et un morphème grammatical (*-ing*). Il est également possible de décomposer les formants dans la langue d'origine (*aqua-* et *plan-*) ou dans la langue d'arrivée, lorsqu'une dérivation, une composition ou une transformation se produit (*anti-aquaplaning*, *aquaplaning-retrofitting*, *aquaplan*, *aquaplaneur*, *aquaplaner*, *aquaplanage-freinage*, etc.). Ce dernier élément permet de détecter les emprunts plus intégrés, et de suivre leur productivité. Un dernier aspect concerne les restrictions morphologiques qui s'appliquent à certaines formations (absence d'un pluriel, limitations des formes verbales, etc.) ; l'ensemble de ces formes générées à partir d'une forme empruntée pourra être repéré automatiquement à l'aide de l'outil *WordList* de *Sketch Engine* ;

- **description lexico-syntaxique (profil combinatoire)** : cette description porte sur le comportement lexico-syntaxique des créations lexicales, qu'il s'agisse des *collocations* ou des *collostructions* saillantes pour la lexie, aboutissant à un *profil combinatoire* (Gries et Divjak 2009) ; le repérage de ses structures lexico-syntaxiques peut être fait sur des bases purement statistiques²; pour expliciter le *profil combinatoire*, il faut combiner les mesures d'association avec des patrons syntaxiques saillants pour la lexie considérée (qui pourront soit être extraits sur des bases statistiques, soit sur la base des schémas syntaxiques attendus pour la lexie étant donné sa catégorie grammaticale) ; pour ce faire, nous pourrions utiliser l'outil *WordSketch* de *Sketch Engine*, ainsi que l'outil *collocations*.
- **description sémantico-distributionnelle (profil distributionnel)** : ce dernier aspect permet d'approcher la sémantique sous-jacente à chaque lexie en identifiant en corpus les lexies dont le comportement combinatoire est similaire, permettant ainsi d'identifier les lexies en relation sémantique avec la lexie étudiée (synonymie, hyperonymie/hyponymie, antonymie, méronymie, etc.). Pour ce faire, des modèles computationnels par embeddings (camemBERT, FlauBERT pour le français) permettent pour un mot donné dans différents contextes phrastiques, d'obtenir les synonymes ou quasi-synonymes, permettant ainsi de suivre les évolutions éventuelles de sens. Dans le cadre de ce travail, cette étude ne sera pas menée, par manque de place.
- **description diatopique, diastratique et diaphasique (profil socio-linguistique)** : le dernier aspect synchronique permet de caractériser les propriétés sociolinguistiques de la lexie des points de vue de son ancrage diatopique (la lexie est-elle limitée à une aire géographique spécifique ?), diastratique (la lexie est-elle propre à un ou des groupes sociaux spécifiques ?) et diaphasiques (la lexie est-elle propre à un certain type de discours, à des situations spécifiques ?) ; cet aspect sera, étant donné les limitations du corpus utilisé, cantonné à l'étude de la diatopie et de la diastratie en français (les textes sont tous des types de discours journalistiques).

2.3 Suivi du cycle de vie des anglicismes et concurrents locaux

Le plan diachronique de l'analyse pourra être traité étant donné l'empan temporel du corpus (2014–2020), afin de caractériser les évolutions globales (à quelle étape du cycle de vie en sommes-nous pour les lexies : émergence, diffusion, adoption ?) et spécifiques (en prenant un certain nombre de lexies déjà ancrées dans l'usage et pour lesquelles il sera possible de tracer les évolutions des propriétés synchroniques). Dans ce cadre, l'évolution comparative des lexies en *-ing* et de concurrents locaux permettra de voir, sur un petit nombre de cas, comment chaque langue gère les emprunts et les nouveautés lexicales issues de l'anglicisation.

² Dans l'espace disponible, nous ne pouvons détailler les mesures utilisables. Nous renvoyons à (Evert et Krenn, 2004) pour une revue des mesures d'attrance collocationnelle.

2.4 Études comparatives

Pour prendre en compte l'aspect comparatif, nous retiendrons 5 lexies, de grande ou moyenne diffusion, qui sont attestées dans les deux langues étudiées, afin d'effectuer une comparaison des destins des lexies en question dans chacune des langues.

3. Résultats et analyse

3.1 Collecte et validation des emprunts à base *-ing/-ink*

Le travail de collecte des formes lexicales dans les corpus JSI a permis de récupérer 9 900 entrées en *-ing* (422 283 occurrences) et 945 entrées en *-ink* (150 332 occurrences) pour le tchèque et 31 473 entrées en *-ing* (3 775 671 occurrences) pour le français. On constate donc dès l'abord le très grand nombre de formations dans les deux langues et une différence notable entre les deux langues, due notamment à un historique des emprunts beaucoup plus ancien en français (Mudrochová, Lazar 2018). Afin de filtrer manuellement parmi ces résultats les seuls emprunts en *-ing/ink*, nous avons réduit la liste des lexies à celles ayant une fréquence supérieure à 5, puis avons filtré manuellement les données afin de ne retenir que les seuls « véritables » emprunts en *-ing/-ink*. L'ensemble des résultats est présenté dans le tableau 1.

Tableau 1 : Lexies en *-ing/-ink* collectées dans les corpus JSI

	fréq. minimale = 1		fréq. minimale = 5			
	nbre de lexies	nbre d'occ.	données brutes		données filtrées	
			nbre de lexies	nbre d'occ.	nbre de lexies	nbre d'occ.
Français	21 077	3 744 101	8 574	3 720 735	539	810 178
Tchèque	8 076 (<i>-ing</i>) 837 (<i>-ink</i>)	404 439 147 314	2 630 (<i>-ing</i>) 234 (<i>-ink</i>)	395 578 146 377	542	44 298

Les données filtrées non retenues appartiennent à trois catégories principales :

- lexies existantes mais ne rentrant pas, à l'évidence, dans le patron recherché (*-ing/-ink*), par exemple *pink*, *link* (tchèque) ;
- emprunts anglais qui, à l'évidence, font partie d'une citation dans le texte ; cela concerne un très grand nombre d'entrées, notamment des termes anglais très fréquents (*being*, *thinking*, etc.)
- coquilles.

Au final, le nombre de lexies retenues (au total 785, dont 296 communes aux deux langues) montre la très grande productivité du patron *-ing/ink*, à la fois en français et en tchèque. Les lexies appartiennent à deux grandes catégories, situées sur un continuum de fréquences :

- **lexies empruntées de longue date ou faux-emprunts intégrés de longue date dans l'une ou les deux langues** (*marketing, holding, happening, trading, meeting, standing, feeling, etc.*) : une fréquence constatée importante et stable sur toute la période, ainsi que la présence de ces lexies dans les dictionnaires de référence, permettent d'identifier ce groupe de lexies ; dans ce corpus, 44 lexies de fréquence supérieure à 100 sont présentes en français et en tchèque.
- **lexies plus récentes et plus ou moins intégrées**, qui seront l'objet principal de notre étude.

On peut essayer de comparer ces listes avec une liste de référence en anglais, afin d'évaluer le taux d'emprunts véritables et le taux de faux-emprunts dans ces listes. Pour ce faire, nous avons évalué, avec le correcteur orthographique Hunspell³, le taux de reconnaissance de l'ensemble des lexèmes retenus (1 037 au total). On aboutit à un taux de 17,07% de lexèmes non reconnus en anglais (134). Cela montre à la fois le très fort pourcentage d'emprunts (651, soit 82,93%), et l'influence du patron en *-ing/ink*, même si une partie des lexèmes non attestés en anglais sont de simples adaptations orthographiques, notamment en tchèque (*inženýring, tréning, kyberstalking, etc.*).

Une autre caractéristique notable des emprunts et faux-emprunts est la présence de sous-patrons productifs : *boarding, surfing, marketing, building, funding, banking, cycling, planning, reading, lifting, washing, monitoring, training, running, sourcing, streaming, scraping, watering, crossing, painting, squatting, schooling, phishing, blogging, checking, ghosting, lighting, shopping, talking, threading, processing, speeding*) ainsi que la présence de préfixes, quasi-préfixes, lexèmes répétés (*micro, aqua, body, crowd, work, anti, home, multi, video*). On constate également des sous-familles liées à des domaines privilégiés (informatique, économie/finances/gestion d'entreprise, pratiques sociales et sportives spécifiques, pratiques des réseaux sociaux).

Évolution de la situation depuis 2013

Afin de comparer les formes attestées en *-ing/-ink* entre les deux langues et l'anglais, nous avons également utilisé, comme lexique de référence, l'ensemble des formes en *-ing/ink* dans les trois corpus comparables Aranea⁴ (Benko 2014). Le diagramme de Venn (figure 1) montre, sans surprise, un nombre bien plus conséquent de lexèmes en anglais et une majorité de lexèmes empruntés à l'anglais (respectivement 46% pour le tchèque et 49% pour le français), mais aussi un nombre important de lexèmes propres à chaque langue (Tchèque : 2103, Français : 1971) et une partie commune entre le tchèque et le français (147). Cependant, par sondage dans le Corpus NOW⁵, on trouve quelques occurrences des lexèmes concernés dans la presse américaine en ligne⁶.

³ <http://hunspell.github.io/>, consulté le 30. 05. 2021.

⁴ Nous avons récupéré la liste des mots en *-ing/ink* pour les trois langues avec l'expression régulière "[\w-]*in[gk]" à partir des versions Minus accessibles à cette adresse : <http://unesco.uniba.sk/guest/index.html>, consulté le 05. 02. 2021.

⁵ <https://www.english-corpora.org/now/>, consulté le 04. 02. 2021.

⁶ Un bon exemple est *home-staging*, présent en français dès les années 2005, alors que la première occurrence dans le corpus NOW est de 2010.

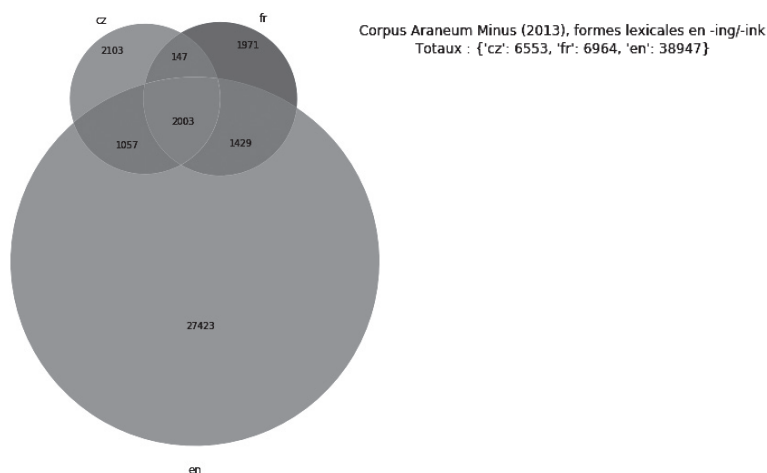


Figure 1 : Diagramme de Venn comparant les lexies en -ing/-ink dans les corpus Aranea 2013

3.2 Description linguistique, sociolinguistique et diachronique des lexies et concurrents locaux

Dans l'espace imparti, il n'est pas possible de présenter exhaustivement les caractéristiques phonologiques, orthographiques et morphologiques (flexion, dérivation) des différentes formations lexicales. On note cependant un certain nombre de régularités dans le mécanisme d'adaptation des emprunts étudiés :

- adaptation phonologique et orthographique surtout pour le tchèque, disposant d'un système phonologique et orthographique plus lointain de l'anglais que le français ;
- un grand nombre de formations restent dans un état de quasi-hapax, comme en atteste, pour les formations à très basse fréquence, des indicateurs de mention (mise entre guillemets, glose à proximité) ;
- la pénétration des formes en -ing est bien plus importante en français qu'en tchèque, en raison de l'historique de cette pénétration, qui a démarré dès la fin de la Seconde guerre mondiale en France, et a connu un regain important avec l'avènement de la communication Internet puis les réseaux sociaux ; c'est à cette époque que la majorité des formations en -ing sont apparues en tchèque; en français, la pénétration aboutit à de nombreux faux emprunts sur la base -ing, en faisant un véritable quasi-suffixe productif ;
- la diffusion des formations se matérialise par quatre caractéristiques :
 - adaptation phonologique / orthographique (essentiellement en tchèque, mais pas de façon systématique, une étude complémentaire devrait être effectuée) ;
 - disparition des emplois en mention ;
 - intégration à la morphologie productive (voir Cartier 2018a et 2019) : apparition sur la base empruntée de nominalisation dénotant l'agent (*busking* > *busker* / *buskeur* ; *shopping* > *shopper*), génération d'une forme verbale (*busking* > *busker* ; *shopping* > *shopper*),

formations dérivées, composées, transformées (*jogging* > *écojogging* > *écojog* ; *running* > *runner* > *run*, *showrunner*, *rétro-running*, *speedrunner*) ;

- une augmentation des fréquences.
- une étude de la concurrence entre le quasi-suffixe *-ing* et des bases locales (*-age*, *-ment*) serait utile pour prendre la mesure de la place exacte de ce premier suffixe en français.

Dans cette section, afin d'étudier plus précisément les caractéristiques des formations en *-ing/-ink*, nous détaillerons 2 lexies spécifiques (*phishing*, *géocaching*), ainsi que deux emprunts en *-ing* productifs (*sharing* et *washing*).

Phishing – Le mot désigne une technique qui est employée par le fraudeur dans le but d'obtenir des renseignements personnels et bancaires (p.ex. numéro de carte, mot de passe, etc.) Étant donné qu'il s'agit d'un acte criminel pénalisé dans les deux pays, les institutions officielles ressentent la nécessité de proposer un équivalent autochtone pour le mot anglais. Dans l'espace francophone, deux équivalents ont été proposés, l'un par la Commission d'enrichissement de la langue française – *filoutage*, l'autre par l'Office québécois de la langue française, *hameçonnage*.

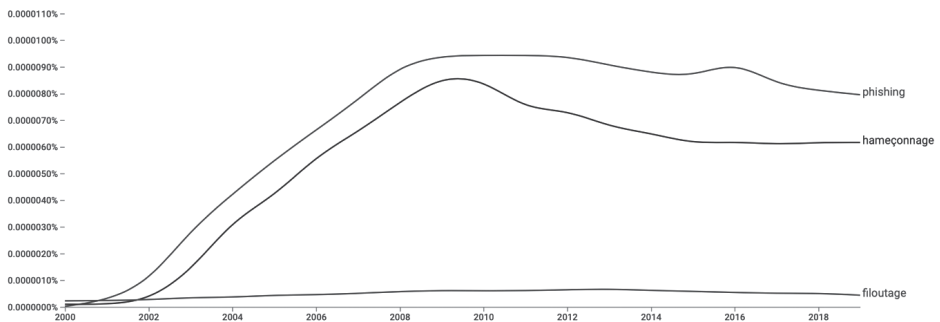


Figure 2 : Graphes d'évolution : *hameçonnage*, *phishing*, *filoutage* (Google Ngrams)

Comme le montre la figure 2 ci-dessus, le seul concurrent réel de l'anglicisme *phishing* est le terme *hameçonnage* proposé en 2004 par l'Office québécois de la langue française. Le site officiel en donne la définition suivante :

Technique de fraude basée sur l'usurpation d'identité, qui consiste à envoyer massivement un message en se faisant passer pour une institution financière ou une entreprise commerciale de renom afin d'induire les destinataires en erreur et de les inciter à révéler des informations sensibles à leur insu.⁷

Le terme *filoutage* proposé par la commission d'enrichissement de la langue française en 2006 semble trop général, désignant toute forme d'escroquerie (Honová, Lazar

⁷ http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8869710, consulté le 01. 02. 2021.

2019 : 22). Le terme *escroquerie par courriel* est aussi déconseillé en raison de son manque de précision. En tchèque, on recourt souvent à l'explication de ce terme « *podvod v rámci internetového bankovníctví – fraude dans le domaine de la banque en ligne* ». Il est évident que ce terme est trop long pour entrer dans l'usage. Le mot *rhybaření*, une variante du mot *rybaření*, est également attesté. En ajoutant la lettre *h* dans le mot *rybaření* (pêche), on fait allusion à la pêche à la ligne (Internet est métaphorisé par l'océan, et les utilisateurs, par des poissons).

Géocaching/Geocaching – Il s'agit d'un loisir qui consiste à utiliser la technique du géopositionnement par satellite (GPS) pour rechercher ou dissimuler des caches ou des géocaches dans divers endroits dans le monde entier. En tchèque, ce mot est prononcé [gɛɔkɛʃɪŋk] et il semble que les locuteurs tchèques essaient de modifier son orthographe pour mieux imiter sa prononciation. Dans l'usage quotidien, on atteste alors couramment la variante tchéquisée *geokešink*. Le mot est employé notamment par les spécialistes qui s'intéressent à cette activité. Il semble intéressant de mentionner que le mot *cache* a aussi subi une modification en transformant son orthographe en *keš*. Comme le montrent les sites spécialisés dans ce domaine, l'usage du mot *keš* est assez répandu. La personne qui pratique cette activité (*geocacher*) est souvent appelée en tchèque comme *geokačer* ou tout simplement *kačer*, ce qui est non seulement une adaptation orthographique, mais aussi un jeu de mot, car *kačer* désigne en tchèque un canard mâle. En français, on retrouve plutôt la variante avec *é* qui semble bien intégrée. Le comité de terminologie de radio Canada a aussi proposé le terme *chasse au trésor GPS*.

Sharing – le mot semble fort productif dans les deux langues en question et il dispose aussi de plusieurs équivalents autochtones *partage/sdílení* . p.ex *code sharing – partage de code/sdílení kódu*, *carsharing – partage des véhicules/sdílení auta*, *knowledge sharing – partage du savoir/sdílení know-how*. En tchèque, le mot *share* fonctionne aussi comme un verbe (*šarovat*) dans les contextes spécifiques, notamment liés à l'informatique. On retrouve aussi la variante qui dispose d'une orthographe tchéquisée – *šerovat*. Le mot *share* peut aussi créer un adjectif comme p.e.x *šarovaný dokument*.

Washing et greenwashing

Pour illustrer plus en détail la pénétration du patron morphologique en *-ing/-ink*, nous focalisons ci-après sur le cas des formations en *washing*. Ce lexème a pris un sens figuré très tôt en anglais, dès 1762 en corpus et dès 1850 pour les attestations lexicographiques⁸, dans le sens de blanchiment (d'une personne ou d'une institution) et déjà dans le sens productif actuel ("*to cover up, conceal, give a false appearance of cleanness to*", Online Etymological Dictionary) mais exclusivement dans le composé *whitewash/whitewashing*. A partir des années 1980–90, on voit apparaître, d'abord en anglais puis, vers les années 2000 en français et un peu plus tard en tchèque, d'autres composés sur la base *washing* avec un double mouvement : d'une part, une extension du paradigme des couleurs (*blackwashing, pinkwashing*, avec mobilisation de la symbolique liée aux couleurs, ici les seins pour la cause du cancer du sein, *redwashing*, qui présente un cas similaire, le rouge

⁸ Voir le dictionnaire étymologique en ligne : https://www.etymonline.com/search?q=whitewash&ref=searchbar_searchhint, consulté le 08. 02. 2021.

désignant l'idéal communiste, *bluwashing* portant pour sa part sur la communication sur des actions en faveur de la planète bleue, *purplewashing* pour la cause féminine) et, d'autre part ou en conséquence du premier mouvement de productivité, une extension à des noms désignant directement le domaine de la pratique frauduleuse (*gender-washing*, *local-washing*, *covid-washing*, *art-washing*, *smart-washing*, *rainbow-washing* etc.), ou même une caractérisation générale de la pratique (*brandwashing*, *purpose-washing*), le nom modifieur n'étant alors plus qu'un spécifieur de la forme en *-ing*. En français, il existe un quasi équivalent sur la même base, *blanchir*, attesté dès les années 1850 (source : TLFi, entrée *blanchir*) s'appliquant à une personne, mais la nominalisation est plus difficilement traçable (aucune entrée dans le TLFi, première attestation en 1995 dans Frantext). L'expression *blanchiment*, dans Google Ngrams, en français, après une première apparition entre 1945 et 1950 (mais exclusivement dans le sens concret), voit son usage décoller après 1985, sans qu'il soit possible de lier cet emploi figuré à *whitewashing* de façon certaine.

La figure 3 détaille le nombre de formations lexicales en *washing* sur la période 2014–2020 dans les trois langues (anglais, français, tchèque), dans les corpus JSI, montrant les formes communes et distinctes.

JSI Lexèmes en washing dans les corpus JSI (2014-2020). Total d'entrées : 1287

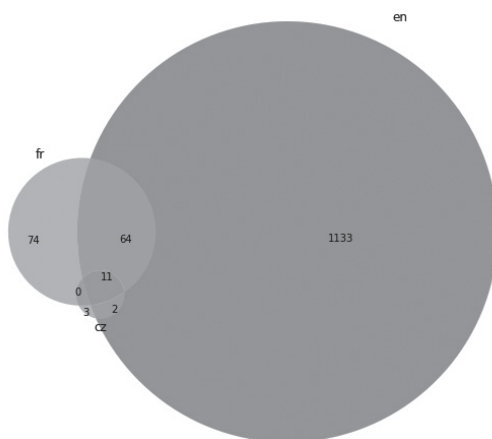


Figure 3 : Diagramme de Venn des lexèmes en *washing* (fr, cz, en)

Sur les 1287 lexèmes au total, la distribution est évidemment largement dominée par les données anglaises (1133 entrées, contre 149 pour le français et 16 en tchèque) avec, pour les données tchèques et françaises, une grande ou écrasante majorité empruntée à l'anglais (75 pour le français et 13 pour le tchèque). On note toutefois la présence de 74 formations propres au français (*agri-washing*, *fabwashing*, *experience-washing*, *fancia-washing*, *truc-washing*, etc.), les 3 propres au tchèque étant en réalité des coquilles (*ncewashing* pour *nicewashing*, *braiwashing* pour *brainwashing* et une création : *auto-brain-washing*, non attesté dans le corpus JSI anglais mais présent sur Google).

11 lexèmes sont partagés par les trois langues : *artwashing*, *blackwashing*, *brain-washing*, *brainwashing*, *greenwashing*, *handwashing*, *pinkwashing*, *sportswashing*, *sportwashing*, *whitewashing* et *washing*⁹.

Dans les lexèmes formés en *washing*, on note également des sous-formations productives : *hand* (53 lexèmes), *brain* (46 lexèmes), *white* (29 lexèmes), et *green* (15 lexèmes). On constate donc la grande productivité, récente (Google Ngrams situe le décollage dans les années 1990 et plus encore après 2000) des formations en *N/ADJ + washing*, en anglais et en français, avec un sens bien défini (le blanchiment d'une réputation par des activités dénotées par *N/ADJ*).

Dans la suite, nous examinons la formation *greenwashing* pour tenter de mieux le caractériser et étudier les éventuels concurrents locaux. Le terme est défini dans le *Grand dictionnaire terminologique* dès 2007 : « Opération de relations publiques menée par une organisation, une entreprise pour masquer ses activités polluantes et tenter de présenter un caractère écoresponsable. ». Selon la version anglaise de Wikipédia¹⁰, le terme a été forgé (sous la forme *green washing*) en 1986 aux États-Unis pour évoquer la pratique consistant à prévoir dans des hôtels des emplacements dans les chambres pour favoriser la réutilisation des serviettes. Il est attesté dès 1991 dans *Europresse* (*The Daily News, Halifax*). Le lexème est une composition entre *green* et la réduction de *whitewashing* (> *washing*) le formant devenant productif. L'office québécois de la langue française a dès 2007 proposé des équivalents français¹¹ : *écoblanchiment* de façon principale, ainsi que *mascarade écologique*, *blanchiment vert* et *verdissage d'image*. De son côté, les commissions terminologiques de l'Académie française ont proposé en 2013¹² *verdissage d'image*, évoquant dans une note également *écoblanchiment* et *blanchiment écologique*. (Dury, 2006) ajoute à cette liste *verdissage*, *maquillage vert*, *mascarade écologique*, *désinformation verte* et *badigeonnage vert*, dont elle trouve des traces dans un corpus constitué ad hoc à partir de sources internet diverses obtenues à partir d'une recherche sur *greenwashing* dans des pages en français. En tchèque, il ne semble pas y avoir d'équivalent, en dehors de l'expression composée *zelené vymývání mozků* (*lavage des cerveaux vert*)

Dans le corpus JSI (2014–2020), nous trouvons, pour le français, les fréquences présentées dans la figure 4 et dans la figure 5 pour ce qui concerne l'évolution.

On constate que *greenwashing* (sous ses trois variantes graphiques *green washing*, *green-washing* et *greenwashing*, cette dernière dominant très largement) est la forme exclusive (tchèque) ou en tout cas écrasante (français, 92,5%), une seule forme concurrente locale étant véritablement utilisée, *écoblanchiment*, mais sans avoir été adoptée.

⁹ *Handwashing* et *sportwashing / sportswashing* font, en réalité, dans les corpus français et tchèque, partie de citations en anglais. Pour ce qui concerne *washing*, il s'agit d'un composant d'emprunts composés (*green*, *social*, *white*, *pink*, *féminisme*, *purpose*, *innovation*, *digital* pour le français, *brain*, *green* et *machine*, postposé, en tchèque).

¹⁰ Voir aussi <https://web.archive.org/web/20150923212726/http://www.dailyfinance.com/2011/02/12/the-history-of-greenwashing-how-dirty-towels-impacted-the-green/> sur l'histoire de la naissance du concept, consulté le 08. 02. 2021.

¹¹ http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8365406, consulté le 08. 02. 2021.

¹² <http://www.culture.fr/franceterme/result?francetermeSearchTerme=%C3%A9coblanchiment&francetermeSearchDomaine=0&francetermeSearchSubmit=rechercher&action=search>, consulté le 08. 02. 2021.

L'évolution diachronique ne fait que préserver les proportions d'emploi entre les variantes. On notera cependant, dans la presse canadienne francophone du corpus, une situation plus équilibrée : *écoblanchiment* (53), *greenwashing* (45), *green washing* (11), *blanchiment écologique* (2) et *mascarade écologique* (1).

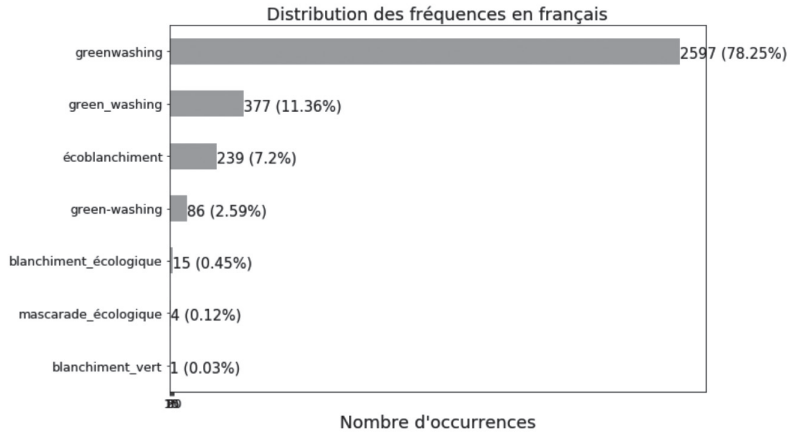


Figure 4 : Distribution des fréquences entre les termes désignant le *greenwashing*

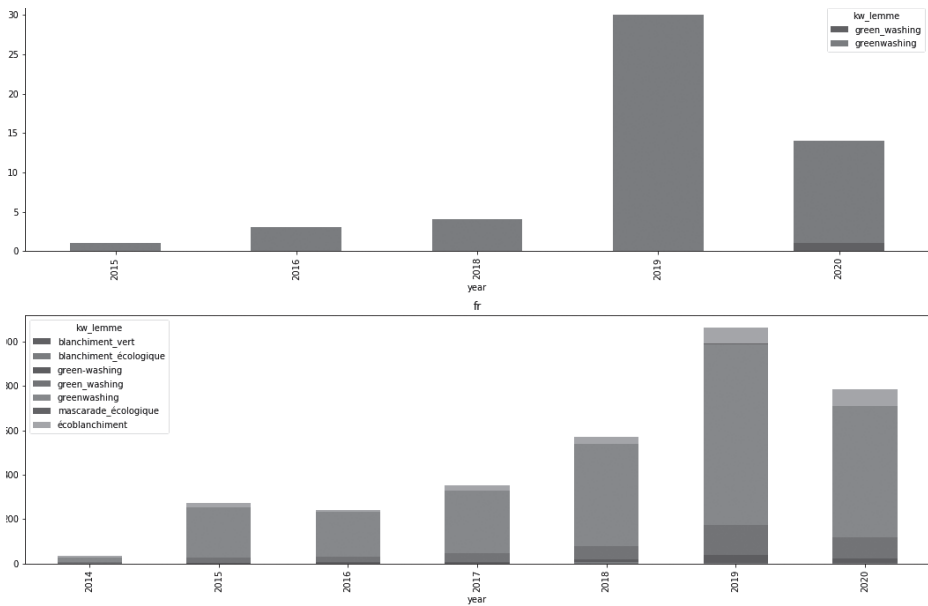


Figure 5 : Évolution de la distribution des fréquences entre les termes désignant le *greenwashing* (français et tchèque)

Conclusion

L'étude proposée ici, encore à un stade préliminaire, a tenté de décrire et d'analyser la pénétration du schéma *N/ADJ(-)ing* en français et en tchèque, en prenant appui sur un corpus diachronique (2014–2020) de presse en ligne et sur une méthodologie d'étude comprenant à la fois une macro-étude, consistant à relever l'ensemble des formes attestées, leur fréquence et l'évolution de leur fréquence en français et en tchèque puis à identifier quelques caractéristiques de ces données, puis une micro-étude, focalisant sur quelques formations, et à en faire une analyse plus détaillée.

Il en ressort quelques enseignements : tout d'abord la très grande productivité, en anglais d'abord, en français et en tchèque ensuite, de cette formation, sous la forme *-ing* ou *-ink* pour le tchèque. Le mouvement a particulièrement décollé après les années 2000 (essor des communications par Internet) et les années 2010 (démocratisation de l'utilisation des réseaux sociaux), et sans relâchement depuis. Les domaines couverts sont larges : le domaine de la finance, de l'informatique, des réseaux sociaux et plus largement les pratiques sociales et professionnelles. En français, nous avons à faire à un quasi-suffixe, avec de nombreux faux-emprunts et un sens distinct des suffixes existants en *-ment* et *-age*, qui mériterait une étude spécifique.

Les formations en *-ing* sont par ailleurs marquées par l'existence de sous-familles elles-mêmes productives, autour de bases comme *boarding*, *surfing*, *marketing*, *building*, *funding*, *cycling*, *washing*, *running*, *sourcing*, etc. montrant là encore l'extension à tous les domaines d'activités. Enfin, il faut noter la différence entre le tchèque et le français, cette première langue étant plus en retrait sur l'intégration qui ne semble porter que sur des emprunts directs.

En français, si on recourt à l'usage de l'anglicisme en *-ing*, il subit plutôt rarement des modifications orthographiques. Au contraire, en tchèque, la transformation de la terminaison *-ing* en *-ink* est fréquente ce qui dénote dans la langue tchèque une meilleure correspondance entre le graphème et le phonème. On peut en déduire que les emprunts à l'anglais circulant déjà depuis longtemps en tchèque subissent souvent une modification orthographique (p.ex. *inženýrink*, *geokešink*) et le locuteur tchèque ne serait pas toujours capable de découvrir leur origine anglaise. Certes, on trouve quelques exemples, mais il s'agit surtout de la créativité des locuteurs natifs (p. ex. *rhybaření*) plutôt qu'une activité organisée par l'Institut de la langue tchèque. Ce n'est pas le cas en France ou au Canada où il existe des institutions spécialisées s'occupant de la veille néologique et proposant des termes autochtones pour éviter les anglicismes. Néanmoins, comme nous le montre le cas de *phishing* et ses équivalents *filoutage* et *hameçonnage*, l'anglicisme reste privilégié dans la presse. Le mot *share* que nous avons analysé dispose aussi d'un équivalent précis *partage/sdílení* et il semble que l'usage de l'anglicisme se limite aux domaines spécifiques comme p.ex. l'informatique. Les formations en *washing* sont particulièrement illustratives des différents phénomènes en jeu : un formant devenu productif en anglais vers 1990 à partir de *whitewashing*, avec d'abord une extension paradigmatique à différentes couleurs et aux symboliques associées, qui ensuite se généralise à la mention nominale de tout nom dénotant une cause éthique. Ces formations se diffusent de manière généralisée en français et dans une moindre mesure en tchèque. En français, nous aboutissons également à la formation d'un grand nombre de faux-emprunts. L'étude des concurrents

locaux pour *greenwashing*, principalement *écoblanchiment*, montre que, en dehors de la zone francophone canadienne, c'est bien l'emprunt qui s'impose largement, avec cependant un emploi faible mais stable d'*écoblanchiment* qui tend à montrer une situation de variation qu'il conviendrait d'étudier de plus près.

BIBLIOGRAPHIE

- Benko, V. « Aranea: Yet Another Family of (Comparable) Web Corpora ». In Sojka, P., Horák, A., Kopeček, I. et Pala, K. (éds) (2014) : *Text, Speech and Dialogue*. 17th International Conference, TSD 2014, Brno, Czech Republic, September 8–12. Proceedings. LNCS 8655. Springer International Publishing Switzerland, pp. 257–264.
- Cartier, E. (2018a). « Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néouvelle ». In Kacprzak, A., Mudrochová, R. et Sablayrolles, J.-F. (éds), *Emprunts en question(s)*. La Lexicothèque, Limoges : Lambert-Lucas, pp. 145–186
- Cartier, E. (2018b) : « Dynamique lexicale des langues : éléments théoriques, méthodes automatiques, expérimentations en français contemporain », document inédit HDR, 237 pages. <https://tal.lipn.univ-paris13.fr/neouvelle/html/data/ecartier/ecartier_inedit_final_09122018.pdf>, consulté le 08. 02. 2021.
- Cartier, E. (2019) : « Neouvelle, plateforme de repérage et de suivi des néologismes en corpus dynamique », *Neologica*, 13, pp. 23–54.
- Crystal, D. (2012) : *English as a Global Language*. Cambridge : Cambridge University Press. <https://doi.org/10.1017/CBO9781139196970>.
- Duběda, T. (2020) : « The Phonology of Anglicisms in French, German and Czech: A Contrastive Approach », *Journal of Language Contact*, 13, 2, pp. 327–350. <https://brill.com/view/journals/jlc/13/2/article-p327_327.xml?rskey=ZRKerY&result=14>, consulté le 08. 02. 2021.
- Dubois, J. (1962) : *Étude sur la dérivation suffixale en français moderne et contemporain : essai d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Paris : Larousse.
- Dury, P. (2006) : « La dimension diachronique en terminologie et en traduction spécialisée : le cas de l'écologie ». *Aspects diachroniques du vocabulaire*, pp. 109–124.
- Evert, S. et Krenn, B. (2004) : « Computational approaches to collocations ». *Introductory course at the European Summer School on Logic, Language, and Information (ESSLLI 2003)*, Vienna.
- Gries, S. T. et Divjak, D. (2009) : « Behavioral profiles: a corpus-based approach to cognitive semantic analysis ». In Evans, BV. et Pourcel, S. (éds) (2009). *New directions in cognitive linguistics*, John Benjamins, pp. 57–75.
- Honová, Z. et Lazar, J. (2019) : « La diffusion des néologismes terminologiques dans l'espace francophone en comparaison avec la situation en République tchèque », *Lingua Viva* 2019, n. 29, pp. 19–28 <https://www.pf.jcu.cz/research/lingua_viva/LV29_2019.pdf>, consulté le 08. 02. 2021.
- Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P. and Suchomel V. (2014) : *The Sketch Engine: ten years on*. *Lexicography*, 1: 7–36, 2014.
- Mudrochová, R. et Lazar, J. (2018) : « La circulation des emprunts néologiques dans le domaine de la mode en tchèque et en français dans une perspective diachronique », *ELAD-SILDA* [en ligne], 1, <<http://revues.univ-lyon3.fr/elad-silda/index.php?id=324>>, consulté le 08. 02. 2021.
- Mudrochová, R. (2017) : « À propos des mots en *-ing* d'origine anglaise issus du dictionnaire le Petit Robert », *Linguistica Pragensia*, 27, 1, pp. 7–19.
- Rejzek, J. (1993) : « K formální adaptaci anglicismů », *Naše řeč*, 76, 1, pp. 26–30, <<http://nase-rec.ujc.cas.cz/archiv.php?art=7106>>, consulté le 08. 02. 2021.
- Svobodová, D. : « K otázce spisovnosti vybraných typů přejatých lexikálních jednotek », [en ligne] <<http://konference.osu.cz/cestina/dok/svobodova-diana.pdf>>, consulté le 08. 02. 2021.
- Trampus, M. et Novak, B. (2004) : « The Internals Of An Aggregated Web News Feed ». *Proceedings of 15th Multiconference on Information Society 2012 (IS-2012)*.

Sitographie

(la ou les langues concernées sont indiquées entre parenthèses, selon la valeur ISO)

Ressources lexicales

(Fr) Le grand dictionnaire terminologique (GDT) : <http://gdt.oqlf.gouv.qc.ca/>

(Fr) Trésor de la Langue Française Informatisé (TLFI) : <http://atilf.atilf.fr/>

(Fr) Wiktionnaire : https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil

(En) Urban Dictionary : <https://www.urbandictionary.com/>

(En) Etymological Dictionary : <https://www.etymonline.com/>

(Fr, En, Cz) Babelnet : <https://babelnet.org/>

Corpus

(En) Corpus *News on the Web* (NOW), Corpus of Contemporary American English (COCA) et Corpus of Historical American English (COHA) : <https://www.english-corpora.org/>

(Fr, En) Europresse : <https://nouveau.europresse.com/Login/> (accès restreint)

(En, Fr, Cz) Archives de l'internet : <http://web.archive.org/>

(Fr) Archives du journal Le Monde : <https://www.lemonde.fr/recherche/> (accès restreint)

(Fr) Frantext : <https://www.frantext.fr> (accès restreint)

(Fr, En, Cz) JSI Timestamped web corpus : <https://www.sketchengine.eu/jozef-stefan-institute-newsfeed-corpus/> (accès restreint)

(Cz) Corpus de référence du tchèque contemporain : <https://www.korpus.cz>

(Fr, En) Google Ngrams : <https://books.google.com/ngrams>

Emmanuel Cartier

Université Sorbonne Paris Nord, LIPN UMR 7030 CNRS

emmanuel.cartier@univ-paris13.fr

Jan Lazar

Université d'Opole/Université d'Ostrava

janlazar@email.cz