

Le corpus heuristique : un outil qui montre mais ne démontre pas

Tobias Scheer



Édition électronique

URL : <http://journals.openedition.org/corpus/210>

DOI : [10.4000/corpus.210](https://doi.org/10.4000/corpus.210)

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 1 décembre 2004

ISSN : 1638-9808

Référence électronique

Tobias Scheer, « Le corpus heuristique : un outil qui montre mais ne démontre pas », *Corpus* [En ligne], 3 | 2004, mis en ligne le 02 décembre 2005, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/corpus/210> ; DOI : <https://doi.org/10.4000/corpus.210>

Le corpus heuristique : un outil qui montre mais ne démontre pas

Tobias SCHEER

« Bases, Corpus et Langage », UMR 6039, CNRS

Résumé : Le présent article met en évidence deux fonctions distinctes du corpus : heuristique et validante. Chacun connaît la seconde : un corpus est utilisé afin de démontrer la plausibilité ou la véracité d'une hypothèse qui a une existence indépendante. A la différence de celle-ci, la première fonction montre, plutôt qu'elle ne démontre. Le linguiste ne cherche plus ici à convaincre ; il cherche à comprendre. Il part à la rencontre de l'inconnu avec un questionnement, mais sans solution, même hypothétique, pour le puzzle dont il a identifié quelques pièces. La vertu du corpus ici est consubstantielle à son existence : le classement. Qui dit corpus dit en effet principe de classement. Le corpus heuristique a ceci de particulier qu'il est éphémère : sitôt ses services rendus, il sera transformé par le linguiste en corpus de validation (qui aura vocation à convaincre le monde). Ce changement d'objectif induit alors une organisation et une présentation très différentes de celles qui ont permis la découverte du fonctionnement linguistique en question. En remontant cette opération, le présent article essaie d'isoler les propriétés invariables des corpus heuristiques. Le propos est illustré par un cas particulier tiré de mon expérience personnelle, *i.e.* les alternances de longueur vocalique dans les itératifs tchèques.

1. Introduction*

Un corpus peut servir deux propos au moins : ou bien il donne à un ensemble de faits une structure qui dégage la vue du linguiste dans une situation où celui-ci poursuit une question mais n'a pas de solution, ou alors il teste la validité empirique

* Ce texte a bénéficié d'une relecture attentive et efficace de la part de Sylvie Mellet.

d'hypothèses qui existent pour des raisons indépendantes. On peut donc parler de corpus heuristiques dans le premier, et de corpus de validation dans le second cas : l'un montre le chemin, l'autre démontre qu'une hypothèse est bonne (ou mauvaise)¹. Je voudrais ici illustrer la fonction heuristique du corpus en phonologie². L'exemple choisi est tiré de ma propre expérience et concerne le domaine slave : la longueur vocalique tchèque en général, et les alternances que l'on trouve au sein des verbes itératifs en particulier. La fonction heuristique du corpus est également illustrée dans la contribution de Marie-José Dalbera-Stefanaggi et Jean-Philippe Dalbera au présent volume.

J'ai choisi de mettre le corpus heuristique sur le devant de la scène parce qu'il est le parent pauvre de la famille : par la nature des choses, il est voué à l'oubli. Car lorsqu'il a rendu service, lorsqu'il a aidé à découvrir, il n'a plus aucun intérêt. Sa seule fonction étant heuristique, il cesse d'exister dès lors que la découverte a été faite, ou au contraire, dès lors qu'aucune découverte n'a été faite et que le projet est avorté. En cas de succès, le linguiste va se dépêcher de transformer son corpus heuristique en corpus de validation : il ne s'agit plus désormais de comprendre, mais de démontrer. La finalité étant tout à fait différente, le corpus initial est remodelé, réaménagé en fonction du nouvel objectif. En résulte l'oubli, voire l'amnésie : on oublie à quoi le corpus a servi lors de l'aventure de la découverte, on oublie quelle forme il avait alors, et on oublie, souvent, en quoi il a été utile et à quel moment précis il a donné des indications déterminantes à la découverte.

Le propos de cet article est donc, aussi, de sortir la fonction heuristique de l'ombre des cabinets et des disques durs afin de montrer, de déterminer son rôle dans la découverte linguistique. Dans les pages suivantes, je me propose de tirer

1. Une variante spécifique de la fonction heuristique est celle où un corpus suggère un questionnement tout autre que celui qui occupait le chercheur au moment de sa constitution.

2. Les fonctions heuristique et validante du corpus, dans un champ d'application différent de la phonologie, ont été mises en perspective dans Mellet (1990 : 108ssq).

Le corpus heuristique : un outil qui montre mais ne démontre pas

d'une étude de cas des enseignements qui soient généralisables à d'autres situations et, peut-être, à d'autres disciplines de la linguistique.

2. Le rôle et les modalités d'intervention du corpus heuristique : le cas des itératifs tchèques

Le sujet d'étude choisi ici est donc la longueur vocalique en tchèque, et plus particulièrement les alternances qui sont observables dans un paradigme précis, celui des verbes itératifs. Ceux-ci ne sont en réalité qu'une pièce d'un puzzle plus large. D'autres catégories de mots tels que les infinitifs, les diminutifs ou encore les comparatifs fonctionnent selon un principe qui est typique pour les langues sémitiques et par conséquent *a priori* étrange pour une langue indo-européenne : le gabarit.

Les faits concernant les diverses catégories de mots mentionnées sont exposés dans les articles d'origine (Scheer 2001a, b, c, 2003, 2004), et il n'est pas utile ici de rappeler leur détail. Il suffira dans un premier temps de comprendre qu'une activité gabaritique se dénonce par l'association nécessaire entre une catégorie morpho-sémantique et un volume consonantique ou vocalique précis (des explications plus détaillées sont fournies en section 10). Dans le cas à l'étude ici, il s'avérera qu'un itératif tchèque doit peser exactement trois unités vocaliques (*i.e.* trois voyelles brèves ou une voyelle brève plus une voyelle longue).

Parmi les multiples catégories gabaritiques du tchèque, j'ai choisi les itératifs parce qu'ils ont été, dans mon histoire personnelle de la découverte du système gabaritique tchèque, la porte d'entrée. C'est seulement lorsque j'ai compris que les itératifs obéissent à un mécanisme gabaritique que les autres catégories mentionnées sont tombées comme des dominos : c'est le point de vue qui crée l'objet, et les itératifs m'ont fourni les lunettes gabaritiques qui seules permettent de comprendre le fonctionnement des autres catégories.

En retraçant le processus de découverte que j'ai vécu dans le cas des itératifs, je montrerai quelle est la fonction du corpus, à quel moment du processus il intervient, quelle est son

atout principal, à quoi il peut servir et où il est sans aucun secours. Sa vertu majeure, je pense, est l'injonction, consubstantielle à son existence, de classer : qui dit corpus dit classification. La classification ne doit pas nécessairement, mais peut, le cas échéant, dégager le brouillard qui empêche le linguiste de voir la réalité linguistique.

Le corpus heuristique est donc un outil qui aide à la découverte – un outil indispensable, oui, mais un outil seulement. Aucun corpus ne s'est jamais substitué, et ne se substituera jamais, à l'intelligence, la perspicacité, le génie et l'expérience du linguiste. Les corpus ne peuvent exister en dehors du raisonnement linguistique : ils ne sont pas une fin en soi. Tout corpus est le résultat d'une activité linguistique préalable : d'un questionnement lorsqu'il est heuristique, d'une hypothèse de travail s'il est validant. En tout état de cause, il fournit du matériel que seul le cerveau humain a, un tant soit peu, structuré et saura interpréter.

Ceci est également vrai pour les corpus qui ont pour seule vocation la sauvegarde d'un patrimoine. Il s'agit ici d'un rassemblement de données sans autre objectif que de les archiver, et de les rendre disponibles pour des exploitations ultérieures non encore identifiées. Or si leur fin est ouverte, leur organisation ne l'est pas. Toute personne chargée de confectionner un corpus patrimonial devra, bon gré mal gré, choisir un principe de présentation, donc de classification. Il est difficile d'imaginer un procédé technique qui échappe à cette règle : un programme par exemple pourrait moissonner Internet selon un principe prédéfini et archiver les résultats par ordre chronologique. Il y a d'abord le principe en question, qui est défini par celui qui moissonne, et constitue donc une activité linguistique préalable, ainsi que la structure de base des faits recueillis. Ensuite, nul ne procédera de la sorte car le résultat serait inutilisable à quelque fin que ce soit, présente ou future, fût-elle patrimoniale. Enfin, il serait probablement déplacé d'appeler l'enregistrement chronologique ainsi obtenu « corpus ».

En somme, donc, le corpus heuristique se trouve, chronologiquement parlant, au milieu du processus de

Le corpus heuristique : un outil qui montre mais ne démontre pas

découverte. C'est une des propriétés qui le distingue du corpus de validation : celui-ci ne sert qu'à démontrer, après le combat de la découverte, qu'une analyse ou une solution déjà acquise est pertinente ou au contraire fausse.

3. L'énigme centenaire de la longueur vocalique Ouest-slave : un terrain miné

Au sein des langues slaves, la famille occidentale (tchèque, slovaque, polonais, kachoube, sorabe) est seule à connaître l'opposition entre des voyelles longues et brèves. Ou plutôt, certaines langues slaves de l'Ouest modernes, le tchèque et le slovaque, ont préservé l'opposition de longueur, commune à la famille entière jadis, qui s'est éteinte ailleurs.

Les alternances tchèques ont une renommée séculaire de bastion imprenable : les voyelles longues et brèves alternent dans tous les sens et apparemment de manière non prédictible. Depuis la fin du 19^{ème} siècle, cet état de fait a été relevé, mais toutes les tentatives de trouver une régularité, diachronique ou synchronique, ont été vaines (Jagić 1894, Belić 1928, Černý, Pedersen 1905, Sedláček 1910, Trávníček 1912, 1921a, b, c, 1925). Vaines à tel point que les slavistes des années 20 ont accepté un mécanisme parfaitement fantaisiste qui ne repose sur rien si ce n'est le désir de pouvoir présenter une solution. Et, aussi, de pouvoir l'enseigner : des générations de slavistes ont dû apprendre par cœur la doctrine de la métatonie fondée par Tadeusz Lehr-Splawiński (1917, 1918, 1926). Ce n'est qu'à partir de la fin des années 50 que la supercherie métatonique a été défaite petit à petit (Kuryłowicz 1952, 1968, Stang 1957, Sadnik 1959, Nonnemacher-Pribić 1961, Shevelov 1964, Kortlandt 1975, 1978a, b, Garde 1976, Šaur 1995, ce dernier étant le seul à faire ce travail pour le tchèque spécifiquement). Le fond (inexistant) et l'histoire de la réception de la métatonie sont exposés en détail dans Scheer (2003).

L'énigme de la longueur vocalique ouest-slave reste donc entière, d'autant que les natifs savent intuitivement que les alternances en question sont tout sauf anarchiques : quelqu'un agite bien les marionnettes selon un scénario précis derrière le paravent de ce qui est visible. Il n'empêche que les grammairistes,

pour l'instant, sont réduites à fournir des listes de mots amorphes où une brève alterne avec une longue dans tel mot, mais non pas dans tel autre mot (c'est le cas par exemple de Trávníček 1948-49 I : 370, 565 et Havránek & Jedlička 1988 : 222ssq, 293ssq).

Je pense que la longueur vocalique tchèque (et probablement ouest-slave) a pu garder son secret si longtemps pour trois raisons. D'abord la cécité synchronique des phonologues pré-structuralistes : la seule cause possible pour tout phénomène phonologique se trouve dans un état antérieur de la langue. Ensuite l'exoticité du mécanisme à l'œuvre : il faut regarder le tchèque à travers des lunettes sémitiques pour voir la réalité – or personne ne songe, *a priori*, que le principe d'organisation fondamental du sémitique puisse être actif dans une langue indo-européenne. Enfin, le seul jugement intuitif d'exemples pris au hasard et l'absence totale de contrôle statistique sur le phénomène en question.

Le but du présent article n'est pas de convaincre le lecteur que la part du lion des alternances de longueur vocalique en tchèque est gérée par un mécanisme gabaritique. C'est ce à quoi je me suis attelé ailleurs (Scheer 2001a, 2003, 2004). L'objet des pages à suivre est plutôt la consubstantialité d'une attitude méthodologique, théorique et statistique qui seule porte au succès. A défaut de l'une de ces composantes, la longueur vocalique tchèque ne s'éclaire pas.

Il est garanti, par exemple, que l'on ne trouvera rien du tout si l'on cherche à expliquer la longueur vocalique tchèque en tournant son regard vers le vieux tchèque, le slave commun, l'indo-européen, les langues slaves de l'Est et du Sud. Ou alors, on trouvera la métatonie. Les alternances de longueur du slave de l'Ouest sont une « invention » de cette branche qui ne doit rien à aucune autre langue slave, ni à un quelconque ancêtre. Elles sont le produit d'une activité dérivationnelle qui relie une base et un objet secondaire. Par conséquent, il s'agit d'un mécanisme synchronique. « Synchronique » ici veut dire, précisons-le, « ayant lieu au moment de la dérivation », ce qui ne renseigne en rien sur le moment où celle-ci a lieu : la dérivation peut être contemporaine au tchèque moderne ou

Le corpus heuristique : un outil qui montre mais ne démontre pas

s'être effectuée dans le passé. La seule chose qui importe est de comprendre que ce n'est pas le temps qui a provoqué les alternances en question, mais une activité grammaticale. L'*a priori* méthodologique néo-grammairien (mais qui irradie encore bien l'espace moderne, surtout dans les pays de l'Est), selon lequel hors diachronie point de salut, condamne donc l'analyse avant de l'avoir entamée.

Par ailleurs, on n'a aucune chance de découvrir le mécanisme gabaritique tchèque si on ne sait pas par avance en quoi il consiste. Il faut donc avoir une attitude théorique qui inclue le fonctionnement gabaritique, et qui plus est conçoit son existence même dans des langues indo-européennes.

Enfin, l'hypothèse gabaritique ne peut émerger qu'à la condition d'étaler devant les yeux de l'analyste un grand nombre d'items que l'on classera selon des critères divers jusqu'à ce que les pièces du puzzle forment une image cohérente. L'appréciation de quelques mots isolés ne produira aucun résultat.

C'est ce dernier aspect que je voudrais développer ici : la confection d'un corpus à des fins heuristiques consiste en un assemblage d'items, lexicaux en l'occurrence, qui partagent une propriété connue, mais dont on ignore le principe classificatoire qui permettra de découvrir leur secret. Chemin faisant, j'exposerai les difficultés artisanales liées au caractère spécifique des objets recherchés qui font qu'une extraction automatique, à partir d'un dictionnaire électronique, est sinon impossible, du moins très difficile et pénible.

4. Il n'y a pas de solution globale

Dans beaucoup de langues où des voyelles longues alternent avec des voyelles brèves, un seul principe contrôle le système entier. Le plus souvent, on rencontre en effet des voyelles longues en syllabe ouverte (tonique), alors que les voyelles brèves apparaissent en syllabe fermée (tonique). Selon que la voyelle est longue ou brève au niveau sous-jacent, on parlera de raccourcissement en syllabe fermée ou au contraire d'allongement en syllabe ouverte (tonique). L'islandais par ex. illustre le dernier cas de figure (p.ex. Gussmann 2002 : 157ssq).

En tchèque, rien de tel. Vue à travers la structure syllabique, cette langue semble présenter aussi bien des cas de raccourcissement en syllabe fermée que des exemples d'allongement en syllabe fermée (*sic*). Les féminins en *-a* par exemple sont bienséants : ils suivent en général le premier modèle : *kráv-a* - *krav* - *krav-ka* « vache NOMsg, GENpl, diminutif NOMsg »³. Mais le phonologue, ahuri, découvre aussi le mouvement inverse : les masculins à finale consonantique semblent au contraire allonger leur voyelle en syllabe fermée ou, si l'on veut, la raccourcir en syllabe ouverte : *nůž* - *nůž-ky* - *nož-e* « couteau NOMsg, ciseaux NOMpl, couteau GENsg ».

Il faut donc se rendre à l'évidence qu'il n'existe pas, en tchèque, de principe général et unique régissant la distribution des voyelles longues et brèves. Et, aussi, que le mécanisme qui contrôle la longueur n'est pas d'ordre syllabique. Il convient donc de rechercher des solutions paradigme par paradigme.

5. La longueur lors de la dérivation itérative

Le paradigme particulier que j'ai choisi afin d'illustrer le propos du présent article est celui des verbes itératifs. Rappelons qu'il ne s'agit là que d'un ensemble de données parmi d'autres qui suivent la logique gabaritique (les comparatifs d'adjectifs et d'adverbes, les diminutifs et les infinitifs font de même). Tous ont en commun le fait que les alternances de longueur sont déclenchées par une activité dérivationnelle (alors que les mêmes objets ne réagissent pas lors de la flexion).

Entrons donc, enfin, dans le vif du sujet. Lorsqu'on compare les verbes itératifs et leurs bases non-itératives à partir

3. Tous les exemples dans cet article sont donnés en orthographe tchèque, qui indique la longueur vocalique de façon systématique : une voyelle surmontée d'un accent aigu (*háček*) ou d'un petit rond (dans le cas de *ů*: *kroužek*) est longue. Elle est brève en l'absence de ces diacritiques. Une apostrophe suivant un *t* ou un *d* indique la palatalité de celui-ci (p. ex. *cit'ovat*).

Le corpus heuristique : un outil qui montre mais ne démontre pas

desquelles ils ont été dérivés, on observe des modifications de longueur dans les deux sens. Cet effet est illustré sous (1)⁴.

(1) le mystère itératif

a. allongement itératif

non-itératif	itératif	glose
sadit	sázet	planter (graine)
skočit	skákat	sauter

b. abrègement itératif

non-itératif	itératif	glose
cítit	cít'ovat	sentir
výšit	-vyšovat	élever, porter à hauteur

On peut s'assurer également que la cause de ces alternances est bien l'activité dérivationnelle : quoi qu'il arrive à un itératif ou à un non-itératif lors de la conjugaison, la longueur de sa voyelle radicale reste stable.

4. Il n'est pas nécessaire de s'attarder beaucoup, dans le cas d'espèce, à la question de savoir quelle catégorie est la source du processus dérivationnel, et quelle en est le fruit : il est évident que les itératifs sont le résultat d'une dérivation qui prend comme argument le verbe simple. Le lecteur observera, tout au long de l'exposé, que les voyelles qui présentent une alternance de longueur ont aussi, parfois, des timbres différents. Ceci ne va pas de soi, bien entendu, mais c'est monnaie courante en tchèque. Je reviendrai à ce phénomène en section 6. Enfin, les langues slaves possèdent un système de préverbes comparable à celui que l'on trouve dans les langues germaniques. Un tiret devant un verbe comme dans *-vyšovat* sous indique que celui-ci n'existe pas en isolation : il est seulement attesté avec un ou des préfixes, qui ne jouent pas de rôle dans la démonstration.

T. SCHEER

- (2) aucune variation de longueur lors de la flexion
a. allongement itératif

sadít	sázet	skočit	skákat
sad-ím	sáz-ím	skoč-ím	skák-ám
sad-íš	sáz-íš	skoč-íš	skák-áš
sad-í	sáz-í	skoč-í	skák-á
sad-íme	sáz-íme	skoč-íme	skák-áme
sad-íte	sáz-íte	skoč-íte	skák-áte
sad-í	sáz-ejí	skoč-í	skák-ají

a. abrègement itératif

cítit	cit'ovat	výšit	po-vyšovat
cít-ím	cit'-uji	výš-ím	po-vyš-uji
cít-íš	cit'-uješ	výš-íš	po-vyš-uješ
cít-í	cit'-uje	výš-í	po-vyš-uje
cít-íme	cit'-ujeme	výš-íme	po-vyš-ujeme
cít-íte	cit'-ujete	výš-íte	po-vyš-ujete
cít-í	cit'-ují	výš-í	po-vyš-ují

Comment se fait-il donc que la même opération dérivationnelle tantôt produit un allongement, tantôt un abrègement ? L'affaire semble mal engagée.

Cela dit, tout phonologue aura le réflexe de vérifier les facteurs habituels qui conditionnent la longueur vocalique : la structure syllabique et l'accent (*cf. supra*). Hélas, le tchèue s'en moque. L'accent élisant siège sur la première syllabe de chaque mot dans cette langue et étant par ailleurs immuable, toutes les voyelles radicales de toutes les formes (itératives ou non) et de tous les verbes sous (1) sont toniques. La dérivation itérative n'y change rien, et pourtant certaines voyelles s'allongent, alors que d'autres s'abrègent. Il en va de même pour la structure syllabique : elle ne joue visiblement aucun rôle puisque toutes les voyelles sous (1), qu'elles s'allongent ou qu'elles s'abrègent, se trouvent en syllabe ouverte aussi bien dans les formes itératives que dans les formes non-itératives.

En désespoir de cause, enfin, on peut vérifier si la modification de la longueur vocalique à l'itératif dépend du timbre de la voyelle en question. La réponse, ici encore, est négative : *sadit* produit *sázet* autant que *klidit* → *-klízet* « nettoyer », *hodit* → *házet* « jeter », *jezdit* → *jíždět* « se déplacer autrement qu'à pied », *pustit* → *pouštět* « laisser tomber » et *myslit* → *-mýšlet* « penser » et ainsi de suite.

A ce stade de l'analyse, on comprendra celui qui jette l'éponge : en dehors du fait que les alternances ont trait à la catégorie itérative et sont déclenchées par sa dérivation, on ne peut déceler aucune régularité.

6. Confection du corpus I – les écueils

Si malgré cela, contre vents et marées, on décide de ne pas abandonner la partie, alors il n'y a plus qu'une seule chose à faire : dans le brouillard et sans boussole, il faut multiplier les exemples. Le seul repère est l'itération : il s'agit de trouver autant de paires non-itératif - itératif que possible.

6.1. *L'extraction automatique : structure des dictionnaires et édition électronique en République Tchèque*

Se pose alors le problème suivant, particulier au corpus à construire. Il s'agit certes d'un corpus lexicographique, mais chaque item consiste en une paire de mots : il y a besoin des deux formes itérative et non-itérative afin de déceler une éventuelle alternance de longueur. Or les dictionnaires ne fonctionnent pas par paires. Du moins n'y a-t-il pas de renvoi réciproque entre les entrées itératives et non-itératives. Le dictionnaire académique en huit volumes qui fait autorité en tchèque (Havránek 1989) renvoie systématiquement au partenaire non-itératif depuis les entrées des verbes itératifs, mais n'en fait pas autant dans la direction inverse. Tel est également le comportement du dictionnaire qui faisait autorité jusqu'à la publication du dictionnaire académique, *i.e.* Trávníček (1952) (un seul volume), et du dictionnaire de référence pour l'orthographe (à l'usage des écoles) Filipec *et al.* (1994) (un seul volume). Néanmoins, ce principe doit être relativisé en fonction de la taille des dictionnaires : Trávníček (1952) et Filipec *et al.* (1994) (et *a fortiori* les autres dictionnaires, bilingues ou non) ne sont pas assez volumineux pour garantir une entrée autonome aux verbes itératifs qui souvent ne sont que la version itérative du verbe non-itératif et ne possèdent aucun contenu lexical propre. Le verbe *cit'ovat* par exemple n'apparaît pas dans la liste alphabétique de Filipec *et al.* (1994). Il faut aller le chercher dans l'article consacré à sa base dérivationnelle *cítit*.

Abstraction faite de cela, aucun des dictionnaires n'utilise un moyen biunivoque pour renvoyer de l'itératif à sa base dérivationnelle. Le renvoi se matérialise en général par « k », qui est une préposition et veut dire « vers, en relation

avec ». Or, cette préposition sert également à une multitude d'autres références croisées, si bien qu'il est impossible d'identifier les seules paires itératives en la ciblant. Pour ce qui est du renvoi à l'itératif au sein de l'article consacré au non-itératif (cas de *cítit - cit'ovat*), la situation est pire car l'itératif apparaît la plupart du temps en tant que simple sous-entrée sans que son caractère itératif ne soit spécifié.

Mais de toute façon ces considérations ne sont que littérature puisqu'une extraction automatique suppose évidemment que les dictionnaires en question soient disponibles sous forme électronique. Or tel n'est pas le cas, toujours pas en 2004. La version électronique de Filipec *et al.* (1994) a été publiée en 1997 (Leda 1997). J'avais commencé à construire le corpus depuis 1996, et croyais dans un premier temps pouvoir m'en servir. Il n'en était rien car le mécanisme de recherche livré avec le dictionnaire ne permet pas d'explorer le corps des articles : on peut rechercher des suites de caractères seulement dans les entrées, et la recherche est encore ici restreinte aux suites initiales de mot. J'ai réussi à extraire du CD la liste des entrées que je possède désormais sous forme informatique malléable, mais le texte des articles est perdu. L'édition électronique de dictionnaires en République Tchèque s'est arrêtée à Leda (1997) : on attend depuis l'informatisation du dictionnaire de l'académie qui, semble-t-il, a été entreprise depuis des années à Brno, mais n'aboutit pas.

En somme, donc, l'absence de lexicographie électronique ou son état anté-diluvien corrompt pour l'instant toute extraction automatique. Mais même au cas où il y aurait des conditions électroniques satisfaisantes, l'extraction automatique serait ponctuée de nombreuses difficultés liées à la structure des entrées lexicographiques.

6.2. Paires itératives synchroniques vs. diachroniques ?

Par ailleurs, la recherche qui doit être entreprise a une composante sémantique, chose qui est plutôt inhabituelle pour un corpus lexicographique. D'ordinaire le phonologue établit une liste de mots selon les seules propriétés de leur Signifiant. Or ici, c'est la dérivation itérative qui produit l'effet étudié. Par

conséquent, les deux mots des paires recherchées doivent être l'un le résultat de la dérivation itérative faite à partir de l'autre.

Cette condition n'est pas anodine car elle pose la question de la diachronie : il existe, en tchèque, des paires de verbes dont le Signifiant invite à penser qu'il s'agit de paires itératives. Or il n'en est rien, tout au moins en synchronie selon le jugement des locuteurs. Le verbe *hledět* par exemple veut dire « regarder » ; le verbe *hlídat* « garder, surveiller » n'en est pas l'itératif dans l'appréciation des locuteurs : il ne veut pas dire « regarder souvent ». Or il est évident que l'on regarde souvent lorsqu'il s'agit de surveiller. S'il n'y a pas de rapport itératif entre *hledět* et *hlídat* en tchèque moderne, nul ne peut douter sérieusement qu'il y en ait eu un dans le passé : *hlídat* est né par dérivation itérative à partir de *hledět*. L'itératif ayant par la suite acquis le statut de lexème, rompant ainsi le lien dérivationnel des deux unités, leurs Signifiés ont pu évoluer séparément. Je reviendrai aux « fausses » paires itératives plus bas en section 9, où leur liste sera également détaillée.

En somme, il s'agit donc de décider si le corpus à construire doit représenter la stricte compétence d'un locuteur moderne, ou s'il doit laisser la porte ouverte à des paires de verbes qui ont entretenu un rapport itératif, ou possiblement itératif, à une époque antérieure, mais l'ont perdu. Le premier réflexe (génératif, je pense, surtout) peut être de dire que le passé ne nous intéresse pas dans cette affaire dans la mesure où les alternances vocaliques que nous appréhendons sont produites par la grammaire synchronique. Or il y a de bonnes chances que rien ne soit plus faux que cela.

Nous touchons ici à une des questions les plus fondamentales de la phonologie des langues slaves en général, et de celle du tchèque en particulier : beaucoup d'alternances telles que les palatalisations ou impliquant la longueur vocalique sont non-naturelles, mais pourtant régulières. Non-naturelles, pour ce qui concerne le cas à l'étude, puisqu'elles engagent des voyelles longues et brèves de timbres différents : [] alterne avec [ii] (*jezdít* - *-jíždět*), [] avec [aa] (*skočit* - *skákat*), [u] avec la diphtongue [w] (*pustít* - *pouštět*). Or nous savons qu'il y a eu, entre le vieux tchèque et le tchèque

moderne, une sorte de Great Vowel Shift où les voyelles longues (mais non pas brèves) non-basses ont été rehaussées d'un niveau, les hautes passant en diphtongue : [] > [uu], [uu] > [w], [] > [ii], [ii] > [j] (p.ex. Lamprecht *et al.* 1986:105ssq)⁵. Aussi bien qu'à l'heure actuelle les brèves du vieux tchèque dont le timbre est resté stable alternent avec les longues dont le timbre a évolué : [] - [ii], [u] - [w] etc. Est-il raisonnable de penser que la grammaire synchronique ait assimilé et fait sienne la mutation de timbre que le temps a produite ? Et qu'elle l'ajoute à la gestion de la longueur propre ?

Mais revenons à notre sujet. Ce n'est certainement pas à l'occasion de la confection d'un corpus que cette question fondamentale sera résolue. Par ailleurs, nous n'avons pas intérêt à ouvrir davantage ici cette boîte de Pandore. A défaut de pouvoir trancher, il est possible au moins d'esquisser les conséquences pour le corpus à construire dans l'une et l'autre des perspectives. S'il était vrai que les alternances de longueur en tchèque moderne ne sont pas le résultat d'un calcul grammatical en ligne mais supposent deux entrées lexicales pour chaque paire longue - brève, alors des paires telles que *hledět - hlídat* sont pertinentes pour l'analyse. Car nous sommes alors à la recherche de l'effet que produit la dérivation itérative au moment précis où elle intervient. Peu importe qu'un itératif soit né l'année dernière, en 1960 ou il y a cinq siècles : si le mécanisme a été le même depuis que les Tchèques dérivent des itératifs, toute paire ainsi produite a droit de cité dans le corpus.

Si en revanche les alternances vocaliques quantitatives et qualitatives sont le fait d'opérations grammaticales synchroniques, aucune paire dont les membres n'entretiennent pas de relation itérative en synchronie ne doit être admise au sein du corpus.

5. Le mouvement est aujourd'hui accompli pour les voyelles d'arrière, mais a « du retard » pour les voyelles d'avant : les formes mutées sont encore rejetées par le standard académique (*spisovná čeština*). Dire *sejr* au lieu de *sýr* « fromage » (<y> transcrit le son [í]) ou *polívka* au lieu de *polévka* « soupe » est parfaitement courant (surtout en Bohême), mais réprimandé par la norme scolaire.

J'ai pris le parti ici de bien recueillir les paires dont le Signifiant laisse croire à une dérivation itérative même si elles n'entretiennent pas de rapport itératif en tchèque moderne. En revanche, je discuterai les deux types de paires séparément : le corpus de 247 paires qui est présenté ci-dessous contient les seules paires où le rapport itératif est synchroniquement décelable. Les paires du type *hledět - hlídat*, au nombre de neuf, seront examinées ensuite, et on verra qu'elles éclairent le corpus synchronique en un point précis.

7. Confection du corpus II – la pratique

Gardant à l'esprit ce qui précède, j'ai rassemblé un corpus de la manière la plus classique qui soit : dépouillement des listes de mots offerts par les grammaires et manuels d'apprentissage de la langue, ainsi que l'observation quotidienne au gré des conversations. Enfin, la pratique aidant, on finit par savoir quels sont les verbes qui ont des chances de produire un itératif – il suffit alors de confronter un locuteur natif avec une liste de verbes candidats. Celui-ci fournira les itératifs correspondants s'ils existent. Il va sans dire que le résultat qui sera discuté plus bas, à savoir un corpus de 247 paires itératives, est imparfait, amendable, ouvert et partiel. Si je devais juger de sa couverture par rapport à l'ensemble du lexique, je dirais, basé sur la seule expérience de chasseur de paires itératives, qu'il en représente à peu près 70%.

J'ai dans un premier temps fait abstraction du Signifié, ce qui a produit une première liste. Les paires dont les membres n'entretiennent pas de rapport itératif dans la langue moderne ont ensuite été identifiés, on y reviendra.

Lorsqu'on est amené à faire une liste mots, il y a toujours un principe classificatoire, fût-il alphabétique. Nul n'aligne 200 mots aléatoirement. Dans le cas des paires itératives, la question suivante se pose naturellement au fur et à mesure que l'inventaire croît : quel type de base permet de fabriquer un itératif, et quel est l'acabit d'un itératif typique ? En d'autres termes, existe-t-il des restrictions sur la source et le résultat de la dérivation ? C'est effectivement le cas.

Afin de comprendre le système à l'œuvre, il est nécessaire de connaître le système verbal tchèque. Un verbe tchèque, sans surprise pour une langue indo-européenne, est fait d'une racine, d'un élément thématique et de désinences. Le slave en général et le tchèque en particulier étant archaïques en comparaison à d'autres langues indo-européennes, les formes finies et non-finies des verbes peuvent être faites sur des thèmes différents. Ainsi, *plav-a-t* « nager, inf. » et toutes les autres formes non-finies (*plav-a-l* participe passé actif etc.), possède une voyelle thématique *-a-*, alors que les formes conjuguées présentent un thème en *-u/ -e* : *plav-u*, *plav-e-š* (« je nage, tu nages », etc. Aussi la tradition grammaticale tchèque est-elle partagée entre les deux écoles correspondantes : celle qui classe les verbes en fonction du thème des formes finies (p.ex. Havránek & Jedlička 1988 : 276ssq), et celle qui utilise le thème des formes non-finies à cette fin (p.ex. Trávníček 1948-49 I : 548ssq). On le verra plus bas, la seule classification pertinente, du moins pour ce qui concerne les itératifs, est la seconde. Elle distingue, à l'infinitif, six classes de verbes, qui apparaissent sous (3).

(3) classes verbales en tchèque		(marqueur infinitif : -t)	
classe	élément thémat.	exemple	glose
1	athém. zéro	nés-ø-t	porter
2	-ou-	tiskn-ou-t	imprimer
3	-e-	lež-e-t	être allongé
	-ě-	vid-ě-t	voir
4	-i-	pros-i-t	prier
5	-a-	děl-a-t	faire
6	-ova-	kup-ova-t	acheter

Lorsqu'on fait le tour des paires itératives, on s'aperçoit que seules les classes trois, cinq et six accueillent des formations itératives. Aucun verbe des classes une, deux et quatre ne semble avoir une valeur itérative. Le mécanisme dérivationnel a donc élu certaines classes de verbe pour verser son fruit, et refuse de se servir des autres classes afin de fabriquer des itératifs. Pourquoi ? Je ne sais répondre à cette question, ni maintenant ni à la fin de cet article ; mais c'est le

seul fait de la poser qui porte en lui le germe de la réponse à la question qui nous intéresse au premier chef : celle concernant la longueur vocalique. Car l'arrangement des paires itératives selon les trois classes verbales qui accueillent les itératifs produit le bon classement, *i.e.* celui qui ordonne les pièces du puzzle de façon à ce que l'observateur ait une chance d'entrevoir son fonctionnement.

8. Comment le corpus fait livrer leur secret aux itératifs

Voici donc sous (4) le tableau récapitulatif des paires itératives que j'ai rassemblées, au nombre de 247 et classées de la manière suivante (ce tableau est un condensé, davantage de détail est disponible en annexe). La colonne d'entrée concerne la base dérivationnelle qui peut comporter soit une voyelle brève, soit une voyelle longue. A partir de cela, la voyelle radicale peut donc ou bien s'abrèger à l'itératif au cas où elle est longue, ou alors s'allonger si est elle brève au départ. Les colonnes deux à quatre indiquent, pour chacune des trois classes verbales qui accueillent les itératifs, si la modification en question (abrègement ou allongement) a eu lieu ou non ($X > 3$ signifie qu'un verbe en provenance d'une classe indéterminée a fait son itératif dans la classe 3). Les lignes munies d'une flèche → indiquent qu'une modification a eu lieu et donne le nombre de paires itératives dans ce cas, alors que les lignes montrant un tiret répertorient les cas où la longueur de la base n'a pas été modifiée. Enfin, la dernière colonne illustre chacune des situations par un couple non-itératif – itératif.

Le corpus heuristique : un outil qui montre mais ne démontre pas

(4) arrangement des paires itératives selon les classes verbales qui accueillent les itératifs

		non-itératifs			itératifs			exemple
		X > 3	X > 5	X > 6	X > 3	X > 5	X > 6	
base	→ VV	49						sadit - sázet
	—	0						—
base brève	→ VV		55					skočit - skákat
	—		13					sedět - sedat
CVC	→ VV					0		—
	—					58		čistit - čist'ovat
base	→ V	2						vrátit - vracet
	—	3						hlásit - ohlášet
base longue	→ V		4					šlápnout - šlapat
	—		7					mávnout - mávat
CVVC	→ V					56		čítit - čít'ovat
	—					0		—
		54	79	114				
			total	247				

Il apparaît d'emblée que les chiffres ne sont pas distribués au hasard. La disproportion la plus spectaculaire est celle qui concerne la classe six : tout itératif qui y est fabriqué abrège la voyelle longue du verbe simple, alors qu'aucune base brève ne s'allonge. La distribution est parfaitement complémentaire. Pour ce qui concerne les itératifs des classes trois et cinq, l'exploitation numérique est moins aisée notamment pour une raison lexicale : on y trouve très peu de bases à voyelle longue. Je n'ai pu rassembler que 5 paires en classe trois, et 11 en classe cinq. Ces chiffres s'opposent à 49 (classe trois) et 68 (classe cinq) bases à voyelle brève. Je ne sais pas si cette disproportion cache un loup – il faut le croire si l'on met en regard l'équilibre de la classe six (58 paires à base brève, 56 à base longue). Ou alors, c'est l'insuffisance du corpus et de ses méthodes de confection, artisanales, qui produisent un mirage statistique.

Quoiqu'il en soit, en travaillant sur le corpus qui est à notre disposition, le faible nombre de bases à voyelle longue permet au mieux de supputer une tendance : moins de verbes s'abrègent qu'il n'y en a qui restent indifférents (2 sur 5 en classe trois, 4 sur 11 classe cinq). En revanche, le verdict numérique

concernant les bases à voyelle brève est sans appel : elles s'allongent massivement. C'est le cas de la totalité des verbes en classe trois (49), et de 55 sur 68 en classe cinq.

On observe donc que la tendance pour les bases à voyelle longue et l'indication claire des bases à voyelle brève s'accordent bien : les itératifs qui sont fabriqués en classes trois et cinq s'allongent, mais ne s'abrègent pas. Voilà un comportement exactement inverse de la classe six où l'on observe un raccourcissement systématique, mais pas d'allongement.

C'est cette complémentarité entre les classes trois et cinq d'une part et la classe six d'autre part qui livre le secret de la longueur vocalique itérative. En voici l'image.

- (5) effet opposé sur la formation itérative des classes 3 et 5
vs. 6
a. un allongement itératif se produit en classes 3 et 5.
b. un abrègement itératif se produit en classe 6.

A partir de cela, la question que tout linguiste se pose mène au trésor : si les classes trois et cinq produisent l'effet inverse de la classe six, quel est donc le point commun des premières, et qui les oppose à la dernière ? La réponse est évidente : les éléments thématiques des classes trois et cinq sont faits d'une seule voyelle (-e-/ -ě-, -a-), alors que la classe six est faite sur -ova- qui présente deux voyelles. Or ceci revient à dire qu'un gros élément thématique provoque l'abrègement (classe 6), alors qu'un petit élément thématique conduit à l'allongement (classes 3 et 5).

On conclut que le résultat itératif, dans tous les cas de figure, a le même nombre d'unités vocaliques, ou de *mores* en langage phonologique technique. Si une voyelle longue en compte deux et une voyelle brève une seule, alors *sázet* et *skákat*, produits par allongement, sont faits de trois mores. Il en va de même pour *cit'ovat* qui a été créé par abrègement. La constante, donc, qui vaut pour tous les itératifs de quelque provenance que ce soit, est le poids vocalique :

- (6) loi des trois mores
un verbe itératif pèse exactement trois mores.

Le corpus heuristique : un outil qui montre mais ne démontre pas

Il s'agit là d'une contrainte dite gabaritique. J'utiliserai cet adjectif ci-dessous, en attendant de plus amples explications en section 10. Le fonctionnement de cette contrainte est illustré sous (7).

(7) illustration : un itératif pèse exactement 3 mores

	base non- itérative	base + élém. thém. itératif	itératif = 3μ	nb
X>3	2μ sad-it	2μ saz-et	allong. sáz-et	49
	3μ hlás-it	3μ ohláš-et	— id.	3
X>5	2μ skoč-it	2μ skak-at	allong. skák-at	55
	4μ mávn-out	3μ máv-at	— id.	7
X>6	3μ cít-it	4μ cít'-ovat	racc. cit'-ovat	56
	2μ číst-it	3μ číst'-ovat	— id.	58
				228

Si une base courte fait son itératif dans une classe verbale à élément thématique bref (classes trois et cinq), le résultat ne pèse pas assez lourd : deux mores seulement. Afin de satisfaire la loi des trois mores itératives, la voyelle radicale est alors allongée (*sázet*, *skákat*). Au cas où le verbe simple est long, en revanche, le compte moraïque y est dès la concaténation et aucun ajustement ne se produit (*ohlášet*, *mávát*). Il se peut également que la concaténation produise des objets qui pèsent trop lourd. C'est le cas de verbes simples longs qui font leur itératif en classe six : le résultat pèse quatre mores et doit donc être diminué afin d'en arriver à trois (*cit'ovat*).

Les paires qui contreviennent à cette régularité sont les suivantes :

(8) les contrevenants : des itératifs qui pèsent 2 mores

	base non- itérative	base + élém. thém. itératif	itératif = 2μ ou 4μ	nb
X>3	2μ CVC-Vt	2μ CVC-et	— CVC-Vt	0
	3μ vrát-it	3μ vrác-et	racc. vrac-et	2
X>5	2μ sed-ět	2μ sed-at	— id.	13
	4μ šláp-n-out	3μ šláp-at	racc. šlap-at	4
X>6	2μ CVC-it	3μ CVC-ovat	allong. CVVC-ovat	0
	3μ CVVC-it	4μ CVVC-ovat	— CVVC-ovat	0
				19

228 des 247 verbes itératifs du corpus, soit 92%, se conforment donc à la règle selon laquelle un itératif pèse exactement trois mores. En revanche, 19 verbes, soit 8%, s'y soustraient. Dans ces conditions, affirmer que la régularité qui a été mise à jour a des chances de refléter la réalité linguistique n'est pas trop s'avancer.

9. Itératifs « diachroniques » et organisation du lexique

Revenons à présent aux paires de type *hledět - hlídat* qui ont toutes les propriétés physiques d'une paire itérative, mais dont les membres n'entretiennent pas de rapport itératif en tchèque moderne. J'ai pu identifier neuf paires de ce type.

(9) liste des paires sans relation itérative synchronique

a. CVC avec allongement

base	"itératif"	glose base	glose "itératif"
kalit	kálet	troubler, brouiller, tremper l'acier	souiller, déféquer, évacuer les excréments
kazit	pře-kážet	gâter, abîmer, détraquer	entraver, gêner, embarrasser, encombrer
kulit	koulet	rouler	rouler
ležet	léhat	être couché	coucher d'habitude, aimer coucher
hledět	hlídat	regarder	garder, surveiller
patřit arch	pátrat	regarder "patřit tváří tvář smrti" affronter la mort	enquêter, rechercher

b. CVC sans allongement

base	"itératif"	glose base	glose "itératif"
mihnout se	mizet	traverser en vitesse	disparaître
chybit	chybět	commettre une faute/ erreur, se tromper	manquer, faire défaut
stavit	stavět	arrêter, inhiber, constiper	s'arrêter

En examinant les Signifiés respectifs de ces paires, le rapport itératif, du moins possiblement itératif, est évident. Il est intéressant d'observer alors que tous les itératifs sous (9b) (mais

Le corpus heuristique : un outil qui montre mais ne démontre pas

non ceux sous (9a)) contreviennent à la loi des trois mores : la dérivation itérative en classe trois, à partir d'une base brève, ne produit pas d'allongement, si bien que les trois itératifs en question ne pèsent que deux mores. Parmi les paires où le rapport itératif n'est plus ressenti à l'heure actuelle, trois sur neuf, soit 33%, refusent donc d'obéir au mécanisme gabaritique. Malgré la pauvreté numérique d'un total de neuf paires qui enjoint prudence, on peut faire valoir un contraste important avec les 8% de contrevenants parmi les paires où le rapport itératif est synchronique. Il y aurait donc une relation entre la synchronicité de la formation itérative et la validité de la régularité gabaritique : si la dérivation est faite en temps réel à chaque fois qu'un locuteur a besoin d'exprimer un sens itératif, alors le résultat dérivationnel observe la loi des trois mores. Si en revanche l'itératif est de provenance lexicale et non pas dérivationnelle, il arrive beaucoup plus fréquemment qu'il se moque de la loi en question.

Ce fonctionnement ne surprend guère : nous le savons, la régularité gabaritique est explicitement une régularité *dérivationnelle* : elle impose un volume d'exactly trois mores aux seuls résultats de la *dérivation* itérative. A défaut de dérivation, il n'y a pas de contrainte gabaritique. Au cas où un mot serait né d'une dérivation itérative il y a longtemps, et qu'il ait été lexicalisé ensuite, si bien qu'aujourd'hui il occupe une entrée lexicale autonome et que les locuteurs s'en emparent par simple accès lexical sans procéder à aucune dérivation, il n'y a aucune raison pour laquelle ce mot devrait observer la loi des trois mores.

Autrement dit, on peut soutenir que tous les itératifs contrevenants, synchroniquement ou diachroniquement parlant, sont lexicalisés. Possédant leur propre entrée lexicale, ils échappent à la contrainte gabaritique. En revanche, ceux qui se conforment à la loi des trois mores naissent d'une réelle activité dérivationnelle synchronique.

Cette perspective est trop belle pour être vraie – aussi est-elle infalsifiable. Infalsifiable au moins par des moyens phonologiques. Les différents statuts lexicaux qu'elle suppose font la prédiction que la gestion en temps réel entre un itératif

gabaritique et un itératif non-gabaritique n'est pas la même. On peut donc espérer qu'une expérience psycholinguistique parviendra à matérialiser ce contraste. Ceci, cependant, va bien au delà de l'ambition du présent article.

Mais une autre solution est également envisageable : rien ne fonctionne plus synchroniquement. Aucun itératif ne naît plus d'aucune dérivation en temps réel. La formation des itératifs a été active dans une période antérieure, et les verbes qui en résultaient ont tous été lexicalisés. L'état que l'on voit actuellement serait alors le corps lexical d'antan sur lequel ne pèse plus aucune restriction gabaritique, et qui par conséquent évolue au gré du vent : certains items lexicaux, 8% pour l'instant, ont modifié la longueur vocalique de leur voyelle radicale qui, dans son état initial, observait la contrainte gabaritique. La masse des mots est inerte et maintient pour l'instant l'état primitif.

On pourra opposer à ce scénario le fait que les tchècophones sont toujours capables à l'heure actuelle de fabriquer des itératifs inédits : la dérivation itérative est vivante et productive. Il n'est pas aisé de trouver des paires itératives dont la jeunesse est certifiée ; un exemple sûr cependant me semble se trouver dans la paire *kliknout* - *klikat* « cliquer (ordinateur) ». La dérivation itérative est donc bien vivante en tchèque moderne. Mais cette paire rend aussi un autre témoignage : rien ne s'allonge plus. La base dérivationnelle non-itérative *kliknout* est brève, et l'itératif est fait dans la classe cinq avec un élément thématique bref, si bien que le résultat brut ne compte que deux mores : *klikat*. La contrainte gabaritique, était-elle active, imposerait alors un allongement, allongement qui n'a pas lieu. Il faudra donc conclure que, oui, la dérivation itérative est vivante actuellement, mais qu'elle n'impose plus aucune restriction gabaritique à son résultat. Ce qui nous ramène au scénario précédent : les deux membres des paires itératives qui attestent d'une contrainte gabaritique sont lexicalisés ; aucun de ces itératifs qui semble modifier la longueur de sa voyelle radicale ne naît plus par activité dérivationnelle.

10. Le mécanisme gabaritique

Après avoir montré qu'en tchèque une certaine catégorie de mots sémantiquement définie, les itératifs, impose (imposait) un volume vocalique précis, trois mores, il reste à situer ce phénomène dans le paysage phonologique général. *Supra* j'ai appelé cette contrainte « gabaritique ». Ce mot est choisi délibérément car il renvoie au fonctionnement typique qui fait la réputation des langues afro-asiatiques en général, et sémitiques en particulier. En effet, la grande différence qui, de l'avis général, départage le sémitique et la plupart des autres langues du monde (parmi lesquelles l'indo-européen) concerne le volume des mots : une racine sémitique a un format qui ne varie guère. Elle comporte normalement trois, parfois quatre consonnes, et exactement deux places vocaliques, dont chacune, mais non les deux, peuvent être non-prononcées. A l'inverse, rien de tel n'est connu en indo-européen : le nombre de consonnes qu'une racine peut héberger n'est pas limité, il n'y a de volume ni minimal, ni maximal. Le nombre de voyelles non plus ne fait l'objet d'aucune restriction particulière.

En arabe classique par exemple, il existe des mécanismes de « remplissage » s'il advient que le lexique ne fournit pas assez de matériel consonantique pour fabriquer une racine. Deux consonnes ne font pas le compte : une racine telle que \sqrt{rm} « jeter » ne pourra faire surface qu'à la condition d'insérer un glide : [ramay-a] « il a jeté ». Une autre stratégie est le dédoublement de la dernière consonne : \sqrt{md} « étendre » apparaît en surface en tant que [madd-a] « il étend » (Chekayri & Scheer 1996, sous presse). Il existe aussi des cas où les deux stratégies ont été appliquées à la même racine. En hébreu biblique par exemple, la racine \sqrt{hm} « agiter (liquide), troubler (qqn) » peut être réalisée en tant que hmm ou hwm (Frajzyngier 1979 : 2).

Certes, il n'est pas coutume en sémitique de rencontrer des contraintes qui portent globalement sur le volume *vocalique* maximal ou minimal d'une racine ou d'une forme grammaticale. Il existe seulement des contraintes sur des voyelles particulières, qui en réalité sont des morphèmes. Ainsi, dans le schème canonique radical $C_1V_1C_2V_2C_3$ de l'arabe classique, V_1

est longue lorsque le locuteur veut exprimer une action réciproque : *katab-uu* veut dire « ils ont écrit », alors que *kaatab-uu* signifie « ils ont écrit l'un à l'autre ».

Dans ce contexte, donc, le mécanisme tchèque présenté est bien d'ordre gabaritique puisqu'il impose, pour une forme morpho-sémantique donnée, un nombre d'objets précis. Il est singulier au sein du concert des mécanismes gabaritiques du fait que son exigence porte sur le volume vocalique et non consonantique. Ceci pose aussi la question du pourquoi : les modifications vocaliques en sémitique ont une fonction claire car elles constituent un morphème (p.ex. V_1 longue égale « sens réciproque »). Or le fait qu'un itératif pèse toujours trois mores au lieu d'en admettre un nombre variable n'a pas vraiment de fonction : les itératifs ne sont pas les seuls verbes qui pèsent trois mores. Dans le cas des autres catégories gabaritiques du tchèque qui ont été mentionnées en section 2, la dysfonctionnalité du mécanisme gabaritique est encore plus évidente : contrairement aux suffixes de classe, le caractère diminutif d'un mot est identifié sans ambiguïté par un suffixe particulier.

La gratuité de l'activité gabaritique en tchèque demande donc à être élucidée. En réalité, je pense qu'elle n'est que le miroir d'une autre particularité du tchèque : en sémitique, toutes les catégories de mots sont gabaritiques. Il n'existe pas de mot qui ait un nombre libre de consonnes ou de voyelles : c'est la raison pour laquelle on parle de « morphologie gabaritique » en sémitique. Or en tchèque, le gabarit ne pèse que sur des catégories sémantiques bien précises, parmi lesquelles les itératifs, les diminutifs, les comparatifs et les infinitifs. En dehors de celles-ci, rien ne contraint le volume vocalique d'un mot : un verbe non-itératif peut peser deux, trois, quatre ou cinq mores sans que la langue n'y voie aucun inconvénient.

Quoi qu'il en soit, l'enseignement typologique à tirer des faits observés est, justement, que les mécanismes gabaritiques ne constituent pas une propriété typologique des langues. On ne peut opposer des langues gabaritiques qui auraient une activité gabaritique, aux langues non-gabaritiques d'où les gabarits seraient absents. Une langue n'est pas

gabaritique ou non-gabaritique, elle est plus ou moins gabaritique. Les langues sémitiques le sont beaucoup, d'autres langues du tout, et le tchèque un peu. Cette perspective est également nourrie par des études qui suggèrent l'existence de mécanismes gabaritiques dans d'autres langues indo-européennes comme par exemple le portugais (Brandão de Carvalho 2003) ou l'allemand (Bendjaballah & Haiden 2003a, b).

11. La fonction du corpus dans la procédure de découverte

Le cas des itératifs tchèques est certes intéressant en soi, et pour la phonologie en général. Cependant, l'aspect qui intéresse le propos du présent article est autre : quel a été le rôle du corpus dans le cheminement qui a abouti au résultat final ? Chaque découverte scientifique a une histoire individuelle qui se joue entre le(s) personne(s) qui cherche(nt) et l'objet de la recherche. J'ai indiqué *supra* comment s'est déroulée celle entre les itératifs et moi. Il y a certes des façons fort variées d'appréhender un objet, et de conduire une recherche. Malgré toutes les bonnes recommandations de la méthodologie et de l'épistémologie, chaque cas demeure individuel, et l'histoire ne se reproduit pas, serait-ce seulement à cause de la dose de hasard qui ponctue le rythme des événements.

Malgré cet état de fait, je pense que le rôle du corpus dans le processus de découverte est sinon stable, du moins récurrent. Au départ il y a un problème, et ce problème est parfaitement indépendant de tout corpus, existant ou futur. Le linguiste s'interroge et commence à travailler avec les moyens du bord : quelques exemples qu'il a sous la main, une hypothèse de travail conditionnée par l'image générale qu'il a de la langue en question, de la théorie qu'il utilise et bien d'autres facteurs encore, qui sont ou ne sont pas « objectifs », et d'ailleurs peuvent parfaitement être d'ordre non-linguistique. Dans le cas des itératifs tchèques, les moyens du bord ne donnent aucun résultat : ils ne permettent même pas de formuler une hypothèse de travail. En revanche, le problème est là, et bien là.

Le corpus entre en scène à ce moment-là, à savoir lorsque la recherche est au point mort, et le linguiste prêt à jeter l'éponge. Il est utile ici d'appréhender non seulement la

chronologie des événements, mais aussi leur durée relative. Si dans le cas des itératifs tchèques les étapes qui ont conduit au point mort se sont faites en très peu de temps (en principe, une soirée suffit), la confection et l'exploitation du corpus se sont mesurées en années. Comme j'ai été réduit à rassembler le corpus de manière artisanale, je l'ai constitué au gré des grammaires auxquelles j'avais accès, et du matériel qui a été apporté par le vent quotidien. Je versais mes paires itératives dans un fichier sans me préoccuper de leur interprétation. Celle-ci n'a repris ses droits que lorsque j'ai estimé avoir assez de matériel en stock.

Il a alors fallu classer : on ne peut pas traiter une masse amorphe de plus de 200 items. Si nul n'échappe à la classification, selon quel principe classe-t-on ? La question des liens dérivationnels se pose alors de manière naturelle : est-ce que les bases dérivationnelles partagent une propriété ? Est-ce qu'il existe un lien entre une classe spécifique de bases et une classe spécifique d'itératifs ? Est-ce que n'importe quel verbe peut engendrer un itératif ? Ce sont ces questions qui produisent automatiquement le principe de classement qui s'avérera déterminant : l'identification des trois groupes verbaux dans lesquels on peut fabriquer des itératifs.

Le classement afférent en main, hélas, je n'ai toujours rien compris. Tout était là, mais il me manquait les lunettes sémitiques. Le corpus, ordonné comme il faut et prêt à rendre son secret, a alors disparu à nouveau sur mon disque dur pendant plusieurs années. Ce n'est que plus tard que je suis revenu aux itératifs parce qu'on m'avait demandé de faire un cours sur la phonologie du tchèque. C'est alors que j'ai compris le fonctionnement gabaritique.

Il est des éléments de cette histoire qui n'ont aucune généralité. Parmi ceux-ci se trouvent, je l'ai dit, la collecte de données artisanale et le fait que ma recherche, bien que visant à construire un corpus lexicographique, concerne des paires de mots dont les membres entretiennent un lien dérivationnel.

En revanche, le fait que le corpus entre en scène lorsque les moyens du bord refusent de parler me semble être une constante dans le processus de découverte. Il en va de même

Le corpus heuristique : un outil qui montre mais ne démontre pas

des vertus inhérentes au corpus qui ont été déterminantes dans le cas des itératifs : qui dit corpus dit classement. *A contrario*, il n'y a pas de classement avec trois exemples et demi. Il n'est peut-être pas toujours vrai que le premier classement est le bon. Mais nul n'échappe au classement, à un classement selon un critère X puis, si celui-ci ne permet pas d'avancer, selon un critère Y et ainsi de suite.

Lorsque le linguiste apprécie un corpus classé, et qu'aucun système caché ne lui apparaisse, ceci peut avoir trois raisons : 1) ou bien le corpus ne renferme aucun secret : il n'y a rien à trouver ; 2) ou bien le corpus n'est pas classé selon le critère qui permet à la réalité linguistique cachée de briller au grand jour ; 3) ou bien tout est là et arrangé de la meilleure façon qui soit, seulement le linguiste n'a pas mis les bonnes lunettes : le déclic ne se produit pas. Dans le premier et le dernier cas, le corpus et l'art de le manier ne pourront jamais rien. Dans le second, oui. Deviner l'arrangement des données qui mettra en lumière la réalité linguistique demande du flair, de l'expérience et, souvent, un peu de chance. C'est encore ici une composante du processus de découverte qui repose sur l'intelligence humaine ; elle peut et doit utiliser les moyens techniques offerts par l'informatique, mais ceux-ci ne pourront jamais la remplacer.

Au début et à la fin du processus de découverte se trouve donc le seul génie du linguiste face à un monde plus ou moins embrouillé : aucun corpus ne crée par lui-même de problème ; aucun corpus ne se substituera au linguiste pour débusquer le secret qu'il renferme. Le corpus intervient, chronologiquement parlant, entre l'activité initiale et finale qui supposent un cerveau humain, une théorie et de l'intuition – tout ce qu'une machine ne peut offrir. Le rôle du corpus est celui d'un conseiller : il suscite un regard nouveau sur les mêmes « faits ». Il peut brouiller ou débrouiller la vue – comme la théorie et l'intuition. L'art de manipuler des corpus consiste justement à trouver l'angle d'attaque qui jettera la bonne lumière

sur les données. La force du corpus est son injonction au classement. Corpus et classement sont consubstantiels⁶.

12. Conclusion

Les propriétés généralisables du cas étudié ont déjà été relevées dans la section précédente. Je voudrais ici compléter ces aspects stables par ceux qui au contraire varient. Les conditions particulières liées à la confection du corpus des itératifs, ainsi que la singularité de l'objet « paire itérative » ont été évoquées. C'est dire la diversité des situations propres à chaque projet d'analyse et à chaque projet de corpus.

On est en droit de déduire de cela qu'il n'existe pas, en phonologie ou ailleurs, de méthode clef en main décrivant « comment construire un corpus ». Cet énoncé est peut-être trop trivial pour qu'il soit d'ordinaire rendu explicite. S'il en fallait une preuve, il suffira de considérer qu'il existe des manuels de statistique, d'informatique et de toute autre sorte d'outil, mais qu'aucun livre n'explique comment il faut s'y prendre lorsqu'il s'agit de confectionner un corpus. Au mieux, trouve-t-on des ouvrages récapitulant les différentes façons d'annoter et de documenter les corpus textuels, écrits ou oraux, et proposant des standards en la matière.

Chaque corpus est une entité individuelle, construite et fonctionnelle. Un corpus n'est pas une fin en soi, jamais⁷. Personne, ni en linguistique ni ailleurs, n'a jamais fabriqué un corpus à la seule fin de son existence. Un corpus existe parce que quelqu'un a poursuivi un dessein. Or, poursuivre un dessein veut dire orienter son activité. On est souvent tenté de penser à « objectivité », « confrontation avec la réalité » ou encore

6. D'où l'importance, pour qu'un corpus puisse être éventuellement réutilisable par d'autres que son créateur, que soient consignés, dans un en-tête ou toute autre forme de commentaire annexe, non seulement les sources du corpus mais aussi les principes qui ont guidé son organisation.

7. Même FRANTEXT, aujourd'hui à la disposition de tout un chacun pour des projets de recherche les plus variés, a d'abord répondu à un objectif précis, à savoir la rédaction du T.L.F. Ceci explique à la fois ses forces et ses faiblesses actuelles.

Le corpus heuristique : un outil qui montre mais ne démontre pas

« empirie » lorsqu'on parle de corpus. Le corpus, en quelque sorte, serait l'éprouvette linguistique dont le jugement ne souffre pas d'exception. C'est faux. Car le corpus ne représente pas la réalité ; il représente la réalité de celui qui l'a construit, et du dessein qui a présidé à sa confection. Le corpus est déjà en lui-même orienté – c'est un objet théorique car il y a eu des choix qui ont été faits avant sa naissance – parfois d'ailleurs à l'insu de leurs auteurs⁸. C'est dire que chaque corpus est un objet unique.

Enfin, il est utile d'ajouter qu'à une échelle plus grande non plus, il n'existe pas d'autoroute de la découverte. La science ne peut être programmée car elle est le produit de l'intelligence humaine. Une vision naïve du processus de la découverte, répandue dans les milieux non-scientifiques, conçoit le scientifique comme un savant en laboratoire et en blouse blanche qui est aidé par de puissantes machines : il leur donne une matière première à analyser, qu'elles transforment en liquide distillé qui livre la réponse. Le succès est au rendez-vous au terme d'une longue série d'expériences et de mètres courants de chiffres produits par les machines. Lorsque la dernière addition de tous les résultats a été faite, la solution apparaîtra.

Or la science bien évidemment ne fonctionne pas comme cela : le positivisme, courant au 19^{ème} et au début du 20^{ème} siècle lorsqu'on croyait qu'il était une simple question de temps et de puissance des machines avant que la nature ne rende tous ses secrets, heureusement n'a plus cours. Une découverte est faite lorsqu'il s'est produit un déclic dans un cerveau humain. Ce déclic est imprévisible, en tout cas n'est pas

8. Même les très grands corpus échantillonnés, tels le Brown Corpus (http://clwww.essex.ac.uk/w3c/corpus_ling/content/corpora/list/private/brown/brown.html) pour l'anglais américain, le LOB Corpus (<http://helmer.aksis.uib.no/icame/lobman/lob-cont.html>) pour l'anglais britannique, le British National Corpus (<http://www.natcorp.ox.ac.uk/>) ou le Corpus National Tchèque (Český Národní Korpus, <http://ucnk.ff.cuni.cz/>), qui prétendent fournir à l'ensemble de la communauté scientifique un échantillon représentatif d'un état de langue, n'échappent pas totalement à cette caractéristique : le choix des textes, des dates d'attestation, des genres et sous-genres organisent la réalité en lui imposant des prérequis d'ordre théorique.

la conséquence mécanique de l'analyse de colonnes de chiffres. L'histoire des sciences et le témoignage de ceux qui y ont contribué montre que le moment crucial où les pièces du puzzle forment un tout arrive souvent bien longtemps après que les faits qui ont nourri la réflexion ont été connus. Et le déclic se produit partout sauf dans un laboratoire : sous la douche, en somnolant (c'est le cas de Karl Verner par exemple, inventeur de la loi germanique qui porte son nom), en contemplant un paysage à travers la fenêtre d'un train, en conduisant une voiture, en regardant la télévision etc.

L'intelligence humaine est aléatoire et capricieuse. Elle est maître d'œuvre de la découverte scientifique : à son début lorsqu'il faut identifier le problème, à la fin lorsqu'il s'agit d'interpréter les résultats de l'expérience (du corpus), et au milieu lorsque l'expérience (le corpus) demande à être arrangée de façon à permettre l'émergence de la réalité. Le corpus joue le rôle d'ingénieur dans ce processus : il dessine des possibles, bannit certains impossibles et rend ainsi la marge de manœuvre du linguiste moins large. Le corpus canalise le caprice du linguiste.

Que faut-il donc conclure ? Qu'il est vain et sans intérêt de produire des corpus en l'absence de projet d'exploitation. Oui. Que le point de vue crée l'objet. Certainement : cette formule saussurienne est on ne peut plus vraie lors de l'utilisation heuristique du corpus. Qu'il n'existe pas de corpus sans théorie. Assurément, ni sans théorie préalable à sa confection, ni sans théorie qui serve à son exploitation. Qu'il ne faut pas tenir compte d'autres projets d'exploitation éventuels lorsqu'on construit un corpus. Non. Rien de ce qui a été dit n'empêche le confectionneur de corpus de s'astreindre à produire un objet qui soit réutilisable à peu de frais.

Annexe

Cette annexe fournit les mêmes informations que le tableau (4), sauf qu'elle propose davantage de détail : les catégories sont ici ventilées en fonction de la classe verbale d'origine (toutes les classes, à l'exception de la sixième en *-ova-*, dérivent des itératifs) (colonne 1) et du timbre de la voyelle

dont la longueur est susceptible d'être modifiée (*i.e.* la voyelle radicale) (colonne 2). En colonnes 3 et 4, la paire itérative qui témoigne de la classe en question apparaît⁹. Enfin, le nombre de paires qui fonctionnent de la sorte est spécifié dans la dernière colonne. La liste complète des paires itératives peut être consultée dans Scheer (2004).

Il est à noter, aussi, que les paires itératives qui présentent une consonne syllabique à la place de la voyelle radicale ont été exclues du corpus. Les quatre paires suivantes sont dans ce cas : *trhnout* - *-trhovat*, *drtit* - *-drcovat*, *-hltit* - *-hlcovat*, *tvrđit* - *-tvrzovat*. La raison en est que le tchèque, contrairement au slovaque, ne fait pas la différence entre des consonnes syllabique brèves et longues. Nous sommes donc dans l'incapacité de savoir si la consonne syllabique des formes itératives a subi un allongement, un abrègement ou encore n'a pas varié.

Enfin, les tableaux qui suivent ne fournissent pas le Signifié des verbes témoin puisque cette information ne paraît pas indispensable et aurait nuit à la lisibilité. La liste intégrale des 247 verbes ainsi que leur signifié peut être consultée dans Scheer (2001b).

9. Les tableaux ci-dessous répertorient les verbes athématiques de la première classe tels que *pit* « boire » ou *lézt* « ramper », qui portent une voyelle longue, en tant que verbes à voyelle brève. La raison en est une autre contrainte gabaritique qui a été mentionnée dans le texte, à savoir celle qui pèse sur les infinitifs. Car en réalité, la longueur visible à l'infinitif des verbes de classe 1 est artificielle : elle est le produit d'un aménagement gabaritique. Les formes non-infinitives de ces verbes témoignent de cela : toutes ont une voyelle radicale brève, *pi-l* participe passé actif, *pij-i*, 1^{ère} sg présent etc. La restriction gabaritique qui frappe les infinitifs est exposée en détail dans Scheer (2001b, c, 2003).

(10) **X > 3 -et, -ět**

a. bases brèves : CVC

1. régulier : allongement itératif

dérivation	alternance	simple	itératif	nombre
1>3	i-í	pít	píjet	3
2>3	i-í	minout	míjet	1
3>3	e-í	hledět	-hlížet	1
	o-á	vonět	-vánět	1
4>3	a-á	sadit	sázet	10
	e-í	jezdit	-jíždět	2
	ě-í	dělit	-dílet	1
	i-í	klidit	-klízet	1
	o-á	hodit	házet	22
	u-ou	pustit	pouštět	6
	y-ý	myslit	-mýšlet	1
				49

2. contre la régularité : pas d'allongement itératif
pas de cas connus**X > 3 -et, -ět**

b. bases longues : CVVC

1. régulier : pas d'abrègement itératif

dérivation	alternance	simple	itératif	nombre
2>3	í-í	-bídnout	-bízet	1
4>3	á-á	hlásit	-hlášet	1
	ou-ou	trousit	-troušet	1
				3

2. contre la régularité : abrègement itératif

dérivation	alternance	simple	itératif	nombre
4>3	á-a	vrátit	vracet	1
	í-i	svítit	-svěcet	1
				2

Le corpus heuristique : un outil qui montre mais ne démontre pas

(11) **X > 5 -at**

a. bases brèves : CVC

1. régulier : allongement itératif

dérivation	alternance	simple	itératif	nombre
1>5	e-é	lézt	-lézat	3
	e-í	mést	-mítat	1
	o-á	moci	-máhat	1
	ø-í	dřít	-dírat	10
2>5	i-í	všimnout si	všímat si	6
	y-ý	poskytnout	poskýtat	2
	e-í	zapomenout	zapomínat	6
	e-ý	vyslechnout	vyslýchat	1
	a-á	chladnout	-chládat	2
	o-á	-hodnout	hádat	1
	e-ou	poslechnout	poslouchat	1
	ø-í	pnout	-pínat	10
2/3>5	e-é	lehnout, ležet	léhat	1
3>5	e-é	letět	létat	1
	i-í	vidět	vídat	1
	ě-í	běžet	-bíhat	1
	y-ý	slyšet	slýchat	1
4>5	o-á	skočit	skákat	2
	u-ou	mluvit	-mlouvat	1
5>5	ø-í	-slat	-sílat	3
				55

2. contre la régularité : pas d'allongement itératif

dérivation	alternance	simple	itératif	nombre
2>5	a-a	padnout	padat	2
	i-i	plivnout	plivat	1
	y-y	chytnout	chytat	1
	e-e	seknout	sekat	5
	ě-ě	běžet	běhat	1
	o-o	bodnout	bodat	2
	u-u	puknout	pukat	1
				13

X > 5 -at

b. bases longues : CVVC

1. régulier : pas d'abrègement itératif

dérivation	alternance	simple	itératif	nombre
1>5	ou-ou	tlouci	-tloukat	1
	ů-ů	růst	-růstat	1
2>5	í-í	líznout	lízat	2
	á-á	mávnout	mávat	3
				7

2. contre la régularité : abrègement itératif

dérivation	alternance	simple	itératif	nombre
2>5	á-a	sáhnout	sahat	1
	í-e	říznout	řezat	1
	á-a	šlápnout	šlapat	2
				4

(12) **X > 6 -ovat**

a. bases longues : CVVC

1. régulier : abrègement itératif

dérivation	alternance	simple	itératif	nombre
2>6	á-a	-přáhnout	-přahovat	4
	í-i	líznout	-lizovat	2
	í-e	říznout	-řezovat	1
	ý-y	dýchnout	-dychovat	1
2/4> 6	ou-u	stoupnout, -stoupit	-stupovat	1
3>6	á-a	sázet	-sazovat	3
4>6	á-a	chválit	-chvalovat	11
	í-i	stínit	-stiňovat	10
	í-ě	navštívit	navštěvovat	1
	ou-u	-loučit	-lučovat	14
	ú-u	úžit	zužovat	1
	ý-y	výšit	-vyšovat	4
5 > 6	á-a	skákat	-skakovat	5
				58

2. contre la régularité : pas d'abrègement itératif
pas de cas connus

Le corpus heuristique : un outil qui montre mais ne démontre pas

X > 6 -ovat

b. bases brèves : CVC

1. régulier : pas d'allongement itératif

dérivation	alternance	simple	itératif	nombre
1>6	i-i	řinout se	-řinovat se	3
	e-e	dechnout	-dechovat	1
	ø-y	hnout	-hybovat	1
2/3> 6	i-i	křiknout, křičet	-křikovat	1
2/4> 6	y-y	chytnout, chytit	-chycovat	1
4>6	a-a	tlačit	-tlačovat	15
	e-e	černit	-čerňovat	7
	ě-ě	měřit	-měřovat	4
	i-i	čistit	-čist'ovat	6
	o-o	hostit	-hošt'ovat	9
	o-a	lomit	-lamovat	1
	u-u	ručit	-ručovat	3
	y-y	sytit	-sycovat	4
				56

2. contre la régularité : allongement itératif
pas de cas connus

Références bibliographiques

La mention suivante à la fin d'une référence indique que celle-ci est disponible en ligne :

WEB : www.unice.fr/dsl/tobias.htm

Belić A. (1928). « Zur westslavischen Akzentlehre », *Zeitschrift für Slavische Philologie* 5 : 146-152.

Bendjaballah, S. & M. Haiden (2003a). « Templatic Architecture », *Recherches Linguistiques de Vincennes* 32 : 157-168.

Bendjaballah, S. & M. Haiden (2003b). « Templatic Inflection in German », *Penn Working Papers in Linguistics* 8 : 29-41.

Brandão de Carvalho J. (2003). « Templatic morphology in the Portuguese verb », in Trudel Meisenburg & Maria Selig (éd) *Nouveaux départs en phonologie : les conceptions sub- et suprasegmentales*. Tübingen : Gunter Narr, 13-32.

- Černý F. (1897-1900). « Studie o české kvantitě », *Listy Filologické* 24+27 : 24 : 343-354, 421-431; 27 : 17-22.
- Chekayri A. & Scheer T. (1996). « The apophonic origin of Glides in the verbal system of Classical Arabic », in Jacqueline Lecarme, Jean Lowenstamm & Ur Shlonsky (éd) *Studies in Afroasiatic Grammar*. La Hague : Holland Academic Graphics, 62-76. WEB.
- Chekayri A. & Scheer T. (in press). « The appearance of glides in Classical Arabic defective verbs », *To appear in Folia Orientalia*.
- Filipec J., Daneš F. e., Machač J. & Mejstřík V. e. (1994). *Slovník spisovné češtiny pro školu a veřejnost*. 2^e édition Praha : Academia.
- Frajzyngier Z. (1979). « Notes on the RRR stems in semitic », *Journal of Semitic Studies* 24 : 1-12.
- Garde P. (1976). *Histoire de l'accentuation slave*. 2 vols, Paris : Institut d'Etudes Slaves.
- Gussmann E. (2002). *Phonology : Analysis and Theory*. Cambridge : Cambridge University Press.
- Havránek B. & Jedlička A. (1988). *Česká mluvnice*. Praha : SPN.
- Havránek B., Belić J., Helcl M., Jedlíček A., Křístek V. & Trávníček F (éd) (1989). *Slovník spisovného jazyka českého*. 8 vols. Praha : Academia.
- Jagić V. (1894). « Mitteilungen zur Wiener Philologenversammlung », *Anzeiger für indogermanische Sprach- und Altertumskunde* 3 : 251-254.
- Kortlandt F. (1975). *Slavic Accentuation : a Study in Relative Chronology*. Lisse : de Ridder.
- Kortlandt F. (1978). « A history of Slavic accentuation », *Lingua* 44 : 67-91.
- Kortlandt F. (1978). « On the history of Slavic accentuation », *Zeitschrift für vergleichende Sprachforschung* 92 : 269-281.
- Kuryłowicz J. (1952). *L'accentuation des langues indo-européennes*. 2nd edition Wrocław 1958 : Ossolineum.
- Kuryłowicz J. (1968). *Indogermanische Grammatik, Band II : Akzent, Ablaut*. Heidelberg : Carl Winter.
- Lamprecht A. t., Šlosar D. a. & Bauer J. (1986). *Historická mluvnice Češtiny*. Praha : SPN.

Le corpus heuristique : un outil qui montre mais ne démontre pas

- Leda (1997). *Slovník spisovné češtiny pro školu a veřejnost. Elektronické vydání*. Praha : LEDA.
- Lehr-Spławiński T. (1917). *Ze studjów na dakcentem słowiańskim*. Kraków : Akademie Umiejętości.
- Lehr-Spławiński T. (1918). « O prasłowiańskiej metatonyi », *Prace Komisji język. Akad. Umiejętości w Krakowie* 2 : 18-24.
- Lehr-Spławiński T. (1926). « Zum westslavischen Akzent », *Zeitschrift für Slavische Philologie* 3 : 364-367.
- Mellet S. (1990). « Quelques réflexions sur l'exploitation statistique des données informatisées », *Les Etudes Classiques* 58 : 105-113.
- Nonnemacher-Pribić E. (1961). *Die baltoslavischen Akzent- und Intonationsverhältnisse und ihr quantitativer Reflex im Slavischen*. Wiesbaden : Harrassowitz.
- Pedersen H. (1905). « Die nasalpräsentia und der slavische akzent », *Zeitschrift für vergleichende Sprachforschung* 38 : 297-425.
- Sadnik L. (1959). *Slavische Akzentuation*. Wiesbaden : Harrassowitz.
- Šaur V. (1995). « K výkladu vývoje české vakalické kvantity », in Petr Karlík (éd) *Pocta Dušanu Šlosarovi. Sborník k 65. narozeninám*. Boskovice : Albert, 91-96.
- Scheer T. (2001a). « The Rhythmic Law in Czech : Vowel-final Prefixes », in Gerhild Zybatow, Uwe Junghans, Grit Mehlhorn & Luka Szucsich (éd) *Current Issues in Formal Slavic Linguistics*. Frankfurt am Main : Lang, 37-48. WEB.
- Scheer T. (2001b). The key to Czech vowel length. Communication présentée au colloque Formal Description of Slavic Languages 4, Potsdam 28-30 novembre 2001. Exemplier WEB.
- Scheer T. (2001c). Čeština ve Varšawě. Exemplier d'un cours de phonologie tchèque à l'université de Varsovie. WEB.
- Scheer T. (2003). « The Key to Czech Vowel Length : Templates », in Petr Kosta, Joanna Blaszcak, Jens Frasek, Ljudmila Geist & Marzena Żygis (éd) *Investigations into Formal Slavic Linguistics*. Frankfurt am Main : Lang, 97-118. WEB.

T. SCHEER

- Scheer T. (2004). « O samohláskové délce při derivaci v češtině », in Zdeňka Hladká & Petr Karlík (éd) *Čeština - univerzálie a specifika 5*. Praha : Lidové noviny, 224-239. WEB.
- Sedláček F. (1910). « K osudu praslovanských slabik dlouhých v češtině », *Listy Filologické* 37 : 22-41.
- Shevelov G. (1964). *A Prehistory of Slavic. The historical phonology of Common Slavic*. Heidelberg : Winter.
- Stang C. (1957). *Slavonic Accentuation*. Oslo : Norske Videnskaps-Akademie.
- Trávníček F. (1912). « K otázce české kvantity za raženou intonaci praslovanských dlouhých vokálů pod přízvukem », *Časopis pro moderní filologii a literatury* 2 : 3-9, 100-104, 196-201.
- Trávníček F. (1921). « De la quantité en tchèque », *Revue des Etudes Slaves* 1 : 204-227.
- Trávníček F. (1921). « Ke kvantitě mužských kmenů na -o-, -jo- a -u- v češtině », *Listy Filologické* 48 : 101-111.
- Trávníček F. (1921). « O kvantitě dvouslabičných aa - a jaa - kmenů s původním přízvukem na koncove a s praslovanskými dlouhými vokály », *Časopis pro moderní filologii a literatury* 7 : 14-18, 73-76, 150-152.
- Trávníček F. (1925). « K vývoji čes. kvantity », in Jiří Horák & Miloslav Hýsek (éd) *Sborník Prací věnovaných profesuro Janu Máchalovi k sedmdesátým narozeninám*. Praha : Klub moderních filologů, 349-357.
- Trávníček F. (1948-49). *Mluvnice Spisovné Češtiny. 2 vol.* Praha : Melantrich.
- Trávníček F. (1952). *Slovník Jazyka Českého*. Fourth edition Praha : Slovanské Nakladatelství.