# La conjugaison tchèque à l'usage des machines et des hommes : quelques réflexions

Ivan ŠMILAUER

Plidam-Inalco, Paris

**Résumé :** Cet article traite quelques questions relatives à la conjugaison tchèque dans une perspective de traitement automatique des langues et de l'enseignement du tchèque en tant que langue étrangère. Après une présentation succincte de la conjugaison, sont traités de manière plus détaillée le classement de types de conjugaison et la génération des formes du présent de l'indicatif et de l'infinitif. Un cadeau spécial pour Patrice Pognan vient à la fin.

**Mots-clés :** grammaire tchèque, conjugaison, TAL

## 1 Introduction

Ce serait porter de l'eau à la rivière (en tchèque *nosit dříví do lesa* « porter du bois à la forêt ») que de rappeler ici la richesse et l'originalité des travaux de Patrice Pognan. Menées de sa hauteur – et ce n'est qu'en second lieu que celle-ci se mesure en centimètres – d'un des meilleurs connaisseurs non natifs du tchèque, ses analyses traitent tantôt la phonologie et la graphie, la morphologie et la syntaxe, la synchronie et la diachronie. Elles mettent en évidence le rapport qu'entretient le tchèque avec d'autres langues slaves et proposent des descriptions linguistiques adaptées au traitement automatique, comme pour l'enseignement.

Le verbe tchèque fait partie des sujets de prédilection de Patrice. Ses travaux initiaux, consacrés à l'extraction automatique du verbe dans les textes médicaux (Pognan, 1972), comme ceux, parmi les plus récents, portant sur la calculabilité des formes verbales slaves à partir des racines historiques (Pognan et Panevová, 2013), témoignent de son intérêt constant pour cet objet linguistique.

Dans ma contribution rédigée en son honneur, je souhaite développer quelques réflexions sur la conjugaison tchèque. Je tenterai d'aborder ce sujet sous deux angles chers à Patrice : le traitement automatique et la didactique du tchèque en tant que langue étrangère. La complémentarité de ces deux points de vue, voire leur identité, du moins pour certains problèmes, est un fil conducteur des travaux de Patrice et représente, à mes yeux, l'une des caractéristiques les plus intéressantes de sa réflexion linguistique. L'exigence d'une rigueur descriptive permettant de rendre le tchèque accessible aux machines et la prise en compte des besoins de ces hommes et de ces femmes dont il a formé plusieurs générations dans ses cours de grammaire à l'Inalco, peuvent être, comme il nous le montre dans ses travaux de recherche et dans sa pratique d'enseignement, tout à fait compatibles.

## 2 Une vue d'ensemble sur la complexité formelle de la conjugaison tchèque

Avant la discussion de quelques points particuliers, rappelons qu'en comparaison avec le français, la flexion verbale tchèque peut paraître moins riche, du moins en considérant le nombre de temps et de modes. Toutefois, pour la catégorie de temps, la complexité morphologique inhérente aux systèmes verbaux des langues romanes y est largement compensée par la catégorie lexico-grammaticale de l'aspect : une partie des fonctions portées par les oppositions temporelles en français sont assurées en tchèque à l'aide de l'aspect perfectif ou imperfectif, faisant partie des traits lexicaux (identifiables hors contexte) du verbe conjugué.

Ainsi, on ne distingue que trois temps morphologiques : le présent (*dělám*), le prétérit (*dělal jsem*) et le futur composé, appelé également périphrastique ou analytique, réservé aux verbes imperfectifs (*budu dělat*). Pour l'expression du futur des verbes perfectifs, on emploie le présent (*udělám*). La forme préfixée du futur des verbes de mouvement (*pojedu*) pourrait être considérée comme un temps morphologique à part entière (on parle du futur synthétique), mais sur le plan flexionnel, leurs désinences personnelles sont identiques au présent. Le nombre de modes verbaux est également moins important, car il n'existe pas de subjonctif en tchèque – ses fonctions sont partagées entre l'indicatif et le conditionnel.

Les formes conjuguées simples sont l'indicatif actif du présent, l'impératif (*dělej*), le gérondif présent (*dělaje*), réservé aux verbes imperfectifs, et le gérondif passé (*udělav*), réservé aux perfectifs (*udělav*), les deux en retrait dans la langue actuelle. L'infinitif (*dělat*) fait également partie des formes simples. Parmi les formes complexes/composées/analytiques, on compte le prétérit (dont l'auxiliaire ne se matérialise pas à la 3$^{e}$ personne), le futur composé, le conditionnel (*dělal bych*) et toutes les formes de la voix passive (*je dělán*, etc.). Le plus souvent, les formes complexes contiennent deux éléments : l'auxiliaire et le participe (*dělal jsem, budu dělat, dělal bych, bude dělán*). Les formes passives à trois éléments (*byl jsem dělán, byl by dělán*) sont moins fréquentes, notamment celles créées à partir des verbes imperfectifs.

Pour l'analyse d'une proposition, il est tout à fait important d'être en mesure d'identifier les différents éléments d'une forme verbale complexe. À la différence du français, les auxiliaires peuvent être placés à une distance parfois assez importante de l'élé-

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

ment sémantiquement plein et rendre ainsi l'analyse plus difficile. De surcroît, pour les verbes essentiellement pronominaux tels que *smát se* « rire », les pronoms réfléchis *se/si* pourraient théoriquement être considérés comme faisant partie des formes verbales. Dans l'exemple (1), *jsme* et *smáli* sont donc les éléments d'une seule forme verbale composée et il serait possible de considérer le pronom *se* comme faisant partie de cette forme verbale complexe.

Exemple 1 :
*Včera jsme se s Patricem v hotelu Kúria v Banské Bystrici hlasitě smáli.*
Hier, à l'hôtel Kúria à Banská Bystrica, nous avons ri fort avec Patrice.

Rappelons ici un fait important qui joue en faveur de la calculabilité de la conjugaison tchèque à partir des formes isolées, hors contexte : les pronoms personnels ne sont pas obligatoires. En règle générale, la forme conjuguée du verbe suffit à elle seule pour identifier la personne.

Le répertoire des éléments participant à la conjugaison d'un verbe contient neuf types de formes qui peuvent être divisées en formes personnelles et impersonnelles (à caractère nominal). Parmi les formes personnelles figurent le présent de l'indicatif actif qui peut, pour le verbe *být* « être », avoir la fonction de l'auxiliaire, tout comme son futur (*budu*) et l'ancienne forme de son aorist, réservée aujourd'hui au conditionnel (*bych*). La dernière forme simple personnelle est celle de l'impératif. Les formes simples impersonnelles sont l'infinitif, le participe passé/actif (*byl*) qui s'accorde en genre et en nombre avec le sujet, tout comme le participe passif (*dělán*) et les gérondifs présent et passé.

La connaissance active des formes de l'infinitif, de l'indicatif du présent, des auxiliaires, du participe passé et de l'impératif fait partie des compétences exigées dans le cadre des niveaux A1-A2 d'après le CECRL (Cvejnová *et al*, 2014), bien évidemment uniquement pour un ensemble de verbes les plus fréquents et/ou les plus utiles.

Rappelons enfin que la définition des différentes classes de conjugaison est articulée essentiellement autour de la notion du *thème* – le morphème suivant la racine et précédant le morphème flexionnel. Deux séries de thèmes se démarquent : les thèmes du présent (*-e-, -ne-, -je-, -í-, -á-*), propres à l'indicatif présent, l'impératif et le gérondif présent ; et les thèmes de l'infinitif (*-Ø-, -nou-, -e-/-ě-, -i-, -a-, -ova-*) que l'on trouve avec quelques modifications aussi dans les formes des participes (on parle aussi du radical du passé). La mise en avant de l'une ou de l'autre série est le facteur principal distinguant les différentes classifications de la conjugaison tchèque, qui doivent, pour être exhaustives, prendre également en compte d'autres facteurs, non paradigmatiques, comme les propriétés morphonologiques de la racine. Dans les pratiques scolaires et linguistiques actuelles, c'est essentiellement le classement d'après le radical du présent qui domine. Il consiste en 5 classes (1re *-e,* 2e *-ne,* 3e *-je,* 4e *-í,* 5e *-á*) qui comportent 14 sous-classes (*nese, bere, peče, maže, umře ; tiskne, mine, začne ; kryje, kupuje ; prosí, trpí, sází ; dělá*), définies d'après des critères supplémentaires.

La racine peut être identique dans toutes les formes verbales, mais elle peut présenter, ce qui se produit pour un grand nombre de verbes, et notamment ceux les plus

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

fréquents, des cas d'allomorphie (p. ex. *br-át – ber-e*), résultat des alternances et des évolutions parfois très anciennes, ou plus rarement, de supplétisme (p. ex. *j-ít – š-el*).

Cette mise en évidence de la composition des formes conjuguées permet de distinguer deux domaines auxquels un système d'analyse et/ou de génération, qu'il soit automatique ou humain, doit se confronter : le traitement (génération/analyse) des éléments simples, qu'ils soient employés dans les formes verbales simples ou complexes, et la connaissance liée à la formation et l'emploi des formes verbales complexes, composées de plusieurs formes, pas forcément attenantes au sein d'une proposition et dont le traitement relève en partie de la syntaxe de surface.

Je n'aborderai ici que la première problématique qui concerne l'analyse et la génération des formes simples relevant du plan purement morphématique, assez complexe. Je vais évoquer quelques questions relatives aux classes de conjugaison, notamment en vue d'une génération automatique au sein d'un système basé sur des règles, et en vue de leur enseignement.

## 3 Quelle classification pour quel public ?

Il existe une multitude d'approches que l'on peut observer dans la formulation des présentations de la conjugaison tchèque. Les disparités sont dues aux divers choix opérés sur les critères morphologiques structurant la classification et la profondeur et l'exhaustivité de l'analyse adjacente. Une des raisons principales des divergences peut être vue dans le public auquel telle ou telle présentation est destinée.

En principe, on peut distinguer quatre types de destinataires d'une description grammaticale : un linguiste, un locuteur natif, un apprenant étranger ou un système de traitement automatique. Les différentes descriptions s'adaptent, de manière plus ou moins volontaire, aux spécificités (et aux attentes) de chaque public.

En laissant du côté un discours purement scientifique, il est clair que la formulation d'une classification à destination du locuteur natif peut s'appuyer sur sa compétence linguistique. Celle-ci lui permet d'utiliser le système sans avoir recours à une classification explicitement formulée. En règle générale, il est en mesure, à partir d'une forme verbale quelconque, de produire le paradigme tout entier ou du moins sa partie la plus vivante.

Pour un locuteur natif, il s'agit donc essentiellement de lui proposer une aide en cas d'hésitation, dès qu'il a besoin de s'adresser à une référence. Pour la langue tchèque, cela peut concerner notamment la relation, assez complexe, qui existe entre la langue standard et le registre « commun » (*obecná čeština*) ou l'emploi des formes qui sortent de l'usage, étant remplacées progressivement par leurs équivalents perçus comme plus « réguliers ».

Pour donner un exemple d'une telle approche, citons la publication de Pravdová et Svobodová (2019) qui s'inspire des sollicitations récurrentes des experts de l'Institut de la langue tchèque de l'Académie des sciences tchèque. Elle cible les phénomènes jugés « difficiles ». Pour la conjugaison, les points traités sont notamment la formation du participe passé des verbes en *-nout* qui peuvent ou non conserver le suffixe de l'infinitif,

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

la distribution des désinences de la 1re personne du sg. *-i* et *-u* et de la 3e personne pl. *-í* et *-ou* précédées d'une consonne molle, la distribution des désinences de la 3re personne du pl. du type *sází*, et une dizaine d'autres points qui posent problème aux Tchèques.

En plus de sa fonction pratique de référence, il est nécessaire de souligner le rôle que joue une description grammaticale dans la consolidation de la compétence linguistique des locuteurs natifs, par l'éducation scolaire. C'est en partant des critères didactiques que le classement basé sur le radical du présent avec cinq classes prévaut dans les grammaires actuelles à la classification par le radical de l'infinitif à six classes, pourtant, par définition, plus détaillée à la base.

De l'autre côté, le besoin qu'un apprenant étranger puisse avoir d'une classification de conjugaison, du moins dans les phases initiales de son apprentissage, est d'une nature tout à fait différente. Dès qu'il dépasse le stade de simple mémorisation de formes toutes faites qui peuvent éventuellement lui suffire pour une communication rudimentaire, il a besoin de règles pour être lui-même en mesure de produire les formes fléchies. Les règles de grammaire sont l'arsenal élémentaire de toute méthode d'apprentissage, mais on entre ainsi également sur le champ de la modélisation formelle, indispensable au traitement automatique.

L'analyse d'une forme peut être effectuée à l'aide d'une segmentation morphématique basée sur une description statique. Pour donner un exemple, la description algorithmique de la conjugaison tchèque mise en place par Osolsobě (2014) vise l'analyse des formes et l'identification de la classe de conjugaison à l'aide des opérations nécessitant la connaissance préalable de toutes les formes paradigmatiques. La génération, en revanche, doit s'appuyer sur des algorithmes qui prennent à l'entrée les différentes composantes et qui fournissent à la sortie la forme fléchie correspondante. Les différences existantes entre les radicaux du présent et de l'infinitif, avec des motivations parfois très archaïques, sont la matière première à l'articulation des telles règles concernant les alternances et l'ajout des suffixes flexionnels.

En lien direct avec la conjugaison tchèque, c'est notamment Komárek (2006), en se référant aux travaux de Jakobson, qui pose la distinction utile entre les classifications descriptives, statiques, et les systèmes génératifs, dynamiques, qui proposent une hiérarchie ordonnée de règles permettant la génération des formes conjuguées. C'était également à ce que pensait Vey (1945) lorsqu'il évoquait la capacité d'une classification de conjugaison d'offrir à l'apprenant un outil pour le « rendement » de formes fléchies, applicable avec le moindre effort.

Aujourd'hui, où il est possible d'avoir recours aux outils fiables qui présentent le paradigme entier de quelconque forme verbale, basés sur des listes complètent de formes, l'on peut se demander quelle est la finalité d'un tel système de règles ? En dehors de l'utilité indéniable de certaines règles pour l'enseignement, ils se montrent indispensables pour le diagnostic des formes erronées, éventuellement produites par l'apprenant, à l'image du diagnostic élaboré pour la déclinaison au sein de la plateforme CETLEF.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

En effet, si nous voulons (et c'est un des projets actuels que mène Patrice), construire des dispositifs d'apprentissage linguistique adaptatifs, nous avons, par exemple, besoin de déterminer automatiquement que les formes comme *kupovám* au lieu de *kupuju* ou *brál* au lieu de *bral* sont des erreurs calculables, nécessitant une intervention didactique ciblée. Pour la première forme, le parcours de l'apprenant serait orienté vers la révision des classes de conjugaison dont l'infinitif est terminé par *-at*, pour la seconde vers le rappel des alternances de longueurs dans les radicaux des participes passés.

## 4 Quelle forme devrait être considérée comme élémentaire ?

La question élémentaire d'ordre morphologique est de savoir quelle forme peut être utilisée comme le point de départ d'une classification destinée à formuler les règles de générations.

La classification par le radical de l'infinitif est entrée dans l'usage avec Dobrovský, on le retrouve chez Gebauer, dans sa grammaire historique comme dans les manuels scolaires de tchèque contemporain, dans les travaux de Trávníček et de manière plus récente dans *Příruční mluvnice češtiny* (Karlík *et al.*, 1996).

La présentation par le radical du présent apparaît pour la première fois dans les travaux du slaviste Leskien et devient la norme, pour la grammaire historique, avec notamment Lamprecht *et al.* (1977), et dans les grammaires et les manuels de langue actuelle avec les travaux de Mazon (1931), Havránek et Jedlička (1963), Poldauf et Špruňk (1968), Šmilauer (1972), Čechová *et al.* (2000), Cvrček *et al.* (2010) ou Štícha *et al.* (2013). Ces travaux servent de base pour la majorité des méthodes à destination des étrangers, procédant à des simplifications plus ou moins importantes.

Néanmoins, deux grammaires récentes dédiées à ce public particulier, rédigées par Naughton (2005) et Esvan *et al.* (2019), proposent le classement articulé autour de la forme de l'infinitif.

La grammaire de Naugthon prend pour départ le classement traditionnel, mais le présente en inversant l'ordre des radicaux (infinitif d'abord) et l'ordre de présentation : type 5 *-at / -á*, type 4 *-it, -et / -í*, type 3 *-ovat / -uje*, 2 *-nout / -ne*, 1 *-st, -zt, -ct / -e*. Chacun de ces types est défini par un ensemble de verbes dont l'indicatif présent est formé de manière régulière selon les classes habituelles. À chaque type s'ajoutent les verbes à radical monosyllabique présentés comme irréguliers dont la terminaison de l'infinitif ne correspond pas aux types définis, mais qui forment le présent de la même manière.

Chez Esvan *et al.*, les classes de conjugaison sont divisées en 4 groupes « réguliers », définis selon la forme de l'infinitif (*-at, -ovat, -nout, -it/-et/-ět*) et 2 groupes de verbes considérés comme irréguliers : ceux à l'infinitif terminé par *-at* avec le présent en *-e* et un second groupe contenant tous les verbes avec un thème long (*-át, -ást, -ést, -éct, -ět, -ít, -out*) et autres.

Dans le cas des grammaires d'apprentissage, la raison d'introduire la conjugaison par la forme de l'infinitif semble évidente. L'infinitif est la forme retrouvée dans le dictionnaire avec une fonction toute particulière de représenter l'unité lexicale dans

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

l'ensemble de ses formes. C'est l'infinitif qu'un apprenant est censé conjuguer dans le cadre des activités d'apprentissage telles que le thème grammatical. Pourtant, étant donné ses fonctions dans des contextes moins artificiels, la position de l'infinitif dans la structuration du lexique mental est plus discutable.

Les premiers énoncés authentiques qu'un apprenant rencontre, ainsi que les phrases-exemples proposées dans les manuels, contiennent essentiellement des verbes conjugués (par ex. *Co to je ? Děkuju. Prosím. Jak se řekne… Co znamená…*, *etc.*). L'infinitif n'est rencontré que par la forme théorique, potentielle. Il est vrai que son apprentissage hors contexte peut participer à la construction de la compétence métalinguistique de l'apprenant. Ce n'est qu'avec l'apprentissage des verbes modaux ou du futur composé, que l'infinitif prend réellement sa place dans des énoncés construits.

Enfin, comme le rappelle Hronová (2009), l'infinitif ne convient pas, car certaines langues maternelles des apprenants du tchèque, comme le bulgare ou l'arabe, ne possèdent simplement pas de forme qui lui équivaut et les verbes sont représentés par une forme du présent de l'indicatif. La tradition exige la même présentation lexicographique pour le latin.

## 5 Quelles sont les règles de génération à mettre en place ?

En partant de la constatation de l'apparition plutôt tardive de l'infinitif dans l'apprentissage, nous pouvons supposer qu'il est plus urgent de maîtriser la conjugaison à l'indicatif du présent, avant de traiter le lien qu'entretient celui-ci avec l'infinitif, puis avec l'impératif, et le lien entre l'infinitif et le participe passé. Faute de place, je n'évoquerai ici que les deux premiers processus, et cela en me limitant sur l'essentiel.

### 5.1 Le présent de l'indicatif

En préliminaire, il est important de mentionner une question d'ordre théorique ayant la répercussion sur la calculabilité d'une forme conjuguée à l'indicatif présent : quel statut accorder aux *thèmes* historiques ? Celui d'un morphème autonome qui a une place dans la segmentation de la forme conjuguée (*nes-e-š, nes-e,* voire *nes*-Ø-*u* comme chez Štícha *et al.* 2013) ? Accepter que dans la perspective synchronique, ils forment un amalgame avec les morphèmes purement flexionnels (*nes-eš, nes-u*) ? Dans tous les cas, la seconde option me semble plus intuitive et plus facile à traiter sur le plan algorithmique.

Il s'agit donc tout d'abord de fournir des règles pour créer une forme de l'indicatif présent à partir d'une autre forme de l'indicatif présent du même paradigme, par exemple *čte* à partir de *čtu.*

L'examen des désinences personnelles montre qu'il est possible, à part quelques exceptions, d'enlever la désinence et de la remplacer par une autre appartenant au paradigme correspondant, donc *-u, -eš, -e, -eme, -ete, -ou* pour les trois premières classes, *-ím, -íš, -í, -íme, -íte, -í* pour la quatrième et *-ám, -áš, -á, -áme, -áte, -ají* pour la cinquième.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

L'introduction des variantes *-i* et *-í* (*děkuji,* voire *hraji, píši*, etc.), dont la distribution relève des questions de registre, peut être, à mon avis (Patrice ne sera pas d'accord…), repoussée à un stade ultérieur. Cela permet de simplifier l'acquisition des désinences dans les premières phases de l'apprentissage sans introduire de distinctions entre les trois premières classes, en parlant tout simplement des verbes de type *-e*.

De la même manière, la désinence *-ejí*, et donc une introduction effective de la distinction entre les types *trpí* et *prosí* d'un côté et *sází* de l'autre, n'est pas nécessaire au début (en tous cas, les deux variantes sont jugées comme équivalentes depuis la réforme de 1993) et peut être présentée seulement avec l'impératif, où le type *sází* se démarque effectivement (*sáz-ej* contre *pros* et *trp*).

Les seuls verbes pour lesquels cette génération nécessite des règles supplémentaires sont le verbe *být (js-em* mais *j-e)*, *chtít* (*chc-i, chc-eš* mais *cht-ějí*), *jíst* (*j-ím* mais *jed-í*) et *vědět* (*v-ím* mais *věd-í*). Il faudrait également traiter le verbe *stonat* qui est le seul à avoir le graphème *-ě-* dans ses désinences (*stůn-ěš* contre *-eš* pour l'ensemble des autres verbes de type -e).

Avec ses règles, on dispose d'un moyen fiable de produire l'ensemble des formes de l'indicatif présent en disposant du radical présent du verbe en question. Pour la classification, il peut être donc suffisant de commencer par l'introduction des trois types définis à l'aide des désinences du présent : *-e*, *-í* et *-á*. Une solution qui va encore plus loin dans la simplification peut être trouvée chez Kastler (1986) qui propose de distinguer deux types de conjugaison définis uniquement en fonction de la terminaison de la première personne du singulier : *-u* et *-m*.

### 5.2 Le présent de l'indicatif et l'infinitif

La génération peut être effectuée dans les deux sens (le présent de l'indicatif <> l'infinitif). En considérant ce qui a été dit plus haut sur la prévalence du présent dans les premières phases de l'apprentissage, les règles de formation de l'infinitif à partir de cette forme pourraient être proposées en premier lieu.

Cette tâche ne serait relativement aisée que pour les verbes de type *dělá – dělat* (avec un traitement particulier pour le verbe *mít* et ceux dont l'infinitif est terminé par *-át*), le type *kupuje – kupovat* et la 2$^{e}$ classe (avec un traitement à part notamment pour *začne – začít, řekne – říct, stát – stane*).

Pour les verbes de la 4$^{e}$ classe, les choix entre les formes en *-it* et *-ět/et* (*chodím – chodit* mais *vidím – vidět*) demanderai l'introduction des listes de verbes avec l'une ou l'autre terminaison, sans parler du verbe *jím – jíst* et les verbes en *-át* (*spím – spát, stojím – stát, bojím se – bát se*) et leurs formes préfixées.

La 1$^{re}$ classe, présentant des nombreuses alternances entre les formes du présent et l'infinitif, serait la plus complexe à traiter. Étant donné qu'il s'agit d'une classe improductive, composée d'un ensemble fermé de verbes, la solution la plus économique serait certainement de lister tout d'abord les différents couples de radicaux, avant de mettre en place des règles basées, pour la plupart, sur la diachronie (approche chère à Patrice).

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

La génération inverse – l'infinitif vers le présent de l'indicatif – présente quelques avantages par rapport à la solution ci-dessus. Tout d'abord, il est plus facile de définir les groupes de verbes dont la conjugaison ne sera pas effectuée par une simple substitution du suffixe de l'infinitif sans incidence sur la forme de la racine. Il s'agit des infinitifs dont le *-t* final est précédé d'une voyelle longue *-í-, -ý-, -á-* (en enlevant éventuellement le préfixe qui peut raccourcir cette voyelle) ou de la diphtongue *-ou-* et par les consonnes *-c-, -s-* et *-z-*. Les seuls verbes dans ce groupe qui forment le présent de manière régulière sont les verbes *mít* (*m-ám*), *dát* (*d-ám*) et *znát* (*zn-ám*) prenant les désinences de la 5e classe, et le type *minout (min-e).* Pour le reste des verbes, il est forcément nécessaire de spécifier le radical du présent, différent par rapport à l'infinitif, et donner éventuellement des règles pour les alternances, dont la plus simple est celle de longueur pour *nést – nese, vézt – veze.*

Les infinitifs les moins ambigus par rapport à leur classe de conjugaison sont ceux terminés par :

*-it* : substitution par *-í,* tous les verbes,

*-et* : substitution par *-í,* tous les verbes sauf *jet – jede*,

*-ět* : substitution par *-í,* sauf *vědět – vím* et aussi *spět – spěje* avec les verbes *přimět, skvět, chvět, pět* qui ne font pourtant pas partie du vocabulaire de base,

*-ovat* : substitution par *-uje,* à part *chovat – chová* et si besoin *kovat – kove* ou *ková*)

*-at* : *cf.* ci-dessous.

La terminaison *-at* demande un traitement plus complexe. Bien évidemment, les verbes les plus représentés sont ceux de la 5e classe, formée par contraction *-a-je-* > *-á-* (*dělat – dělá*). À cela, il faut rajouter les verbes de la 1e classe du type *mazat – maže* et *brát – bere* comme *chápat – chápe, ukázat – ukáže, plavat – plave,* etc. Il est particulièrement important de traiter la préfixation qui raccourcit la voyelle finale des verbes appartenant à d'autres classes comme *přespat – přespí* et *dostat – dostane.*

Les alternances de quantité vocalique dues à la préfixation peuvent présenter un problème lié à la génération à partir de l'infinitif. Sans avoir de connaissance approfondie sur ce procédé morphologique, un apprenant ayant appris la conjugaison du verbe *brát* peut se retrouver démuni devant le verbe *sebrat* et le considérer comme un verbe de la 5e classe. Un système de génération automatique doit forcément rendre compte de cette alternance qui n'est liée qu'aux verbes terminés en *-át,* à part ceux de la 3e classe (*hrát, přát*, etc.) ou le *-á* long résulte de l'ancienne contraction. Un cas intéressant où la préfixation n'a pas d'incidence sur la forme du radical, mais modifie le type de conjugaison, sont les deux exemples bien connus de *krást – krad-e* mais *u-krást – u-krad-ne* et *lézt – lez-e* et *nalézt – na-lez-ne.*

## 6 Comment ordonner la présentation des différentes classes ?

Les classements traditionnels présentent d'abord les verbes les plus irréguliers de la 1re classe ce qui peut paraître assez décourageant pour un apprenant, comme le constate par exemple Kastler, 1986, p. 145. En effet, pour des raisons liées à la diachronie, les

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

classes régulières et/ou productives ne sont positionnées qu'en queue du classement. Nous avons vu que Naughton et Esvan *et al.* adaptent cette présentation dans le but de mettre en valeur ces types avant de traiter les verbes de la 1re classe comme des exceptions.

Pour participer à la discussion qui doit forcément prendre en compte les aspects liés à la communication (quels verbes sont les plus utiles pour un apprenant afin de pouvoir communiquer en tchèque) et de fréquence (quels sont les verbes qu'il est censé rencontrer le plus souvent dans les énoncés authentiques), il pourrait être intéressant d'avoir un aperçu sur la répartition des cinq classes traditionnelles en fonction de leur fréquence.

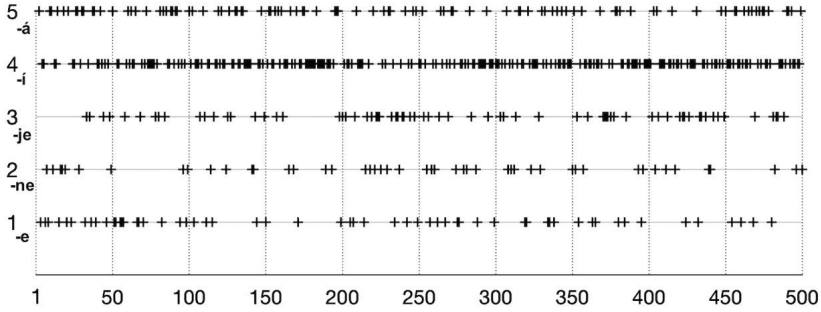



Fig. 1 : La répartition des classes de conjugaison pour les 500 verbes les plus fréquents d'après le *Frequency Dictionary of Czech* (2011).

La figure 1 montre l'importance relative des différentes classes de conjugaison au sein des 500 verbes les plus fréquents présentés dans le *Frequency Dictionary of Czech* (Čermák *et al.*, 2013). La valeur numérique de l'abscisse désigne le rang du verbe en question (le plus fréquent est celui à la valeur 1) représenté par une croix et placé sur une des cinq lignes correspondantes aux classes de conjugaison.

Je laisserai ce graphique sans commentaires tout en attirant l'attention sur la quantité de verbes de la 4e classe et une représentation relativement faible de la 3e classe.

## 7 En guise de conclusion

Les éléments de réponses apportées aux différentes questions soulevées ci-dessus ne sont forcément que parcellaires et très lacunaires. Leur fonction est surtout d'inviter à la discussion. Ce serait un très grand plaisir pour moi, comme à chaque fois quand nous nous rencontrons, de pouvoir en parler avec Patrice. Un travail beaucoup plus minutieux sera nécessaire avant de proposer un outil fonctionnel, pour le traitement automatique comme pour l'enseignement.

Pour terminer, un cadeau…

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

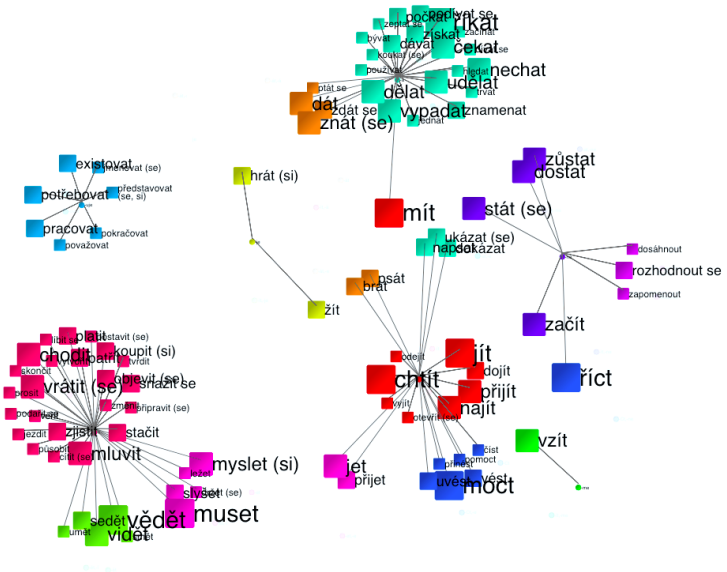



Fig. 2 : Une visualisation des verbes les plus fréquents en fonction de leur conjugaison.

Non pas sans rappeler un bouquet de fleurs d'anniversaire, la figure 2 est une visualisation du système verbal tchèque articulé autour de la conjugaison des verbes les plus fréquents (la position des nœuds et des arêtes) et de leur fréquence (la taille des nœuds et des étiquettes). Les différents nœuds sont connectés par la terminaison de la 3$^{e}$ personne du présent de l'indicatif et se positionnent sur le plan en deux dimensions en fonction de la terminaison de leur infinitif.

Ce graphique a été généré automatiquement à l'aide d'un outil en ligne mise en place par Pierre Magistry que je remercie vivement.

Bon anniversaire, Patrice, et au plaisir de poursuivre nos projets !

## Bibliographie

CVEJNOVÁ, Jitka *et al.* (2014), *Referenční popis češtiny pro účely zkoušky z českého jazyka pro trvalý pobyt v ČR, úrovně A1 a A2,* Národní ústav pro vzdělávání, Praha.

CVRČEK, Václav *et al.* (2010), *Mluvnice současné češtiny*, Karolinum, Praha.

ČECHOVÁ, Marie *et al.* (2000), *Čeština – řeč a jazyk,* ISV, Praha.

ČERMÁK, František *et al.* (2013), *A frequency dictionary of Czech : core vocabulary for learners,* Routledge, New York.

ESVAN, François *et al.* (2019), *Grammatica ceca : Fonetica, morfologia e sintassi con esercizi e soluzioni,* Hoepli, Milan.

HAVRÁNEK, Bohuslav et JEDLIČKA, Alois (1963), *Česká mluvnice.* SPN, Praha.

HRONOVÁ, Karla (2009), *Čeština jako cizí jazyk : Vstupní kurs a základní gramatika,* Didakta Praha.

KARLÍK, Petr at al. (1996), *Příruční mluvnice češtiny,* Nakladatelství lidové noviny, Praha.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260

KASTLER, Claude (1986), *La langue tchèque : grammaire tchèque pratique et raisonnée,* Ophrys, Paris.

KOMÁREK, Miroslav (2006), *Příspěvky k české morfologii,* Periplum, Olomouc.

LAMPRECHT, Arnošt *et al.* (1977), *Historický vývoj češtiny : hláskosloví, tvarosloví, skladba,* SPN, Praha.

MAZION, André (1931), *Grammaire de la langue tchèque.* Honoré Champion, Paris.

NAUGHTON, James (2005), *Czech : An Essential Grammar,* Routledge, London – New York.

OSOLSOBĚ, Klára (2014), *Česká morfologie a korpusy,* Karolinum, Praha.

POGNAN, Patrice (1972), *Analyse morphosyntaxique de textes médicaux tchèques. Extraction automatique du verbe,* Thèse de doctorat de 3$^{e}$ cycle. Université de Paris 3.

POGNAN, Patrice et PANEVOVÁ, Jarmila (2013), « Génération automatique de lexèmes slaves à partir de leurs racines historiques. Une des bases de l'enseignement multilingue des langues slaves de l'Ouest (Nord et Sud) », *Linguistica,* n° 52, p. 59-75.

POLDAUF, Ivan et ŠPRUNK, Karel (1968), *Čeština jazyk cizí : mluvnice češtiny pro cizince,* SPN, Praha.

PRAVDOVÁ, Markéta *et al.* (2014), *Akademická příručka českého jazyka,* Academia, Praha.

ŠMILAUER, Vladimír (1972), *Nauka o českém jazyku,* SPN, Praha.

VEY, Marc (1945), *Le classement du verbe en tchèque littéraire contemporain,* St. Austin & Sons, Hertford.

ŠTÍCHA, František *et al.* (2013), *Akademická gramatika spisovné češtiny,* Academia, Praha.

Published in : JAMBOROVA LEMAY, Diana ; PEREIRA, Christophe ; ŠMILAUER, Ivan (dir.), *Des langues calculables à l'homme incalculable. Hommage à Patrice Pognan*, Editions des archives contemporaines, ISBN : 9782813004260, doi : 10.17184/eac.9782813004260