Dictionnaire langue maternelle - langue étrangère

à visée “d’apprentissage de la langue”

Gregor PERKO(1), Patrice POGNAN(2)

(1) Maître de conférences de français à l’Université de Ljubljana,
(2) Professeur émérite à l’INALCO

Résumé : Nous proposons des traitements informatiques qui permettent de réaliser des dictionnaires doublement étiquetés (langue maternelle - langue étrangère) à partir d’une base de données lexicale langue étrangère avec traduction dans la langue maternelle ou même d’un dictionnaire langue étrangère - langue maternelle. Un tel dictionnaire à finalité d’apprentissage de la langue a déjà été proposé pour la méthode d’apprentissage du slovaque dans le cadre d’un contrat européen ALPCU. Nous présentons ici la réalisation en cours d’un dictionnaire français - slovène à partir du dictionnaire slovène - français de Gregor Perko.

Mots-clés : français, slovène, slovaque, traitements automatiques dictionnairiques, dictionnaire doublement étiqueté.

Beaucoup d’entre nous ont eu, pendant leurs études secondaires, des difficultés avec l’apprentissage des langues. A une époque déjà lointaine, il existait une merveilleuse méthode, « prouvée scientifiquement », qui consistait à mettre un magnétophone sous son oreiller la nuit. L’apprentissage sans le moindre effort pendant son sommeil était vraiment tentant. Hélas ! La méthode a disparu, et pour cause ! Il n’y a pas de secret et l’apprentissage conscient d’une langue demande des efforts et du travail[1].

Mais on peut, au moins, rendre cet apprentissage plus confortable grâce à des outils adéquats. Et le dictionnaire que nous proposons souhaiterait participer à ce confort d’apprentissage.

Lorsque l’on apprend une langue étrangère ou lorsque l’on traduit ou on écrit dans une langue étrangère, on a – dans un premier temps – besoin de chercher ou vérifier les

[1]Ces travaux ont été réalisés dans le cadre du PHC PROTEUS 2018 n° 40061YD.

équivalents lexicaux à choisir au mieux dans la langue étrangère, mais ce premier choix fait, on peut se trouver, notamment chez l'apprenant, dans la nécessité de vérifier les conditions de fonctionnement de ces mots étrangers : caractéristiques grammaticales en particulier les modèles de flexion, l'environnement syntaxique,… et ceci tant dans les grammaires que dans les dictionnaires langue étrangère – langue maternelle.

L'informatique permet d'envisager des opérations auxquelles le lexicographe n'est pas encore habitué.

Il y a quelques années déjà, nous avions conçu, dans le cadre d'un contrat européen ALPCU, réalisé à l'INALCO, un dictionnaire français-slovaque « à double visée »[2].

La méthode multimédia d'apprentissage du slovaque comprenait un dictionnaire pour débutants slovaque – français à partir duquel a été construit le dictionnaire français – slovaque à visée d'apprentissage. Originellement, c'est le manque de temps pour créer un lexique français – slovaque qui nous a mené vers l'idée d'utiliser le dictionnaire slovaque – français qui avait été construit par l'intermédiaire d'une base de données en Access.

Au sein de cette base, chaque entrée (mot) est abondamment documentée. En premier lieu, les mots sont répartis entre dérivation nominale et dérivation verbale. Y sont associées différentes catégories lexicales. Ceci permet de traiter aisément les substantifs verbaux (dérivation verbale, catégorie lexicale de substantif) et les différents adjectifs issus de formes verbales diverses (participes et gérondifs). A ces indications s'ajoute une référence aux modèles morphologiques de Diana Lemay : de 1 à 12 pour les modèles de flexion substantivale, 13 à 15 pour la flexion des adjectifs et de 16 à 36 pour les conjugaisons.

Un verbe (*cf.* la vue ci-dessous) est décrit par sa dérivation verbale, sa catégorie lexicale de verbe, son modèle flexionnel (ici, le 27 pour le verbe « baliť »), par son aspect imperfectif. On donne également le verbe de l'autre aspect (ici, perfectif : « zabaliť »), mais aussi le ou les verbe(s) réfléchis en « sa (accusatif) » ou « si (datif) ». Chaque verbe reçoit également son cadre verbal valenciel (ici, +A | do +G = quelque chose | dans quelque chose).

Un substantif recevra une dérivation nominale, une catégorie de substantif, un modèle de flexion, un genre, mais aussi d'éventuelles formes particulières, par exemple de génitif pluriel. Si le substantif est composé, les lexèmes en composition sont donnés (très généralement deux). Il en va de même pour les autres catégories non verbales.

Dans tous les cas, et dans la mesure du possible, les synonymes et antonymes sont indiqués, un sens français générique est proposé, puis les différents sens attestés.

[2]Baranová, Křečková, Lemay, Pognan : « Découvrir et pratiquer le slovaque » Projet européen ALPCU [lingua II], L'Asiathèque, Paris, 2007.

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

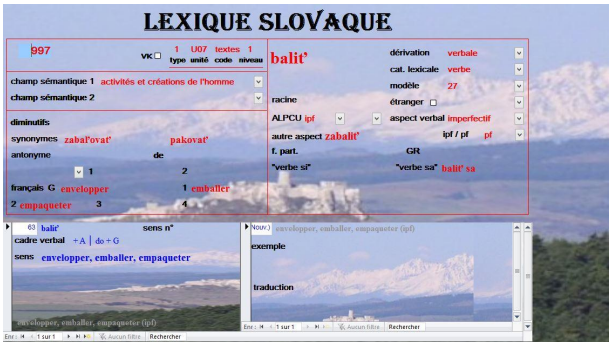


Fig. 1 : Extrait de l'interface de la base de données lexicales slovaque – français

Le contenu de cette base est ensuite extrait et mis en page dans Word pour fournir le lexique slovaque français ci-dessous :

predpredaj 4 *m i* ■ réservation de billets à l'avance, location de places
predstavenie 11 *n* ■ représentation, spectacle
predstavit' 27 *pf* ● (+ D l) + A ■ présenter qqn (à qqn) *(ipf predstavovať)* *(→ predstavit'sa)*
predstavit' sa 27 *pf* ● + D ■ se présenter *(ipf predstavovať sa)*
predstavovať 26 *ipf* ● (+ D l) + A ■ présenter qqn (à qqn) *(pf predstaviť)* *(→ predstavovať sa)*
predstavovať sa 26 *ipf* ● + D ■ se présenter *(pf predstaviť sa)*
predtým *adv* ■ avant, auparavant
prechádzať 16 *ipf* ● odkiaľ kam / cez + A ■ passer, franchir, traverser *(pf prejsť)* *(→ prechádzať sa)*
preložiť 27 *pf* ● + A | do + G ■ traduire *(ipf prekladať)*
premiéra 5 *f* ◆ G pl –miér ■ première
prenajať 23 *pf* ● + A ■ louer, donner en location *(ipf prenajímať)*
prenajať si 23 *pf* ● + A ■ louer, prendre en location *(ipf prenajímať si)*
prenajímať 16 *ipf* ● + A ■ louer, donner en location *(pf prenajať)*
prenajímať si 16 *ipf* ● + A ■ louer, prendre en location *(pf prenajať si)*
prepáčiť 27 *pf* ● + D l + A ■ pardonner, excuser : *(loc)* ***prepáčte*** *pardon, excusez-moi ; (loc)* ***prepáč*** *pardon, excuse-moi*
prerušiť 27 *pf* ● + A ■ interrompre, rompre *(ipf prerušovať)*

Fig. 2 : Extrait du lexique slovaque-français

La réalisation directe d'un lexique français-slovaque étant fortement compromise par le manque de temps avant la fin du contrat, le contenu du dictionnaire slovaque-français a été exploité informatiquement en retournant les informations. Si l'on se reporte à la vue ci-dessus (fig. 2), à partir de « preložiť », on pourra créer une entrée française « traduire ». Nous pourrons alors obtenir une entrée telle que :

traduire : preložiť / prekladať

Il devient alors quasi-évident que la conservation des informations morphologiques et morphosyntaxiques du dictionnaire langue étrangère – français est un avantage considérable. Nous obtenons une entrée doublement étiquetée :

traduire *v. tr.* : preložiť 27 (*pf*) / prekladať 16 (*ipf*) +A | do +G

Ainsi, le verbe transitif « traduire » peut se traduire en slovaque par les verbes preložiť perfectif de modèle de conjugaison 27 et prekladať imperfectif de modèle 16, tous deux ayant pour cadre verbal : +A | do +G (A = accusatif et G = génitif).

Cela nous a permis de publier le lexique français-slovaque :

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782



Fig. 3 : Extrait du lexique français-slovaque

Nous pensons qu'un dictionnaire organisé de cette manière est de nature à apporter à celui qui le consulte – que ce soit pour traduire ou pour rédiger – les informations dont il a besoin pour effectuer la tâche prévue. L'usage d'un tel dictionnaire, surtout s'il est répété, renforce, à notre avis, les processus d'apprentissage de la langue en luttant contre la paresse d'aller consulter une grammaire et/ou le dictionnaire converse.

C'est ce qui explique que depuis 2007 toutes nos bases de données langue étrangère – français sont équipées d'une table de « renversement » intégrée comme on peut le voir sur la vue suivante :

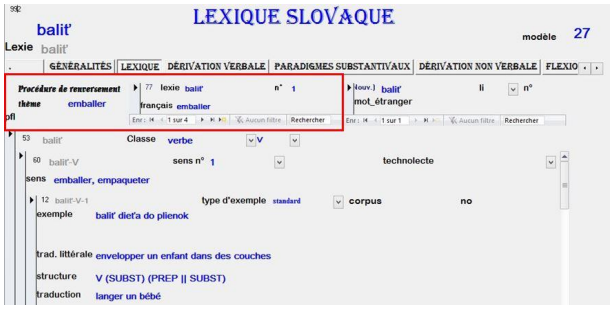


Fig. 4 : Partie lexicale avec table de « renversement » de nos bases de données langue étrangère-français

Dans le cadre de notre coopération franco-slovène, nous avons en plus de la réalisation de monographies portant sur la morphologie du slovène, entrepris différents types de travaux, notamment un dictionnaire de valences verbales et un dictionnaire français-slovène « à visée didactique » sur le modèle du slovaque. Ces projets ont nécessité des travaux préparatoires.

C'est le dictionnaire slovène – français de Gregor Perko que nous avons choisi pour pouvoir extraire et exploiter le vocabulaire.

Pour le slovaque, les opérations ont été menées à partir d'une base de données dédiée à la construction du dictionnaire, mais ici, au contraire, nous partons d'un existant non prévu pour une telle construction.

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

En effet, en retournant le contenu des entrées, nous conservons les informations grammaticales portées dans le dictionnaire, mais ici dans le dictionnaire slovène – français l'entrée slovène n'est pas étiquetée étant destinée à un public slovène censé posséder les connaissances morphologiques correspondantes. Nous ne possédons d'informations que pour les correspondants français, en particulier le genre. Nous serons donc amenés à compléter les informations pour la partie slovène.

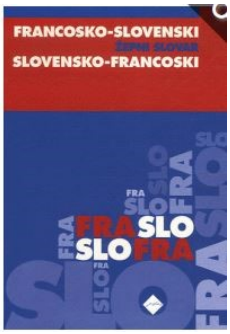


Fig. 5 : Dictionnaire de Gregor PERKO

Nous avons repris les fichiers qui ont servi à construire le dictionnaire.

Ces fichiers ont été réalisés en Word, un fichier par caractère alphabétique. Les entrées en slovène sont en caractères gras et il n'existe aucun séparateur entre l'entrée slovène et le(s) correspondant(s) français.

bábica grand-maman ž; (*med.*) sage-femme ž
bacíl bacille *m*
báger excavateur *m*
baháč vantard *m*
bahàv fanfaron
bájen fabuleux
bájka conte *m*

Fig. 6 : Forme originelle du dictionnaire

Nous avons programmé Word à l'aide de son langage de programmation VBA pour insérer, grâce à la présence du gras, un séparateur (« : ») au-delà de l'entrée mise en relief (couleur bleue). Nous pourrons dès à présent manipuler par programme les parties slovène et française.

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

bábica : grand-maman ž; (*med.*) sage-femme ž
bacíl : bacille *m*
báger : excavateur *m*
baháč : vantard *m*
bahàv : fanfaron
bájen : fabuleux
bájka : conte *m*

Fig. 7 : Insertion automatique d'un séparateur

Le fichier Word obtenu est transformé avec le transcodeur Microsoft intégré en fichier .txt en UTF-8.

Ce fichier constitue le début de la chaîne de traitements que nous allons appliquer aux 29 fichiers alphabétiques issus du dictionnaire originel.

Fichier Edition Format Affichage ?
bábica : grand-maman ž; (med.) sage-femme ž
bacíl : bacille m
báger : excavateur m
baháč : vantard m
bahàv : fanfaron
bájen : fabuleux
bájka : conte m

Fig. 8 : Fichier brut (UTF-8) base des traitements

La construction d'un dictionnaire français – slovène à visée d'apprentissage commence réellement ici. Le programme inverse les parties slovène et française. Cette dernière devient l'entrée. La partie française est déjà renseignée, portant les étiquettes grammaticales adéquates écrites en slovène.

žága : scie ž ; (podjetje) scierie ž

Le programme double les informations grammaticales du français énoncées en slovène par celles qui leur correspondent en français :

ž f/ž

Les opérations exécutées par le programme sont visualisées dans le schéma suivant :

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

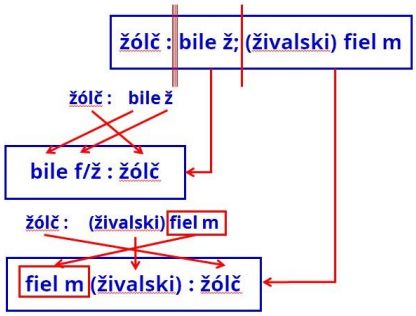


Fig. 9 : Analyse de la ligne et construction des entrées françaises

Chaque ligne du fichier traité est vérifiée et complétée manuellement de la manière suivante :



Fig. 10 : Ajouts manuels pour certaines lignes

Le résultat est un fichier transformé suivant les procédés exposés ci-dessus, mais encore dans l'ordre alphabétique du fichier slovène d'origine, ce que l'on peut vérifier d'après la partie droite de chaque ligne :

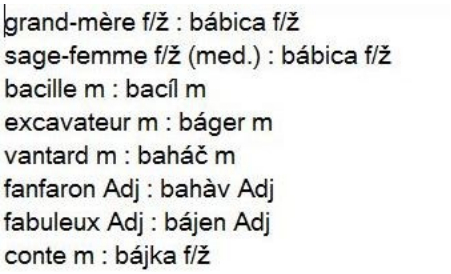


Fig. 11 : Fichier originel transformé

Ainsi sont transformés, l'un après l'autre, les fichiers obtenus à partir du dictionnaire slovène-français. Au-delà, c'est la construction d'un nouveau dictionnaire unique, d'un seul tenant (un seul fichier) qui débute.

Les différents fichiers alphabétiques complets sont assemblés dans un fichier commun. Ce fichier est ensuite classé alphabétiquement. Nous donnons un exemple ci-dessous :

Dans quelque temps, lorsque toutes les entrées seront vérifiées, complétées et triées, il nous restera à faire un programme qui réduira les entrées dupliquées, par exemple :

abcès m : abscés m

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

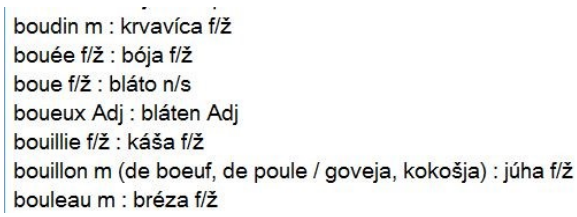


Fig. 12 : Fichier résultant français-slovène doublement étiqueté

abcès m : búla f/ž

en une seule entrée avec plusieurs correspondants slovènes :

abcès m : abscés m, búla f/ž

En retournant le dictionnaire langue étrangère vers la langue maternelle, on reste tributaire du dictionnaire d'origine, notamment pour la masse lexicale et les relations à la civilisation portée par la langue étrangère. La conception d'un dictionnaire langue maternelle – langue étrangère n'est pas acquise. On peut, en effet, se trouver en présence de « trous lexicaux », vraisemblablement d'autant plus nombreux que les langues sont éloignées l'une de l'autre. Des aménagements manuels au dictionnaire obtenu informatiquement seront nécessaires, mais le gain de temps de réalisation globale est important. Nos estimations semblent indiquer que l'on divise la durée au moins par deux.

Voici un type d'aménagement que nous serons amenés à réaliser de manière importante :

Des entrées telles que :

sabot m : cókla f/ž

sabot m : kopíto n/s

réduites automatiquement en :

sabot m : cókla f/ž, kopíto n/s

devront être travaillées ultérieurement de manière manuelle, peut-être organisées dans un ordre différent et avec des annotations d'ordre sémantique :

sabot m : kopíto n/s, cókla f/ž

« kopíto » étant un sabot animal, « cókla » un sabot chaussure en premier lieu, mais aussi avec un sens technique dérivé, nous devons transformer l'entrée de la manière suivante :

sabot m : kopíto n/s (animal/žival), cókla f/ž (chaussure/čévelj)

∼ d'arrêt : cókla.

Ce qui est important est que nous obtenons ainsi un dictionnaire qui permet à un apprenant francophone de chercher à partir du français les correspondants slovènes pour

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782

lesquels il trouve immédiatement les informations grammaticales sans avoir à les chercher dans une grammaire ou un dictionnaire slovène – français. Toutes les étiquettes grammaticales étant bilingues, le dictionnaire est accessible aux deux communautés linguistiques.

Bibliographie

Baranová Elena, Křečková Vlasta, Lemay Diana, Pognan Patrice (2007) : *Découvrir et pratiquer le slovaque*. Projet européen ALPCU [lingua II], livre et DVD, L'Asiathèque, Paris.

Perko Gregor (2007) : *Francosko-slovenski in slovensko-francoski žepni slovar* (Dictionnaire de poche français-slovène et slovène-français). Cankarjeva Zalozba, Ljubljana.

Pognan Patrice, Schlamberger Brezar Mojca, Perko Gregor. (2016) : Des corpus aux bases de données en langue slovène – des chaînes de traitement informatique pour la préparation d'applications (lexiques et dictionnaires, grammaires, manuels, analyse automatique) in *Enseignement/apprentissage des langues et pratiques numériques émergentes*. Editions des Archives contemporaines. Paris.

Schlamberger Brezar Mojca, Perko Gregor, Pognan Patrice (2015) : *Les bases de la morphologie du slovène pour locuteurs francophones*. Filosofska Fakulteta, Univerza v Ljubljani, Ljubljana. 192p. (ISBN 978-961-237-741-0 (zv. 1))

Szende Thomas (dir.) (1999) : *Dictionnaires bilingues. Méthodes et contenus (Actes de la 1re Journée d'étude sur la lexicographie bilingue)*, 1999, Editions Honoré Champion, Collection « Études de lexicologie, lexicographie et dictionnairique », Paris.

Szende Thomas (dir.) (2000) : *Approches contrastives en lexicographie bilingue* (Actes des 2es Journées d'étude sur la lexicographie bilingue), Editions Honoré Champion, Collection « Études de lexicologie, lexicographie et dictionnairique, Paris.

Szende Thomas (dir.) (2006) : *Le français dans les dictionnaires bilingues*, Editions Honoré Champion, Paris.

Taïfi Miloud (2017) : "*Dictionnaire raisonné berbère – français. Parlers du Maroc*" (1223p), Pognan Patrice : index "*Formes de mots berbères – racines*" (166p) et index "*Significations françaises – racines berbères*" (319p). IRCAM, Rabat.

Published in : BERK, Cybèle (dir.), *Dictionnaires et apprentissage des langues*, Editions des archives contemporaines, ISBN : 9782813003782, doi : 10.17184/eac.9782813003782