

Analyse qualitative et quantitative des erreurs de flexion : le cas du locatif singulier en tchèque

Ivan Šmilauer

► **To cite this version:**

Ivan Šmilauer. Analyse qualitative et quantitative des erreurs de flexion : le cas du locatif singulier en tchèque. Analyse des erreurs commises par des francophones apprenant une langue éloignée, Feb 2016, Paris, France. Pré-actes du Colloque international de linguistique Analyse des erreurs commises par des francophones apprenant une langue éloignée. <hal-01444983>

HAL Id: hal-01444983

<https://hal-inalco.archives-ouvertes.fr/hal-01444983>

Submitted on 24 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse qualitative et quantitative des erreurs de flexion : le cas du locatif singulier en tchèque

Ivan Šmilauer
EA 4514 PLIDAM, INALCO
ivan.smilauer@inalco.fr

Cette étude traite des erreurs commises par des apprenants francophones du tchèque dans des exercices de déclinaison proposés en ligne. Notre but est de dresser un tableau de la diversité des formes erronées produites dans le domaine particulier du locatif singulier des substantifs et d'établir les erreurs les plus fréquentes.

1. Déclinaison tchèque

L'apprentissage du tchèque, langue slave appartenant à la famille des langues indo-européennes, est relativement complexe, notamment à cause de sa morphologie flexionnelle nominale. Dans ce système linguistique, la fonction syntaxique d'un substantif est exprimée par une forme fléchie portant la valeur de l'un des sept cas (nominatif, génitif, datif, accusatif, vocatif, locatif, instrumental). Les formes casuelles, constituées du radical du substantif et de la désinence, forment un paradigme appartenant à un des quatorze types de déclinaison, avec des dizaines de sous-types et de nombreuses exceptions. Enfin, de fréquentes alternances morphologiques se produisent sur le radical des substantifs en fonction de la désinence choisie (notamment la palatalisation de la consonne finale du radical et l'alternance de la longueur vocalique des radicaux monosyllabiques).

En respectant la dichotomie linguistique classique de la forme et de la fonction, nous pouvons émettre l'hypothèse que pour un apprenant en situation de production, la difficulté se situera à deux niveaux : d'abord, (1) savoir choisir la fonction syntaxique adéquate pour un substantif dans son contexte, puis, en fonction de ce choix abstrait, (2) savoir produire la forme casuelle correspondante.

2. Données étudiées

Les productions étudiées proviennent d'une base de données constituée à l'aide de la plateforme CETLEF (accessible librement à l'adresse <http://www.cetlef.fr>, cf. Šmilauer 2008, Šmilauer 2010) qui propose des exercices de déclinaison tchèque, la tâche de l'apprenant étant de décliner le lemme d'un substantif en fonction de son contexte dans une phrase à trous.

Au total, la base de données contient à ce jour quelque 40 000 productions produites par plus de 600 apprenants dans 58 exercices. Pour les besoins de la présente étude, un sous-corpus a été constitué contenant 3 534 productions issues de 5 exercices (49 tâches différentes) ciblés sur le locatif singulier des substantifs produites, en fonction de l'exercices, entre 64 et 82 apprenants.

Le dépouillement de ces données est facilité par une procédure automatique d'analyse des erreurs, effectuée à l'aide des techniques de traitement automatique des langues (e.g. Heift et Schulze 2007). En effet, chaque forme requise au sein d'une tâche est annotée par une étiquette morphosyntaxique comportant les informations sur le type de déclinaison, le cas, le nombre et le genre et, éventuellement, l'alternance morphologique à effectuer. Dans la version actuelle du dispositif, il n'y a pas d'annotation d'ordre syntaxique. Ces informations permettent d'établir une description formelle de la forme erronée (voir plus bas) et d'automatiser certaines étapes de l'analyse.

3. Question de méthodologie : description et explication des erreurs

En respectant la prémisse fondamentale de l'analyse des erreurs dans sa version classique (cf. Corder 1981), nous distinguons entre la description de l'erreur, basée sur des critères purement formels, et l'explication des causes probables de l'erreur.

Au niveau de la description, nous utilisons le terme **erreur de flexion** qui désigne ici toute forme déviante par rapport à la forme nominale requise au sein d'une tâche de déclinaison. Inspirés par les travaux sur la modélisation psycholinguistique de la morphologie dans l'apprentissage d'une L2 en tant qu'acquisition progressive des règles combinatoires appliquées sur les morphèmes lexicaux et grammaticaux, assorties d'un ensemble d'idiosyncrasies (e.g. Pirelli 2007, Gor 2010), nous supposons qu'une erreur de flexion peut être décrite comme une combinaison erronée du radical du substantif avec la désinence casuelle. D'après cette hypothèse, la forme erronée peut être le résultat (a) du choix d'une désinence casuelle inappropriée, (b) du non-respect des alternances morphologiques et/o (c) de la violation des règles de graphie à l'écrit (ou de la prononciation à l'oral).

Il s'en suit qu'en disposant d'un modèle formel de la déclinaison, il est possible de recréer artificiellement une forme erronée à partir du lemme d'un substantif, tout en contrôlant les valeurs des catégories grammaticales affectées par cette erreur. En comparant ces valeurs avec celles de la forme requise, nous obtenons une description de la forme erronée en mentionnant le nom de la catégorie affectée : une erreur de cas, de genre, de nombre, de variante de désinence, de type de déclinaison, d'alternance et/ou de graphie.

Deux problèmes surgissent avec cette approche : (1) plusieurs catégories peuvent être affectées en même temps (par exemple une erreur de cas et de nombre) ce qui n'est pas en soit très problématique, contrairement au cas (2), où une forme erronée peut avoir, à cause de l'homonymie très fréquente des désinences casuelles, plusieurs descriptions possibles (par exemple erreur de cas ou erreur de nombre). Dans cette situation, il est nécessaire d'accéder à l'étape de l'explication de l'erreur et choisir la description la plus probable par la mise en évidence des facteurs intra- et inter-linguistiques à travers le prisme des processus cognitifs mis en œuvre pour la constructions de la variété de l'apprenant (analogie, généralisation, interférences, e.g. Klein 2012).

4. Analyse des erreurs au locatif singulier des substantifs

Sur les 3534 productions, un quart (24,9 %) sont des productions erronées dont la grande majorité (79,6 %) sont des erreurs de flexion. Concernant le nombre de descriptions possibles, 69,7 % n'en possèdent qu'une seule, les 31,3 % restant ont deux ou plusieurs descriptions.

Les formes qui n'ont pas été diagnostiquées comme erreur de flexion (19,3 % des formes erronées) sont notamment des **fautes de frappes** ou des formes présentant des différences dans le radical du substantif, interprétables linguistiquement mais souvent assez aléatoires.

La plus grande partie des erreurs de flexion (38,4 %) sont des erreurs qui n'atteignent pas les catégories grammaticales (cas, nombre, genre, type de déclinaison) mais qui touchent la graphie et/ou la génération de la forme fléchie (la réalisation des alternances morphologiques et le choix de la variante appropriée de la désinences) que nous appelons également les erreurs locales. Le premier type est représenté par les erreurs de **diacritique**, liées aux contraintes du médium électronique, nécessitant un clavier paramétré pour la saisie des caractères diacrités tchèques. Bien qu'un clavier virtuel soit proposé aux apprenants, ce type d'erreur, est le plus fréquent et de plus, il peut contaminer des formes erronées des types différents. Les erreurs dans la génération de la forme fléchie sont les **erreurs d'alternance** (la dernière consonne du radical doit être palatalisée en combinaison avec la désinence -'e, p.ex. *matk + a > matka* mais *matk + 'e > matce*). En suite viennent les erreurs reflétant distribution relativement aléatoire des désinences -'e et -u dans les types non-palatalisé du masculin inanimé et du neutre ainsi que la distribution des désinence -u/-i et -ovi pour les masculins

animées, causant des **erreurs de variantes de désinences**. Ces erreurs apparaissent de manière massive dans les formes susceptibles d'être atteintes, cependant leur pourcentage dans l'ensemble des productions n'est pas significatif.

Après les erreurs locales viennent les **erreurs de cas** (12,1 %). En effet, sur le plan syntaxique, les fonctions du locatif peuvent être facilement confondues avec celles de l'accusatif qui est souvent précédé par les mêmes prépositions exprimant des circonstances spatio-temporelles similaires ou avec celles du génitif, utilisé dans la traduction tchèque des syntagmes prépositionnels français contenant la préposition *de*. Les occurrences de ces erreurs peuvent être trouvées dans le corpus, notamment celles où transparait une interférence du français impliquant que la préposition *de* demanderait un génitif. Néanmoins, la plupart des erreurs de cas n'ont pas d'explication transparente d'ordre syntaxique, la raison pour laquelle l'apprenant choisit une forme casuelle erronée reste obscure.

Les erreurs **de genre** et/ou **de type de déclinaison** représentent 10,7 %. Les types de déclinaisons tchèque sont déterminés principalement par la nature de la terminaison du substantif (type vocaliques ou consonantiques, types au radical terminé par une consonne palatalisée ou non-palatalisée) et de son genre grammatical (masculins animés et inanimés, féminins et neutres), leur attribution à un substantif inconnu n'est pas déterministe et représente naturellement une source d'erreur. De nombreuses erreurs de genre qui sont observées dans le corpus peuvent être expliquées par une interférence du français, le substantif en question ayant un genre différent dans les deux langues.

Les **erreurs de nombre** représentent 6 % des erreurs de flexion et reflètent la situation où l'apprenant donnerait un locatif pluriel au du locatif singulier, probablement en lisant mal l'énoncé de l'exercice.

Parmi les erreurs à **plusieurs descriptions possibles**, les plus nombreuses (23,7 %) sont celles qui peuvent être interprétées soit comme des erreurs de cas, soit comme des erreurs de genre et/ou de type de déclinaison. Choisir la description appropriée relève déjà du domaine de l'explication et nécessite de prendre en compte la perspective de l'apprenant et les opérations d'ordre cognitif mises en place.

Bibliographie

- Corder S. Pit, 1981, *Error Analysis and Interlanguage*, Oxford University Press.
- Gor Kira, 2010, Introduction. beyond the obvious: Do second language learners process inflectional morphology? *Language Learning*, 60(1):1-20.
- Heift T., Schulze M., 2007, *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge, UK.
- Klein Wolfgang, 2012, A Way to Look at Second Language Acquisition. in Watorek M., Benazzo S., Hickmann M. (éds), *Comparative Perspectives on Language Acquisition*. Multilingual Matters, p. 23-36.
- Pirelli Vito, 2007, Psycho-computational issues in morphology learning and processing : An ouverture. *Lingue e Linguaggio*, VI(2):131-138.
- Šmilauer Ivan, 2008, Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison. *The Prague Bulletin of Mathematical Linguistics*, 90, p. 33-56.
- Šmilauer Ivan, 2010, Exploitation d'un outil ELIAO dans la recherche sur l'acquisition de L2, in *Actes du colloque AcquisiLyon'09*, p. 150-154.