



Genre des substantifs en tchèque : l'ambiguïté de ses marqueurs formels et le diagnostic automatique des erreurs du point de vue de son acquisition par les apprenants francophones

Ivan Šmilauer

► To cite this version:

Ivan Šmilauer. Genre des substantifs en tchèque : l'ambiguïté de ses marqueurs formels et le diagnostic automatique des erreurs du point de vue de son acquisition par les apprenants francophones. Colloque des doctorants et des jeunes chercheurs (COLDOC09) "Ambiguïté dans les sciences du langage", Jun 2009, Nanterre, France. hal-01375644

HAL Id: hal-01375644

<https://inalco.hal.science/hal-01375644>

Submitted on 20 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.

Genre des substantifs en tchèque

l'ambiguïté de ses marqueurs formels
du point de vue de son acquisition par les apprenants francophones
et le diagnostic automatique des erreurs



Ivan ŠMILAUER
LaLIC–CERTAL, INALCO (Paris)
smilauer@cetlef.fr
session poster COLDOC'09

Ambiguïté morphologique du point de vue d'un apprenant de LE

Les effets de l'**ambiguïté**, définie dans un modèle linguistique stratificationnel par la **multitude de fonctions prises par un élément de niveau n au niveau n+1** (cf. Sgall et al. 1986) peuvent être également étudiés dans les productions langagières des apprenants d'une langue étrangère. Nous l'illustrons par l'étude des phénomènes d'ambiguïté entre le niveau phonologique-graphémique et le niveau morphologique du tchèque qui résultent en **erreurs de genre** dans les productions des apprenants francophones.

Dans la plupart des cas, la relation entre un substantif et son genre grammatical ne présente pour un locuteur natif aucun caractère d'ambiguïté car cette information fait partie de son stock « naturel » de connaissances lexicales. La situation est différente pour un apprenant étranger confronté à un substantif au genre inconnu. Celui-ci se trouve devant deux solutions : rechercher et intégrer l'information sur genre « en dur » dans son vocabulaire ou essayer de trouver et d'appliquer des règles basées sur la forme qui permettrait d'attribuer le genre sans recours à un lexique. En se servant de cette seconde technique, l'apprenant se trouve nécessairement heurté aux limites des règles établies dont la surgénéralisation produit des erreurs.

C'est ainsi que nous considérons une erreur de genre comme le résultat de l'attribution d'une fonction inadéquate à une forme ambiguë par rapport aux règles dont se sert l'apprenant. De ce point de vue, nous estimons que les substantifs sont plus ou moins ambigus par rapport au genre.

Le tchèque possède quatre valeurs pour la catégorie grammaticale de genre : masculin animé et inanimé, féminin, neutre. A l'aide des critères formels et sans recours à une connaissance d'ordre lexical, il est possible de définir des règles pour l'attribution du genre en fonction des marqueurs situés à la périphérie droite des lexèmes (la désinence casuelle vocale du féminin ou la terminaison consonnantique du radical). Ces marqueurs permettent avec un degré variable de certitude d'identifier le genre des substantifs : le nominatif singulier terminé par "-o" désigne sans ambiguïté un substantif neutre, le nominatif singulier terminé par "-a" désigne le plus souvent un féminin, mais aussi, marginalement, un masculin etc.

Dans une production d'apprenant, une erreur de genre peut se manifester (a) par une **erreur d'accord** des éléments dépendants de ce substantif, (b) par

l'**attribution d'un type paradigmatique erroné** qui se manifeste lors de la déclinaison du substantif en question. Par exemple, le substantif masculin "stůl" (table) peut être interprété par l'apprenant en tant que féminin car un substantif terminé par la consonne "l" peut être masculin, mais également féminin, voir par exemple "sůl" (sel). Ainsi, cette analyse erronée peut mener à des productions du type (a), par exemple "velká stůl", qui pourrait être traduit littéralement comme « grand table » ; ou à des erreurs de type (b), par exemple le génitif singulier "stoli" d'après le type de déclinaison consonnantique féminin correspondant, au lieu de "stolu" d'après le type correct masculin. C'est sur ce deuxième type d'erreur que nous nous concentrons ici.

L'étude de l'ambiguïté des marqueurs formels du genre est intéressante également du point de vue pédagogique car cette propriété est en relation directe avec le degré de **diagrammaticité** du substantif (caractère indexical de sa construction formelle par rapport à sa signification) qui peut faciliter le traitement cognitif du lexème par l'apprenant.

Nous apportons ici des éléments pour étudier, dans un cadre expérimental spécifique (exercices de déclinaison sur une plate-forme ELAO), les erreurs de genre produites par des apprenants francophones et le rôle de l'ambiguïté des marqueurs formels des substantifs. Nous supposons que les erreurs de genre les plus fréquentes devraient être celles qui ont lieu dans les substantifs formellement ambigus par rapport à la catégorie du genre.

En nous basant sur une **étude quantitative**, nous présentons d'abord l'ambiguïté des marqueurs de genre en tchèque. Ensuite, nous introduisons l'application CETLEF.fr (un outil d'enseignement du tchèque assisté par ordinateur avec un diagnostic automatique des erreurs) qui sert pour la collecte et l'analyse des productions langagières des apprenants.

Finalement, nous présentons les données recueillies à l'aide de CETLEF.fr et nous proposons un **algorithme modélisant l'attribution du genre** à un substantif par les choix d'ordres formels, sémantiques et lexicaux effectués par l'apprenant pendant l'analyse de la forme au genre inconnu. Cet algorithme pourrait servir à des fins pédagogiques mais également pour l'affinement du diagnostic automatique existant au sein de CETLEF.fr.

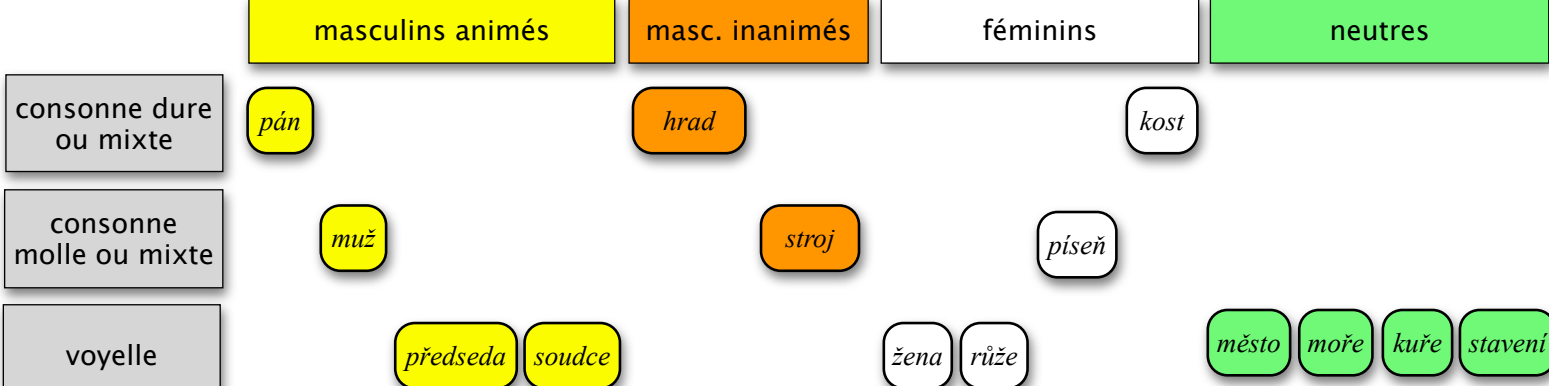
Déclinaison et genre grammatical en tchèque

- langue slave occidentale
- flexion nominale** très riche : sept cas (nominatif, génitif, datif, accusatif, vocatif, locatif, instrumental), quatre genres (masculins animé et inanimé, féminin, neutre), deux nombres (singulier, pluriel) avec des résidus du duel
- nombreux paradigmes** de déclinaison avec des sous-types et des exceptions.
- des **variantes** (doublets, triplets) de registre et/ou de fonction.
- alternances **vocaliques** et **consonnantiques** du radical lors de la déclinaison.
- système vocalique** avec une opposition phonologique de longueur (quantité vocalique).
- système consonnantique** relativement complexe distinguant trois grandes classes de consonnes :
 - consonnes dures : c, č, k, g, r, d, t, n
 - consonnes molles : b, f, l, m, p, s, v, z
 - consonnes mixtes : h, š, j, ň, ř, ž, x
- grâce à la flexion, l'ordre des mots est géré par la structure informationnelle de la phrase.

Exemple d'un paradigme neutre	Exemple d'un paradigme féminin	Exemple d'un paradigme masculin
nominal genit datif accusatif vocatif locatif instrumental	single pluriel růž růžek růžce růžky růžce růžky růžce růžky růžce růžky	single pluriel stůl stůlů stůlům stůly stůle stůle stůle stůle stůle stůle stůle

Paradigmes de la déclinaison nominale tchèque et le genre grammatical

Les grammaires traditionnelles reconnaissent 14 types de déclinaison nominale. Le type paradigmatique d'un substantif (l'ensemble de désinences s'attachant à son radical) est donné par son genre grammatical et par la nature morpho-phonologique de la périphérie droite (terminaison) de son lemme. Chaque type est représenté par un mot modèle.



Remarque : Au sein de chaque type, il existe de nombreux sous-types, définis par les différences par rapport au mot modèle. Cependant, ces différences touchent rarement la forme du lemme qui est déterminée pour l'identification du type paradigmatique (il existe notamment des glissements entre les types féminin -řice et -žice et il ne serait pas utile de les prendre en compte ici).

Erreur de genre sur CETLEF.fr

CETLEF.fr est un outil d'enseignement de langue assisté par ordinateur (ELAO) proposant des exercices de déclinaison tchèque : la tâche de l'apprenant est de créer la forme fléchie d'un lemme en fonction de son contexte syntaxique au sein d'une proposition donnée. Une erreur commise dans une telle tâche est appelée **erreur de déclinaison**. Erreur de genre est un des types d'une telle erreur.

Une erreur de déclinaison est considérée non pas comme un phénomène aléatoire mais comme le résultat d'une activité succombant à des règles d'ordre linguistique et cognitif. Le diagnostic automatique des erreurs, implémenté sur cette plate-forme, est basé sur cette hypothèse. Il a nécessité l'élaboration d'un modèle formel spécifique de la déclinaison contenant un classement des types paradigmatiques et des règles pour la réalisation des alternances morphématiques.

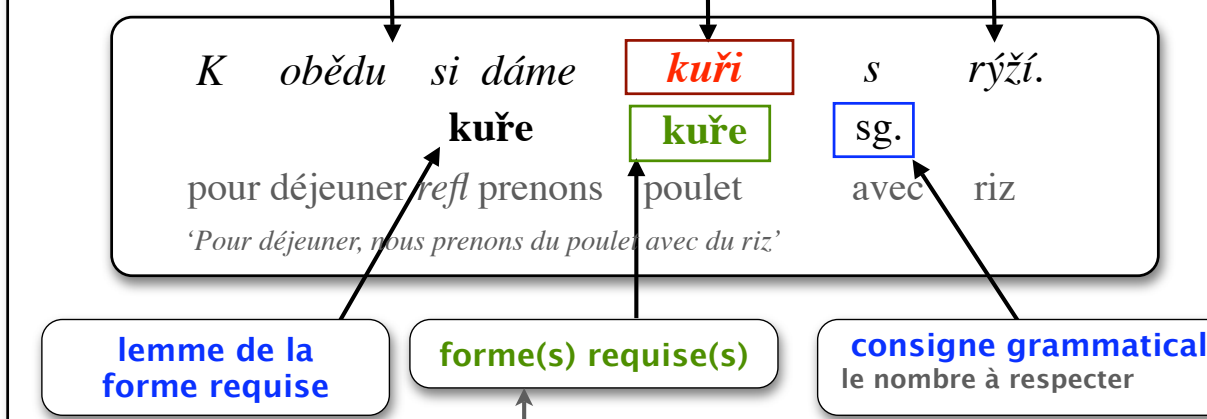
CETLEF.fr illustre les possibilités d'un riche matériel morphologique et des techniques de TAL employées dans un outil d'enseignement de langue assisté par ordinateur, voir par exemple (Heift & Schulze 2007).

Du point de vue d'une recherche sur l'acquisition de langue étrangère, CETLEF.fr sert comme un outil pour la compilation d'un corpus d'erreurs. Par rapport aux productions libres, l'analyse des erreurs recueillies au sein des exercices grammaticaux permet un meilleur contrôle sur le volume de données pertinentes, car les productions doivent nécessairement contenir les phénomènes qui ont été établis comme l'objet de l'investigation.

Tâche de déclinaison au sein d'un exercice à trous

Exemple : « K obědu si dáme (kuře, sg., s rýží. »

Terminologie employée :



Annotation morphologique de la forme requise

(1) lemme de la forme requise, (2) catégorie lexicale, (3) type morphologique (4) paradigme, (5) cas, (6) genre, (7) nombre, (8) alternance éventuelle

Exemple : l requisi 1 lemme 1 tagLex 1 tagMorph 1 pdgm 1 cas 1 num 1 gen 1 alt 1

kuře | kuře | subst | N | kr | acc | sg | n | sans |

Remarque : L'annotation est assignée automatiquement à la forme requise à l'aide d'un générateur morphologique rudimentaire et vérifiée par l'auteur des exercices.

Une tâche de déclinaison présentée sur CETLEF.fr

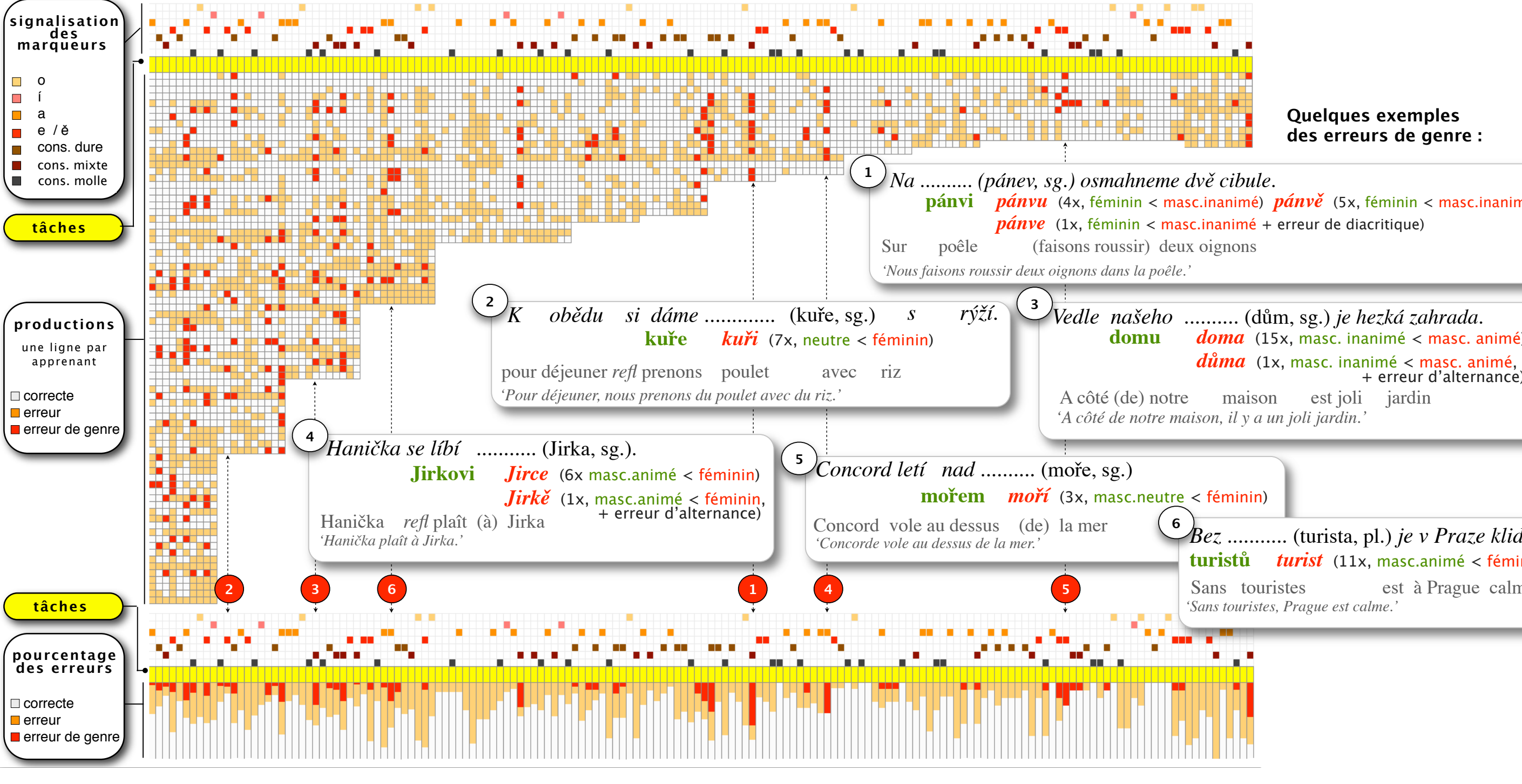
Jana má hodného muže
K obědu si dáme kuře s rýží.
Dobře znám své
Večer Honza nepracuje, ale poslouchá

La correction de la tâche avec le diagnostic

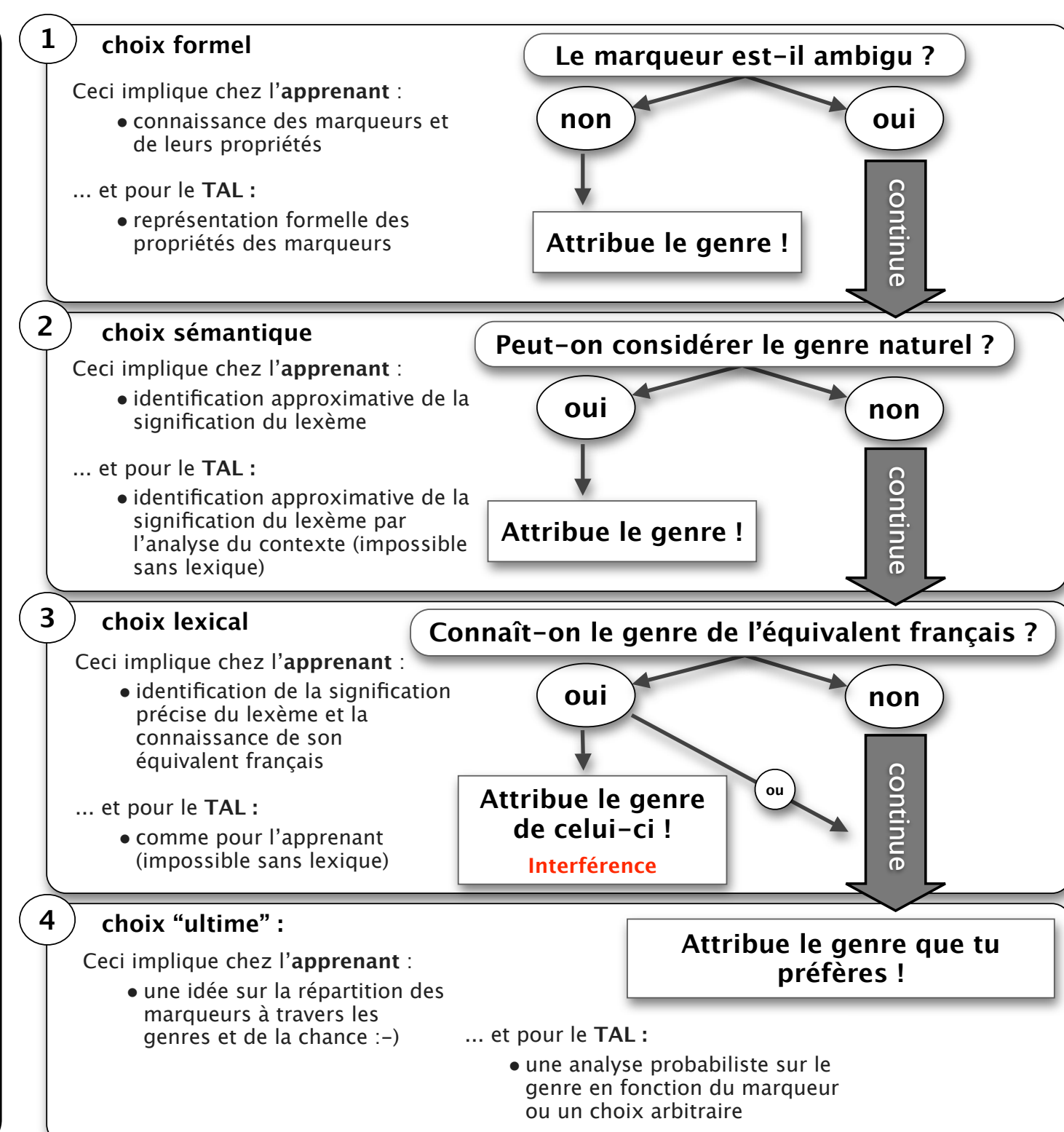
Jana má hodného muže
K obědu si dáme kuře kuře s rýží.
Dobře znám své
Večer Honza nep...
Le cas et le nombre sont corrects mais peut être que vous vous trompez de genre et que vous déclinez le mot d'après le type rýže.

Erreurs de genre dans les productions recueillies

Le graphique ci-dessous présente les erreurs de genres commises par les apprenants dans 162 tâches de déclinaison au sein de 16 exercices sur CETLEF.fr. A l'aide de cette affiche, possible grâce à la base de donnée et au diagnostic des erreurs de genre et leur relation avec la nature du marqueur formel du lemme de la production requise. Avec plus de données recueillies dans le futur, il sera possible de mieux dégager des tendances générales.



Algorithme pour l'attribution du genre



Bibliographie

- Allerton, D. J., Tschichold, C. et Wieser, J., éditeurs (2005). *Linguistics, Language Learning and Language Teaching*. Schwabe, Basel.
- Čermák, F. & Křen, M. (2004). *Frekvenční slovník češtiny*. Nakladatelství Lidové Noviny, Praha.
- Čermák, F. & Křen, M. (2005). *New generation corpus-based frequency dictionaries: The case of Czech*. International Journal of Corpus Linguistics, 10:453-467.
- Čermák, F. (1997). *Czech national corpus: A case in many contexts*. International Journal of Corpus Linguistics, 2(2):161-197.
- Heift, T. & Schulze, M. (2007). *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge, UK.
- Hlaváčková, J. & Savický, P. (2002). *Measures of word commonness*. Journal of Quantitative Linguistics, 9(3): 215-213.
- Klanten, R. éditeur (2009) *Data Flow. Design graphique et visualisation d'information*. Thames & Hudson, Paris.
- Sgall, P., Hajičová, E. & Panevová, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Praha - Amsterdam.
- Šmilauer, I. (2008). *Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison*. The Prague Bulletin of Mathematical Linguistics 90:33-56.