HAL
open science

# Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison

Ivan Šmilauer

## To cite this version:

Ivan Šmilauer. Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison. Linguistique. INALCO; Université Charles de Prague, République tchèque, 2008. Français. NNT : . tel-01373141

**HAL Id: tel-01373141**

**https://inalco.hal.science/tel-01373141**

Submitted on 28 Sep 2016

**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.

**Institut National des Langues et Civilisations Orientales**
LALIC–CERTAL
Laboratoire Langages, Logiques, Informatique, Cognition
Centre d'Études et de Recherche en Traitement Automatique des Langues
&
**Univerzita Karlova v Praze**
Filozofická fakulta
Ústav teoretické a komputační lingvistiky

IVAN ŠMILAUER

# Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison

Thèse de doctorat

études tchèques, option traitement automatique des langues
filologie, matematická lingvistika

Directeurs :

Prof. PhDr. Eva Hajičová, DrSc. Prof. PhDr. Patrice Pognan, DrSc.

2008

Cette thèse a été réalisée dans le cadre d'une co-tutelle de doctorat entre l'Institut National des Langues et Civilisations Orientales, Paris (France) et l'Université Charles de Prague (République tchèque).

La thèse a été soutenue le 29 novembre 2008 avec la mention : *très honorable avec les félicitations du jury*.

## Jury

**Thomas Szende**, INALCO, Paris (président du jury)

**Milan Hrdlička**, Filozofická fakulta, Univerzita Karlova, Praha (rapporteur)

**Sylviane Cardey-Greenfield**, Faculté des Lettres et Sciences Humaines, Université de Franche Comté, Besançon (rapporteur)

**Jarmila Panevová**, Matematicko-fyzikální fakulta, Univerzita Karlova, Praha (membre du jury)

**Patrice Pognan**, INALCO, Paris (directeur de recherche)

**Eva Hajičová**, Filozofická fakulta, Univerzita Karlova, Praha (directeur de recherche)

## Remerciements

*Je voudrais remercier très chaleureusement les deux directeurs de ma thèse, Eva Hajičová et Patrice Pognan, pour leur accueil, leurs conseils, leur soutien, le temps qu'ils m'ont consacré et leur confiance en mon travail.*

*Je remercie Mme Sylviane Cardey et M. Milan Hrdlička d'avoir accepté d'être rapporteur de ma thèse.*

*Je suis reconnaissant à Marie-Anne Moreaux et Serge Fleury pour leur enseignement en traitement automatique des langues et pour leur aide. Je voudrais également remercier František Čermák, Jarmila Panevová et Vladimír Petkevič, avec qui j'ai eu l'occasion de parler de mon travail.*

*Merci à Colette et Francis Vandevelde ainsi qu'à Philippe Gontier pour leur accueil pendant la rédaction de ce travail. Enfin, merci à Nicolas Aubry, Olivier Bondeelle, Harold Friedman, Paul Grundy et Martin Svášek. Merci à Marlène pour tout.*

*Děkuji svým rodičům za jejich lásku a podporu.*

# Table des matières


**1 Introduction** **1**

**2 Cadre méthodologique** **5**

2.1 Tchèque langue étrangère . . . . . 6

2.1.1 Besoin d'une méthodologie du TLE . . . . . 6

2.1.2 Recherche sur le TLE - des cas concrets . . . . . 7

2.1.3 Présentation didactique de la déclinaison du tchèque . . . . . 9

2.2 Acquisition d'une langue étrangère . . . . . 11

2.2.1 Revue de la discipline . . . . . 11

2.2.2 Sources de données . . . . . 14

2.3 Enseignement des langues assisté par ordinateur . . . . . 19

2.3.1 Traitement automatique des langues dans ELAO . . . . . 20

2.3.2 Correction et diagnostic des erreurs . . . . . 21

2.4 Plateforme CETLEF . . . . . 23

2.4.1 Présentation générale de CETLEF . . . . . 23

2.4.2 Présentation des notions et des termes employés . . . . . 24

**3 Annotation morphologique** **31**

3.1 Préliminaires . . . . . 31

3.1.1 Introduction à la déclinaison du tchèque . . . . . 31

3.1.2 Utilité de l'annotation linguistique . . . . . 34

3.1.3 Couverture de l'annotation . . . . . 36

3.1.4 Morphologie automatique . . . . . 43

3.1.5 Modèle de la déclinaison dans CETLEF . . . . . 46

3.2 Catégories lexicales et types morphologiques . . . . . 47

3.3 Catégories morphologiques . . . . . 52

3.3.1 Catégorie de cas . . . . . 52

3.3.2 Catégorie de nombre . . . . . 53

3.3.3 Catégorie de genre . . . . . 53

3.4 Paradigmes de déclinaison . . . . . 56

3.4.1 Définition des paradigmes . . . . . 56
3.4.2 Fichier *pdgm.xml* . . . . . 65
3.4.3 Déclinaison nominale . . . . . 69
3.4.4 Déclinaison adjectivale . . . . . 78
3.4.5 Déclinaison mixte . . . . . 78
3.4.6 Déclinaison pronominale . . . . . 79
3.4.7 Déclinaison numérale . . . . . 80
3.4.8 Fréquence des types paradigmatiques . . . . . 80
3.5 Alternances . . . . . 83
3.5.1 Sens de l'alternance . . . . . 83
3.5.2 Vérification de la réalisation des alternances . . . . . 83
3.5.3 Alternances consonantiques . . . . . 85
3.5.4 Alternances vocaliques . . . . . 88
3.5.5 Fichier *alt.xml* . . . . . 98
3.6 Procédure *AlterneRadical* . . . . . 102
3.6.1 Description . . . . . 102
3.6.2 Tests algorithmiques des alternances . . . . . 105
3.7 Procédure *Annote* . . . . . 111
3.7.1 Description . . . . . 111
3.7.2 Annotation des lemmes connus . . . . . 113
3.7.3 Annotation des lemmes inconnus . . . . . 115

**4 Diagnostic des erreurs** **117**
4.1 Hypothèses sur les erreurs de déclinaison . . . . . 117
4.2 Définition des types d'erreurs . . . . . 120
4.2.1 Définition de l'univers U . . . . . 120
4.2.2 Forme requise et production erronée dans U . . . . . 122
4.2.3 Interprétation morphologique . . . . . 122
4.2.4 Attributs de graphie . . . . . 123
4.2.5 Exemple d'interprétations . . . . . 124
4.2.6 Erreur d'après l'attribut atteint . . . . . 126
4.2.7 Erreur par rapport au paradigme de la forme requise . . . . . 127
4.2.8 Diagnostic morphologique d'une production erronée . . . . . 135
4.2.9 Plausibilité d'une interprétation morphologique . . . . . 138
4.3 Procédure *Diagnostic* . . . . . 139
4.3.1 Description . . . . . 139
4.3.2 Structure de la procédure . . . . . 141

4.4 Message d'erreur . . . . . 141
4.4.1 Attribut `erreur` . . . . . 142
4.4.2 Attributs morphologiques . . . . . 143
4.4.3 Attribut `spec` . . . . . 143
4.4.4 Attributs concernant les alternances . . . . . 145
4.4.5 Attribut `dia` . . . . . 145
4.4.6 Attribut `var` . . . . . 145
4.4.7 Exemples . . . . . 146
4.5 Traitement non morphologique . . . . . 147
4.5.1 Tests sur les caractères . . . . . 148
4.5.2 Tests sur les chaînes . . . . . 148
4.6 Traitement morphologique . . . . . 150
4.6.1 Lecture locale . . . . . 151
4.6.2 Lecture verticale . . . . . 151
4.6.3 Lecture horizontale interne . . . . . 151
4.6.4 Lecture horizontale externe . . . . . 151
4.6.5 Procédure *AlterneHypothèse* . . . . . 152
4.6.6 Exemple . . . . . 154
4.7 Filtrage des interprétations . . . . . 156
4.7.1 Exemple du groupement des interprétations . . . . . 157
4.7.2 Filtrage des interprétations locales . . . . . 157
4.7.3 Filtrage des interprétations verticales . . . . . 158
4.7.4 Filtrage des interprétations horizontales internes . . . . . 161
4.7.5 Filtrage des interprétations horizontales externes . . . . . 161
4.7.6 Exemple . . . . . 162
4.8 Formatage du diagnostic . . . . . 163
4.8.1 Composante 1 : Traduction des attributs `erreur` et `spec` . . . . . 164
4.8.2 Composante 2 : Traduction des attributs concernant les alternances . 166
4.8.3 Composante 3 : Traduction des attributs `dia` et `var` . . . . . 166
4.8.4 Exemples . . . . . 167

**5 Évaluation** **171**
5.1 Enquête publique . . . . . 171
5.1.1 Présentation de l'enquête . . . . . 171
5.1.2 Caractéristique des apprenants et leurs scores . . . . . 172
5.1.3 Présentation des erreurs recueillies . . . . . 173
5.2 Enquête préliminaire . . . . . 186

5.2.1 Présentation de l'enquête . . . . . 186
5.2.2 Caractéristique des apprenants et leurs scores . . . . . 187
5.2.3 Présentation des erreurs recueillies . . . . . 187
5.2.4 Comparaison avec l'enquête publique . . . . . 189
5.2.5 Illustration d'une analyse d'erreurs par tâche . . . . . 189

**6 Implémentation** **195**
6.1 Application Web dynamique . . . . . 195
6.1.1 Techniques de programmation employées . . . . . 196
6.1.2 Architecture de CETLEF . . . . . 197
6.2 Base de données . . . . . 198
6.2.1 Rappels formels et définitions . . . . . 198
6.2.2 Tables dans la base de données `cetlef` . . . . . 200
6.2.3 Table `apprenants` . . . . . 202
6.2.4 Table `exercices` . . . . . 206
6.2.5 Table `taches` . . . . . 209
6.2.6 Table `requis` . . . . . 211
6.2.7 Table `lexique` . . . . . 214
6.2.8 Table `prod_exercices` . . . . . 215
6.2.9 Table `prod_taches` . . . . . 217
6.2.10 Illustration des requêtes . . . . . 220
6.3 Interface utilisateur CETLEF . . . . . 225
6.3.1 Plateforme auteur . . . . . 226
6.3.2 Plateforme apprenant . . . . . 228

**7 Conclusion** **237**

**A CETLEF Plateforme auteur** **241**

**B CETLEF Plateforme apprenant** **249**

**C Paradigmes** **269**
C.1 Flexion nominale . . . . . 269
C.2 Flexion adjectivale . . . . . 301
C.3 Flexion mixte . . . . . 309
C.4 Flexion pronominale . . . . . 313
C.5 Flexion numérale . . . . . 324

**D Alternances** **329**

D.1 Alternances consonantiques . . . . . 329
D.2 Alternances vocaliques . . . . . 330

**Bibliographie** **335**

# Chapitre 1

# Introduction

## Objectif

L'objectif de ce travail est d'effectuer une analyse des erreurs commises par les apprenants francophones dans des exercices de déclinaison du tchèque. La description de ces erreurs doit être formulée de manière à ce qu'elle puisse être utilisée dans un module de diagnostic automatique travaillant par la comparaison d'une forme erronée avec la forme requise, au sein d'un exercice à trous. Ce diagnostic est implémenté sur une plateforme d'enseignement de langue assisté par ordinateur et génère un message de retour aux apprenants sur leurs erreurs.

## Erreur de déclinaison

La langue tchèque avec sa flexion nominale très riche – sept cas, quatre genres, un nombre important de types de déclinaison – offre un matériau intéressant du point de vue de l'acquisition de la morphologie. Une erreur de déclinaison se manifeste par un écart entre une ou plusieurs formes requises dans un certain contexte syntaxique et la forme erronée produite par l'apprenant. Pour exemple, voici trois erreurs de ce type produites par des apprenants au sein d'un exercice à trous :

(1) ***Aničce** *bolí* *hlava.*
Aničku acc.sg.f.
Anne fait mal tête
'Anne a mal à la tête' (1.175.24)[1]

[1]Chaque production authentique citée dans ce travail sera accompagnée par son identifiant (enquête.tâche.apprenant) dans la base de données contenant les erreurs, voir les détails dans le chapitre *Implémentation*, p. 195.

L'exemple (1) montre un emploi erroné de la forme du datif singulier du nom propre *Aničkaf*. L'alternance $k > c$, obligatoire devant la désinence $-e$ sur un radical terminé par une vélaire, a été effectuée. Néanmoins, l'emploi du datif est incorrect dans ce contexte : c'est l'accusatif qui doit être employé. La raison de cette erreur devrait donc être cherchée sur le plan syntaxique car l'explication la plus évidente est le non respect, sous l'influence du français, de la valeur de la catégorie du cas du complément d'objet imposé par la rection du verbe *bolet*.

(2) *V muzeu je mnoho vzácných* ***díl.***
děl gen.pl.n.
Dans musée est beaucoup précieux œuvres
'Dans le musée, il y a beaucoup d'œuvres précieuses.' (0.17.15)

Dans l'exemple (2), la valeur de la catégorie du cas demandée dans la construction ($mnoho + N_{gen}$) a été respectée, cependant la forme est incorrecte, car l'alternance *í* > *ě*, qui a lieu dans le génitif pluriel du substantif $dílo_n$, n'a pas été effectuée. Nous pouvons donc voir à l'origine de cette erreur un dysfonctionnement au niveau phonologique et morphologique.

(3) *Za* ***vesem*** *bylo obilné pole.*
vsí inst.sg.f.
Derrière village était de blé champ
'Derrière le village, il y avait un champ de blé' (0.80.5)

L'erreur dans l'exemple (3) est le résultat d'une confusion dans l'attribution du modèle paradigmatique au substantif $ves_f$. En effet, l'apprenant respecte la valeur du cas imposée par la préposition ($za + N_{inst}$), car c'est la valeur de cette catégorie exprimée par le morphème $-em$ qui s'attache au radical des substantifs masculins animés / inanimés ou des substantifs neutres, néanmoins cette terminaison n'appartient pas au paradigme flexionnel féminin d'après lequel doit être décliné ce lexème.

## Hypothèses sur les erreurs

Les exemples ci-dessus illustrent qu'une erreur de déclinaison peut être une manifestation de dysfonctionnements sur des plans linguistiques différents. Dans le cadre du diagnostic automatique, nous allons limiter l'analyse des erreurs uniquement sur des critères morphologiques car nous estimons que c'est l'étape qui doit précéder la mise en évidence de niveaux linguistiques supérieurs, et l'interprétation de ces erreurs en terme d'interférences entre la langue natale et la langue étrangère.

L'hypothèse sous-jacente au diagnostic des erreurs est que les erreurs de déclinaison sont calculables, c'est-à-dire qu'une forme erronée peut être générée automatiquement à partir du lemme de la forme requise à l'aide des moyens formels de la déclinaison : dans les exemples ci-dessus, nous avons vu qu'une forme erronée peut être une forme casuelle inappropriée, une forme sans la réalisation d'une alternance nécessaire ou une forme créée avec une désinence n'appartenant pas au paradigme de la forme requise. Pour pouvoir reconnaître ces formes dans les exercices, nous devons élaborer un modèle de la déclinaison contenant le classement des désinences dans des paradigmes ainsi qu'un ensemble de règles pour la réalisation des alternances vocaliques et consonantiques.

## Plateforme CETLEF

La réalisation de la plateforme CETLEF[2] représente la partie pratique de notre thèse. Il s'agit d'une application Web dynamique avec une base de données relationnelle pour la récolte des données issues des apprenants.

Ce dispositif est d'une utilité directe pour l'apprenant. L'intégration de la sortie du diagnostic est un élément qui permet de lui fournir un retour immédiat sur ses erreurs. L'utilisation de l'annotation morphologique pour l'affichage des informations grammaticales dans une perspective didactique donne à l'apprenant une présentation explicite du système de la déclinaison du tchèque.

Les productions des apprenants recueillies, annotées morphologiquement, représentent un matériau précieux pour un large éventail d'études et d'expériences menées dans l'optique de l'acquisition du tchèque langue étrangère. La collecte de données directement au format numérique est une alternative intéressante aux enquêtes de terrain ou à la constitution de corpus de productions libres à partir de textes manuscrits.

Dans son aspect linguistique, le travail a suscité l'élaboration d'un modèle formel de la déclinaison du tchèque, adapté à un traitement automatique du diagnostic des erreurs, et à l'annotation des formes dans les exercices. Enfin, la description des erreurs devrait élucider quelques aspects de l'acquisition de la déclinaison du tchèque par les francophones.

## Plan de la thèse

Les points de départ de notre travail sont présentés dans le chapitre *Cadre méthodologique*. Nous présentons d'abord un état de la recherche portant sur le tchèque langue étran-

[2]Connaître, Comprendre, Corriger les Erreurs en Tchèque Langue Étrangère pour les Francophones, disponible pour les apprenants sur `http://www.cetlef.fr`.

gère. Nous nous situons après dans le contexte des recherches en acquisition d'une langue étrangère, et nous établissons le lien entre ce domaine et le domaine de l'enseignement des langues assisté par ordinateur, qui justifiera les bases de notre travail.

Après une revue des approches de la description de la déclinaison tchèque, un modèle de la déclinaison adapté à nos besoins est introduit dans le chapitre *Annotation morphologique*. Ce modèle, contenant un classement détaillé des paradigmes de déclinaison et des alternances, est à la base d'une annotation linguistique dont l'objectif est de spécifier les productions des apprenants, de servir les procédures automatiques au sein du diagnostic des erreurs, et d'être utilisé pour une présentation de la déclinaison dans un but didactique.

Le chapitre *Diagnostic des erreurs* met en évidence les hypothèses que nous prenons sur la nature des erreurs et leurs motivations, linguistiques ou autres. Nous établissons des types d'erreurs qui peuvent être calculées automatiquement et nous proposons une procédure de diagnostic automatique, basée sur la génération de formes déclinées et sur leur comparaison à une production erronée.

Le chapitre *Évaluation* apporte des résultats d'une étude pilote visant à tester le dispositif avec des données authentiques recueillies par deux enquêtes différentes. Nous présenterons la couverture du diagnostic, son efficacité et ses faiblesses.

L'architecture de la plateforme CETLEF et l'intégration des procédures automatiques au sein de ce système sont présentées dans le chapitre *Implémentation*. Nous illustrons ensuite le fonctionnement de cette plateforme du point de vue d'un enseignant ou d'un expérimentateur ainsi que du point de vue de l'apprenant.

# Chapitre 2

# Cadre méthodologique

Ce chapitre présente d'abord l'état de la recherche sur le tchèque langue étrangère. La présentation didactique de son système casuel fait l'objet de nombreuses discussions, sans que, pour le moment, une étude détaillée sur les difficultés rencontrés par les apprenants soit entreprise. L'ambition de ce travail est de proposer un outil pour pouvoir effectuer une étude de ce genre, ce qui nécessite son ancrage dans le domaine de l'acquisition de langues étrangères et la présentation des procédés de recherche qui y sont employés.

L'analyse des erreurs est un moyen privilégié pour l'identification des difficultés des apprenants dans leur acquisition : une erreur est considérée non pas comme un phénomène aléatoire mais comme le résultat d'une activité succombant à des règles d'ordre cognitif. La problématique des sources de données pour cette analyse est abordée par la suite. La technique de la compilation des corpus contenant les productions langagières des apprenants (corpus d'apprenants) est confrontée à la méthode employée dans CETLEF qui collecte les données à partir des exercices grammaticaux.

La décision de publier les exercices sur une plateforme Web exige nécessairement une introduction dans la problématique de l'enseignement des langues assistés par ordinateur. Nous allons mentionner l'intégration des techniques du traitement automatique des langues ainsi que la problématique du diagnostic des erreurs et d'un retour sur les production erronées. La dernière section de ce chapitre présente le dispositif CETLEF dans une perspective de synthèse des différentes problématiques introduites.

## 2.1 Tchèque langue étrangère

Bien que l'enseignement pratique du tchèque langue étrangère (TLE) soit d'une tradition relativement riche[1], ce n'est qu'à partir des années 1980 que des publications théoriques traitant ce sujet commencent à apparaître. Après des travaux préliminaires, incitant à la constitution d'un champ de recherche autonome dont l'objet serait une méthodologie spécifique pour l'enseignement du TLE, des participations diverses – en forme de monographies ou d'articles dans des revues spécialisées – commencent à se multiplier aujourd'hui. Nous proposons une vue d'ensemble des travaux existant en nous arrêtant sur ceux qui touchent la présentation didactique de la déclinaison.

### 2.1.1 Besoin d'une méthodologie du TLE

Il est généralement admis que la description traditionnelle de la grammaire tchèque destinée aux locuteurs natifs n'est pas adéquate pour les besoins des étudiants étrangers, voir par exemple Millet (1968), Poldauf et Špruňk (1968), Čermák (1985), Hronek (1985), Hronová (1993), Malinovský (1995), Hrdlička (2002), Nekula (2007). Les problèmes principaux sont vus dans la présentation de la matière grammaticale et du lexique, dans la hiérarchisation des différents phénomènes dans l'enseignement et dans la présentation de la scission entre les deux registres principaux du tchèque – le tchèque écrit / littéraire et le tchèque parlé / commun.

Un recueil d'articles concernant les différentes facettes du TLE est proposé dans Hrdlička (2002). L'auteur traite de nombreux problèmes pratiques et théoriques du domaine tout en soulignant le retard tchèque par rapport aux didactiques des autres langues. Hormis le manque d'une méthodologie du TLE, Hrdlička déplore également l'absence d'études théoriques qui seraient basées sur les notions développées en acquisition d'une langue étrangère, telles que *interlangue*, *transfert*, *interférence*, etc.

Une quinzaine d'années avant le recueil de Hrdlička (2002), une conférence sur la problématique a été organisée à l'Université Charles de Prague, voir Tax (1985). Concernant la description grammaticale, Hronek (1985) réclame la définition d'un ensemble d'éléments grammaticaux et lexicaux qui seraient présentés comme la base du système linguistique tchèque et qui seraient définis dans le cadre de l'approche fonctionnelle de l'école de Prague. Il mentionne l'utilité de la linguistique formelle mathématique qui apporte des outils pour une telle description – la définition explicite des algorithmes de l'analyse et de la synthèse utilisés pour le traitement automatique des langues pourraient trouver une application pra-

[1] Pour une synthèse, voir Hrdlička (2002), p. 111–112.

tique dans l'enseignement. Dans le même esprit, Čermák (1985) souligne l'importance de la dichotomie pragoise du *centre* et de la *périphérie* du système linguistique pour une présentation didactique. Dans cette perspective, le centre serait un ensemble restreint de moyens fréquents et réguliers, et devrait représenter le point de départ pour une pédagogie efficace. La même perspective est adoptée pour l'estimation du vocabulaire actif et passif que devrait posséder un apprenant à un niveau donné. Čermák énumère les principales qualités que devrait avoir une description pédagogique du TLE : simplicité, cohérence, hiérarchisation, sélection des faits linguistiques en fonction des besoins de communication propres aux apprenants étrangers. Le rôle des différents plans linguistiques dans la description devrait être équilibré, la vision fonctionnelle du système devrait être observée à tous les niveaux. De ce point de vue, la présentation didactique de la morphologie flexionnelle nominale doit être envisagée en vue de sa fonction d'exprimer les relations syntaxiques et non pas comme une fin en soi, indépendamment du système.

### 2.1.2 Recherche sur le TLE - des cas concrets

L'élaboration d'une méthode d'enseignement implique nécessairement un choix de méthodologie qui détermine la présentation de la matière à enseigner et à apprendre. Il existe par exemple de nombreuses méthodes de tchèque élaborées avec l'intention de servir au public francophone, voir par exemple Milde et kol. (1948), Ducháček (1949), Lyer (1963), Vlasák *et al.* (1967), Ducháček (1971), Confortiová et Michálková (1972), Man (1986), Lipková et Kissling (1991), Čechová et kol. (1993), Hobzová (1993), Hubáčková (1994), Spilar (1994), Kastler (1995), Hrdlička et Hrdličková (1997), Palíková (2001), Froulíková (2002). Des formulations explicites des critères servant à la structuration didactique de ces méthodes restent cependant assez rares. Nous allons présenter quelques unes ci-dessous.

L'objectif de Hronová (1993) est de fournir une image simplifiée de la grammaire tchèque pour les besoins d'un enseignement sans le recours à une langue tierce. Cet objectif l'amène à la recherche d'une restructuration de la description grammaticale du tchèque, pour qu'il puisse être enseigné et appris avec plus de facilité dans l'objectif de satisfaire en premier lieu les besoins des apprenants en communication quotidienne. L'enseignement de variantes du tchèque parlé, dans les positions où celles-ci rendent le système plus régulier et plus prévisible, est considéré comme un moyen de simplification de la grammaire pédagogique. En ce qui concerne la hiérarchisation de la matière, elle se base sur la fréquence des différentes unités lexicales et des éléments de grammaire dans l'usage général. D'après le témoignage de Hronová, les résultats de cette méthode sont positifs, car elle fournit aux étudiants des moyens qui leur permettent de communiquer, plus ou moins efficacement, dès les premières semaines de leur apprentissage.

Hrdlička (2002, pages 74–77) réclame plutôt une présentation complète de la grammaire sans simplifications, qu'il considère déformatrices. Le défi est de trouver le moyen d'améliorer la présentation grammaticale sans la rendre incomplète ou réduite. Dans Hrdlička (2000), l'auteur propose une présentation didactique des prépositions, traditionnellement jugées comme une matière difficile. Son point de départ est la présentation verticale de la déclinaison. Une telle présentation réside dans l'introduction de paradigmes casuels dans leur intégralité. Il favorise cette approche à la présentation horizontale, qui propose au contraire l'introduction progressive de moyens formels qui servent à exprimer, à travers tous les types paradigmatiques, le sens et la fonction des unités pour une valeur donnée de la catégorie de cas. Le choix de la présentation verticale implique la présentation parallèles des prépositions dans des ensembles établis sur la base de critères sémantiques, par exemple les expressions de temps, de lieu, etc. À part l'usage des prépositions exprimant des relations spatiales et temporelles, une attention particulière est accordée à la présentation des prépositions dans leur usage abstrait pour exprimer la cause, le but, la condition etc.

Un important travail collectif, Hádková *et al.* (2005), Bidlas *et al.* (2005), Adamovičová *et al.* (2005), partagé entre les centres de Prague (Ústav bohemistických studií, Filozofická fakulta Univerzity Karlovy) et d'Olomouc (Katedra bohemistiky, Filozofická fakulta Univerzity Palackého), représente l'élaboration du contenu lexical, grammatical et communicationnel des niveaux de référence A1, A2, B1 et B2, pour le tchèque élaboré d'après le Cadre européen commun de référence pour les langues CECRL (2001). Il s'agit d'un guide d'enseignement proposant la définition des compétences que doit posséder un apprenant à un certain niveau. Cette définition, élaborée sur la base d'un schéma commun pour toutes les langues européennes, est très détaillée et elle recouvre tous les plans linguistiques mis en exercice par l'apprenant. Concernant la déclinaison, les auteurs proposent une présentation horizontale du système casuel avec le parcours suivant : nominatif, accusatif, génitif, locatif, vocatif, datif, instrumental.

Pour citer des travaux dont l'objet est le système verbal, nous pouvons mentionner par exemple Millet (1968), qui traite le problème de la distinction entre les verbes déterminés et indéterminés en tchèque, par exemple *jít* vs. *chodit*. Cette dichotomie est considérée comme un des points problématiques et sa présentation doit être comprise d'après Millet comme une partie du système aspectuel du tchèque dont les difficultés principales sont soulignées par Hrdlička (2005).

La place du tchèque commun dans l'enseignement est intensément débattue – voir Adam (1996), Bermel (2000), Confortiová (1993), Cvrček (2007), Hrdlička (2002), Lukášová (2007), Nekula (2007). La question principale est de trouver un équilibre entre le tchèque littéraire et le tchèque parlé en fonction des besoins de communication de l'apprenant. La focalisa-

tion uniquement sur l'un de ces deux registres est jugée inadéquate car une telle approche ne fournit pas à l'apprenant tous les outils dont il a besoin pour communiquer. De la même façon, un amalgame entre ces deux registres peut être perçu comme une variante déformée de la langue. Malgré ces remarques, l'intégration de certaines formes parlées dans les premières phases de l'apprentissage est parfois considérée comme un élément qui peut rendre le système plus régulier et faciliter ainsi son acquisition, voir Hronová (1993), Nekula (2007).

Un événement important pour le TLE a été la constitution de l'Association des enseignants du tchèque langue étrangère (Asociace učitelů češtiny jako cizího jazyka) qui organise des rencontres d'enseignants dont le fruit est la publication d'actes[2] contenant des interventions sur des sujets variés, classés généralement en trois parties : (1) questions théoriques et méthodologiques, (2) expériences avec l'enseignement du tchèque à des apprenants de langues maternelles diverses, portant sur des problèmes linguistiques ou socio-culturels, (3) présentation de méthodes d'enseignement, conseils pratiques, illustration d'activités dans la classe, etc.

### 2.1.3 Présentation didactique de la déclinaison du tchèque

Étant donné que l'objet de notre étude sont les erreurs dans la déclinaison, nous allons maintenant étudier de plus près les travaux qui traitent ce sous-système linguistique du tchèque. Il s'agit notamment du chapitre sur la flexion nominale dans la monographie de Poldauf et Špruňk (1968) *Cizí jazyk čeština* et d'un article récent de Nekula (2007) intitulé *Systém a úzus – K výuce české deklinace se zřetelem k substantivům*. Ces deux travaux proposent des moyens divers de réduction de la complexité du système de la déclinaison, tel qu'il est présenté dans les grammaires pour les locuteurs natifs. Les quarante ans qui séparent ces deux publications et la proximité dans l'approche des auteurs montrent que cette problématique est toujours d'actualité[3].

L'ouvrage de Poldauf et Špruňk a l'ambition d'être un outil de travail pour les étudiants avancés et pour leurs professeurs en raison de l'inadéquation des grammaires élaborées pour les tchécophones. L'objectif des auteurs est « d'analyser la langue tchèque comme elle est perçue par les yeux et les oreilles d'un étranger, lecteur ou auditeur »[4]. Le but est de développer chez l'apprenant la capacité d'une analyse formelle pour l'aider à identifier les propriétés grammaticales et paradigmatiques des unités linguistiques. Cette approche est intéressante

[2]Voir Ryndová *et al.* (2005), Ryndová *et al.* (2006), Čemusová et Holá (2007).

[3]Il faut noter que Nekula ne cite pas le travail de Poldauf et Špruňk, même si son approche est assez semblable à celle de ses prédécesseurs.

[4]Poldauf et Špruňk (1968), p. 416.

pour nous, car c'est justement cette capacité qui permet à l'apprenant de créer et d'employer correctement les formes déclinées.

Concernant les paradigmes nominaux, les auteurs distinguent quatre formes à la place des sept formes traditionnelles exprimant la catégorie du cas en tchèque (nominatif, génitif, datif, accusatif, vocatif, locatif, instrumental). Cette réduction est possible grâce à l'homonymie de certaines terminaisons qui expriment plusieurs valeurs de la catégorie du cas. Il s'agit de ces quatre formes : (1) la forme de base (B) exprimant soit le sujet, soit le complément d'objet et regroupant les formes du nominatif et de l'accusatif qui sont, d'après les auteurs, les mêmes pour une grande partie des substantifs ; (2) la forme du génitif (G) qui est spécifique pour chaque type ; (3) la forme du datif (D) qui regroupe le datif et le locatif pour les mêmes raisons de similitude que le nominatif et l'accusatif ; la (4) la forme de l'instrumental (I). Le vocatif est considéré comme une « forme appellative » qui est présentée à part car, à la différence des autres formes, elle n'a pas de fonction syntaxique dans la phrase.

Cette présentation permet de réduire le nombre de formes à apprendre et rend le système moins complexe. L'inconvénient de cette approche est la nécessité de l'introduction postérieure des formes qui sortent de ce cadre unifié : il ne s'agit pas uniquement des exceptions mais des ensembles entiers des substantifs qui sont traités traditionnellement comme des phénomènes réguliers. Ainsi, les formes accusatives des substantifs qui ne sont pas égales à leur forme du nominatif (les masculins animés *pán* × *pána* au singulier, *páni* × *pány* au pluriel ; le type *muž* × *muže* au singulier, *muži* × *muže* au pluriel, le type féminin *žena* × *ženu* au singulier) sont présentées comme des idiosyncrasies, ne correspondant pas à la structure des quatre formes amalgamées, bien qu'elle représentent le centre du système substantival du tchèque.

La même problématique est développée dans Nekula (2007) qui souligne la complexité de la déclinaison du tchèque telle qu'elle est perçue par les apprenants étrangers. Ceux-ci se plaignent du nombre de types de déclinaisons, de l'existence de nombreuses exceptions et des variantes. L'auteur se focalise sur les possibilités de la simplification de ce système et examine les différentes solutions. Il s'agit d'après lui de : (1) l'introduction successive des cas dans une vision horizontale par opposition à la présentation verticale ; (2) des modifications de la présentation des paradigmes casuels qui serait différente de celle dans les grammaires pour les autochtones ; (3) le choix d'un registre de la langue avec un système plus simple (intégration des variantes parlées qui reflètent les tendances d'unification des moyens formels exprimant les mêmes significations grammaticales) ; (4) une limitation de la variabilité dans la déclinaison en choisissant les variantes plus récentes du point de vue de l'évolution historique du système ; (5) une limitation de la variabilité par rapport à l'usage réelle, vérifié dans un corpus représentatif de la langue.

Parmi ces outils de simplification de la déclinaison, les plus adéquats d'après Nekula sont la limitation du nombre de paradigmes, au moins temporaire, en écartant les types qui sont peu représentés dans l'usage ou qui ne correspondent pas au contenu lexical des programmes pédagogiques (il s'agit surtout de la déclinaison des mots d'origine étrangère). Une autre possibilité est une présentation alternative des paradigmes. Cette approche se heurte cependant à des problèmes de compatibilité avec la présentation traditionnelle : l'apprenant est confronté non seulement à un nouveau système linguistique mais également à une nouvelle manière de le présenter, différente de ce qu'il a pu rencontrer pendant ses études d'autres langues étrangères. Le choix d'une variante simplifiée de la langue, qui serait rendue plus transparente grâce à l'introduction de moyens formels appartenant aux différents registres (tchèque parlé, tchèque écrit) n'est pas considéré comme une approche adéquate. En effet, les fonctions communicationnelles des deux variantes du tchèque sont différentes, et l'emploi d'un amalgame est considéré comme un handicap pour l'apprenant.

## 2.2 Acquisition d'une langue étrangère

Le domaine de l'acquisition d'une langue étrangère (Second Language Acquisition) est une discipline développée depuis les années 1950 dans une perspective interdisciplinaire entre la linguistique, la psychologie, la pédagogie et la didactique. L'étude des productions langagières des apprenants est généralement admise comme un moyen méthodologique qui peut élucider le processus d'acquisition et apporter des résultats applicables dans la pratique, voir par exemple Klein (1989), Besse et Porquier (1991), Gaonac'h (1991), Ellis (1994, 1997).

L'approche contrastive, la méthode dominante dans les débuts de la discipline, utilisait la comparaison de la langue maternelle de l'apprenant avec la langue cible pour la prédiction des causes de difficultés d'apprentissage. Les vérifications des hypothèses élaborées dans son cadre ont montré que les difficultés d'apprentissage d'une langue étrangère dépassent largement ce cadre et touchent à des mécanismes linguistiques et cognitifs plus complexes. Les investigations menées directement sur le matériau des productions langagières se sont montrées comme nécessaires et elles ont participé au changement de perspective dans la considération de ces données.

### 2.2.1 Revue de la discipline

Avant de présenter les étapes principales du développement du domaine, il sera utile de distinguer avec Klein (1989) les facteurs déterminant le processus d'acquisition : les fac-

teurs extra-linguistiques (impulsion à apprendre, accès à la langue) et les facteurs intra-linguistiques (capacités linguistiques de l'apprenant). L'impulsion à apprendre représente les motivations déterminant l'attitude de l'apprenant envers son apprentissage, l'accès à la langue détermine la façon de l'apprendre et les possibilités de la pratiquer. La distinction entre l'acquisition guidée et non guidée est en relation avec ces deux facteurs. L'acquisition guidée s'effectue par l'intermédiaire d'une description explicite de la langue formulée à cet effet. L'acquisition non guidée consiste en une exposition spontanée de l'apprenant à la langue sans interventions pédagogiques ou métalinguistiques. Cette distinction peut rendre compte de la nuance sémantique entre les termes *acquisition* et *apprentissage* d'une langue étrangère qui sont souvent utilisés comme synonymes. Le terme acquisition rend compte du processus d'ordre linguistique et cognitif de l'appropriation d'une langue qui peut s'effectuer naturellement par l'acquisition non guidée ou dans le cadre d'un enseignement ; le terme apprentissage est lié plutôt à l'exposition guidée de l'apprenant à la langue par l'intermédiaire d'une méthode et il désigne une activité consciente.

La catégorie des facteurs intra-linguistiques est largement définie comme un ensemble de capacités de l'apprenant à apprendre et à utiliser une langue étrangère. Ces capacités peuvent varier en fonction de l'âge, du niveau de la connaissance métalinguistique du système de sa langue maternelle ou des autres langues, etc. Les études des facteurs intra-linguistiques appartiennent déjà à la dimension de l'étude du processus d'acquisition et des production langagières.

La théorie la plus discutée est connue sous le nom de l'*analyse contrastive*, voir Lado (1957). Son présupposé principal est que l'acquisition d'une langue étrangère est déterminée par les structures de la langue maternelle et qu'elle peut être expliquée sur la base du mécanisme général *stimulus – réponse – renforcement*. L'idée que l'enseignement des langues le plus efficace devrait être basé sur la comparaison de la langue maternelle de l'apprenant était apparue déjà dans des travaux antérieurs[5], néanmoins c'était Lado qui a renforcé sa conception par une base théorique solide, inspirée par le courant béhavioriste dans la linguistique américaine de la première moitié du siècle dernier. Skinner (1957) supposait que le langage est un ensemble d'habitudes comportementales. Lado considérait l'acquisition d'une langue étrangère comme le développement d'un comportement langagier influencé par les habitudes et comportements langagiers déjà maîtrisés, provenant de la langue maternelle. Cette influence devait se manifester par un transfert d'habitudes positif pour les structures où la langue maternelle de l'apprenant étaient similaire à la langue étrangère, et par un transfert négatif dans les cas où les deux langue étaient différentes.

[5] Voir Besse et Porquier (1991), pp. 200–201

La version dite « forte » de l'analyse contrastive de Lado supposait que les points similaires des deux langues sont faciles à apprendre et les points différents, donnant lieu au transfert négatif et donc aux erreurs, représentent les sources de difficultés que la comparaison systématique des deux langues devrait prévoir et expliquer. Les travaux empiriques tentant de vérifier la validité de cette hypothèse l'ont mise en cause sur de nombreux points : les erreurs prévues par l'analyse contrastive ne se produisaient pas ou rarement, les locuteurs de langues maternelles typologiquement très différentes produisaient les même erreurs aux mêmes endroits, les erreurs qui ont été prévues et expliquées en tant qu'interférence se produisaient même pendant l'acquisition de la langue maternelle par les enfants natifs, etc.[6]

Ces recherches ont rapidement relevé le fait que ni l'analyse contrastive, ni la notion behavioriste de transfert, ne peuvent servir d'outil universel pour rendre compte ou prévoir des erreurs rencontrées pendant le processus d'acquisition. La conséquence la plus importante de l'échec de l'analyse contrastive « a priori » devrait être vue dans la focalisation de l'intérêt scientifique sur les productions des apprenants en elles mêmes. La recherche des causes des erreurs a été reprise dans les années 1970 et 1980 par les courants nouveaux dans la psycholinguistique inspirée par la psychologie cognitive ou néo-behavioriste, voir par exemple Doca (1981), Slama-Cazacu (1981)[7]. L'approche contrastive modifiée et moins radicale (car, malgré tout, l'influence de la langue maternelle est indéniable sur certains aspects de l'acquisition) a su apporter des résultats significatifs, voir Ragusich (1977), Bautier-Castaing (1977), Giacobbe (1990), Giacobbe (1992).

Les premières analyses des erreurs effectuées directement sur des corpus de productions des apprenants sont déjà apparues comme un complément de l'analyse contrastive et ceci pour les langues maternelles qui n'étaient pas suffisamment décrites pour pouvoir accéder à cette analyse[8]. Tandis que les approches basées sur le behaviorisme considéraient les erreurs dans les productions comme un défaut qui devrait être éliminé par l'application des résultats d'une analyse contrastive adéquate, les erreurs sont désormais appréhendées comme une manifestation de fausses hypothèses sur le fonctionnement de la langue étrangère qui sont construites inévitablement par l'apprenant et qui font partie de sa compétence linguistique provisoire, appelé l'*interlangue*[9], voir Selinker (1972).

L'interlangue est défini comme un système linguistique autre que celui de la langue étrangère mais qui, quel que soit le stade d'apprentissage que l'on appréhende, en comporte certaines composantes[10]. Ces caractéristiques principales sont l'instabilité, la perméabilité, la

[6]Voir par exemple Corder (1981), Dulay *et al.* (1982), Hawkins (2001).

[7]Pour une revue synthétique voir Gaonac'h (1991), pp. 84–121.

[8]Voir par exemple Champion (1974).

[9]Pour désigner la notion de interlanguage des termes divers sont aussi utilisés : système approximatif, compétence transitoire, langue de l'apprenant, dialecte idiosyncrasique ou système approché.

[10]Besse et Porquier (1991), pp. 216–239.

simplification et le caractère évolutif. L'objectif des études portant sur les interlangues est de les « décrire à travers les activités langagières qui les manifestent, pour en caractériser les spécificités, les propriétés et les modalités de leur développement »[11].

Les analyses des erreurs visent non seulement une analyse des performances des apprenants mais aussi une analyse des mécanismes linguistiques et cognitifs qui influencent ces performances. L'analyse des erreurs est donc un procédé complexe dont l'intérêt principal est de décrire et d'expliquer les erreurs pour mieux comprendre les processus et les stratégies d'acquisition des langues étrangères afin d'améliorer l'enseignement, voir Porquier (1977). Les erreurs les plus étudiées sont celles qui sont faites dans les structures grammaticales de la langue étrangère mais aussi dans l'emploi du lexique, des aspects phonétiques, etc. Les erreurs recensées sont classées dans des groupes établis selon des critères divers qui rendent compte, le plus souvent, de leur nature formelle, des règles linguistiques enfreintes par les erreurs ou des mécanismes psycholinguistiques qui aboutissent à ces erreurs[12]. Ces classements servent a posteriori à des évaluations statistiques. Une analyse des erreurs doit donc être basée sur un large corpus de productions d'apprenants, suffisamment représentatif, et être collectées et traitées de manière cohérente[13].

### 2.2.2 Sources de données

Les recherches sur l'acquisition d'une langue étrangère peuvent être menées par l'étude de différents types de données. D'après Ellis (1994) (voir fig. 2.1), les productions langagières des apprenants ne sont en effet qu'une des sources possibles à côté des données recueillies par des techniques expérimentales faisant appel à l'introspection de l'apprenant, son jugement métalinguistique sur les énoncés ou sur sa propre compétence[14].

Dans notre travail, nous nous situons dans le domaine des études de la production et de la compréhension. Dans ce cadre, l'analyse des erreurs peut être effectuée soit dans les textes issus de la production dite « naturelle » ou « libre », soit des productions « sollicitées » dans le cadre d'une étude clinique ou expérimentale.

L'avantage de la *production libre* est l'authenticité des données qui reflètent l'emploi effectif de la langue par l'apprenant dans un contexte libre ou semi-restreint par des facteurs pédagogiques (par exemple, la rédaction d'une dissertation au cours d'un examen). La qualité de ces données est fortement influencée par la situation dans laquelle elles sont produites

[11] Idem p. 216.

[12] Voir par exemple le classement des erreurs dans Champion (1974), Doca (1981). Dans le milieu tchèque, une analyse erreurs des apprenants de l'anglais a été effectuée par Dušková (1969).

[13] Porquier (1977), p. 41.

[14] Cette technique est fréquemment utilisée dans le cadre des travaux basée sur le concept chomskyan de *grammaire universelle*, voir Hawkins (2001), White (2003).

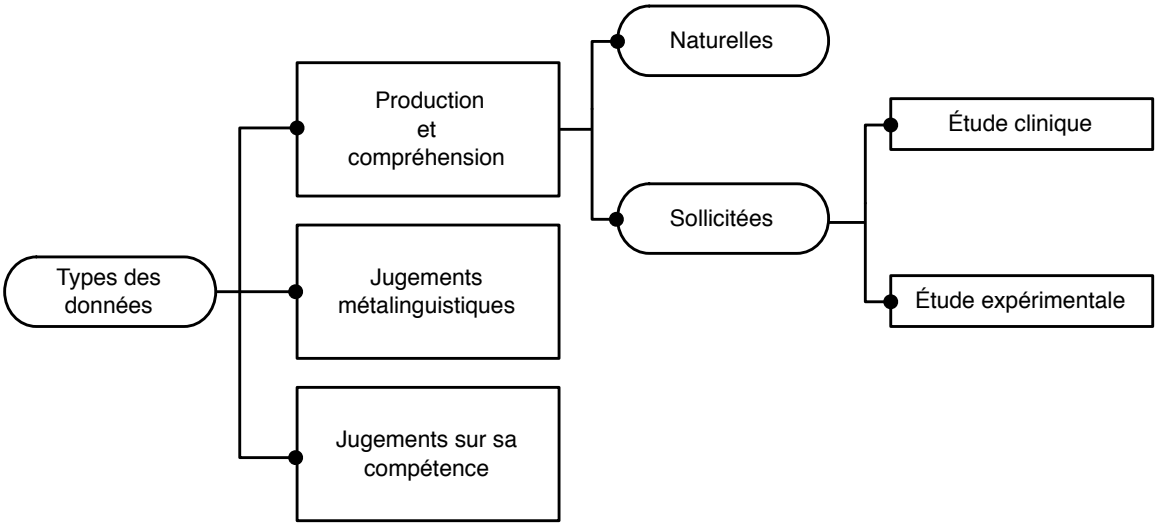



FIG. 2.1 – Types de données d'après Ellis (1994), p. 670

(le stress, les outils à la disposition de l'apprenant, etc.). L'inconvénient principal est une collecte de données coûteuse en temps et effort. Dans la plupart des cas, les données sont orales ou manuscrites et doivent être d'abord traitées manuellement, soit pour une étude directe, soit pour leur retranscription dans un format numérique. Les données collectées de cette façon peuvent également être peu représentatives pour l'étude d'un certain phénomène souhaité, établi au départ d'une enquête, mais qui n'est pas assez fréquent dans les productions des apprenants.

Au contraire, les *données sollicitées* dans un cadre expérimental, permettant de mieux contrôler les facteurs situationnels, doivent nécessairement contenir les phénomènes spécifiques qui ont été établis comme l'objet de l'investigation. La nature des tâches demandées pour la production de ces données peut varier des opérations définies dans un cadre stricte (par exemple la manipulation des structures grammaticales, la création des formes) jusqu'à des activités langagières relativement libres (par exemple les réponses à des questions, la description d'une image, etc.). Néanmoins, l'authenticité de ces données peut être contestée, ainsi que leur ambition de refléter l'état réel de la compétence de l'apprenant.

Les données recueillies à l'aide du dispositif CETLEF, qui traite des productions issues des exercices grammaticaux, appartiennent au second groupe. Avant d'argumenter ce choix avec les avantages et les inconvénients qu'il implique dans l'analyse des erreurs dans la

déclinaison du tchèque, nous allons présenter le courant d'études portant sur les productions libres rassemblées dans des corpus électroniques.

#### 2.2.2.1 Corpus d'apprenants

À la fin des années 1980, une nouvelle impulsion a été donnée aux recherches dans l'acquisition d'une langue étrangère grâce aux progrès de la linguistique de corpus. Effectivement, les principaux défauts des méthodes les plus répandues précédemment sont aujourd'hui vus dans la faible représentativité de leur résultats. La volonté d'utiliser une large base matérielle pour rendre les analyses des productions des apprenants plus crédibles était le motif principal pour la collection des corpus électroniques (Learner Corpora) qui commencent à émerger depuis une quinzaine d'années, voir Granger (1998), Granger *et al.* (2002), Pravec (2002), Tono (2003).

Un corpus d'apprenants est défini comme une collection électronique de données linguistiques authentiques produites par des apprenants dans une certaine situation qui est déterminée par des variables comme le type de texte, le niveau de la maîtrise de la langue, la langue maternelle de l'apprenant, etc. Du point de vue de la maîtrise de la langue cible, les corpus contenant des textes produits par des apprenants au même niveau sont appelés « statiques », ceux qui contiennent des productions recueillis pendant différentes phases de l'apprentissage sont appelés « longitudinaux ».

Les corpus d'apprenants, anglophones pour la plupart[15], sont développés dans le milieu commercial ou académique et la plupart d'entre eux sont compilés en Europe, en Chine et au Japon. Dans le milieu commercial, ce sont les maisons d'éditions Longman et Cambridge qui ont collecté des corpus comptant plusieurs millions de mots et qui ont servi à l'élaboration des dictionnaires et des méthodes d'apprentissage ciblés sur les problèmes des apprenants : Longman Dictionary of Contemporary English (2003) et Cambridge Advanced Learner's Dictionary (2003). Dans le milieu académique, les corpus sont plus nombreux mais aussi plus variables en taille qui varie entre quelques dizaines de milliers de mots (50 000 dans le Montclair Electronic Language Database) et plusieurs millions (25 000 000 mots pour le corpus compilé à Honk Kong university of Science & Technology)[16]. Un des corpus les plus importants dans le milieu académique est ICLE (International Corpus of Learner English) compilé au Centre for English Corpus Linguistics à l'Université de Louvain sous la direction de Sylviane Granger[17] qui contient des essais en anglais (2 millions de mots)

[15]Pour le français, il existe le corpus FRIDA (French Interlanguage Database), voir Granger (2003), Kraif et Ponton (2007). Pour le tchèque, il n'y a pas de corpus de ce type actuellement.

[16]Pour plus de détails voir Pravec (2002).

[17]Voir Granger (2004).

provenant des apprenants d'une dizaines de langues maternelles différentes. La plupart des corpus contiennent les essais provenant des apprenants intermédiaires ou avancés.

Granger (2004) distingue trois types de recherches effectuées dans le cadre de la discipline, appelée CLC (Computer Learner Corpora research) : (1) les travaux sur le cadre méthodologique et analytique, concernant principalement les aspects techniques de la conception des corpus ; (2) des études contrastives des interlangues des différents apprenants et de leur rapport à la norme native de la langue en cours d'acquisition ; (3) les analyses des erreurs assistées par ordinateurs et leurs interprétations, voir par exemple Dagneaux *et al.* (1998).

En comparaison avec les méthodes utilisées préalablement dans les recherches sur l'acquisition d'une langue étrangère, l'atout principal des ces corpus est le volume de donnée et la possibilité des recherches automatisées. Par contre, la plupart des textes, qui sont des copies manuscrites, doivent être saisi au format électronique manuellement, ce qui est relativement exigeant au niveau du temps. Cet inconvénient est cependant équilibré par la possibilité d'addition des métadonnées dans le corpus électronique, qui peuvent porter des informations diverses, caractérisant le texte produit et son auteur.

L'annotation linguistique des corpus d'apprenants représente un défi important, dépassant la discipline CLC et impliquant la problématique du traitement automatique des langues, voir Granger (2003), Izumi *et al.* (2005). En effet, les productions langagières doivent être annotées non seulement de la même manière comme les corpus de textes contenant la langue « ordinaire » (l'ajout des étiquettes morphologiques), mais ce sont également les erreurs commises par les apprenants qui doivent être identifiées et, si possible, classées automatiquement. L'utilisation des outils existants pour l'annotation et pour la correction des textes standard (les correcteurs d'orthographe et de grammaire) ne sont pas suffisants pour cette tâche, voir par exemple Cornu (1994), Johannessen *et al.* (2002). La plupart des annotations des erreurs sont actuellement intégrées dans les corpus manuellement.

Granger *et al.* (2001) présentent un projet d'annotation des erreurs dans le corpus d'apprenants French Interlanguage Database (FRIDA). Les erreurs sont annotées manuellement à l'aide d'un outil qui permet d'ajouter dans le texte des balises caractérisant l'erreur. Chaque erreur est annotée par une étiquette spécifiant le domaine de l'erreur, établi d'après le niveau linguistique où l'erreur se manifeste : graphie, morphologie, grammaire, lexique, syntaxe, registre, style, ponctuation et faute de frappe. L'erreur de chaque domaine peut être spécifiée par une étiquette portant une information plus détaillée. Par exemple, les erreurs de grammaire sont classées directement à l'aide des catégories grammaticales atteintes par l'erreur : *classe <CLA>, auxiliaire <AUX>, genre <GEN>, mode <MOD>, nombre <NBR>, personne <PER>, temps <TPS>, voix <VOI>, euphonie <EUF>*. Chaque forme erronée est étiquetée en fonction de sa catégorie lexicale et de ses propriétés grammaticales. Par

exemple, pour la catégorie lexicale de nom, les spécifications sont : *commun simple <NOM>, commun composé <NOC>, commun complexe <NOL>, nom propre <NOP>*.

Les auteurs soulignent qu'une telle annotation, bien que coûteuse en temps nécessaire pour son intégration dans le corpus, peut servir non seulement les recherches sur l'acquisition mais également contribuer au développement des outils automatiques pour la correction et l'annotation des erreurs[18].

#### 2.2.2.2 Exercices grammaticaux

Une source alternative de données langagière produites par les apprenants d'une langue étrangère sont les exercices grammaticaux. Leur fonction est à la fois de renforcer l'acquisition des compétences grammaticales et de permettre leur contrôle et évaluation[19].

Un exercice grammatical est constitué d'un ensemble de tâches qui demandent à l'apprenant de produire une forme, une structure ou d'accomplir une autre action motivée grammaticalement dans un cadre défini par l'énoncé de l'exercice. Son objectif n'est pas de simuler une situation de communication naturelle, comme cela peut être le cas dans d'autres activités en classe (le dialogue, les jeux de rôle), mais d'exposer l'apprenant à un élément de la grammaire explicitement défini afin de le préparer à son emploi dans les énoncés libres.

Une typologie des exercices grammaticaux peut être établie dans un premier temps sur la base de la nature de la tâche à accomplir[20].

- un **exercice de répétition** consiste dans la simple répétition des structures grammaticales sans que l'apprenant ait à les modifier.
- un **exercice à trous** propose à l'apprenant de choisir ou de produire une unité faisant partie d'un paradigme et de l'insérer dans une certaine position – le « trou » – dans la structure d'un énoncé en fonction de critères syntaxiques.
- un **exercice structurel** consiste dans la reproduction de structures grammaticales en remplaçant les unités de la structure modèle par d'autres unités.
- un **exercice de reformulation** consiste dans le remplacement d'une structure par une autre. C'est le moins restreint des quatre types d'exercices du point de vue de l'activité langagière de l'apprenant.

Dans un second temps, il est possible de distinguer, en fonction des éléments de grammaire qui font partie de l'exercice, les exercices de conjugaison, de déclinaison, d'accord, de passivation, etc.

[18] Pour un tel projet basé sur FRIDA, voir L'haire et Vandeventer-Faltin (2003).
[19] Voir par exemple Bubeníková *et al.* (1975), Besse et Porquier (1991), Heift (2003).
[20] Voir Besse et Porquier (1991), p. 124.

Par rapport à un corpus de productions libres, le recueil de données produites dans des exercices ciblés sur une compétence spécifique peut apporter plus rapidement des données pertinentes : l'avantage des données langagières ainsi recueillies est le rapport entre leur volume et l'information qu'elles contiennent. Les informations sur l'emploi d'une certaine structure, obtenues relativement rapidement à l'aide des exercices, seraient beaucoup plus éparses dans un corpus traditionnel et le nombre de leurs occurrences serait proportionnel à sa taille.

Les productions libres des apprenants sont naturellement affectées par la volonté d'utiliser les structures et le vocabulaire que l'apprenant estime maîtriser suffisamment bien pour pouvoir s'en servir dans la communication, voir par exemple Porquier (1977). Cette façon de contourner des sources potentielles d'erreurs est la cause principale de la disproportion dans l'utilisation de certains éléments linguistiques dans le discours des apprenants. Bautier-Castaing (1977) parle des « stratégies d'évitement » et il en distingue trois types :

- la production d'un énoncé aussi minime et simple que possible ;
- le remplaçement d'éléments jugés difficiles par des structures plus accessibles ;
- le suremploi d'éléments considérés par l'apprenant comme suffisamment acquis.

L'avantage des exercices est qu'elle permettent de focaliser l'attention de l'apprenant sur les tâches sollicitant des structures qu'il aurait tendance à éviter dans une production libre. Une erreur dans ces tâches orientées peut contenir une information précieuse sur l'appréhension de l'apprenant du système linguistique et peut dévoiler une facette spécifique de son interlangue. Cependant, il s'agit d'une activité artificielle qui met en jeu non seulement les compétences de production mais également les compétences de compréhension.

Pour un exercice à trous, qui est le type d'exercice choisi pour la collecte des productions langagière au sein de CETLEF, la tâche de l'apprenant n'est pas uniquement de produire une forme morphologique sur une position donnée dans une phrase mais également d'analyser son contexte syntaxique pour le choix des valeurs appropriées des catégories grammaticales de la forme requise.

## 2.3 Enseignement des langues assisté par ordinateur

L'enseignement ou l'apprentissage des langues assisté par ordinateur (ELAO ou ALAO, CALL pour Computer Assisted Language Learning) est un domaine pluridisciplinaire dont l'objet est l'intégration des outils informatiques dans l'enseignement des langues[21]. Ces outils sont considérés plutôt comme un complément de l'enseignement traditionnel qu'une

[21]Pour des revues synthétiques sur la discipline, voir par exemple Levy (1997), Nerbonne *et al.* (1998), Cameron (1999), Hanson-Smith (2003), Demaizière (2007).

alternative à celui-ci, voir Bertin (2001). Ils sont censés aider l'apprenant dans son apprentissage en lui proposant des activités variées, liées aux possibilités du support électronique, en le rendant plus autonome par rapport aux activités en classe. Une application ELAO typique contient du contenu multimédia, des exercices de différents types et éventuellement des fiches de grammaire spécifiant la matière abordée.

### 2.3.1 Traitement automatique des langues dans ELAO

En fonction de la perspective adoptée – celle d'un utilisateur d'un tel outil, ou celle de son concepteur – de nombreuses disciplines sont mobilisées dans ELAO : la linguistique appliquée, la didactique des langues, la psychologie et les sciences cognitives, l'informatique, l'interaction humain-machine, l'intelligence artificielle, traitement automatique des langues, etc. Avec le développement de l'informatique et l'accessibilité croissante de ses produits au cours des trente dernières années, la plupart des technologies ont été testées ou utilisées pour des objectifs d'enseignement des langues dans le cadre académique ou industriel. Cette problématique est devenue le centre d'intérêt de nombreux chercheurs, participant à des activités d'associations spécialisées comme EUROCALL[22], CALICO[23], IALLT[24] et d'autres.

L'intégration des techniques de traitement automatique des langues (TAL) dans ELAO est la problématique du domaine appelé parfois l'enseignement de langues *intelligemment* assisté par ordinateur (ELIAO, Intelligent CALL – ICALL). D'après Karttunen (1986), Zock (1996), Nerbonne (2003) et d'autres, l'enseignement assisté par ordinateur est un domaine idéal pour la vérification des fonctionnalités des techniques de TAL, car la tâche d'assister un apprenant dans son apprentissage implique virtuellement tous les objectifs visés par cette discipline. Nerbonne (2003), p. 680, énumère les technologies qui sont ou qui ont été appliquées dans les différentes applications de ELIAO :

- analyse morphologique (lemmatisation, génération de formes pour les exercices grammaticaux et pour faciliter l'accès au dictionnaire) ;
- analyse syntaxique (correction des erreurs dans les productions des apprenants, visualisation de la structure de la phrase) ;
- emploi de corpus électroniques (source du matériau linguistique authentique) ;
- alignement de corpus bilingues (assistance à la traduction) ;
- traduction automatique ;

[22] `http://siglp.eurocall-languages.org/`
[23] `https://calico.org/`
[24] `http://www.iallt.org/`

- analyse et synthèse de la parole (entraînement de la prononciation et la compréhension) ;

La nécessité de traiter des entrées langagières non standard, contenant des erreurs dues aux spécificités de l'interlangue des apprenants, représente une exigence supplémentaire portée sur les outils de TAL dans l'ELAO. De ce point de vue, la recherche en acquisition d'une langue étrangère peut apporter de précieuses informations sur le fonctionnement du système linguistique intermédiaire de l'apprenant, que les techniques de TAL devraient modéliser, voir Nerbonne (2003), Schulze (2008). De la même façon, et sous l'angle opposé, des tentatives de formalisation des propriétés des erreurs entreprises pour un diagnostic automatique peuvent être utiles pour l'appréhension de certains aspects de l'acquisition.

Les premiers outils ont été développés au début des années 1980 dans le cadre de recherches sur l'intelligence artificielle[25]. L'objectif de ces outils était de guider l'apprenant dans son parcours, de lui proposer des activités pédagogiques et de lui fournir un retour adéquat sur ses compétences et ses erreurs. Après une période d'enthousiasme, générale pour le champ de l'intelligence artificielle à cette époque, ce sont des approches plus sobres, basées notamment sur la recherche fondamentale en TAL, qui ont gagné du terrain à partir des années 1990.

Dans cette perspective, la problématique la plus urgente est la correction des productions des apprenants et la génération d'un retour approprié en fonction du diagnostic de l'erreur identifiée dans une production, voir Heift et Schulze (2003, 2007).

### 2.3.2 Correction et diagnostic des erreurs

Des méthodes de correction peuvent être appliquées soit sur des productions libres, soit sur des productions provenant de tâches fermées comme dans les exercices grammaticaux. Pour le traitement des productions libres, différentes techniques sont expérimentées pour adapter les correcteurs orthographiques et grammaticaux, destinés à l'usage universel, afin qu'ils puissent prendre en compte des spécificités des textes produits par des apprenants étrangers.

Une des techniques utilisées se base sur le relâchement des contraintes imposées par les règles de la grammaire implémentée dans un parseur (analyseur syntaxique), voir par exemple Faltin (2003), L'haire et Vandeventer-Faltin (2003). Ce relâchement permet de continuer l'analyse d'une phrase malgré la rencontre de structures illicites dont les caractéristiques servent pour un diagnostic de l'erreur. Un correcteur de grammaire française, basé

[25] Voir Last (1989), Bruillard (1997), Dodigovic (2005).

sur ce principe, a été implémenté au sein du projet européen FreeText[26]. Les erreurs d'accord (erreur de nombre ou de genre) peuvent être corrigées efficacement avec cette technique.

Une autre méthode d'analyse syntaxique intègre des « mal-rules » : règles modélisant directement les constructions déviantes produites par les apprenants. Ces règles sont implémentées dans la grammaire du parseur, et si les structures correspondantes sont reconnues dans les énoncés des apprenants, l'erreur est interprétée sur la base de la spécification liée à ces règles, voir par exemple Schneider et McCoy (1998) et Fortmann et Forst (2004).

Par rapport à l'imperfection actuelle des outils disponibles pour la correction des productions libres, Holland et Kaplan (1995), Kraif *et al.* (2004), Tschichold (2006) estiment nécessaire l'adoption d'une approche « pédagogiquement responsable », favorisant l'emploi de techniques de base qui sont suffisamment bien maîtrisées pour réduire le bruit ou le silence à la sortie du traitement. Ces imperfections, qui peuvent être acceptables pour certaines applications dans leur usage « non pédagogique », se révèlent particulièrement perturbantes pour un apprenant au sein d'un didacticiel.

Un exemple d'application ELAO basée sur une technique simple de correction est présenté dans Desmet (2006). Cette application est une plateforme ELAO publiée sur le Web[27] proposant des exercices grammaticaux divisés en trois groupes en fonction de la nature des tâches qu'ils contiennent : des tâches fermées, des tâches semi-ouvertes et des tâches ouvertes. Les tâches fermées peuvent avoir une seule réponse possible (par exemple dans le cadre d'exercices à trous, de questionnaires à choix multiple, etc.). Les tâches ouvertes sont non restreintes dans le nombre de réponses possibles ou probables (par exemple la rédaction sur un certain sujet). Dans les tâches semi-ouvertes, le nombre de réponses possibles est limité, l'activité type dans une tâche semi-ouverte est la modification de structures grammaticales à partir de phrases prédéterminées ou la traduction basique.

La méthode de *approximate pattern matching*, utilisée dans la correction des productions issues des tâches semi-ouvertes, n'est pas à proprement parler une méthode basée sur un traitement linguistique, mais une technique comparant la chaîne produite par l'apprenant avec un ensemble de chaînes proposées comme des réponses correctes possibles. Après cette comparaison, des marques sont insérées dans la production de l'apprenant pour lui signaler l'absence d'un mot dans la chaîne, la présence inappropriée d'un mot dans la chaîne ou une erreur sur un mot.

Par exemple, pour une tâche de traduction de l'anglais vers le français, les deux composants de la tâche sont spécifiés de cette manière :

- la phrase à traduire : *In the evening, my sister Mary often watches the French television.*

[26] Voir `http://www.latl.unige.ch/freetext/`
[27] Voir `http://www.kuleuven-kortrijk.be/franel`.

- la solution possible avec ses alternatives entre crochets : [Le soir, ma soeur Marie regarde [souvent / fréquemment] la [télé / télévision] française.] / [Ma soeur Marie regarde [souvent / fréquemment] la [télé / télévision] française, le soir.]

Un exemple d'une production erronée et de sa correction, contenant les symboles XXX pour un mot contenant une erreur, (XXX) pour un mot en trop et (...) pour un mot manquant :

- la production de l'apprenant : *Le soir, ma soer regarde souvent à la télé français.*
- la correction : *Le soir, ma XXX (...) regarde souvent (XXX) la télé XXX.*

Avec ce retour, l'apprenant est invité à corriger sa production et la solution correcte lui est montrée après le second essai.

L'avantage de cette technique, permettant d'attirer l'attention de l'apprenant sur ses erreurs, réside dans sa fiabilité, cependant toutes les réponses possibles doivent être explicitement spécifiées, ce qui peut devenir problématique, voir impossible pour des tâches moins restreintes, par exemple la reformulation des énoncés, une traduction des phrases plus complexes, etc. Le retour sur la production erronée est néanmoins assez basique et ne met pas en relief les propriétés linguistiques des productions erronées.

Dans la perspective de l'emploi des techniques de TAL pour la correction des erreurs dans un outil ELAO, nous estimons qu'un diagnostic des erreurs issues des exercices grammaticaux à trous pourrait être effectué par des procédés relativement simples basés sur la génération morphologique. Ce diagnostic pourrait être motivé linguistiquement en interprétant l'écart formel entre la forme requise dans une tâche de l'exercice et la production erronée de l'apprenant.

## 2.4 Plateforme CETLEF

Dans cette section, l'outil CETLEF est présenté selon les différents angles qui viennent d'être exposés et qui justifient l'intérêt d'un recueil de données langagières à l'aide des exercices grammaticaux au sein d'une application ELAO publiée sur le Web.

### 2.4.1 Présentation générale de CETLEF

En considérant les avantages et les inconvénients de la compilation de corpus avec des productions libres des apprenants, nous avons décidé de développer une application Web qui permet de collecter les données au sein des exercices grammaticaux contenant des tâches de déclinaison. Les formes requises dans de telles tâches peuvent être facilement accompagnées

par une annotation morphologique, ce qui peut être utilisée pour la recherche des données et pour l'intégration des traitements automatiques.

Le cadre restreint des exercices de déclinaison permet l'intégration d'un module de diagnostic des erreurs qui peut être basé sur des techniques rudimentaires du traitement automatique des langues. L'idée générale de ce diagnostic est qu'une forme erronée produite par l'apprenant peut être calculée automatiquement et les propriétés formelles de cette production peuvent servir comme l'explication de l'erreur. Nous estimons que le développement d'un tel diagnostic peut apporter des observations précieuses sur la nature des erreurs dans la déclinaison.

Les exercices de déclinaison représentent un cadre limité du point du vue des compétences mises en jeu par l'apprenant car la production des formes fléchies dans leur cadre est d'une certaine façon une activité artificielle. Nous estimons néanmoins, qu'une analyse détaillée des erreurs dans ces formes qui soit basée uniquement sur leur propriétés morphologiques, est un bon point de départ pour l'analyse des productions libres et des erreurs au niveaux linguistiques plus élevés.

Finalement, une application Web peut servir non seulement comme un dispositif d'acquisition des données mais aussi comme un outil d'apprentissage orienté vers l'apprenant : les exercices servent pour son entraînement et le diagnostic de ses erreurs peut l'aider dans son apprentissage.

### 2.4.2 Présentation des notions et des termes employés

Dans ce qui suit, nous allons présenter les notions et les termes principaux concernant l'application CETLEF, employés dans notre travail : *tâche*, *lemme*, *consigne grammaticale*, *forme requise*, *annotation*, *production*, *production erronée*, *diagnostic*, *exercice*, *enquête*, *plate-forme auteur*, *plate-forme apprenant*, *base de données*.

#### 2.4.2.1 Tâche

Une *tâche* est une phrase dans laquelle l'apprenant doit décliner un lexème. Ce lexème est représenté par son lemme et il est accompagné par une consigne grammaticale. Le terme tâche peut être également compris comme l'action de décliner le lexème donnée afin de produire la ou les formes requises, demandées par le contexte syntaxique de la phrase.

La structure d'une tâche avec une production erronée, telle qu'elle est présentée dans le texte de cette thèse, est la suivante :

*contexte gauche* ***production erronée*** *contexte droit.*
forme requise cas nombre genre
traduction littérale
'traduction libre' (nº enquête.identifiant de la tâche.identifiant de l'apprenant)

Exemple :

*Mařenka nese plný koš* ***houby*** *a směje se.*
hub gen.pl.f.
Marie porte plein panier champignons et rit *refl*
'Marie porte un panier plein de champignons et elle rit' (1.21.36)

La même tâche, telle qu'elle est présentée à l'apprenant sur CETLEF, est présentées ci-dessous :

Sur la plate-forme apprenant, tous les mot tchèques dans une tâche sont traduits en français et présentés dans un lexique qui est à la disposition de l'apprenant afin qu'il puisse se concentrer uniquement sur les aspects grammaticaux.

#### 2.4.2.2 Lemme

Le *lemme* d'un lexème est sa forme canonique (la forme de dictionnaire). Pour les substantifs, il s'agit du nominatif singulier (par exemple *pán*) ; pour les adjectifs et les pronoms exprimant le genre, il s'agit du nominatif singulier masculin (par exemple *mladý, on*) ; pour les pronoms qui n'expriment pas le genre, il s'agit de la forme du nominatif avec la valeur de la catégorie du nombre correspondante (par exemple *já, se, vy*).

Le lemme de la forme requise est proposée à l'apprenant au sein d'une *tâche* : il s'agît de la forme qui doit être déclinée. Elle est systématiquement accompagnée par sa traduction française pour les mêmes raisons que tous les mots dans les tâches.

#### 2.4.2.3 Consigne grammaticale

Une *consigne grammaticale* est l'information qui spécifie le nombre et éventuellement le genre de la forme requise au sein d'une tâche, par exemple *pl.* dans la tâche présentée ci-dessus. Cette spécification est nécessaire sauf dans les phrases où les valeurs des catégories

du nombre et de genre pourraient être déduites grâce aux règles d'accord, par exemple dans la tâche *Do třídy vcházejí ... (student)*.

#### 2.4.2.4 Forme requise

Une *forme requise* est la forme lexicale qui doit être produite dans une tâche donnée à partir du lemme, en fonction du contexte syntaxique et de la consigne grammaticale. Dans le cadre d'une tâche, il peut y avoir au maximum 3 formes requises. Plusieurs formes requises sont possibles dans les situations suivantes :

- les formes requises possibles sont des variantes fonctionnelles d'une forme casuelle, par exemple *páni*$_{nom.pl.m.}$ × *pánové*$_{nom.pl.m.}$ ; *svou*$_{acc.sg.f.}$ × *svoji*$_{acc.sg.f.}$ × *svojí*$_{acc.sg.f.}$ pour le pronom possessif *svůj*.
- les formes requises possibles sont régies par un mot qui peut avoir deux rections différentes pour exprimer approximativement la même signification, par exemple *Petr je učitel*$_{nom.sg.m.}$ × *učitelem*$_{inst.sg.m.}$ (*Pierre est enseignant*) ; *Dosáhnout úspěch*$_{acc.sg.i.}$ × *úspěchu*$_{gen.sg.i.}$ (*Obtenir un succès*).
- les formes requises possibles expriment deux significations différentes en fonction de la rection choisie de l'élément régisseur, par exemple *létat nad moře*$_{acc.sg.n.}$ × *mořem*$_{inst.sg.n.}$ ; *smést prach pod skříň*$_{acc.sg.f.}$ × *skříní*$_{inst.sg.f.}$.
- les formes requises possibles ont deux fonctions syntaxiques différentes, par exemple *maminka mě*$_{acc.sg.}$ *poslala pro nákup* × *maminka mi*$_{dat.sg.}$ *poslala pro nákup* (*Ma mère m'a envoyé faire des courses* × *Ma mère a envoyé quelqu'un pour faire des courses pour moi*.).

Chaque forme requise au sein d'une tâche peut être accompagnée par un commentaire explicatif qui précise la nature de la variante ou les spécificités de l'emploi des rections différentes.

#### 2.4.2.5 Annotation

Une *annotation* de la forme requise est l'ensemble des informations grammaticales qui la caractériset : le lemme, la catégorie lexicale, le type morphologique, le type paradigmatique, le cas, le nombre, le genre et une éventuelle signalisation d'une alternance morphématique. La définition de cette annotation est présentée dans le chapitre *3 Annotation morphologique*, pp. 31–115.

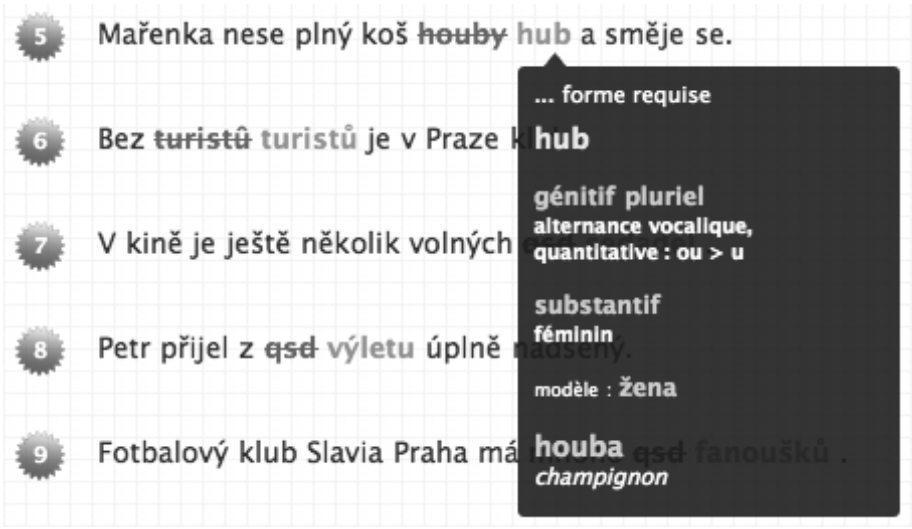
L'objectif de l'annotation de la forme requise est triple : (1) servir pour la recherche des productions dans la base de données en fonction des critères linguistiques ; (2) être publiée

sur la plate-forme apprenant et spécifier explicitement les propriétés morphologiques de la forme requise ; (3) être utilisée pendant le diagnostic automatique des erreurs.

Pour donner un exemple, l'annotation d'une forme requise dans la base de données (voir le chapitre *6 Implémentation*, section 6.2.6, pp.211–214) a cette forme :

| requis | lemme | tagLex | tagMorph | pdgm | cas | num | gen | alt |
|---|---|---|---|---|---|---|---|---|
| hub | houba | subst | N | zn | gen | pl | f | ou > u |

La même annotation, intégrée sur la plate-forme apprenant (voir le chapitre *6 Implémentation*, section 6.3.2, pp.228–236), est présentée ci-dessous :




La forme requise dans une tâche peut être annotée manuellement, ce qui serait problématique dans le cas des productions libres, stockées dans un corpus. Vu le nombre de tâches qui sont proposées dans le cadre d'une enquête, une annotation manuelle ne représente pas un travail inabordable. Pour aider l'auteur à annoter une forme requise, un module d'annotation semi-automatique lui est proposé (voir le chapitre *2 Annotation morphologique*, pp. 111–115).

#### 2.4.2.6 Production

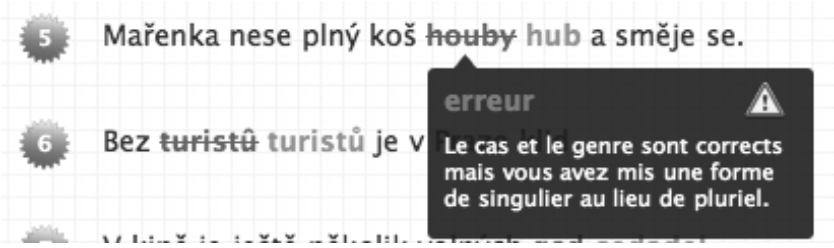
Une *production* est la forme produite par un apprenant dans le cadre d'une tâche. Pour que la production soit considérée comme correcte, elle doit correspondre au moins à une des formes requises.

#### 2.4.2.7 Production erronée

Si une production ne correspond pas à aucune des formes requises au sein d'une tâche, elle est appelée *production erronée*.

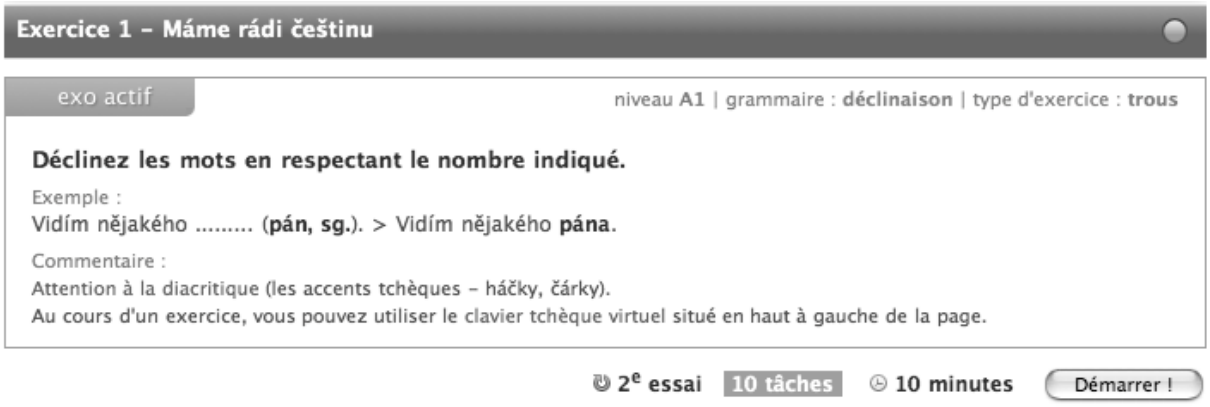
#### 2.4.2.8 Diagnostic

Un diagnostic est le message généré automatiquement (voir le chapitre *4 Diagnostic des erreurs*, pp. 117–169) qui propose une spécification de la nature de l'erreur, si la production ne correspond à aucune des formes requises. Un exemple du diagnostic d'une production erronée, tel qu'il est affiché sur la plateforme apprenant, est présenté sur cette image :




#### 2.4.2.9 Exercice

Un exercice est un ensemble de tâches accompagné (1) par un énoncé, (2) par un exemple d'une tâche et par (3) un commentaire éventuel. Un exemple des ces informations est présenté ci-dessous :




Un exercice publié sur la plateforme apprenant est composé en moyenne de 10 tâches et l'apprenant doit effectuer cet exercice en un temps limité.

#### 2.4.2.10 Enquête

Une enquête est un ensemble d'exercices publiés sur la plateforme apprenant de CETLEF. Les exercices dans l'enquête sont accessibles un par un, c'est-à-dire qu'il y a un seul parcours prédéfini pour tous les apprenants. Ceci permet de recueillir des données d'une façon cohérente car tous les apprenants effectuent les mêmes tâches dans l'ordre établi par l'auteur de l'enquête. Les apprenants peuvent renoncer à continuer dans les exercices, si leur difficulté dépasse le niveau de leur maîtrise du tchèque.

#### 2.4.2.11 Plate-forme auteur et plate-forme apprenant

La plate-forme auteur est l'interface de CETLEF qui permet la création des exercices de déclinaison (voir le chapitre *6 Implémentation*, pp. 226–228). La modification, la suppression et l'ajout des exercices sont possibles à tout moment.

La plate-forme apprenant est l'interface de CETLEF, accessible publiquement sur le Web, qui propose à l'apprenant des exercices de déclinaison (voir le chapitre *6 Implémentation*, pp. 228–236). Les erreurs dans les exercices sont corrigées et leur diagnostic est généré automatiquement. Pour pouvoir accéder à cette plate-forme, l'apprenant doit remplir un formulaire d'enregistrement. Les informations fournies spécifient chaque apprenant par des informations d'ordre sociolinguistique qui peuvent aider pour l'interprétation des ses productions (voir le chapitre *6 Implémentation*, pp. 202–206).

Des fonctionnalités supplémentaires sur la plateforme apprenant permettent d'ajouter à l'outil une utilité dépassant le cadre des exercices. Les apprenants remplissent leur propre base de connaissances en progressant dans les exercices. Cette base contient le lexique, les modèles de déclinaison, les cas et les alternances rencontrés et peut servir comme un aide dans l'apprentissage. Un forum de discussion est également mis à la disposition des apprenants pour favoriser leur échanges.

#### 2.4.2.12 Base de données

Toutes les données utilisées dans l'application CETLEF sont stockées dans une base de données relationnelle (voir le chapitre *6 Implémentation*, pp. 198–225) : les exercices, les informations sur les apprenants et leurs productions. La base de données permet une recherche de données très variée en fonction de toutes les informations au sein de la base.

# Chapitre 3

# Annotation morphologique

Dans ce chapitre, un modèle de la déclinaison du tchèque est réalisé afin d'établir un répertoire d'étiquettes morphologiques (annotation), assignées aux formes requises dans CETLEF : la catégorie lexicale, les catégories morphologiques, les types paradigmatiques et éventuellement les alternances. Les choix théoriques et techniques déterminant l'organisation de ce modèle seront argumentés. Les procédures automatiques de génération des formes casuelles, utilisées pendant l'annotation manuelle de la forme requise et pour la génération du diagnostic automatique sont présentées à la fin de ce chapitre.

## 3.1 Préliminaires

Dans cette section, les bases du système de la déclinaison tchèque sont présentées avant l'introduction de la notion de l'annotation linguistique. La question de couverture d'une telle annotation est abordée par la suite. Enfin, les principes du modèle de la déclinaison, employé pour l'annotation et dans les procédures automatiques sur CETLEF, sont établis sur la comparaison avec les modèles existant.

### 3.1.1 Introduction à la déclinaison du tchèque

La déclinaison (flexion nominale) du tchèque est un système très riche : sept cas, quatre genres, deux nombres, quelques dizaines de types de déclinaison substantivale, adjectivale, pronominal et numérale, une importante homonymie et synonymie des désinences casuelles. Cette situation est une des plus complexes parmi toutes les langues slaves, voir Skalička (1951), p. 65. L'état synchronique est le résultat d'un développement guidé par des tendances différentes (les changements dans le système phonologique, l'analogie, etc.), voir Skalička

(2004), p. 210–248, qui ont remodelé l'ancien système, basé sur les thèmes vocaliques et consonantiques, en une structure dominée principalement par la catégorie grammaticale du genre.

Avant d'exposer ce système plus en détail, il est nécessaire de présenter la structure morphologique d'une forme lexicale en tchèque. Nous reprenons à cet effet les définitions des notions et des concepts principaux telles qu'elles sont présentées dans Komárek (1978), Petr *et al.* (1986b), Čechová *et al.* (2000), Karlík *et al.* (2002) :

- la **forme lexicale** (*slovní tvar, word form*) est un syntagme dans lequel la composante grammaticale (les morphèmes grammaticaux) détermine la composante lexicale, appelée la *base morphologique*.
- la **base morphologique** (*tvarotvorný základ, le radical, stem*) est soit un seul morphème (la racine), soit un complexe de morphèmes de types différents (la racine avec les affixes dérivationnels).
- les **morphèmes** sont les plus petites unités significatives de la langue. Deux types de morphèmes sont distingués : *racines* et *affixes*. Les affixes peuvent être *dérivationnels* (préfixes, suffixes) ou *flexionnels* (les désinences casuelles ou verbales, appelées également le *formant morphologique*, *tvarotvorný formant*). Une réalisation (concrétisation) d'un morphème est appelée un *morphe*. Dans le cas de la racine et des affixes dérivationnels, on parle des morphèmes *lexicaux*, dans le cas des affixes flexionnels, on parle des morphèmes *grammaticaux*.
- du point de vue de la structure sémantique, un morphème grammatical est un ensemble de **sèmes**, unités morphologiques significatives élémentaires, voir également Skalička (1935), Sgall (1960), Sgall (1967). Ainsi, un morphème grammatical « accumule » plusieurs significations.
- les **paradigmes morphologiques**, liés à une certaine catégorie lexicale (les substantifs, les adjectifs, les pronoms, les verbes), sont des systèmes de morphèmes grammaticaux qui déterminent les bases morphologiques dans les formes lexicales. Les *types morphologiques* sont des formes typiques des paradigmes morphologiques.
- les paradigmes morphologiques sont divisés en deux classes : **déclinaison** (flexion nominale) et **conjugaison** (flexion verbale). Les sèmes exprimées dans la flexion nominale sont les valeurs de la catégorie grammaticale du cas (nominatif, génitif, datif, accusatif, vocatif, locatif, instrumental), du genre (masculin animé, masculin inanimé, féminin, neutre) et du nombre (singulier, pluriel).
- les **alternances** sont les changements du radical qui réalisés pendant la flexion. Les réalisations formellement différentes des morphèmes alternés sont des *allomorphes*.

D'après la nature phonologique de la composante du radicale alternée sont distinguées les alternances vocaliques et consonantiques.

Sur la base des définitions exposées ci-dessus, nous allons donc considérer dans ce travail la structure d'une forme lexicale déclinée (forme casuelle) comme étant composée de deux parties : le **radical** (base morphologique, tvarotvorný základ, stem) et la **désinence** (morphème grammatical, tvarotvorný formant) dont la signification est composée de **sèmes** exprimant le cas, le genre et le nombre. Les valeurs de la catégorie du cas, exigées par la rection de l'élément régissant, expriment les fonctions syntaxiques des formes fléchies ; les valeurs des catégories du genre et du nombre sont soit déterminées par les facteurs sémantiques et lexicaux, soit imposées par les contraintes de l'accord grammatical. En fonction de la nature morphonologique du radical et de la désinence, les **alternances** sont réalisées dans certaines formes casuelles.

En fonction de la catégorie lexicale du lexème, il est possible de distinguer entre la déclinaison nominale, adjectivale, pronominale et numérale. Chaque déclinaison possède plusieurs types paradigmatiques. Les paradigmes substantivaux sont classés d'après trois facteurs :

- le **genre grammatical** qui permet de distinguer entre les types de déclinaison masculins animés et inanimés, féminins et neutres ;
- la **nature phonologique** de la périphérie droite (finale) du radical (la partie du mot à laquelle s'attache la désinence casuelle) qui permet d'établir des types « durs » (*pán, hrad, předseda, žena, město, kost*), avec le radical terminé par les consonnes dures (*h, ch, k, r, d, t, n*) ou mixtes (*b, f, l, m, p, s, v, z*) ; des types « mous » (*muž, stroj, soudce, růže, píseň, kuře, moře*), avec le radical terminé par une consonne molle (*č, ř, š, ž, c, j, d', t', ň*) ou mixte ; et des types « contractés » (*stavení, Jiří*).
- la **périphérie du lemme** qui permet de classer les substantifs en types consonantiques (*pán, hrad, muž, stroj, píseň, kost*) et vocaliques (*předseda, soudce, žena, růže, město, moře, kuře, stavení*).

Ainsi, on distingue le type masculin animé dur consonantique *pán*, le type masculin inanimé, dur et consonantique *hrad*, le type masculin animé, mou et consonantique *muž* et ainsi de suite.

La déclinaison adjectivale est divisée en type dur (*mladý*) et mou (*jarní*). Il existe également une déclinaison spécifique pour les adjectifs possessifs (*otcův, matčin*) qui manifestent des traits de déclinaison substantivale et adjectivale à la fois. Les déclinaisons pronominales et numérales sont un ensemble de paradigmes particuliers d'après lesquels sont déclinés des groupes fermés de lexèmes, souvent avec un haut niveau de supplétivisme.

Pour illustration, le paradigme des formes fléchies du substantif *dívka* (*fille*, décomposition en radical et désinence : *dívk* + *a*), avec des alternances dans le datif et le locatif singulier (palatalisation) ainsi que dans le génitif pluriel (-*e*- intercalaire), est le suivant :

| | singulier | | pluriel | |
|---|---|---|---|---|
| nominatif | *dívk* | *–a* | *dívk* | *–y* |
| génitif | *dívk* | *–y* | *dívek* | *–#* |
| datif | *dívc* | *–e* | *dívk* | *–ám* |
| accusatif | *dívk* | *–u* | *dívk* | *–y* |
| vocatif | *dívk* | *–o* | *dívk* | *–y* |
| locatif | *dívc* | *–e* | *dívk* | *–ách* |
| instrumental | *dívk* | *–ou* | *dívk* | *–ami* |

### 3.1.2 Utilité de l'annotation linguistique

L'annotation linguistique, notion issue du traitement automatique des langues et de la linguistique de corpus, enrichie des données textuelles dans un corpus par des informations d'ordre linguistique. L'utilité d'une telle annotation est évidente car les métadonnées élargissent le répertoire de critères pour la recherche de l'information. L'annotation peut être ajoutée automatiquement ou manuellement. Dans le premier cas, elle peut être considérée comme un produit dérivé des diverses techniques de traitement automatique des langues car la fiabilité de l'annotation dans un corpus est souvent un indice important de la qualité des méthodes d'analyse qui la produisent. L'autre possibilité, plus coûteuse mais généralement plus fiable, est la saisie manuelle de l'annotation par des équipes d'annotateurs humains qui suivent des consignes élaborées dans le cadre d'une certaine théorie linguistique.

Les deux corpus majeurs disponibles pour le tchèque, Corpus National Tchèque (ČNK)[1] et Prague Dependency Treebank[2], sont des exemples de corpus annotés. Le premier contient un étiquetage morphologique généré par l'analyse automatique de Hajič (2004). Le second contient également une annotation aux niveaux syntaxique et sémantique, ajoutée manuellement ou par des procédures semi-automatiques, d'après la conception théorique pragoise *description générative fonctionnelle*[3]. Pour d'autres langues, il existe des corpus annotés comme Penn Treebank pour l'anglais ou Tiger pour l'allemand.

Dans les années 1990, une discussion critique sur l'annotation des corpus[4] a été suscitée notamment par les questions de fiabilité, souvent faible, ainsi que par l'opacité de certains

[1] Voir Čermák (1997), l'accès au ČNK est possible sur `http://www.korpus.cz`.
[2] Voir Hajič *et al.* (2001), `http://ufal.mff.cuni.cz/pdt2.0/`.
[3] Voir Sgall (1967), Sgall *et al.* (1986)
[4] Pour un résumé de cette discussion dans Mitkov (2003), p. 453.

cadres théoriques à la base de l'annotation. En effet, l'interprétation de l'annotation peut être considérée comme biaisée par l'approche choisie. Aujourd'hui, il est entendu qu'un corpus annoté est un outil précieux pour un linguiste. Cependant, le répertoire des étiquettes, les règles de leur attribution aux unités dans le texte, et la théorie linguistique sous-jacente, doivent être clairement exposés[5]. La description du répertoire des étiquettes et la spécification de la terminologie utilisée sont des outils nécessaires de « traduction », qui permettent de se servir d'une annotation dans des travaux sur des fonds théoriques divers.

Dans le cadre de CETLEF, une annotation linguistique accompagne les formes requises dans les tâches d'exercices grammaticaux : la forme à décliner est accompagnée par la valeur de la catégorie lexicale, des catégories morphologiques, par son type paradigmatique et par l'indication d'une éventuelle alternance. Le rôle de l'annotation dans CETLEF est triple :

- Accompagner les productions des apprenants par des métadonnées linguistiques qui seraient utiles pour leur analyse. Comme dans le cas des corpus annotées, la première motivation est de **faciliter la recherche de l'information** qui peut se baser sur les étiquettes linguistiques. Ainsi, il sera possible de rechercher, par exemple, toutes les erreurs dans les substantifs appartenant à un certain type paradigmatique dans une forme casuelle spécifique ; toutes les occurrences d'une certaine alternance, etc.
- Servir pour le **diagnostic automatique des erreurs**. L'indication des valeurs des catégories grammaticales, du type paradigmatique et de l'alternance est une information cruciale pour l'interprétation automatique d'une erreur de déclinaison. Un modèle formel de la déclinaison du tchèque, qui puisse être implémenté dans une procédure automatique, est nécessaire pour effectuer cette tâche.
- Être affichée, dans un format adapté pour une **présentation didactique**, sur la plateforme apprenant de CETLEF. Pendant la correction des exercices et lors de l'avancement de l'apprenant dans une enquête, les informations grammaticales sont affichées pour l'aider à intégrer la structure grammaticale et les propriétés formelles de la déclinaison du tchèque[6].

Dans le cadre de ce travail, nous avons élaboré une annotation spécifique, adaptée aux objectifs établis : décrire de manière la plus exhaustive les erreurs de déclinaison reflétant le processus de l'acquisition de la déclinaison du tchèque en tant que langue étrangère. Avant de présenter ce modèle qui, nous allons nous pencher sur l'étendue du matériau linguistique qui doit être couvert par cette annotation.

[5] Voir Hajičová *et al.* (2002b,a), Hajičová (2006). Pour un exemple de spécification du répertoire des étiquettes morphologiques, voir Hana *et al.* (2005).

[6] A propos de l'emploi des outils du TAL pour l'intégration des informations grammaticales explicites dans les plateformes d'enseignements des langues intelligemment assistés par ordinateur voir par exemple Zock (1994), Dokter *et al.* (1998), Nerbonne et Smit (1996), Nerbonne et Dokter (1999). L'utilisation de l'annotation linguistique du PDT dans une application pédagogique est présentée dans Kučera (2006).

### 3.1.3 Couverture de l'annotation

Comme pour toute description grammaticale, l'ambition des différents systèmes d'annotation est de couvrir le nombre le plus large d'éléments linguistiques[7]. Ces systèmes sont destinés à être appliqués à des textes dans un univers ouvert où tout énoncé dans une langue donnée pourrait être annoté avec des étiquettes élaborées à cet effet. Cette approche implique un entretien et un élargissement continuel du répertoire des étiquettes. En effet, il existe nécessairement des éléments périphériques auxquels les systèmes d'étiquetage automatiques ou les annotateurs sont confrontés au fur et à mesure, et qui n'étaient pas prévus lors de la conception du répertoire initial des étiquettes.

Dans le cas d'une analyse des erreurs dans la déclinaison du tchèque par des apprenants étrangers, la situation est quelque peu différente. Le matériau linguistique, exposé aux apprenants dans les premiers stades de l'apprentissage n'est pas choisi de façon arbitraire mais représente un sous-ensemble défini sur la base de critères didactiques et pédagogiques. Les éléments de langue doivent être choisis en fonction de leur complexité, de leur fréquence dans le discours, de leur emploi dans une situation communicationelle spécifique etc., pour qu'une assimilation optimale des différents éléments soit assurée.

Ainsi, l'annotation morphologique nécessaire pour l'étude de la déclinaison, pourrait être élaborée sur un ensemble fermé de moyens linguistiques, qui serait défini par la spécification de la matière lexicale et grammaticale pour un certain niveau de la maîtrise du tchèque. Il ne serait donc pas nécessaire de couvrir toute la complexité de la flexion nominale du tchèque mais uniquement une partie de ce système qui devrait suffire à un apprenant sur un niveau donné. Le dispositif CETLEF pourrait dans ce cas utiliser un système avec une couverture restreinte mais suffisante pour son objectif. La viabilité d'une telle approche sera examiner plus bas en comparant la définition didactique proposée pour le tchèque dans Hádková *et al.* (2005), Bidlas *et al.* (2005), avec le matériau linguistique authentique en univers ouvert.

#### 3.1.3.1 Cadre européen commun de référence

Une spécification des compétences linguistiques, pragmatiques et communicationnelles est proposée dans le *Cadre européen commun de référence pour les langues*, CECRL (2001) et son adaptation pour le tchèque dans SERRJ (2001), Hádková *et al.* (2005), Bidlas *et al.* (2005), Adamovičová *et al.* (2005). Ces travaux proposent une échelle commune pour servir d'évaluation des compétences des apprenants. La maîtrise d'une langue européenne est évaluée sur une échelle de six niveaux : A1 – élémentaire, A2 – élémentaire avancé, B1 –

[7] Voir par exemple Osolsobě (1996), Hajič et Hladká (1997), Hajič (2004), Hajičová *et al.* (2004).

intermédiaire, B2 – intermédiaire avancé, C1 – supérieur, C2 – supérieur avancé. Une définition concise des deux premiers niveaux est proposée dans CECRL (2001), p. 25 :

- **Niveau A1 - Élémentaire** Peut comprendre et utiliser des expressions familières et quotidiennes ainsi que des énoncés très simples qui visent à satisfaire des besoins concrets. Peut se présenter ou présenter quelqu'un et poser à une personne des questions la concernant – par exemple, sur son lieu d'habitation, ses relations, ce qui lui appartient, etc. – et peut répondre au même type de questions. Peut communiquer de façon simple si l'interlocuteur parle lentement et distinctement et se montre coopératif.
- **Niveau A2 - Élémentaire avancé** Peut comprendre des phrases isolées et des expressions fréquemment utilisées en relation avec des domaines immédiats de priorité (par exemple, informations personnelles et familiales simples, achats, environnement proche, travail). Peut communiquer lors de tâches simples et habituelles ne demandant qu'un échange d'informations simple et direct sur des sujets familiers et habituels. Peut décrire avec des moyens simples sa formation, son environnement immédiat et évoquer des sujets qui correspondent à des besoins immédiats.

Cette définition est nécessairement large et laisse beaucoup d'espace pour son application pratique, surtout au niveau grammatical, qui dépend des propriétés de la langue en question. D'après les recommandations du Cadre européen et des différentes approches didactiques[8], les phases initiales de l'apprentissage doivent porter sur les aspects communicationnels de la langue, employée dans des situations modèles auxquelles un apprenant débutant est confrontées. La présentation explicite de la grammaire devrait être réduite au strict minimum. Cependant, les auteurs dans Hádková *et al.* (2005) apportent des arguments pour l'introduction de la grammaire dès les premières phases de l'enseignement du tchèque. La raison principale est justement son caractère flexionnel qui nécessite davantage d'attention et d'explication grammaticale qu'un système avec une flexion beaucoup plus faible, comme l'anglais ou le français.

Pour le niveau A1, il est proposé, en adoptant l'approche de la présentation horizontale de la déclinaison, d'introduire d'abord les formes du nominatif, de l'accusatif et du génitif singulier et pluriel. Une grande partie des types paradigmatiques nominaux, adjectivaux et pronominaux est couverte pendant cette étape. Une connaissance active des formes casuelles restantes (datif, vocatif, locatif, instrumental) n'est pas exigée, sauf pour les pronoms personnels dans certaines constructions, comme *je mi zima* (*j'ai froid*) pour le datif, *mluvit o něčem* (*parler de quelque chose*) pour le locatif, *s někým* (*avec quelqu'un*) pour l'instrumental, et pour les formes courantes de vocatif comme *pane, slečno, studente*. Les apprenants de niveau A1 doivent également connaître les alternances telles que *pes* > *psa, dům* > *domu, žák* > *žáci* etc. L'image de la déclinaison se complète dans la définition du niveau A2 dans Bidlas *et al.* (2005). Une présentation détaillée des types paradigmatiques (idem, p. 156–161) suppose une connaissance approfondie de la déclinaison avec ses particularités les plus

[8]Voir par exemple Besse et Porquier (1991), Martinez (2004).

fréquentes et, par principe, elle correspond à la présentation des paradigmes dans les grammaires traditionnelles.

Bien que la spécification des éléments morphologiques pour les niveaux A1 et A2 soit assez explicite, elle n'est pas formulée de façon à ce qu'elle puisse servir comme base d'une annotation formellement adéquate. Les définitions des deux niveaux sont cependant dotées d'un lexique (nous allons l'appeler *lexique A1A2*) qui devrait fournir du matériau nécessaire à la vérification de la présence des types paradigmatiques et aider ainsi à clore l'espace des moyens que nous devrons étiqueter.

Après que les différents types de déclinaison qui font partie de l'annotation linguistique sur CETLEF auront été présentés, la fréquence des différents types paradigmatiques dans le lexique A1A2 sera étudiée et comparée avec les données équivalentes, disponibles pour la langue employée dans des textes authentiques, voir la section 3.4.8, p. 80. Avant de faire ceci, le lexique A1A2 est examiné du point de vue de la fréquence de ses unités afin d'apporter une réponse sur l'adéquation d'une annotation basée sur un ensemble d'éléments linguistiques restreints.

#### 3.1.3.2 Lexique A1A2

Chacun des deux volumes Hádková *et al.* (2005) et Bidlas *et al.* (2005) contient une liste de mots classés alphabétiquement. Il s'agit d'un vocabulaire dont la connaissance est estimée nécessaire pour les apprenants aux niveaux A1 ou A2.

Concernant la matière lexicale exposée dans les premières phases de l'apprentissage d'une langue étrangère, il est généralement admis qu'elle devrait contenir les lexèmes les plus employés, les plus utiles dans la communication et autant que possible réguliers. Autrement dit, les phénomènes centraux de la langue devraient être présentés avant les phénomènes périphériques – peu fréquents ou irréguliers. Un des critères les plus immédiats pour définir les éléments centraux du lexique est la fréquence des unités dans les textes authentiques d'une langue donnée, voir Nation (2001). Une des motivations principales de la création des dictionnaires de fréquence est d'ailleurs de servir pour la conception de matériaux pédagogiques, voir *Frekvence slov, slovních druhů a tvarů*, Jelínek *et al.* (1961) et *Frekvenční slovník češtiny*, Čermák et Křen (2004).

Certaines méthodes ou collections d'exercices pour le tchèque langue étrangère ont été élaborées à partir de listes de mots classés d'après leur fréquence, voir par exemple Čermák *et al.* (1993), Čermák et Holub (2005). Cette approche a des avantages et des inconvénients : il est indéniable que dans un texte quelconque, il est plus probable de trouver les éléments les plus fréquents dans la langue en général (par exemple *člověk, stát, dělat*) que des éléments

peu fréquents (par exemple *rozedma, kramle*). Cependant la fréquence des unités est naturellement influencée par la composition du corpus dans lequel elle a été calculée. Ainsi, il est possible qu'en respectant strictment la fréquence des unités dans un certain corpus, les matériaux pédagogiques se trouvent orientés vers des problématiques (par exemple la politique, l'économie pour un corpus de textes journalistiques) qui ne sont pas toujours conformes aux besoins d'un apprenant débutant.

La compilation du lexique A1A2 pour le tchèque a été dirigée par des facteurs communicationnels et pragmatiques communs pour toutes les langues, établis dans CECRL (2001), et qui vise à couvrir les besoins immédiats d'un apprenant. Le critère de la fréquence des éléments n'est pas mentionné. Nous nous proposons donc d'examiner le lexique A1A2 du point de vue de la fréquence des unités qu'il contient. S'il existe une correspondance entre les différentes unités dans ce lexique et leur fréquence dans un corpus de langue générale, la pertinence de l'approche quantitative serait vérifiée par un argument indépendant : la liste des éléments choisis sur des critères extra-linguistiques correspondrait à une liste, établie sur la fréquence des éléments. Dans le cas idéal, le lexique A1A2 contenant à peu près 2700 mots serait donc égal à la liste des 2700 mots tchèques les plus fréquents.

Afin de déterminer s'il existe une telle correspondance, il est nécessaire d'attribuer à chaque mot du lexique A1A2 une valeur $V$ exprimant son ordre dans une liste $L$, ordonnée d'après la fréquence décroissante des unités dans un corpus donné. Ensuite, il faut classer les mots du lexique A1A2 en ordre croissant d'après la valeur $V$ et comparer la suite de ces valeurs pour les 2700 unités du lexique A1A2 avec la même suite pour les premiers 2700 éléments de la liste $L$, qui est nécessairement une suite incrémentée d'un point pour chaque élément – pour le mot le plus fréquent, la valeur $V$ égale à 1, pour le deuxième mot le plus fréquent, la valeur $V$ égale à 2, pour le troisième à 3 etc. Dans le cas d'une correspondance complète, la valeur $V$ du mot de l'ordre 2700 dans lexique A1A2 serait égale à la valeur $V$ du mot de l'ordre 2700 dans la liste $L$, c'est-à-dire 2700, ce qui impliquerait que cette valeur est nécessairement la même pour toutes les unités précédentes.

Pour pouvoir effectuer cette expérience, *Frekvenční slovník češtiny* (FSČ), Čermák et Křen (2004) a été utilisé. Il s'agit d'un dictionnaire de fréquence, disponible en format numérique sur CD-ROM, élaboré à partir des textes contenus dans le Corpus national tchèque. Il contient les 50 000 mots communs les plus fréquents, dont la fréquence a été calculée dans le SYN2000 (corpus représentative du tchèque écrit contemporain contenant 100 millions de mots). Le dictionnaire FSČ offre un classement d'après la fréquence absolue, c'est-à-dire d'après le nombre total d'occurrence d'une unité lexicale donnée, mais également d'après la fréquence moyenne réduite – Average Reduced Frequency (ARF), qui est une mesure plus intéressante pour notre objectif, voir Hlaváčová et Savický (2002). La valeur ARF, considé-

rée comme plus fiable, est basée évidemment sur la fréquence des unités mais elle reflète également la répartition des occurrences d'une certaine unité dans l'ensemble du corpus, ce qui permet de diminuer l'influence des textes avec une fréquence trop importante d'une unité spécifique, employée uniquement dans un contexte donné.

Si la liste de 50 000 mots de FSČ est ordonnée d'après la valeur ARF, l'ordre des différents mots dans cette liste est exprimé par la valeur appelée *rank ARF* – elle est égale à 1 pour le lexème le plus fréquent, à 2 pour le deuxième lexème le plus fréquent etc. Dans les calculs suivants, cette liste va correspondre à la liste *L*, introduite ci-dessus. Pour donner un exemple de l'organisation de cette liste, une dizaines de mots classée d'après ARF à partir de l'unité sur la 961$^{e}$ position (rank ARF) d'après la fréquence est présentée ici : *961 student, 961 študent, 962 probíhat, 963 potíž, 964 občanský, 965 stavět, 966 novinář, 967 zboží, 968 instituce, 969 tiskový, 970 tvorba, 971 bohatý*

Le chiffre correspond à la valeur du rank ARF et d'après la conception choisi dans FSČ, cette valeur est égale pour les différentes variantes d'un seul lexème.

Il est maintenant nécessaire d'extraire le lexique A1A2, disponible à l'origine dans un document en format PDF (Portable Document File) et le formater en XML pour faciliter les manipulations informatiques. Ce fichier a ensuite été traité par un script Perl qui a attribué à chaque unité sa valeur du rank ARF, lue dans la liste *L*. Voici un extrait du fichier résultant, contenant les lexèmes de A1A2, ayant comme point de départ l'unité avec le rank ARF 961, comme dans l'exemple précédent : *961 student, 964 občanský, 966 novinář, 967 zboží, 972 dívat se, 974 poslat, 975 únor, 977 červenec, 979 červený, 986 radost*

En observant la différence entre les deux extraits, nous pouvons constater que le lexique A1A2 ne contient pas la variante *študent* de l'élément avec le rank ARF égale à 961, il ne contient ni l'élément 962 *probíhat* et 963 *potíž*, il contient l'élément 964 *občanský* etc. – cette comparaison révèle rapidement qu'il y a des unités présentes dans la liste *L* et absentes dans le lexique A1A2. Ainsi, la valeur du rank ARF des unités du lexique A1A2 n'est pas égale à leur ordre – la plus grande valeur du rank ARF n'est pas égale au nombre d'unités, comme nous l'avons supposé plus haut pour le cas idéal, mais il dépasse largement la limite de 2700 mots les plus fréquents dans le FSČ, voir fig. 3.1.

Il est évident que le lexique A1A2 contient des lexèmes qui sont beaucoup moins fréquents qu'un ensemble de lexèmes correspondant qui serait délimité uniquement sur la base de la fréquence – la courbe correspondrait à la flèche pointillée sur la fig. 3.1.

La valeur du rank ARF monte d'une façon régulière jusqu'aux alentours de la valeur 10 000, atteinte autour du lexème de l'ordre 2 000 dans le lexique A1A2, les 700 éléments restants sont des lexèmes repartis sur une plage entre 10 000 et 50 000 de la valeur ARF. Les derniers lexèmes entre 40 000 et 50 000 sont les suivants : *40381 nastydlý, 40631 nahustit,*

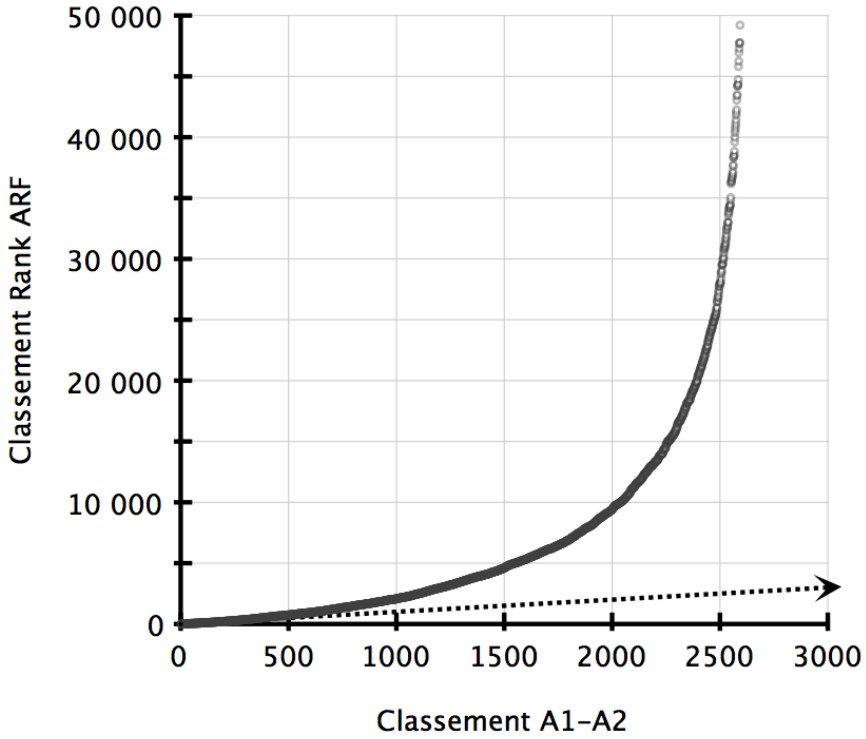



FIG. 3.1 – Évolution du rank ARF dans le lexique A1A2

*40884 decilitr, 41109 zubařka, 41402 optimistka, 41806 muslimka, 42013 běloška, 43037 svíčkový, 43498 teroristka, 44205 přistýlka, 44223 vegetariánka, 44290 navečeřet se, 44401 uprchlice, 44743 indiánka, 45814 informatik, 47282 zavináč, 47669 dramatička, 47735 konzulka, 47792 nekuřačka, 49209 hláskovat*.

En consultant ces mots, il est possible d'identifier les différents domaines thématiques qui doivent être couverts d'après les recommandations du Cadre commun telle que les professions (*zubařka, informatik, dramatička, konzulka*), la nourriture (*svíčkový, navečeřet se, zavináč, decilitr*), la santé (*nastydlý, nekuřačka*), la religion (*muslimka*), etc. Une grande partie de ces lexèmes sont des formes féminines dérivées à partir de formes masculines – ces formes sont généralement plus fréquentes parce que le genre masculin est l'élément non marqué dans l'opposition masculin / féminin.

Pour illustrer encore le rapport entre le lexique didactique et la fréquence des lexèmes dans l'usage général, la fig. 3.2 montre la courbe décroissante indiquant le nombre de formes du lexique A1A2 dans des intervalles de 100 point définis par la valeur du rank ARF. Il serait

également possible de parler de la saturation des intervalles de 100 points établies sur cette valeur.

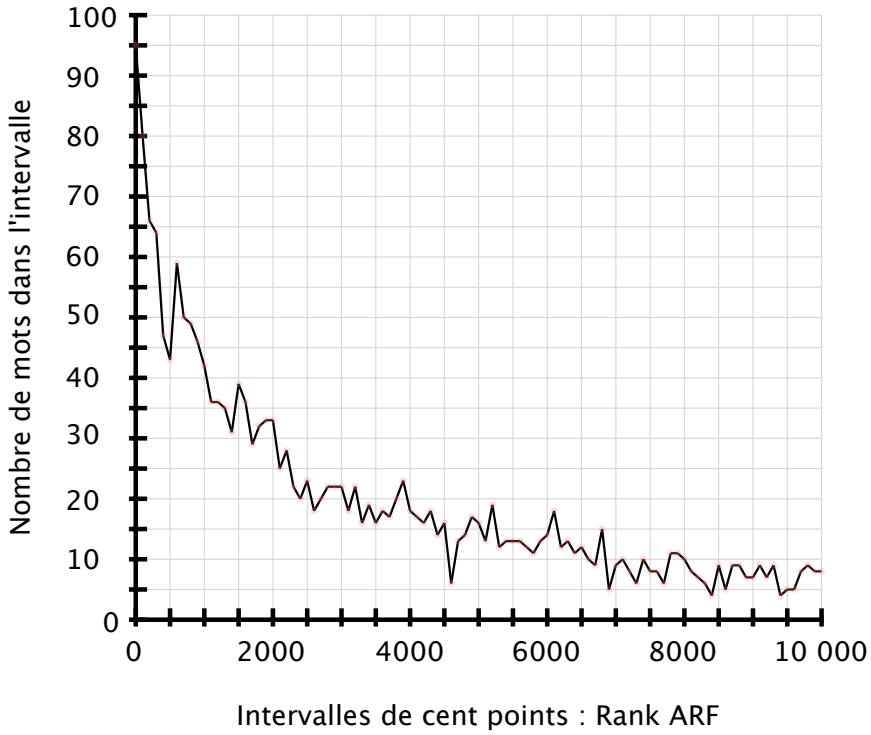



FIG. 3.2 – Saturation des intervalles de 100 points dans A1A2

Nous pouvons lire sur ce graphique que 95 lexèmes parmi les 100 premiers mots dans la liste *L* du FSČ sont présents dans le lexique A1A2 – ce n'est pas étonnant car les lexèmes les plus fréquents sont des mots grammaticaux employés indépendamment du contenu du texte. Pour les mots entre 100 et 200 points de la valeur ARF, il y 80 mots qui sont présents dans le lexique A1A2, pour l'intervalle 200 – 300, le nombre d'éléments baisse à 66, pour l'intervalle 300 – 400 à 64 etc. Pour les mots autour de la valeur 9 000 et 10 000 du rank ARF, il y a à peu près entre 5 et 10 mots dans chaque intervalle. Dans le cas idéal, c'est-à-dire si le lexique A1A2 correspondait exactement à la liste définie sur la fréquence, la courbe serait droite avec une seule valeur sur l'ordonnée égale à 100.

Grâce à ces mesures, il est évident que le matériau linguistique présenté dans les définitions des niveaux A1 et A2, bien qu'il soit considéré comme basique de point de vue des fonctions communicationnelles d'un apprenant étranger, ne correspond pas strictement à une liste de mots établie uniquement sur la fréquence : il existe des lexèmes très fréquents qui ne sont pas présents dans le lexique A1A2, et inversement, certains lexèmes peu fréquents dans les textes communs figurent dans le lexique A1A2. Une étude détaillée sur les motivations

des différents écarts sortirait largement du cadre de notre travail, et serait peu utile, car la composition des lexiques didactiques est par principe arbitraire : elle dépend des intentions de l'auteur d'un tel lexique, du public pour lequel il est destiné, etc.

L'idée exposée au début de cette section – celle de restreindre l'annotation sur un ensemble fermé d'éléments linguistiques définis par une conception didactique ou par la fréquence des unités – ne se montre donc pas raisonnable. Il sera nécessaire de viser la couverture la plus large et de concevoir un système d'annotation des types paradigmatiques de manière à ce que les nouveaux éléments puissent être facilement introduits au fur et à mesure en fonction des besoins révélés « sur le terrain ».

### 3.1.4 Morphologie automatique

L'ambition des approches formelles en linguistique est de formuler une description explicite des éléments de la langue et de leurs relations. Cette description doit être basée sur des critères testables et elle doit permettre une modélisation du système linguistique suffisamment adéquate pour qu'elle puisse être utilisée dans des applications d'analyse et de génération automatique.

Les travaux sur la modélisation formelle du tchèque ont été commencés dans les années 1950 et ils ont cristallisés dans une approche théorique : *description fonctionnelle générative*, voir Sgall (1967), Sgall *et al.* (1986), Sgall *et al.* (1988), Sgall (2006). Cette théorie relie la linguistique générative avec les principes de la linguistique fonctionnelle, élaborée dans le cadre structuraliste de l'école de Prague. Ses traits principaux sont : (1) la description de la structure de la phrase à l'aide d'une procédure générative ; (2) une approche stratificationnelle qui distingue des niveaux hiérarchiques de description allant du niveau phonétique au niveau tectogrammatique (sémantique) ; (3) la syntaxe de dépendance, le rôle central du verbe dans la structure de la phrase et l'importance de la notion de valence ; (4) l'accent sur la représentation de la structure sémantique de la phrase comprenant les informations sur le dynamisme discursif reflété dans la division en thème et rhème.

Le traitement de la morphologie est la condition nécessaire pour pouvoir accéder aux traitements des niveaux supérieurs (syntaxe, sémantique). Dans Sgall (1967), la description de la déclinaison du tchèque est conçue dans la perspective générative. La langue est modélisée comme une hiérarchie de niveaux qui possèdent chacun des unités spécifiques et une syntaxe propre. La génération des phrases est effectuée par la retranscription des unités entre les niveaux voisins. La relation entre un morphe, qui est une unité du niveau phonologique, et le morphème, qui est une unité du niveau morphématique, est donc d'ordre fonctionnel : le morphe représente le morphème, le morphème est la fonction du morphe.

Des morphes différents peuvent avoir la même fonction, dans ce cas, il s'agit de la synonymie (par exemple *–a, –e* pour exprimer le génitif singulier du masculin animé dans *pán–a* × *muž–e*). Au contraire, un morphe peut exprimer plusieurs morphèmes, il s'agit donc ici de l'homonymie (par exemple la désinence *–e* dans le vocatif singulier du masculin animé *pan–e*, et dans le génitif singulier du masculin animé *muž–e*). Les alternances, effectuées pendant la traduction des unités morphématiques vers le niveau phonétique, sont gérées par un ensemble de règles contextuelles, appliquées sur les radicaux et les désinences.

Les travaux sur le traitement automatique de la morphologie tchèque, voir par exemple Cejpek *et al.* (1982), Kirschner (1983), Pognan (1972, 1988, 1992) sont largement inspirés par ce modèle. Une revue des problèmes pratiques, liés à l'implémentation de cette approche dans des applications concrètes, est présentée dans Bémová et Králíková (1988). D'après les auteurs, l'analyse morphologique effectuée sur la base des types paradigmatiques et des règles pour la réalisation des alternances est plus effective qu'un traitement basé sur les inventaires des formes lexicales, car un tel modèle est universel et ne nécessite pas la maintenance laborieuse d'un dictionnaire. L'inconvénient de cette approche réside principalement dans les difficultés liées à la classification des paradigmes, des variantes dans la déclinaison et des alternances. Cette classification exige une description beaucoup plus explicite que celle contenue dans les grammaires traditionnelles.

Pour contourner les problèmes liés à la modélisation linguistique, les outils actuels d'analyse morphologique automatique se basent sur des mécanismes travaillant avec des dictionnaires de radicaux et des listes de terminaisons élaborés sur des critères plus convenables pour un traitement informatique. Des règles linguistiques, qui seraient appliquées sur les combinaisons des morphèmes, ne sont pas employées, et les phénomènes liés aux alternances sont traités par des techniques diverses. La conception de ces outils est motivée par les besoins d'efficacité au niveau informatique et la nécessité du traitement de larges volumes de données. Nous allons présenter deux de ces outils.

Le premier est un analyseur morphologique avec une désambiguïsation probabiliste, développé par Hajič (2004)[9]. Une des versions de ce taggeur a été utilisée pour l'annotation morphologique de ČNK et de PDT. Le principe de l'analyse est basé sur la segmentation de chaque mot en deux parties : la racine et la terminaison. En bref, la racine est la partie d'un mot qui ne change pas dans les différentes formes lexicales, tandis que la terminaison est la partie qui varie. Chaque occurrence de mot est lu à partir de la fin (la périphérie droite). La chaîne de caractères finaux est comparée systématiquement avec une liste des terminaisons dans un lexique morphologique, contenant les racines et les terminaisons avec un identifiant du paradigme flexionnel associé à chaque unité. Si la terminaison du mot analysé est

[9]Voir également sur `http://ufal.mff.cuni.cz/pdt/Morphology_and_Tagging/Morphology/index.html`

trouvée dans la liste, la chaîne restante à gauche est considérée comme radical. Pour que la concaténation du radical et de la terminaison puisse être considérée comme une forme lexicale existante du tchèque, le radical supposé doit être également retrouvée dans le lexique et les informations grammaticales des deux éléments doivent correspondre. Si c'est le cas, une étiquette morphologique et le lemme, associé au radical dans le lexique, sont imprimés à la sortie. Pendant l'étape suivante, la sortie est désambiguïsée (l'élimination des étiquettes inadéquates) à l'aide des calculs probabilistes.

L'étiquette morphologique est composée de 15 positions qui peuvent contenir les valeurs appartenant à des ensembles définis en fonction de la nature de l'information grammaticale décrite : la catégorie lexicale et sa spécification, le genre, le nombre, le cas, le genre du possesseur, le nombre du possesseur, la personne, le temps, le grade, la négation, la variante. Pour donner un exemple, voici l'étiquette pour la forme lexicale *prezident* :
`NNMS1-----A----`
Différentes positions peuvent être occupées ou rester libres en fonction de la catégorie lexicale du lexème annoté. Étant donnée la combinatoire possible des valeurs dans différentes positions, il existe environ 3500 étiquettes possibles pour le tchèque : un corpus relativement petit de 1,8 millions de mots contient plus que 2000 étiquettes différentes.

Dans le système de Hajič, les terminaisons sont classées dans des ensembles paradigmatiques, conçu sur la base des paradigmes traditionnels. La différence réside dans la prise en compte des alternances. Pour les alternances régulières, les parties alternées font directement partie de la terminaison et non pas du radical, comme c'est le cas dans une approche motivée linguistiquement. Par exemple, le substantif *kočka* est traditionnellement segmenté en radical *kočk–* (avec un allomorphe alterné *kočc–* devant la désinence *–e*) et la désinence correspondante. Dans le lexique de Hajič, la racine est *koč* et l'ensemble de désinences contient les terminaisons *–ka, –ky, –ku, –ce,* etc.

Cette conception des paradigmes casuels augmente considérablement leur nombre car ils sont définis non seulement par les désinences mais également par la périphérie droite du radical qui alterne. Dans Hajič (2004), il y a plus de 200 types paradigmatiques pour la déclinaison, au lieu de 14 types traditionnels. Les mots contenant des alternances irrégulières ou des alternances dans la racine (par exemple *sníh* > *sněhu*), sont considérés formellement comme une terminaison qui s'attache à une racine vide et toutes les formes de leur paradigme doivent être inscrites dans le lexique.

Le second outil, Ajka[10], présenté dans Sedláček et Smrž (2001), est un analyseur morphologique qui implémente la description algorithmique de la morphologie de Osolsobě (1996). Le concept linguistique sous-jacent à cet analyseur est basé sur la segmentation du lexème

[10]Une interface publique de l'analyseur est disponible sur `http://nlp.fi.muni.cz/projekty/ajka/`.

en radical et terminaison, conformément à l'approche traditionnelle de Komárek (1978), Petr *et al.* (1986b) et d'autres. Dans l'approche de Osolsobě, le radical peut être ensuite décomposé en racine et l' « intersegment » qui représente la partie du mot qui change à cause des alternances. Cela permet, comme chez Hajič, de rester dans une procédure de génération et d'analyse des formes casuelles effectuée par simple concaténation des différents éléments. Pour l'exemple donné ci-dessus, la segmentation du substantif *kočka* est la suivante : la racine *koč* + l'intersegment *k* + la terminaison *a* ; pour la forme alternée, la segmentation serait la racine *koč* + l'intersegment *c* + la terminaison *e*.

Les lexèmes avec des alternances à l'intérieur du radical sont segmentés de la même façon, ce qui donne comme résultat, par exemple pour le lexème *hůl* avec une alternance vocalique *hol–e*, la segmentation suivante : la racine *h* + l'intersegment *ůl* + la terminaison ∅, le radical *h* + l'intersegment *ol* + la terminaison *e*. Les paradigmes définis sur la base des couples *intersegment—terminaison* nécessitent l'élaboration d'un classement très détaillé et encore plus large que dans le cas de Hajič : Osolsobě (1996) indique 370 types paradigmatiques pour les substantifs, Osolsobě *et al.* (1998) en utilisent 380.

Les deux modèles présentés sont efficaces au niveau informatique dans un univers ouvert. Cependant, les solutions choisies pour contourner la représentation explicite des alternances, qui sont des règles applicables en fonction du contexte, ne sont pas, d'après notre avis, adéquates pour être utilisées dans un système simulant l'activité langagière d'un locuteur humain. C'est pour cette raison que, dans le cadre de ce travail, un modèle de la déclinaison autonome a été élaboré dans CETLEF. Dans la partie suivante, nous allons présenter ce modèle qui est basé sur des listes de désinences et des règles contextuelles pour la réalisation des alternances.

### 3.1.5 Modèle de la déclinaison dans CETLEF

Pour pouvoir diagnostiquer les erreurs de déclinaison, nous devons disposer d'un générateur de formes déclinées qui puisse travailler par la concaténation du radical et de la désinence, accompagnée par une éventuelle réalisation d'une alternance.

Le diagnostic est effectué par la comparaison de la forme erronée, avec un ensemble de formes qui sont générées sur la base des propriétés formelles du système de la déclinaison thèque : une erreur de cas est identifiée si la forme erronée correspond à une autre forme casuelle du lexème en question ; une erreur de nombre est identifiée si la forme erronée correspond à la forme fléchie exprimant le nombre opposé ; une erreur de l'alternance est identifiée si la forme erronée correspond à la concaténation du radical et de la désinence sans que l'alternance soit effectuée, etc.

Pour faire ceci, il faut disposer d'un mécanisme de génération des formes fléchies dont le fonctionnement serait un modèle des activités de l'apprenant afin qu'il puisse de produire également les mêmes erreurs. Nous supposons que les erreurs de déclinaison peuvent être induites par une défaillance dans une ou plusieurs des opérations suivantes[11] : (1) choix des valeurs de la catégorie du cas, du nombre et du genre ; (2) classement du lexème dans le paradigme approprié et le choix de la désinence correspondante aux valeurs des catégories grammaticales ; (3) réalisation, si cela est nécessaire, des alternances.

Le mécanisme doit donc pouvoir générer non seulement un paradigme « correct », qui suffit pour l'identification des erreurs de cas et de genre (par exemple **kočku* au lieu de *kočky* pour le cas, **kočky* au lieu de *koček* pour le nombre), mais il doit être également capable de générer des formes erronées pour l'identification des erreurs dans l'alternance (**kočke* au lieu de *kočce*) et des erreurs de genre ou dans l'attribution du paradigme de déclinaison (par exemple **kočků* au lieu de *koček*).

Les composants nécessaires pour la génération des formes fléchies sont donc (1) une classification des types paradigmatiques établie exclusivement sur la base des désinences et (2) une classification des alternances ainsi que des règles de leur réalisation.

Le fonctionnement du générateur des formes casuelles est exposée à la fin de ce chapitre[12]. Dans les sections suivantes, nous allons présenter les éléments qui sont nécessaires pour son fonctionnement et qui font partie de l'annotation linguistique des formes requises dans les exercices sur CETLEF : la catégorie lexicale, le type morphologique, les catégories grammaticales (cas, nombre, genre), les paradigmes de déclinaison et les alternances.

## 3.2 Catégories lexicales et types morphologiques

La classification traditionnelle des lexèmes tchèques en dix catégories lexicales (substantifs, adjectifs, pronoms, numéraux, verbes, adverbes, prépositions, conjonctions, particules, interjections) est basée sur trois critères[13] :

- le critère **sémantique** d'après lequel les catégories lexicales sont appréhendées comme réalisation des concepts cognitifs, logiques et philosophiques. On distingue les catégories sémantiquement autonomes (substantifs, adjectifs, verbes, adverbes) et des catégories complémentaires (pronoms, numéraux, interjections) et dépendantes (prépositions, conjonctions, particules).

[11] Voir le chapitre *4 Diagnostic des erreurs*.
[12] Voir la section 3.6, p. 102.
[13] Voir Petr *et al.* (1986b), Havránek et Jedlička (2002), Karlík *et al.* (1995) et d'autres.

- le critère **syntaxico-fonctionnel** qui prend en compte le rôle des mots dans la structure de la phrase et du texte (sujet, objet, prédicat, épithète, attribut, circonstance, mot outil).
- le critère **morphologique** qui permet de classer les mots d'après la nature de leur flexion (ou par l'absence de flexion).

Un « mode d'emploi » algorithmique pour l'attribution d'une catégorie lexicale à un mot est présenté dans Čechová *et al.* (2000), p.91, voir la fig. 3.3. Ce classement, conçu dans

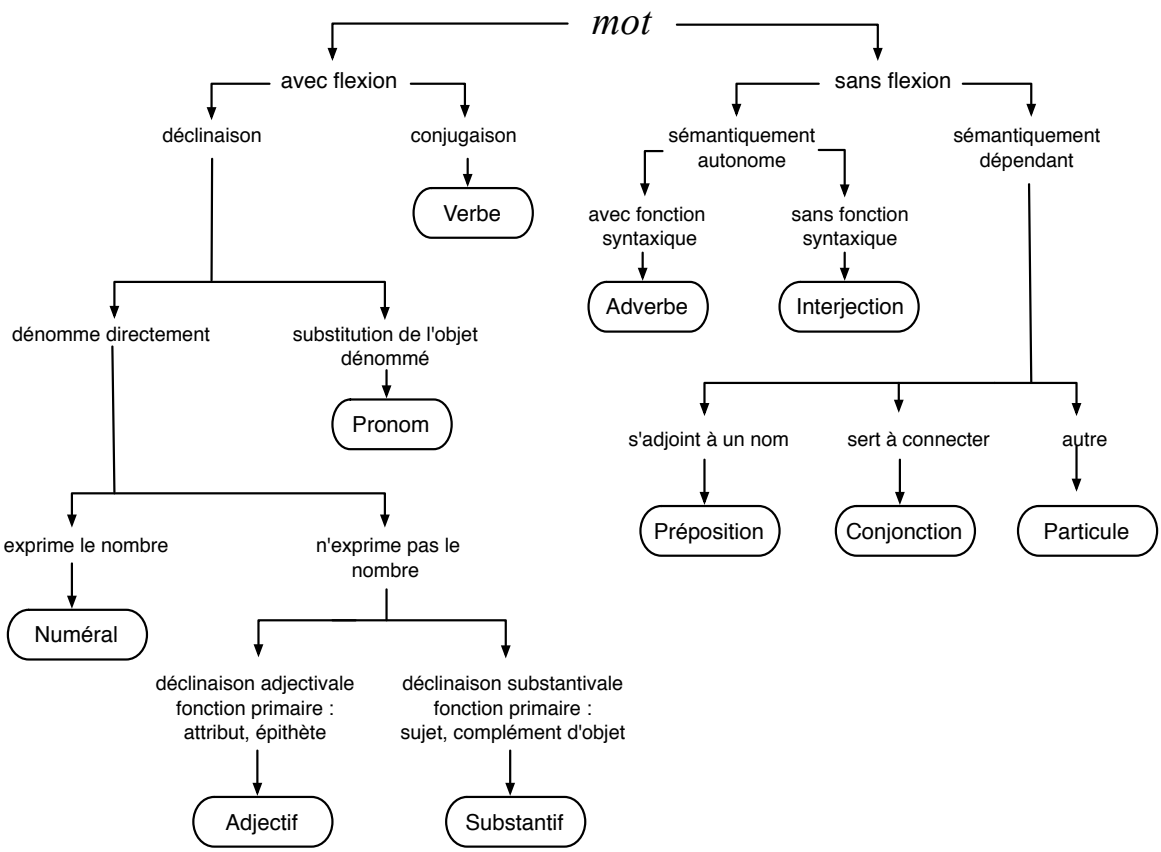



FIG. 3.3 – Catégories lexicales d'après Čechová et al. (2000, p. 91)

la perspective du fonctionnelle, peut être considéré comme une approche souple et cohérente, surtout si l'ambition de la description grammaticale est d'enregistrer tous les aspects sémantiques et fonctionnels du lexique, voir Komárek (1956). En même temps, il représente nécessairement un amalgame de trois critères différents, qui peut mener à des interprétations diverses suivant le critère choisi. Par exemple *každý* (*chacun*) est désigné du point de vue de sa fonction comme un pronom, mais morphologiquement, il s'agit d'un adjectif décliné d'après le type adjectival dur *mladý*. De la même façon, le lexème *vedoucí* (*dirigeant*) peut

appartenir à deux catégories différentes en fonction de sa fonction syntaxique et de sa valeur sémantique : s'il fonctionne comme un épithète (*vedoucí pracovník*) et qu'il exprime une qualité, il est considéré comme un adjectif ; s'il ne détermine pas un autre substantif et qu'il a une fonction de sujet ou de complément d'objet, en exprimant une entité (*Vedoucí přišel. Hledám vedoucího.*), il est considéré comme un substantif. La prise de fonction d'une catégorie lexicale par un mot d'une autre catégorie est appelé une *transposition lexicale*, voir Karlík *et al.* (2002), p. 117. Dans cette conception, entérinée dans Petr *et al.* (1986b), les catégories lexicales sont donc considérées comme perméables et ouvertes.

La possibilité de classer un mot dans plusieurs catégories crée des ambiguïtés du point de vue d'un traitement automatique. Par exemple Hlaváčková (2001), p. 67, montre les problèmes avec l'annotation classique des catégories lexicales dans un corpus oral : un parseur rudimentaire qui cherche des groupes nominaux, identifie la suite *cestující cestující* (il s'agit de la répétition dans un discours parlé) comme une suite *adjectif* + *substantif*. Des problèmes de nature différentes sont présentés dans Jelínek *et al.* (1961), p. 23, qui présentent les difficultés liées au double classement des unités lexicales lors des travaux sur un dictionnaire de fréquences. Cependant, comme le montre par exemple Mey (1966), l'intégration des critères syntaxiques peut être utile pour l'analyse automatique. En effet, l'information sur la fonction syntaxique d'un élément est déjà portée par sa catégorie lexicale et ne doit pas être cherchée dans le contexte.

La dominance du critère morphologique pour l'étiquetage dans CETLEF se montre évident compte tenu de l'objectif de notre travail. La solution adoptée pour l'intégration des informations contenues dans la catégorie lexicale réside donc dans la séparation des critères employés traditionnellement pour l'attribution d'une seule étiquette et dans la définition de deux catégories autonome :

- Le premier groupe que nous appelons le **type morphologique**, classe les lexèmes strictement en fonction de leur forme morphologique et des propriétés de leur flexion. Par exemple les lexèmes *mladý, hajný, který, čtvrtý* seront classés ensemble car leur déclinaison est identique (déclinaison adjectivale) ; dans l'approche traditionnelle, ils seraient désignés respectivement comme un adjectif, un substantif, un pronom et un numéral.
- Le second groupe, pour lequel nous reprenons l'appellation **catégorie lexicale**, prend en compte les critères syntaxiques et fonctionnels. Les règles d'attribution des différentes valeurs de cette catégorie suivent les recommandations des grammaires traditionnelles présentées ci-dessus.

L'existence des deux étiquettes différentes permet d'effectuer des recherches indépendantes en fonction des deux critères : il peut être utile de regrouper tous les lexèmes apparte-

nant à une certaine déclinaison sans prendre en compte leur catégorie lexicale et vice-versa. Dans l'annotation de la forme requise, le type morphologique est enregistré comme valeur de l'attribut `tagMorph`, la catégorie lexicale comme valeur de l'attribut `tagLex`.

La liste des valeurs dans l'attribut `tagMorph` avec une description et des exemples est présenté ci-dessous :

| `tagMorph` | *description* | *exemples* |
|---|---|---|
| `N` | déclinaison nominale | *pán, žena, stavení ; rád ; teplo ; sto, tisíc, pětka, milión, ...* |
| `A` | déclinaison adjectivale | *mladý, jarní ; hajný, vstupné, který, každý, čí, můj, ...* |
| `M` | déclinaison adjectivale mixte | *otcův, matčin* |
| `P` | déclinaison spécifique pronominale | *já, ty, on, ten, náš, ...* |
| `C` | déclinaison spécifique numérale | *dva, pět, kolik, málo* |
| `X` | sans déclinaison | *blond, nealko, ...* |

Dans la présentation des paradigmes de déclinaison (voir p. 56), les différents types morphologiques seront spécifiés comme des ensembles homogènes de différents types paradigmatiques.

Les valeurs de l'attribut `tagLex` sont les suivantes :

| `tagLex` | *description* | *exemples* |
|---|---|---|
| `subst` | fonctions d'un substantif | *pán, žena ; hajný, vstupné ; vedoucí, plzeňské ; atašé ; ČR ...* |
| `adj` | fonctions d'un adjectif | *mladý ; otcův ; blond, ...* |
| `pron` | fonctions d'un pronom | *já ; který, čí ; jeho, ...* |
| `num` | expressions du nombre | *dva ; jeden ; tisíc, milión, ...* |

La correspondance basique – non marquée – entre les valeurs du type morphologique et de la catégorie lexicale est *substantif – déclinaison nominale* et *adjectif – déclinaison*

*adjectivale* ou *déclinaison adjectivale mixte*. Pour les pronoms, la situation est plus complexe car à côté de la correspondance *pronom – déclinaison pronominale*, il y a une grande partie de pronoms qui se déclinent comme des adjectifs. Du point de vue historique, la déclinaison adjectivale, appelée également la *déclinaison composée*, est le résultat d'une agglutination d'une forme nominale avec un pronom déictique, il y a donc de nombreuses ressemblances entre les deux types de déclinaison. Du point de vue synchronique, une union des deux types et une classification plus adéquate génétiquement n'aurait pas d'utilité pratique. La classe la plus hétérogène est celle des numéraux, rassemblés en un groupe sur la base de leur fonction d'exprimer le nombre. A côté de la déclinaison particulière numérale *dva, pět, kolik, málo*, tous les types de déclinaison sont représentés.

Les incompatibilités possibles entre la catégorie lexicale et le type morphologique sont affichées dans le tableau suivant :

| | subst | adj | pron | num |
|---|---|---|---|---|
| N | | | | *tři, čtyři ; sto, tisíc, milión, miliarda* |
| A | *hajný, vedoucí, vstupné* | | *který, čí, můj, týž, sám* | *druhý, první* |
| P | | | | *jeden* |
| C | | | | |
| M | | *otcův, matčin* | | |
| X | *atašé* | *blond* | *jeho, jejich* | |

**Remarque :** L'axe horizontal contient les valeurs de l'attribut `tagLex`, l'axe vertical les valeurs de l'attribut `tagMorph`.

Dans la présentation didactique des catégories lexicales sur la plateforme apprenant, l'accent est mis sur les aspects morphologiques. L'introduction de nouveaux modèles pour la déclinaison des « adjectifs substantivés » ou des « substantifs adjectivaux » est supposée comme une complexification de la description et cette approche est écartée. Dans le cas des lexèmes où les valeurs des deux attributs ne sont pas compatibles, la nature de cet écart est

spécifiée en procédant systématiquement de la forme vers la fonction. Ainsi, les lexèmes comme *vstupné, hajný, vedoucí, příbuzný*, ne sont pas présentés comme des substantifs mais comme des formes qui sont morphologiquement des adjectifs et qui ont acquis, au cour de l'évolution historique (pour *vstupné, hajný*) ou dans un contexte donné (pour *vedoucí, příbuzný*), une fonction de substantif.

## 3.3 Catégories morphologiques

Les catégories morphologiques dans le système nominal du tchèque sont le *cas*, le *genre* et le *nombre*. Concernant l'annotation dans CETLEF, la place et le traitement de la catégorie lexicale qui peut être considérée comme la catégorie morphologique la plus générale, vient d'être exposée dans la section précédente. Pour un substantif ou un pronom personnel, la valeur de la catégorie du cas est imposée par des critères syntaxiques, la valeur du nombre est imposée par le contenu cognitif exprimé (un ou plusieurs objets) et la valeur du genre par des critères sémantiques et lexicaux. Pour un adjectif ou un pronom relatif ou interrogatif, ces valeurs sont imposées par le membre régissant du syntagme à l'aide des règles d'accord.

### 3.3.1 Catégorie de cas

Sgall *et al.* (2003), p. 17–18, proposent une classification des cas tchèques, élaborée suivant deux axes : (1) les cas dont la fonction primaire est syntaxique et les cas dont la fonction primaire est sémantique ; (2) les cas qui dépendent habituellement des verbes et les cas qui dépendent habituellement des substantifs. Ainsi, les cas peuvent être classés en trois groupes :

- syntaxique, dépendant du verbe : *nominatif, accusatif, datif* ;
- syntaxique, dépendant du substantif : *génitif* ;
- sémantique, dépendant du substantif : *locatif, instrumental* ;

Du point de vue fonctionnel, le *vocatif* est situé en dehors du système casuel car il a une fonction appellative. Il présente également des particularités formelles car c'est uniquement en vocatif que l'on peut observer la réalisation des alternances *k* > *č, h* > *ž, c* > *č* (*člověk* > *člověče, bůh* > *bože, otec* > *otče*) qui sont caractéristiques de la dérivation (par exemple *lék* > *léčit, záloha* > *založit, tvůrce* > *tvůrčí*).

Dans l'annotation de la forme requise, le cas est enregistré comme valeur de l'attribut `cas`. Les valeurs possibles de cet attribut sont les suivantes :

| cas | *description* |
|---|---|
| nom | nominatif |
| gen | génitif |
| dat | datif |
| acc | acusatif |
| voc | vocatif |
| loc | locatif |
| inst | instrumental |

### 3.3.2 Catégorie de nombre

Le tchèque distingue le *singulier* et le *pluriel* qui reflètent la distinction sémantique *'un objet'* × *'deux ou plusieurs objets'*. Dans certaines formes des appellations de parties paires du corps humain (par exemple *rukama, očima*), il existe des résidus de l'ancien duel, qui n'est plus vivant dans le tchèque actuel. Certains lexèmes ont un paradigme défectif par rapport à la catégorie du nombre. La manifestation de cette défectivité est double :

- les **pluralia tantum** sont des substantifs dénombrables dont le paradigme est limité uniquement au pluriel (par exemple *nůžky, dvířka*). Il s'agit d'une restriction formelle car les formes hypothétiques singulières (**nůžka, *dvířko*) ne sont pas attestés dans le lexique. Dans notre classement des types paradigmatiques, les pluralia tantum sont classés dans des types autonomes.
- les **noms collectifs** (par exemple *ptactvo, hmyz, listí*) sont des substantifs non dénombrables pour lesquelles la réalisation des formes du pluriel est limitée par des raisons sémantiques. Dans notre classement des types paradigmatiques, les collectifs sont considérés comme les membres des paradigmes correspondants non défectifs.

Dans l'annotation de la forme requise, le nombre grammatical est enregistré comme valeur de l'attribut `num`. Les valeurs possibles de cet attribut sont les suivantes :

| num | *description* |
|---|---|
| sg | singulier |
| pl | pluriel |

### 3.3.3 Catégorie de genre

Le tchèque distingue le *masculin animé* et *inanimé*, le *féminin* et le *neutre*. Comme nous l'avons vu, cette catégorie figure parmi les principes organisateurs de la déclinaison

du tchèque. L'attribution du genre grammatical à un substantif peut être dirigée par sa relation avec le contenu cognitif qu'il désigne : le genre naturel des êtres vivants correspond dans la plupart des cas à leur genre grammatical. Pour les objets non vivants, cette relation est arbitraire et il est nécessaire de prendre en considération des critères formels. Pour un substantif inconnu, sa périphérie droite permet souvent de déterminer, avec un degré de certitude variable, son genre. Par exemple, un substantif fictif comme *mlaba* sera avec le plus de probabilité un féminin du type *žena*, mais cela peut être également un masculin du type *předseda* ou, dans un cas extrême, un neutre d'origine grecque décliné d'après le type *drama* (*mlaba, mlabatu, mlabatem, ...*).

Pour des séries de substantifs comme $srdce_n$, $dozorce_m$, $opce_f$ ou $topič_m$, $bič_i$, $tyč_f$, il est impossible de déterminer le genre uniquement en ayant recours à leurs propriétés formelles, car celui-ci est nécessairement inscrit dans le lexique. Dans ce cas, le sème du genre d'un substantif, exprimé par la désinence casuelle, serait plutôt un reflet de la propriété lexicale du lexème dans son ensemble. Ainsi, la catégorie du genre est souvent désignée comme une catégorie morpho-lexicale. La situation est différente pour les désinences des adjectifs et de certains pronoms, car le sème du genre est ici une réelle variable (*mlad–ý, –á, –é*).

Une présentation de la déclinaison du tchèque élaborée dans cette perspective est proposée dans Poldauf et Špruňk (1968), p. 48 :

| **périphérie du substantif** | **masculins** | **neutres** | **féminins** |
|---|---|---|---|
| *-a* | uniquement des personnes *předseda* | uniquement les emprunts au grec *drama* | toujours *žena* |
| *-o* | | toujours : *město* | |
| *-e* | uniquement des personnes *soudce* | parfois *moře* | parfois *růže* |
| *-í* | uniquement des personnes *vedoucí* | les autres après masculins et féminins | uniquement des personnes *vedoucí* |
| *consonnes dures* *-h, -ch, -k, -r, -d, -t, -n* | toujours *pán, hrad* | | seulement 'čtvrt', 'nit', 'smrt' |
| *-st* | seulement 'host', 'most', 'prst' | | toujours *kost* |
| *consonnes molles* *-č, -ř, -š, -ž, -j, -c, -ď, -ť, -ň* | les personnes *muž* | | parfois *píseň* |
| | autres parfois *stroj* | | |
| *consonnes mixtes* *-b, -f, -l, -m, -p, -s, -v, -z* | les personnes *pán / muž* | | parfois *píseň / kost* |
| | autres parfois *stroj / hrad* | | |
| *-ý* | toujours *mladý* | | |
| *-é* | | toujours *mladé* | |
| *-á* | | | toujours *mladá* |

TAB. 3.1 – Types de déclinaison et le genre d'après Poldauf et Šprunk (1968, p. 48)

Les auteurs donnent un outil pour l'identification du genre et du type paradigmatique d'un substantif par la considération de sa périphérie droite et par la considération des informations supplémentaires, comme le caractère de l'objet désigné (une personne × un objet) ou l'origine du mot. Ils montrent les exceptions et les ambiguïtés possibles dans l'attribution d'une classe pour les lexèmes dont les propriétés formelles ne le permettent pas (voir les encadrés sur le schéma).

Dans l'annotation de la forme requise, le nombre grammatical est enregistré comme valeur de l'attribut gen. Les valeurs possibles de cet attribut sont les suivantes :

| gen | *description* |
|---|---|
| m | masculin animé |
| i | masculin inanimé |
| f | féminin |
| n | neutre |

## 3.4 Paradigmes de déclinaison

Comme nous l'avons déjà vu, les formes déclinées sont générées en tchèque par deux opérations distinctes : le choix de la désinence et l'exécution d'une éventuelle alternance. L'amalgame de ces deux opérations peut être efficace au niveau informatique, mais il n'est pas adéquat du point de vue linguistique – notre système de génération doit modéliser le plus fidèlement possible l'activité de l'apprenant, qui procède, comme nous le supposons, et comme nous le constatons sur des erreurs commises, par l'application successive des différentes opérations. Ainsi nous avons besoin d'une définition de paradigmes casuels qui serait basée sur des listes de désinences, telles qu'elles sont attachées au radical du lexème décliné. La réalisation d'une alternance est guidée par des règles applicables en fonction de la désinence et ne peut être calculée qu'après l'attribution de celles-ci.

### 3.4.1 Définition des paradigmes

Avant d'introduire la classification de types paradigmatiques dans CETLEF, nous présentons quelques définitions de la notion *paradigme*, élaborées sur des critères formels. La définition de la structure d'un paradigme est donnée par Stump (2001), p. 13 :

> "The structure of paradigms in given language is determined by the inventory of morphosyntactic properties avaiable in that language. Given a lexem L of category C, the structure of L's paradigm is determined by the set S of morphosyntactic properties appropriate to C and by the co-occurence restrictions on these properties : for each maximal consistent subset of S, there is a corresponding cell in the paradigm of L."

Dans le cas de la déclinaison du tchèque, la structure du paradigme d'un lexème L appartenant à une des catégories C (substantifs, adjectifs, pronoms, numéraux) est donc déterminé

par les propriétés morphosyntaxiques qui sont les valeurs de la catégorie du cas, du genre et du nombre.

Mel'čuk (1993), pp. 355–362, travaille dans sa théorie Sens–Texte avec deux acceptations du terme *paradigme*. Le *paradigme*$_1$ est défini ainsi :

> "Soit un lexème L de la langue *L*, appartenant à la classe distributionnelle K. Nous appelons *paradigme*$_1$ du lexème L la liste qui contient toutes les combinaisons possibles en *L* de tous les grammèmes caractérisant les lexèmes de la classe K(L) et telle que pour chaque combinaison de grammèmes, cette liste spécifie le(s) lexe(s) de L qui l'exprime(nt)."

Il n'est pas nécessaire d'exposer ici la terminologie de Mel'čuk pour comprendre que dans le cas de la déclinaison, le *paradigme*$_1$ est l'ensemble de toutes les formes casuelles d'un lexème. Pour représenter le *paradigme*$_1$, on utilise la *matrice paradigmatique* où chaque case est associée à une combinaison particulière de grammèmes (dans notre terminologie, il s'agit des *sèmes*), et on met dans la case la forme lexicale correspondante.

Les propriétés formelles d'un *paradigme*$_1$ constituent le *type de paradigme*$_1$ qui est défini par : a) l'organisation grammaticale du *paradigme*$_1$, i.e. l'ensemble des combinaisons de grammèmes qui y sont représentées ; b) la défectivité du *paradigme*$_1$, i.e. le nombre de cases vides dans la *matrice paradigmatique* ; c) les moyens formels utilisés pour construire tous les lexes du *paradigme*$_1$ - sont les moyens morphologiques : les affixes et les modifications (les alternances, les reduplications, etc.), les schémas accentuels, ainsi que les moyens non morphologiques, utilisés dans la construction des formes analytiques. La seconde acceptation du terme *paradigme* chez Mel'čuk est donc *paradigme*$_2$ qui veut dire le *type de paradigme*$_1$. Ainsi, il est clairement distingué entre l'appellation de l'ensemble des formes fléchies pour un certain lexème et les propriétés de cet ensemble. Il est possible de dire que deux lexèmes différents avec deux *paradigmes*$_1$ différents, ont le même *paradigmes*$_2$ – les deux lexèmes sont fléchis de la même manière, selon un seul type.

Pour la représentation de paradigmes casuels dans le modèle de la déclinaison employé dans CETLEF, la distinction entre les notions *paradigmes*$_1$ et *paradigmes*$_2$ est strictement observée. La seule différence avec la conception de Mel'čuk est que les types de paradigmes – *paradigmes*$_2$ – sont définis uniquement par les affixes (les désinences casuelles) et non pas par les modifications représentées par les alternances.

#### 3.4.1.1 Paradigmes de désinences casuelles

L'ensemble des désinences qui s'attachent à un lexème pour créer ses formes casuelles (son *paradigme*$_1$) est appelé dans ce travail le *paradigme de désinences casuelles*. Voici la définition formelle de cette notion :

**Préliminaires** Soit l'ensemble $L$ contenant tous lexèmes qui peuvent être déclinés (qui porte la catégorie grammaticale du cas) : $L = \{socha, mech, ono, ...\}$. Soit l'ensemble $T$ contenant toutes les désinences casuelles du tchèque standard ou du tchèque commun : $T = \{\#, e, i, a, ami, ama, ech, ...\}$. Soit l'ensemble $C$ de tous les triplets $c$ possibles contenant des valeurs de catégories de cas, de nombre et de genre : $C$ = {(nom, sg, m), (gen, sg, m), (inst, pl, f), ... }.

**Définition** Un *paradigme de désinences casuelles* est donc défini comme un ensemble $P$ tel que $P \subset T \times C$. Il s'agit d'un ensemble fini contenant des couples $(t, c) \mid t \in T \ \& \ c \in C$ : $P$ = {(*e*, (nom, sg, m) ), (*e*, (gen, sg, m) ), (*i*, (dat, sg, m) ), ...}

**Exemple** Soit le substantif *pán* appartenant à l'ensemble $L$. L'ensemble $P$ pour ce lexème est donc : $P$ = {(#, (nom, sg, m)), (*a*, (gen, sg, m)), (*u*, (dat, sg, m)), (*ovi*, (dat, sg, m)), (*a*, (acc, sg, m)), (*e*, (voc, sg, m)), (*u*, (loc, sg, m)), (*ovi*, (loc, sg, m)), (*em*, (inst, sg, m)), (*i*, (nom, pl, m)), (*ové*, (nom, pl, m)), (*ů*, (gen, pl, m)), (*ům*, (dat, pl, m)), (*um*, (dat, pl, m)), (*y*, (acc, pl, m)), (*i*, (voc, pl, m)), (*ové*, (voc, pl, m)), (*ech*, (loc, pl, m)), (*y*, (inst, pl, m)), (*ama*, (inst, pl, m))}. Cet ensemble contient les désinences employées dans les formes casuelles du substantif *pán* et les valeurs des catégories grammaticales qui leur sont assignées. Cet ensemble est représenté sur la matrice suivante :

| cas | num | gen | term |
|---|---|---|---|
| nom | sg | m | # |
| gen | sg | m | a |
| dat | sg | m | u |
| dat | sg | m | ovi |
| acc | sg | m | a |
| voc | sg | m | e |
| loc | sg | m | u |
| loc | sg | m | ovi |
| inst | sg | m | em |

| cas | num | gen | term |
|---|---|---|---|
| nom | pl | m | i |
| nom | pl | m | ové |
| gen | pl | m | ů |
| dat | pl | m | ům |
| dat | pl | m | um |
| acc | pl | m | y |
| voc | pl | m | i |
| voc | pl | m | ové |
| loc | pl | m | ech |
| inst | pl | m | y |
| inst | pl | m | ama |

Il peut exister des éléments dans l'ensemble P avec une valeur identique des trois catégories grammaticales et des désinences distinctes. Pour l'exemple ci-dessus, il s'agit des éléments (*u*, (dat, sg, m)) × (*ovi*, (dat, sg, m)) ; (*u*, (loc, sg, m)) × (*ovi*, (loc, sg, m)) ; (*i*, (nom, pl, m)) × (*ové*, (nom, pl, m)) ; (*i*, (voc, pl, m)) × (*ové*, (voc, pl, m)) ; (*ům*, (dat, pl, m)) × (*um*, (dat, pl, m)) ; (*y*, (inst, pl, m)) × (*ama*, (inst, pl, m)).

Pour les besoins de la classification formelle des variantes dans le cadre de CETLEF, il a été décidé de distinguer uniquement deux types de variantes : les variantes fonctionnelles et les variantes de registre.

- **variantes fonctionnelles** sont des variantes équivalentes du point de vue du registre. L'une ou l'autre variante est choisie en fonction de divers facteurs d'ordre sémantique, stylistique, rythmique ou autre, par exemple loc. sg. du substantif *hrad* : *na hrad–ě* × *o hrad–u* ; nom. sg. f. du pronom *tvůj* : *tv–á* × *tv–oje* ; dat. sg. du substantif *pán* : *pan–u* × *pán–ovi* ; gen. sg. *tábor–a* × *tábor–u* ; loc. sg.*v mlék–u* × *v mléc–e* etc.
- **variante de registre** est une variante non standard, provenant du tchèque parlé, par exemple l'inst pl. du substantif *pán* : *pán–y* × *pán–ama* ; nom. sg. m. de l'adjectif *mladý* : *mlad–ý* × *mlad–ej* ; dat. pl. du substantif *hrad* : *hrad–ům* × *hrad–um*, etc.

Dans ce travail, nous ne spécifions pas la nature des différentes variantes car cette problématique dépasse largement la cadre de ce travail. Une présentation générale de la variation dans la morphologie ainsi qu'une esquisse des critères pour leur classement est donnée dans Petr *et al.* (1986b), pp. 266–276. Les auteurs distinguent les *variantes équivalentes* qui ont une distribution identique, et les *variantes divergentes* qui diffèrent soit dans leur valeur stylistique, soit dans leur valeur sémantique. En effet, la complexité de la variation dans la morphologie tchèque a plusieurs sources et ce sujet est beaucoup discuté en bohémistique, voir par exemple Sgall et Hronek (1992), Šimandl (2000), Sgall et Panevová (2004), Prošek (2007). L'une des sources principales est la relation déséquilibrée entre la codification et l'usage, causant les hésitations des locuteurs natifs à choisir des moyennes formelles appropriées. Sur la plateforme apprenant, une spécification plus détaillée d'une certaine variante peut toujours être ajoutée dans le cadre d'une tâche donnée. Par exemple, dans une tâche comme *Petr hledá v ...* (*korpus, sg.*) « *Pierre cherche dans un corpus.* », les deux variantes de la forme requise *korpuse* et *korpusu* seront indiquées comme correctes, la forme *korpusu* sera en plus accompagnée d'une notice, indiquant que c'est la meilleure forme des deux.

Dans la représentation formelle, l'attribut `var` porte l'information sur la variante :

| `var` | *description* | *exemple* |
|---|---|---|
| `fnct` | variante fonctionnelle | *na hradě*$_{fnct}$ × *o hradu*$_{fnct}$ |
| `reg` | variante de registre | *ženami* × *ženama*$_{reg}$ |

S'il s'agit des variantes fonctionnelles, la valeur *fnct* de l'attribut `var` est assignée à toutes les variantes qui partagent les mêmes valeurs des catégories grammaticales. S'il s'agit d'une variante de registre, uniquement la variante non standard est marquée par la valeur *reg* de l'attribut `var`. Pour être complet, l'ensemble $P$ doit donc contenir les informations sur la nature des variantes, pour l'exemple cité plus haut, il sera donc défini : $P$ = {(#, (nom, sg, m)), ($a$, (gen, sg, m)), ($u_{fnct}$, (dat, sg, m)), ($ovi_{fnct}$, (dat, sg, m)), ($a$, (acc, sg, m)), ($e$, (voc, sg, m)), ($u_{fnct}$, (loc, sg, m)), ($ovi_{fnct}$, (loc, sg, m)), ($em$, (inst, sg, m)), ($i_{fnct}$, (nom, pl, m)), ($ové_{fnct}$, (nom, pl, m)), ($ů$, (gen, pl, m)), ($ům$, (dat, pl, m)), ($um_{reg}$, (dat, pl, m)), ($y$, (acc, pl, m)), ($i_{fnct}$, (voc, pl, m)), ($ové_{fnct}$, (voc, pl, m)), ($ech$, (loc, pl, m)), ($y$, (inst, pl, m)), ($ama_{reg}$ , (inst, pl, m))}.

La représentation tabulaire du paradigme des désinences casuelles a donc cette forme :

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | a |
| dat | sg | m | fnct | u |
| dat | sg | m | fnct | ovi |
| acc | sg | m | | a |
| voc | sg | m | | e |
| loc | sg | m | fnct | u |
| loc | sg | m | fnct | ovi |
| inst | sg | m | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | i |
| nom | pl | m | fnct | ové |
| gen | pl | m | | ů |
| dat | pl | m | | ům |
| dat | pl | m | reg | um |
| acc | pl | m | | y |
| voc | pl | m | fnct | i |
| voc | pl | m | fnct | ové |
| loc | pl | m | | ech |
| inst | pl | m | | y |
| inst | pl | m | reg | ama |

Les désinences de cet ensemble s'attachent au radical d'un lexème $L$ pour créer son paradigme casuel en provoquant éventuellement des alternances. Par exemple, la combinaison nom. pl. *zedník* + *i* nécessite la réalisation de l'alternance consonantique $k > c$ à la périphérie du radical en contact avec la désinence. L'information sur la réalisation de l'alternance n'est pas exprimée explicitement dans la représentation du paradigme des désinences casuelles mais elle fait partie des règles appliquées sur l'ensemble des combinaisons *radical* + *désinence*. L'application de ces règles est une étape postérieure dans la génération du paradigme casuel d'un lexème donné.

#### 3.4.1.2 Définition des types et des sous-types paradigmatiques

Le type paradigmatique d'un lexème est traditionnellement déterminé en fonction de son genre grammatical, de la périphérie de son lemme (type consonantique ou vocalique), et

éventuellement, en fonction de sa forme au génitif singulier. Chaque type est désigné par un lexème modèle (lexème prototypique). Ainsi nous parlons des types *pán, hrad, žena,..* qui englobent des ensembles de substantifs déclinés de la même façon : par exemple, le substantif *zlost* est du type paradigmatique *kost*, car c'est un substantif féminin terminé par une consonne dure et la désinence dans le génitif singulier est *–i* ; le substantif *most* est du type paradigmatique *hrad*, car c'est un substantif masculin inanimé terminé par une consonne dure et la désinence de son génitif singulier est *–u* ; etc. Notre approche développe cette notion et essaye d'y apporter des précisions nécessaires pour la génération et l'analyse automatique des formes déclinées.

Comme c'est défini plus haut, il existe un ensemble de désinences appelé **paradigme de désinences casuelles** pour chaque lexème appartenant à l'ensemble $L$ (portant la catégorie grammaticale du cas). Nous allons dire, que deux lexèmes avec le même paradigme de désinences casuelles sont de même type paradigmatique, indépendamment des éventuelles alternances. L'acceptation traditionnelle de cette notion est cependant plus large car elle permet de parler du même type paradigmatique pour les substantifs dont le paradigme de désinences casuelles n'est pas complètement identique. Ainsi, il est possible de dire que les substantifs *les* ou *rybník* sont déclinés d'après le type *hrad* mêmes si leurs paradigmes de désinences casuelles sont différents sur plusieurs points : dans le génitif singulier *hrad-u* × *les-a*, *rybník-a* et dans le locatif pluriel *hrad-ech* × *rybníc-ích, les-ích*.

Cette situation, acceptable dans les grammaires pour les autochtones, est nécessairement une source d'erreurs et pour les étrangers et pour le traitement automatique. Les apprenants, sans la connaissance des particularités dans la déclinaison d'un certain lexème, appliquent l'ensemble des désinences établies pour le type et peuvent créer ainsi des formes incorrectes. La situation est la même pour la génération automatique des formes d'après des types de déclinaison prédéfinis qui peuvent ne pas être appropriés pour un lexème particulier.

Les différences entre les paradigmes du même type sont souvent employées pour la granulation de la classification par l'établissement des sous-types, surtout lorsqu'il y a un nombre important de lexèmes présentant la même différence dans la déclinaison. Dans d'autres cas, lorsqu'il s'agit uniquement des lexèmes avec des particularités isolées, on parle des exceptions. En pratique, un type paradigmatique englobe souvent deux ou plusieurs ensembles de lexèmes possédant des paradigmes de désinences casuelles distincts.

Nous allons donc redéfinir les notions utilisées pour la manipulation des paradigmes de déclinaison, afin d'éviter tout ambiguïté du point de vue de l'apprenant et du point de vue de la génération automatique. En bref, nous définissons un type paradigmatique comme un ensemble de sous-types qui sont à leur tour des ensembles de lexèmes avec les paradigmes de désinences casuelles strictement identiques. Parmi ces sous-types, il y en a un qui est choisi

pour représenter le type. Un lexème décliné d'après ce sous-type est choisi en tant qu'un aide mnémotechnique.

- **sous-type paradigmatique** est un ensemble de lexèmes avec le même paradigme de désinences casuelles, par exemple *občan, křest'an, venkovan*. Chaque sous-type peut être désigné par un lexème modèle qui a la fonction d'une aide mnémotechnique, par exemple *občan* ;
- **type paradigmatique** est un ensemble de sous-types dont le genre grammatical, la périphérie droite du lemme et, éventuellement, la forme du génitif singulier sont identiques, par exemple les sous-types *pán, hoch, občan* sont les éléments d'un type. Ce type est désigné par un lexème représentatif qui appartient à un des sous-types, par exemple *pán* ;
- **sous-type modèle** est un sous-type paradigmatique choisi parmi les autres pour représenter le type paradigmatique. Il s'agit souvent du sous-type contenant le plus grand nombre de lexèmes.

En fonction du nombre d'unités dans un certain sous-type, nous pouvons distinguer des sous-types ouverts et des sous-types fermés :

- **sous-type ouvert** contient un nombre de lexème non défini. Il s'agit des sous-types contenant des lexèmes formés par des suffixes dérivationnels identiques ou tout simplement des sous-types contenant un nombre significatif d'éléments.
- **sous-type fermé** contient un nombre de lexèmes défini. Il s'agit des sous-types établis sur la base d'une seule lexèmes où sur un ensemble fermé d'unités. Chaque sous-type fermé doit pouvoir être énuméré en extension.

Cette spécification permet de manipuler d'une façon unique les notions de *sous-type* et d'*exception*. Les critères pour le choix de la valeur de cet attribut restent libres. Théoriquement, il est possible de fermer tous les sous-types par une énumération de tous les éléments qu'ils contiennent, en pratique cette distinction a un rôle significatif pour la présentation didactique de la déclinaison sur CETLEF : les sous-types ouverts peuvent être généralement définis par des règles applicables aux lexèmes avec les mêmes propriétés formelles, les sous-types fermés sont des exceptions qui doivent être apprises par coeur.

#### 3.4.1.3 Nature de l'écart entre les sous-types

Les différents sous-types dans le cadre d'un type sont définis par des différences dans leurs paradigmes de désinences casuelles. Pour décrire cette différence, nous utilisons la

comparaison du paradigme de désinences casuelles du sous-type modèle avec les autres sous-types. La nature de cet écart peut être exprimée en termes de différence de deux ensembles. Dans la représentation formelle, le couple *désinence – valeurs des catégories grammaticales* qui fait la différence est marqué par une valeur caractérisant la nature de l'écart (l'attribut `ET` – écart du type modèle).

Le principal intérêt de signaler explicitement la nature de l'écart entre les sous-types est la simplification de l'attribution d'un certain sous-type à un lexème donné pendant l'annotation manuelle. En général, l'identification du type ne pose pas de problème à l'annotateur – il s'agit d'une compétence enseignée dès l'école primaire –, la difficulté réside souvent dans l'attribution des sous-types moins fréquents ou des sous-types fermés : la spécification des cas et des désinences est un indice qui permet de faciliter cette tâche. Pour les traitements automatiques dans le diagnostic des erreurs, la spécification de l'écart sert pour la génération du message de retour approprié.

Nous reprenons ici la définition formelle du paradigme de désinences casuelles qui est un ensemble de couples $(t, c)$ où $t$ est une désinence casuelle, avec un attribut éventuel signalant son rôle de variante fonctionnelle ou de variante de registre, et $c$ est un triplet de valeurs des catégories du cas, du nombre et du genre. Nous définissons un ensemble $S_m$ qui est le paradigme de désinences casuelles du sous-type modèle, et l'ensemble $S_t$ qui est le paradigme de désinences casuelles du sous-type comparé avec le modèle. Sur la base de la différence entre ces deux ensembles, nous pouvons distinguer les écarts suivants :

- **ajout d'une désinence** : pour un triplet $c$, le sous-type $S_t$ contient une désinence $t$, désignée comme une variante, fonctionnelle ou de registre, que le sous-type modèle $S_m$ ne contient pas. La valeur de l'attribut `ET` est `A`.
- **remplacement d'une désinence** : pour un triplet $c$, le sous-type $S_t$ contient une désinence $t$ que le sous-type modèle $S_m$ ne contient pas. La valeur de l'attribut `ET` est `R`.
- **exclusion d'une désinence** : pour un triplet $c$, le sous-type modèle $S_m$ contient une désinence $t$ que le sous-type $S_t$ ne contient pas. La valeur de l'attribut `ET` est `X`.
- **changement du registre d'une variante** : pour un triplet $c$, la valeur de l'attribut `var` d'une désinence $t$ n'est pas identique dans le sous-type modèle $S_m$ et le sous-type $S_t$. La valeur de l'attribut `ET` est `V`.

Au cas où l'écart de deux désinences, qualifié comme un remplacement, soit causé simplement par les variantes orthographiques *e* × *ě* (par exemple dat. sg. *žen–ě* × *holc–e*) et *i* × *y* (par exemple acc. pl. *předsed–y* × *paňác–i)*, l'attribut `ET` porte les valeurs `Re` ou `Ri` en fonction de la voyelle affectée. Cette situation est spécifiée dans l'annotation, car le plan

linguistique affecté par la différence des deux sous-types n'est pas le plan morphologique mais le plan graphique (orthographique).

Un résumé schématique des différents écarts avec des exemples est présenté dans le tableau suivant :

| ET | *description* | *exemple* |
|---|---|---|
| A | ajout d'une désinence | gen. sg. *hrad–u* × *tábor–u, tábor–a* |
| R | remplacement d'une désinence | loc. pl. *pán–ech* × **hoch–ech* > *hoš–ích* |
| Re | remplacement *e* × *ě* | nom. sg. *růž–e* × **sukn–e* > *sukn–ě* |
| Ri | remplacement *i* × *y* | gen. sg. *žen-y* × **skic–y* > *skic–i* |
| X | exclusion d'une désinence | loc. sg. *hrad-ě* × **vzor–ě / e* |
| V | changement du registre d'une variante | nom. pl. *pán–i / –ové* × *občan–i / –ové* |

La représentation complète du paradigme de désinences casuelles d'un certain sous-type a donc par exemple la forme suivante :

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | m | | | # |
| gen | sg | m | | | a |
| dat | sg | m | fnct | | u |
| dat | sg | m | fnct | | ovi |
| acc | sg | m | | | a |
| voc | sg | m | | R | u |
| loc | sg | m | fnct | | u |
| loc | sg | m | fnct | | ovi |
| inst | sg | m | | | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | fnct | | i |
| nom | pl | m | fnct | | ové |
| gen | pl | m | | | ů |
| dat | pl | m | | | ům |
| dat | pl | m | reg | | um |
| acc | pl | m | | | y |
| voc | pl | m | fnct | | i |
| voc | pl | m | fnct | | ové |
| loc | pl | m | | R | ích |
| loc | pl | m | | A | ách |
| inst | pl | m | | | y |
| inst | pl | m | reg | | ama |

Ce tableau représente le paradigme de désinences casuelles du sous-type *hoch*, défini par rapport au sous-type modèle *pán* du type *pán*. Il se distingue par le remplacement de la désinence *–e* par *–u* dans le vocatif singulier (*pan–e* × *hoch–u*), par le remplacement de la désinence *–ech* par *–ích* dans le locatif pluriel et par l'ajout d'une variante de registre *–ách* pour le même cas.

### 3.4.2 Fichier *pdgm.xml*

Pour pouvoir être utilisées dans des traitements automatiques, les informations concernant les types paradigmatiques qui viennent d'être exposées, doivent être inscrites dans un format lisible par ordinateur. Le format XML[14] a été choisi pour sa maniabilité et pour la facilité du traitement des données dans ce format. Les critères formels que nous avons établis pour leur définition servent d'éléments de structuration de ce ficher. Il contient les types paradigmatiques, les ensembles de terminaisons qui leur sont assignés et des informations complémentaires exploitées dans des différents traitements sur CETLEF.

#### 3.4.2.1 Rappel technique

Le langage XML (eXtensible Markup Language) est un méta-langage destiné au traitement de documents. Il permet de structurer les données à l'aide de balises et des attributs personnalisables – ils ne sont pas définis à l'avance, comme c'était le cas pour le prédécesseur du XML, le langage international de documentation normalisé SGML (Standard Generalized Markup Language), mais ils sont choisis par l'auteur du document en fonction de la nature des données encodées. La seule restriction est le respect de la syntaxe XML dont les règles principales sont l'existence d'un seul élément racine pour chaque document et la correspondance des balises ouvrantes et des balises fermantes.

Pour donner un aperçu de la structure d'un document formaté avec XML, nous allons présenter ici les notions de base de ce langage :

- une **balise ouvrante** est une suite de caractères entourée par des chevrons, par exemple `<balise>`.
- une **balise fermante** est est une suite de caractères entourée par des chevrons avec une barre oblique au début, par exemple `</balise>`.
- un document XML contient un ou plusieurs **éléments**, qui sont définis par une balise ouvrante, une balise fermante et le contenu entouré par ces deux balises, par exemple `<balise>contenu</balise>`.
- une balise peut posséder un ou plusieurs **attributs** qui doivent être situés à l'intérieur de la balise après son nom ; la valeur de l'attribut est située entre des guillemets, par exemple `<balise attribut=‘‘valeur’’>`.
- un **élément vide** est signalé par une balise qui est ouvrante et fermante à la fois, par exemple `<balise/>`.

[14] Voir `http://www.w3.org/XML/`.

La complexité d'un document XML est déterminée par la propriété récursive des éléments qui peuvent à leur tour contenir un ou plusieurs éléments. La seule condition est que les balises de deux éléments distincts ne se chevauchent pas : il n'est pas possible d'écrire `<baliseA>contenu A<baliseB></baliseA>contenu B</baliseB>` mais il est nécessaire d'écrire `<baliseA>contenu A</baliseA><baliseB>contenu B</baliseB>`.

Un fichier DTD (Document Type Definition) peut être assigné à un document XML. Il spécifie les noms des éléments et des attributs, les restrictions sur leur contenu et l'organisation hiérarchique des éléments. Un document XML est *bien formé*, s'il observe les règles de la syntaxe XML ; il est *valide*, s'il observe les règles établies dans un fichier DTD par l'auteur du document.

L'avantage de la structuration des données avec le format XML est la facilité de la création et de la modification d'un tel fichier. Grâce aux noms des balises qui reflètent leur contenu, les éléments sont facilement repérables « à l'œil nu », ce qui permet un contrôle des données et leur modification directement dans le fichier. La prise en charge du format XML est parfaitement assurée par le langage PHP (le langage de programmation utilisé pour le développement du CETLEF) : les données en XML sont facilement accessibles pour les traitements linguistiques sur CETLEF ainsi que pour leur affichage sur les pages HTML[15].

#### 3.4.2.2 Structure du fichier *pdgm.xml*

Le fichier *pdgm.xml* (voir un extrait de ce fichier sur la fig. 3.5, p. 68) reflète l'organisation des types paradigmatiques qui viennent d'être présentés. Sa structure ainsi que les valeurs permises de certains attributs sont définis dans le fichier *pdgm.dtd* (voir la fig. 3.4).

L'élément racine `<root>` contient plusieurs éléments `<main>`, chaque élément `<main>` contient plusieurs éléments `<type>`. La valeur de l'attribut `@tagMorph` des éléments `<type>` frères est identique. Ainsi, chaque élément `<main>` contient tous les types paradigmatiques qui appartiennent à un certain type morphologique. Par exemple, pour la déclinaison nominale, l'attribut `@tagMorph` de l'élément `<type>` a la valeur *N* ; pour la déclinaison adjectivale, l'attribut `@tagMorph` de l'élément `<type>` a la valeur *A*) et ainsi de suite. Chaque élément `<type>` est donc défini par la valeur de l'attribut `@tagMorph` ainsi que par la valeur de l'attribut `@modèle` qui contient le mot modèle défini pour chaque type paradigmatique (par exemple *pán, předseda, jarní, ten*).

Chaque élément `<type>` contient un ou plusieurs éléments `<soustype>` qui sont définis par les attributs `@mnemo` et `@pdgm`. L'attribut `@mnemo` contient un lexème qui sert pour marquer

[15]Pour plus de détails sur les moyens de programmation utilisés pour le développement du CETLEF, voir le chapitre *5 Implémentation*, pp. 195–236.

```
<!ELEMENT root (main+)>
<!ELEMENT main (type+)>
<!ELEMENT type (soustype+)>
<!ATTLIST type modele CDATA #REQUIRED
               tagMorph   ( N | A | M | P | C | X ) #REQUIRED>
<!ELEMENT soustype (term+, ex)>
<!ATTLIST soustype mnemo CDATA #REQUIRED
                   pdgm  CDATA #REQUIRED>
<!ELEMENT term (#PCDATA)>
<!ATTLIST term  cas ( nom | gen | dat | acc | voc | loc | inst ) #REQUIRED
                num ( sg | pl ) #REQUIRED
                gen ( m | i | f | n | s ) #REQUIRED
                ET  ( A | R | Re | Ri | X | V ) #IMPLIED
                var ( fnct | reg ) #IMPLIED>
<!ELEMENT ex (u+)>
<!ATTLIST ex    ouvert  ( 1 | 0 ) #REQUIRED
                A1A2    ( 1 | 0 ) #REQUIRED>
<!ELEMENT u (#PCDATA)>
```

FIG. 3.4 – Fichier *pdgm.dtd*

(comme une aide mnémotechnique) un sous-type paradigmatique donné, par exemple *pán, hoch, husita, úterý* etc. ; l'attribut `@pdgm` contient un identifiant unique de chaque sous-type, par exemple *pn, pn_h, pd_h, st_Ri*.

Comme il est montré dans la section 3.4.1, p. 56, le choix d'un sous-type pour représenter un type paradigmatique entier est, en principe, arbitraire – ce sous-type est appelé le modèle d'un type paradigmatique, par exemple le sous-type *pán* pour le type paradigmatique désigné traditionnellement comme *pán*, le sous-type *předseda* pour le type paradigmatique désigné traditionnellement comme *předseda*, etc.

Dans le fichier *pdgm.xml*, le sous-type modèle est distingué d'abord par la valeur de son attribut `@mnemo`, égale à la valeur de l'attribut `@modele` de l'élément `<type>` qui est son père dans la hiérarchie des éléments. Le second signe distinctif pour faire la différence entre le sous-type modèle et les autres sous-types est la forme de la valeur de l'attribut `@pdgm` :

- pour le sous-type modèle, cette valeur contient uniquement deux lettres qui sont, dans la plupart des cas, les deux premières consonnes du nom du type paradigmatique, par exemple *pn* pour *pán*, *zn* pour *žena* etc.
- pour un sous-type « ordinaire », cette valeur contient les deux lettres du sous-type modèle et une spécification du sous-type précédée par le symbole _ , par exemple *pn_h*

```
<MAIN>
   <TYPE modele="pán" tagMorph="N">
      <SOUSTYPE mnemo="pán" pdgm="pn">
         <TERM cas="nom" num="sg" gen="m">#</TERM>
         <TERM cas="gen" num="sg" gen="m">a</TERM>
         <TERM cas="dat" num="sg" gen="m" var="fnct">u</TERM>
         <TERM cas="dat" num="sg" gen="m" var="fnct">ovi</TERM>
         <TERM cas="acc" num="sg" gen="m">a</TERM>
         <TERM cas="voc" num="sg" gen="m">e</TERM>
         <TERM cas="loc" num="sg" gen="m" var="fnct">u</TERM>
         <TERM cas="loc" num="sg" gen="m" var="fnct">ovi</TERM>
         <TERM cas="inst" num="sg" gen="m">em</TERM>
         <TERM cas="nom" num="pl" gen="m" var="fnct">i</TERM>
         <TERM cas="nom" num="pl" gen="m" var="fnct">ové</TERM>
         <TERM cas="gen" num="pl" gen="m">ů</TERM>
         <TERM cas="dat" num="pl" gen="m">ům</TERM>
         <TERM cas="dat" num="pl" gen="m" var="reg">um</TERM>
         <TERM cas="acc" num="pl" gen="m">y</TERM>
         <TERM cas="voc" num="pl" gen="m" var="fnct">i</TERM>
         <TERM cas="voc" num="pl" gen="m" var="fnct">ové</TERM>
         <TERM cas="loc" num="pl" gen="m">ech</TERM>
         <TERM cas="inst" num="pl" gen="m">y</TERM>
         <TERM cas="inst" num="pl" gen="m" var="reg">ama</TERM>
         <EX ouvert="1" A1A2="1">
            <U>pán</U>
            <U>ministr</U>
            <U>prezident</U>
            <U>autor</U>
            <U>premiér</U>
            <U>bratr</U>
            <U>pes</U>
         </EX>
      </SOUSTYPE>
      <SOUSTYPE mnemo="hoch" pdgm="pn_h">
         <TERM cas="nom" num="sg" gen="m">#</TERM>
         <TERM cas="gen" num="sg" gen="m">a</TERM>
         <TERM cas="dat" num="sg" gen="m" var="fnct">u</TERM>
         <TERM cas="dat" num="sg" gen="m" var="fnct">ovi</TERM>
         <TERM cas="acc" num="sg" gen="m">a</TERM>
         <TERM cas="voc" num="sg" gen="m" ET="R">u</TERM>
```

FIG. 3.5 – Extrait du fichier *pdgm.xml*

pour *hoch* qui est un sous-type du type *pán*, *mz_u* pour *učitel* qui est un sous-type du type *muž* etc.

Chaque élément `<soustype>` contient plusieurs éléments `<term>` et un seul élément `<ex>`. L'ensemble des éléments `<term>` représente le paradigme casuel spécifique pour chaque sous-type : le contenu de l'élément `<term>` est une désinence casuelle, par exemple *–a, –ě, –ými* etc. Les attributs de l'élément `<term>` sont :

- les catégories grammaticales de cas (attribut `@cas`), de nombre (attribut `@num`) et de genre (attribut `@gen`).
- les attributs optionnels `@var` et `@ET`.

La valeur de l'attribut `@var` spécifie la nature d'une variante si deux ou plusieurs désinences ont les valeurs identiques de toutes les catégories grammaticales. La valeur de l'attribut `@ET` spécifie la nature de l'écart entre le sous-type actuel et le sous-type modèle. Naturellement, l'attribut `@ET` ne peut pas figurer dans les éléments `<term>` d'un sous-type modèle.

L'élément `<ex>` contient un ou plusieurs éléments `<u>` dont le contenu sont des exemples de lexèmes appartenant à un sous-type donné. Ces exemples ont été choisi en fonction de leur fréquence ou de leur caractère typique. Quand c'était possible, ils ont été choisis dans le lexique A1A2. L'élément `<ex>` est spécifié par deux attributs.

- La valeur binaire de l'attribut `@ouvert` spécifie la nature du sous-type. Dans le cas d'un sous-type ouvert, uniquement quelques lexèmes typiques déclinés d'après ce sous-type sont indiqués. S'il s'agit d'un sous-type ouvert, tous les lexèmes appartenant à ce sous-type sont énumérés.
- La valeur binaire de l'attribut `@A1A2` indique la présence des lexèmes appartenant à un certain sous-type dans le lexique A1A2, voir la section suivante.

L'annexe C (p. 269) donne une présentation succincte des types paradigmatiques générée automatiquement à partir du fichier *pdgm.xml*.

Dans les sections suivantes, les cinq types de déclinaison utilisés dans le cadre de CETLEF – *nominale, adjectivale, mixte, pronominale* et *numérale* – sont présentées avec les définitions exhaustives des types et des sous-types paradigmatiques, basées sur la nature de l'écart entre le sous-type modèle d'un certain type et des autres sous-types. Les marques entre les parenthèses (par exemple `pn_h`, `zn_Re`, `ps`) sont les identifiants des différents sous-types. Ces identifiants sont les valeurs de l'attribut `@pdgm`.

## 3.4.3 Déclinaison nominale

La déclinaison nominale (la valeur de l'attribut `tagMorph` est N) contient 15 types : *pán, hrad, muž, stroj, předseda, soudce, žena, růže, píseň, kost, město, moře, kuře, stavení, klima*.

### 3.4.3.1 Type *pán* (`pn`)

D'après le type ***pán*** (`pn`) (voir l'annexe C.1, p. 269–272) sont déclinés les substantifs masculins animés, dont le radical est terminé par une consonne dure ou mixte, le nom. sg. est exprimé par la désinence zéro *–#*, éventuellement par un suffixe d'origine étranger (*–ius, –o*) ; le gén. sg. est exprimé par la désinence *–a*. Le modèle de ce type contient des variantes fonctionnelles dans le dat. et loc. sg. (*–u* × *–ovi*), et dans le nom. et voc. pl. (*–i* × *–ové*) et des variantes de registre dans le dat. pl. *–um* et l'inst pl. –ama.

Le sous-type ouvert ***hoch*** (pn_h) est défini par le voc. sg. *–u* (R) et loc. pl. *–ích* (R), accompagné de sa variante de registre *–ách* (A), qui s'attachent aux radicaux terminé par *–h, –ch, –k, –g*. La terminaison *–u* est employée pour empêcher la palatalisation qui serait provoquée par la terminaison du modèle (*–e*). La désinence *–e* exprimant le voc. sg. est néanmoins conservée dans les sous-types fermés *člověk* (pn_c) (défini par l'exclusion du pluriel – X) et *bůh* (pn_b) (qui reprend cependant la terminaison *–ích* (R) pour le loc. pl.) où elle provoque l'alternance du radical (*k* > *č, h* > *ž* respectivement). La désinence *–u* (R) dans le voc. sg. définit le sous-type fermé *syn* (pn_s).

Le sous-type ouvert ***občan*** (pn_o) est défini par le nom. et voc. pl. *–é* (R), les désinences du modèle *–i, –ové* (V) deviennent des variantes de registre. Dans le sous-type fermé *host* (pn_h), la désinence *–é* (A) s'ajoute en tant que variante fonctionnelle ainsi que la désinence du gén. pl. *–í* (A). Le sous-type fermé *manžel* (pn_m) prend la désinence *–é* (R) dans le nom. et voc. pl., il est défini en plus par la présence des terminaisons du type masculin animé mou *muž*(mz) dans le voc. sg. *–i* (R), loc. sg. *–i* (A), acc. pl. *–e* (A), loc. pl. *–ích* (A) et inst. pl. *–i* (A).

Le sous-type ouvert ***génius*** (pn_g) est défini par le suffixe *–us* (R) qui prend formellement la place de la désinence zéro du modèle dans le nom. sg. ; il n'admet pas la désinence *–i* (X) dans le nom. et voc. pl., il remplace les désinences du modèle dans le acc. pl. par *–e* (R) et dans inst. pl. par *–i* (R), et il permet une variante fonctionnelle dans le loc. pl. *–ích* (A).

Le sous-type ouvert ***džigolo*** (pn_g) est défini par le suffixe *–o* (R) dans le nom. sg. et dans le voc. sg., la désinence *–i* (X) n'est pas admise dans le nom. et voc. pl.

#### 3.4.3.2 Type *hrad* (hd)

D'après le type ***hrad*** (voir l'annexe C.1, p. 272–275) sont déclinés les substantifs masculins inanimés, dont le radical est terminé par une consonne dure ou mixte ; le nom. sg. est exprimé par la désinence zéro *–#*, éventuellement par un suffixe d'origine étranger *–us* et le gén. sg. est exprimé par *–u* ou *–a*. Le modèle de ce type contient des variantes fonctionnelles dans le loc. sg. (*–u* × *–ě*) et des variantes de registre dans le dat. pl. *–um* et l'inst pl. –ama.

Le sous-type ouvert ***čas*** (hd_Re) est défini uniquement par la variante orthographique de la désinence *–ě* × *e* (Re) dans le loc. sg., le radical des substantifs appartenant à ce type sont terminées par *–s, –z, –l, –r, –h, –ch, –k, –g*.

Le sous-type ouvert ***vzor*** (hd_v) n'admet pas la désinence *–ě* × *e* (X) dans le loc. sg., la seule désinence possible est *–u*.

D'après le sous-type ouvert ***zámek*** (hd_z) sont déclinés des substantifs avec le radical terminées par les vélaires *–k, –h, –ch, –g*. Comme dans le sous-type *hoch* pour les substantifs

animés, le voc sg. est exprimé par *–u* (R) et le loc. pl. par *–ích*, avec une variante de registre *–ách* (R). La variante fonctionnelle *–ě* × *e* (X) dans le loc. sg. n'est pas admise.

Le sous-type ouvert ***ostrov*** (`hd_o`) est défini par la désinence *–a* (R) dans le gen. sg. qui repousse la désinence du modèle *–u* (V) dans un rôle de variante de registre. Le sous-type ouvert ***rybník*** (`hd_r`) est défini par l'union des critères pour la définition du type *zámek* et du type *ostrov* avec la variante *–e* dans le loc. sg. Le sous-type fermé *červen* (`hd_m`) a uniquement la désinence *–a* dans le gen. sg. et il n'admet pas la variante *–ě* dans le loc. sg. Le sous-type fermé *tábor* (`hd_t`) est défini par l'union des critères pour le sous-type *čas* et *ostrov*. Le sous-type fermé *les* (`hd_l`) est défini de la même manière que le sous-type *tábor* avec l'ajout d'une variante fonctionnelle *–ích* dans le loc. pl. Le sous-type fermé *sklep* (`hd_s`) est égal au sous-type *les* à part la désinence du loc. sg. qui conserve la forme du modèle -ě.

Le sous-type ouvert ***kámen*** (`hd_k`) est défini par la désinence *–i* (R) dans le voc. sg. et par l'ajout des variantes fonctionnelles dans le gen. sg. *-e* (A), dat. et loc. sg. *-i* (A). Le sous-type fermé *den* (hd_d) est défini par désinence *–e* dans le gen. sg. (R) et loc. sg. (Re) et par l'ajout des variantes fonctionnelles dans dat., loc., voc. sg. *–i* (A), et dans nom., acc., voc. pl. *–i* (A), gen. pl. *–í* (A) a dat. pl. *–ím* (A).

Le sous-type ouvert ***virus*** (`hd_u`) est défini par le suffixe *–us* (R) qui prend formellement la place de la désinence zéro du modèle dans le nom. sg. et acc. sg., la variante *–ě* dans le loc. sg. n'est pas admise.

Les deux sous-types ouverts ***šaty*** (`hd_a`) a ***hradčany*** (`hd_h`) sont définis par l'exclusion du singulier (X), il s'agit des pluralia tantum. Le type *hradčany* prend la désinence zéro (R) pour exprimer le gen. pl.

#### 3.4.3.3 Type *muž* (`mz`)

D'après le type ***muž*** (`mz`) (voir l'annexe C.1, p. 277–278) sont déclinés les substantifs masculins animés, dont le radical est terminé par une consonne molle ou mixte ; le nom. sg. est exprimé par la désinence zéro *–#* et le gén. sg. est exprimé par *–e*. Le modèle de ce type contient des variantes fonctionnelles dans le dat. et loc. sg (*–i* × *–ovi*), et dans le nom. et voc. pl. (*–i* × *–ové*) et des variantes de registre dans le dat. pl. *–um* et dans l'inst pl. *–ema*.

Le sous-type ouvert ***učeň*** (`mz_Re`) est défini uniquement par la variante orthographique de la désinence (Re) dans le gén., acc. sg. *–ě*, inst. sg. *–ěm*, acc. pl. *–ě* et dans la variante de registre inst. pl. *–ěma*.

Le sous-type ouvert ***učitel*** (`mz_u`) est défini par la désinence *–é* (R) dans le nom. et voc. pl. qui repousse la variante fonctionnelle *–ové* du modèle au rôle d'une variante de registre et par une possible variante de registre dans l'inst pl. *–ama* (A). Le sous-type fermé *přítel*

(mz_p) est défini de la même façon à part la désinence zéro (R) du gén. pl., la désinence du modèle –*ů* devient une variante de registre.

Le sous-type ouvert ***otec*** (mz_o) se distingue uniquement par le voc. sg. –*e* (R). Le sous-type fermé *rodič* (mz_r) est défini par la désinence –e (R) dans le nom. et voc. pl. Le sous-type fermé *kůň* (mz_k) est défini sur la base du même critère que le sous-type *učeň*. Il a en plus la désinence -*ě* (R) dans le nom. pl. et des variantes fonctionnelles dans le gén. pl. –*í* (A), dat. pl. –*ím* (A) et inst. pl. –mi (A) avec une variante de registre –*ma* (A).

#### 3.4.3.4 Type *stroj* (sj)

D'après le type ***stroj*** (sj) (voir l'annexe C.1, p. 279–275) sont déclinés les substantifs masculins inanimés, dont le radical est terminé par une consonne molle ou mixte ; le nom. sg. est exprimé par la désinence zéro –# et le gén. sg. est exprimé par la désinence –*e*. Le modèle de ce type contient des variantes de registre dans le dat. pl. –*um* et l'inst pl. –*ema*.

Le sous-type ouvert ***stupeň*** (sj_Re) est défini uniquement par la variante orthographique de la désinence dans le gén., acc. sg. –*ě* (Re), inst. sg. –*ěm* (Re), acc. pl. –*ě* (Re) et la variante de registre inst. pl. –*ěma* (Re).

Le sous-type ouvert ***punčocháče*** (sj_p) est défini par l'exclusion du singulier (X), il s'agit des pluralia tantum. Le sous-type fermé *peníze* (sj_n) est défini de la même façon, il y a en plus une différence dans le gen. pl. # (R) et dans l'inst pl. –*mi* (R) et la variante de registre –*ma* (R).

#### 3.4.3.5 Type *předseda* (pd)

D'après le type ***předseda*** (pd) (voir l'annexe C.1, p. 280–281) sont déclinés les substantifs masculins animés, dont le nom. sg. est exprimé par la désinence –*a* et le gén. sg. par –*y*. Le modèle de ce type contient des variantes de registre dans le dat. pl. (–*um*) et l'inst pl. (-*ama*).

Le sous-type ouvert *sluha* (pd_s) est défini par la désinence –*ích* (R) dans le loc. pl. Le sous-type ouvert *husita* (pd_h) est défini par la désinence –*é* (A) dans le nom. et voc. pl. et par l'ajout d'une variante de registre –*i* (A) dans les mêmes cas. Le sous-type ouvert *paňáca* (pd_p) est défini par la désinence –*ích* (R) dans le loc. pl. et par la variante orthographique dans le gén. sg., acc. et inst. pl. –*i* (Ri) ou lieu de –*y*.

#### 3.4.3.6 Type *soudce* (sc)

D'après le type ***soudce*** (sc) (voir l'annexe C.1, p. 281) sont déclinés les substantifs masculins animés, dont le nom. sg. est exprimé par la désinence *–e* et le gén. sg. par *–e*. Le modèle de ce type contient des variantes fonctionnelles dans le dat. et loc. sg (*–i* × *–ovi*), le nom. et voc. sg (*–i* × *–ové*) et des variantes de registre dans le dat. pl. (*–um*) et l'inst pl. -ema. Dans notre classification, il n'y a pas de sous-type pour *soudce*.

#### 3.4.3.7 Type *žena* (zn)

D'après le type ***žena*** (zn) (voir l'annexe C.1, p. 282–284) sont déclinés les substantifs féminins, dont le nom. sg. est exprimé par la désinence *–a* et le gén. sg. par *–y*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. (*–ama*).

Le sous-type ouvert ***holka*** (zn_h) est défini uniquement par la variante orthographique de la désinence (Re) dans le dat. et loc. sg. *–e* (Re). De la même façon est défini le sous-type ouvert ***skica*** (zn_s), qui remplace également la désinence *–y* par *–i* (Ri) dans le dat. et loc. sg. et dans le nom., acc. et voc. pl. Dans le sous-type ouvert ***Stáňa***, uniquement les désinences *–y* sont remplacés par *–i* (Ri).

Le sous-type fermé *dcera* (zn_d) est défini par la désinence *–i* (R) dans le dat. et loc. sg., la variante orthographique de la désinence du modèle *–e* (V) devient une variante de registre. Le sous-type fermé *idea* (zn_i) a une déclinaison spécifique dûe à la périphérie vocalique de son radical – gen. sg. *-je* (A), dat. et loc. sg. *-ji* (R), inst. sg. *jí* (A), nom., acc. et voc. pl. *-je* (A), gen. pl. *-jí* (R), dat. pl. *-jím* (A), loc. pl. *-jích* (A) et inst. pl. *-jemi*.

Les deux sous-types fermés *noha* (zn_n) et *ruka* (zn_r) ont la même propriété que le sous-type *holka*, ils sont en plus définis par des désinences conservées de l'ancien duel – gen. et loc pl. *–ou* (A), inst. pl. – *ama* (R) et une variante de registre dat. pl. *–oum* (A) pour *noha* ; nom., acc. et voc. pl. *–e* (R), gen. et loc. pl. *–ou* (R), inst. pl. – *ama* (R) et une variante de registre dat. pl. *–oum* (A) pour *ruka*.

Le sous-type ouvert ***kalhoty*** (zn_k) est défini par l'exclusion du singulier (X), il s'agit des pluralia tantum.

#### 3.4.3.8 Type *růže* (rz)

D'après le type ***růže*** (rz) (voir l'annexe C.1, p. 284–287) sont déclinés les substantifs féminins, dont le nom. et gen. sg. sont exprimés par les désinences *–e* ou *–ě*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. (*–ema*).

Le sous-type ouvert ***sukně*** (`rz_s`) est défini uniquement par la variante orthographique de la désinence *–ě* (Re) dans le nom., gen. et voc. sg. et nom., acc. voc., inst. pl.

Le sous-type ouvert ***žákyně*** (`rz_z`) a les mêmes propriétés comme le sous-type *sukně*, il est défini en plus par une possible variante fonctionnelle *–#* (A) dans le gen. pl. Le sous-type ouvert ***ulice*** (`rz_u`) est défini par la désinence zéro *–#* (R) dans le gen. pl. qui est la seule désinence possible ; le sous-type ouvert ***Lidice*** (`rz_l`) a la déclinaison du pluriel identique avec *ulice* mais il s'agit d'un pluralium tantum. Le sous-type fermé *Velikonoce* (`rz_v`) est identique au sous-type *Lidice*, il se distingue par le dat. pl. *–ům* (R). Le sous-type ouvert ***definice*** (`rz_d`) a également cette désinence dans le gen. pl. mais il admet la désinence du modèle *–í* (V) comme une variante de registre.

Le sous-type ouvert ***brýle*** (`zn_b`) est défini par l'exclusion du singulier (X). La situation est la même pour le sous-type fermé ***dveře*** (`zn_d`) qui se distingue en plus par la désinence du inst. pl. *–mi* (R) et par sa variante de registre *–ma* (R).

#### 3.4.3.9 Type *píseň* (`ps`)

D'après le type ***píseň*** (`ps`) (voir l'annexe C.1, p. 287–289) sont déclinés les substantifs féminins, dont le radical est terminé par une consonne molle ou mixte, le nom. sg. est exprimé par la désinence zéro *–#* et le gén. sg. par la désinence *–ě*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. *–ěma*.

Le sous-type ouvert ***tvář*** (`ps_Re`) est défini uniquement par la variante orthographique de la désinence *–e* (Re) dans le gen. sg. et dans le nom., acc. et voc. pl., ensuite par *–emi* (Re) dans l'inst pl. et par sa variante de registre *–ema* (Re). Le sous-type ouvert ***čtvrť*** (`ps_c`) est défini par l'ajout de la variante fonctionnelle dans le gen. sg. et dans le nom., acc. et voc. pl. *–i* (A). Le sous-type ouvert ***ocel*** (`ps_o`) a les mêmes propriétés que le sous-type *tvář*, il est en plus défini par l'ajout de la variante fonctionnelle *–i* (A) dans le gen. sg. et dans le nom., acc. et voc. pl. Le sous-type ouvert ***lod'*** (`ps_l`) est défini par les mêmes critères que le sous-type *čtvrť*, il prend en plus une variante fonctionnelle *–mi* (A) dans l'inst pl., accompagnée par sa variante de registre *–ma* (A).

Les sous-types ouverts ***kleště*** (`ps_k`) et ***housle*** (`ps_h`) sont définis par l'exclusion du singulier, le sous-type *housle* est défini en plus par les variantes orthographiques identiques au paradigme pluriel du sous-type *tvář*.

#### 3.4.3.10 Type *kost* (kt)

D'après le type ***kost*** (kt) (voir l'annexe C.1, p. 289–291) sont déclinés les substantifs féminins, dont le radical est terminé par une consonne dure, moins souvent par une consonne molle ou mixte, le nom. sg. est exprimé par la désinence zéro *–#* ; le gén. sg. est exprimé par *–i*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. *–ma*.

Le sous-type ouvert ***pěst*** (kt_p) est défini par l'ajout d'une variante fonctionnelle aux dat. et loc. pl. *–ím* (A). Dans le sous-type ouvert ***odpověď*** (kt_o), la désinence du dat. et loc. pl. *–ím* (R) remplace celle du modèle. Le sous-type ouvert ***noc*** (kt_n) est défini par les mêmes critères que le sous-type *odpověď*, il est en plus défini par la désinence du inst. pl. *–emi* (R) et sa variante de registre *–ema* (R). Le sous-type ouvert ***moc*** (kt_m) a les mêmes propriétés que le sous-type *pěst*, il est en plus défini par la désinence *–emi* (R) dans l'inst pl. et par sa variante de registre *–ema* (R).

Les trois sous-types fermés ***lidé*** (kt_l), ***čtyři*** (kt_4) et ***tři*** (kt_3) sont des pluralia tantum. Le sous-type *lidé* est défini par le nom. et voc. pl. *–é* (R) qui repousse la désinence du modèle *–i* dans le rôle d'une variante de registre. Le sous-type *čtyři* est défini par l'ajout d'une variante de registre dans le gén. pl. *–ech* (A) ; il en est également pour le sous-type *tři* qui se distingue par les désinences *–emi* (R) dans l'inst pl. et par sa variante de registre *–ema* (R).

#### 3.4.3.11 Type *město* (mt)

D'après le type ***město*** (mt) (voir l'annexe C.1, p. 292–295) sont déclinés les substantifs neutres, dont le nom. sg. est exprimé par la désinence *–o*, éventuellement par un suffixe d'origine étranger (*-um*), et le gén. sg. par la désinence *–a*. Le modèle de ce type contient des variantes fonctionnelles dans le loc. sg (*–u* × *–ě*) et des variantes de registre dans le dat. pl. *–um* et l'inst pl. *–ama*.

Le sous-type ouvert ***tělo*** (mt_t) est défini uniquement par la variante orthographique de la désinence dans le loc. sg. *–ě* (Re). Il en est de même pour le sous-type ouvert ***lečo*** (mt_l) qui est en plus défini par la variante orthographique de la désinence à l'inst. pl. *–i* (Ri).

Le sous-type ouvert ***božstvo*** (mt_b) est défini par l'exclusion de la variante fonctionnelle *–ě* (X) dans le loc. sg. Le sous-type ouvert ***echo*** (mt_e) a la même propriété comme le sous-type *božstvo*, mais il est en plus défini par la désinence *–ách* (R) dans le loc. pl. Le sous-type ouvert ***středisko*** (mt_s) est défini par les mêmes propriétés que le sous-type *echo*, cependant la désinence *–ách* (A) dans le loc. pl. est une variante fonctionnelle et elle figure avec la

désinence *–ích* (R). Le sous-type ouvert ***břicho*** (`mt_b`) est défini uniquement par la désinence *–ách* (R) dans le loc. pl.

Les sous-types ouverts ***ústa*** (`mt_u`) et ***dvířka*** (`mt_d`) sont des pluralia tantum, *dvířka* se distingue en plus par la désinence du loc. pl. *–ách* (R).

Le sous-type ouvert ***video*** (`mt_v`) est défini par l'exclusion de la variante fonctionnelle *–ě* (X) dans le loc. sg., par la désinence *–í* (R) dans le gen. pl., par l'ajout d'une variante fonctionnelle au dat. pl. *–ím* (A) et au loc. pl. *–ích* (A) et par la variante orthographique de la désinence de l'inst. pl. *–i* (Ri).

Le sous-type fermé *koleno* (`mt_k`) est défini par l'ajout d'une variante fonctionnelle dans le loc. pl. *–ou* (A), accompagné d'une variante de registre *–ách* (A). Le sous-type fermé *prso* (`mt_k`) a les mêmes propriétés que le sous-type *koleno*, il a en plus une variante orthographique de la désinence du loc. sg. *–e* (Re). Le sous-type fermé *oko* (`mt_o`) est défini par l'exclusion de la variante fonctionnelle *–ě* (X) dans le loc. sg. et par les désinences nom., acc., voc. pl. *–i* (R), gen. pl. *–í* (R), dat. pl. *–ím* (R) avec une variante de registre *–im* (A), dat. pl. *–ích* (R) et inst. pl. *–ima* (R) avec une variante de registre *–imi* (A). Le sous-type fermé *sto* (`mt_0`) est défini par l'ajout d'une variante fonctionnelle dans le nom., acc. et voc. pl. *–ě* (A).

Le sous-type ouvert ***centrum*** (`mt_c`) est défini par le suffixe *–um* (R) dans le nom., acc. et voc. sg., la variante fonctionnelle *–ě* (X) dans le loc. sg. n'est pas admise. Le sous-type ouvert ***muzeum*** (`mt_m`) a les mêmes propriétés que le sous-type *centrum*, il se distingue en plus par la désinence *-í* (R) dans le gen. pl., par la désinence *–ím* dans le dat. pl. qui repousse la désinence du modèle *–ům* (V) au rôle d'une variante de registre, par le loc. pl. *–ích* (R) et par la variante orthographique de la désinence de l'inst. pl. *–i* (Ri).

#### 3.4.3.12 Type *moře* (`mr`)

D'après le type ***moře*** (`mr`) (voir l'annexe C.1, p. 296–297) sont déclinés les substantifs neutres, dont le nom. et gen. sg. sont exprimés par les désinences *–e* ou *–ě*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. *–ema*.

Le sous-type ouvert ***letiště*** (`mr_l`) est défini par la désinence *–ě* (Re) dans le nom., gen., acc. et voc. sg. et nom., acc. voc., pl., par la désinence *–ěm* (Re) dans l'inst sg. et dans la variante de registre dans l'inst pl. *–ěma* (Re) et par la désinence zéro *–#* (R) dans le gén. pl. Le sous-type fermé *vejce* est défini par la désinence zéro *–#* (R) dans le gén. pl. Le sous-type fermé *poledne* est défini par la désinence *–ách* (R) dans le loc. pl., par la désinence *–y* (Ri) dans l'inst pl. et par son variante de registre *–ama* (R).

#### 3.4.3.13 Type *kuře* (kr)

D'après le type ***kuře*** (kr) (voir l'annexe C.1, p. 297–298) sont déclinés les substantifs neutres, dont le nom. sg. est exprimé par la désinence *–e* ou *–ě* et le gén. sg. par la désinence *–ete* ou *–ěte*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a des variantes de registre dans le dat. pl. (*–atum*) et dans l'inst pl. *–atama*.

Le sous-type ouvert ***mládě*** (kr_m) est défini uniquement par la variante orthographique de la désinence *–ě* (Re) dans le nom., acc. et voc. sg., *–ěte* (Re) dans le gén. sg., *–ěti* (Re) dans le dat. et loc. sg. et *–ětem* (Re) dans l'inst sg. Le sous-type fermé *dítě* a les mêmes désinences au singulier ; au pluriel, ce substantif est féminin et il est décliné d'après le type *kost* – nom., acc. et voc. pl. *–i* (R), gen. pl. *–í* (R), dat. pl. *–em* (R), loc. pl. *–ech* (R), inst. pl. *–mi* (R) avec une variante de registre *–ma* (R).

#### 3.4.3.14 Type *stavení* (st)

D'après le type ***stavení*** (st) (voir l'annexe C.1, p. 298) sont déclinés les substantifs neutres, dont le nom. et gen. sg. sont exprimés par la désinence *–í*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans l'inst pl. *–íma*.

Le sous-type fermé *úterý* est défini uniquement par le remplacement des désinences *–í, –ím, –ích, –ími, –íma* par leur variantes orthographiques *–ý, –ým, –ých, –ými, –ýma*

#### 3.4.3.15 Type *klima* (kl)

D'après le type ***klima*** (kl) (voir l'annexe C.1, p. 299) sont déclinés les substantifs neutres d'origine étrangère, dont le nom. sg. est exprimé la désinence *–a* et le gen. sg. par *–at–u*. Le modèle de ce type n'a pas de variantes fonctionnelles ; il a une variante de registre dans le dat. pl. *–at–um*.

#### 3.4.3.16 Type *rád* (rd)

D'après le type ***rád*** (rd) (voir l'annexe C.1, p. 300) sont déclinés les formes nominales des adjectifs (par exemple *rád, jist, nemocen,* ...). Il s'agit d'un paradigme défectif qui ne contient que des formes pour exprimer le nom. sg. et pl. et acc. sg. et pl. pour les quatre genres.

### 3.4.4 Déclinaison adjectivale

La déclinaison nominale (la valeur de l'attribut `tagMorph` est A) contient 5 types : *mladý, můj, týž, sám, jarní*.

D'après le type ***mladý*** (`md`) (voir l'annexe C.2, p. 301–302) sont déclinés les adjectifs dont le radical est terminé par une consonne dure ou mixte, le nom.sg. est exprimé par les désinences *–ý, –ý, –á, –é* respectivement pour le masculin animé et inanimé, le féminin et le neutre ; le gén. sg. est exprimé par *–ého, –ého, –é, –ého* respectivement pour les quatre genres. Ce type ne contient pas de variantes fonctionnelles ; il contient des variantes de registre liés à la transformation *ý* > *ej* et *é* > *ý*, à l'élimination de la quantité vocalique (*ym* au lieu de *ým*) et à la présence de la voyelle *–a* dans la désinence de l'inst. pl. (*–ýma*) pour les quatre genres.

Le type fermé *můj* (`md_m`) (voir l'annexe C.2, p. 303–304) a des ressemblances formelles avec le type *mladý* mais il est défini comme un type à part à cause de nombreuses spécificités (voir l'annexe C.2, p. 301–302). La situation est la même pour les types fermés *týž* (`md_t`) (voir l'annexe C.2, p. 306–307) et *sám* (`md_s`) (voir l'annexe C.2, p. 305–306).

D'après le type ***jarní*** (`jn`) (voir l'annexe C.2, p. 308–308) sont déclinés les adjectifs dont le radical est terminé par une consonne molle ou mixte, le nom.sg. est exprimé par la désinence *–í* respectivement pour le masculin animé et inanimé, le féminin et le neutre ; le gén. sg. est exprimé par *–ího, –ího, –í, –ího* respectivement pour les quatre genres. Ce type ne contient pas de variantes fonctionnelles ; il contient des variantes de registre liés à l'élimination de la quantité de la voyelle *í* dans le désinence du dat. et loc. pl. de tous les genres *–ím* et à la présence de la voyelle *–a* dans la désinence de l'inst. pl. (*–ýma*).

### 3.4.5 Déclinaison mixte

La déclinaison nominale (la valeur de l'attribut `tagMorph` est M) contient 2 types : *otcův, matčin*.

D'après le type ***otcův*** (`ot`) (voir l'annexe C.2, p. 309–310) sont déclinés les adjectifs possessifs dérivés à partir des substantifs masculins par le suffixe *–ův–/–ov–* ; le nom. sg. est exprimé par les désinences *–#, –#, –a, –o* respectivement pour le masculin animé et inanimé, le féminin et le neutre ; le gén. pl. est exprimé par *–a, –a, –y, –a* respectivement pour les quatre genres. Ce type ne contient pas de variantes fonctionnelles ; les variantes de registre sont les mêmes comme pour le type *mladý* dans l'inst. sg. du masculin animé/inanimé et du neutre et dans tous les cas de tous les genres du paradigme pluriel sauf le nom., acc. et voc.

pl. du féminin et du masculin où il n'y a pas de variantes de registre ; le neutre prend dans le nom., acc. et voc. pl. la désinence masculine *–y* en tant que variante de registre.

D'après le type ***matčin*** (`mt`) (voir l'annexe C.3, p. 310–312) sont déclinés les adjectifs possessifs dérivés à partir des substantifs féminins à l'aide du suffixe *–in–* ; le nom. sg. est exprimé par les désinences *–#, –#, –a, –o* respectivement pour le masculin animé et inanimé, le féminin et le neutre ; le gén. sg. est exprimé par *–a, –a, –y, –a* respectivement pour les quatre genres. Ce type ne contient pas de variante fonctionnelle ; la distribution des variantes de registre est identique au type *otcův*.

### 3.4.6 Déclinaison pronominale

La flexion pronominale (la valeur de l'attribut `tagMorph` est M) est un ensemble de paradigmes fermés contenant les formes casuelles des pronoms personnels, démonstratifs, interrogatifs et relatifs qui ne sont déclinés d'après aucun type de flexion nominale ou adjectivale. Les pronoms appartenant à cette flexion pourraient être divisés en trois groupes :

- les pronoms avec une flexion spécifique (des occurrences du supplétivisme ou des paradigmes défectifs) : ***já*** (`ps_ja`) (voir l'annexe C.4, p. 313), ***my*** (`ps_my`) (voir l'annexe C.4, p. 313), ***ty*** (`ps_ty`) (voir l'annexe C.4, p. 314), ***vy*** (`ps_vy`) (voir l'annexe C.4, p. 314), ***se*** (`ps_r`) (voir l'annexe C.4, p. 315).
- les pronom classés traditionnellement comme un ensemble décliné d'après le type *ten* : ***ten*** (`tn`) (voir l'annexe C.4, p. 315–316), ***kdo*** (`tn_k`) (voir l'annexe C.4, p. 316).
- les pronoms classés traditionnellement comme un ensemble décliné d'après le type *náš* : ***náš*** (`ns`) (voir l'annexe C.4, p. 317–317), ***co*** (`ns_c`) (voir l'annexe C.4, p. 318), ***nic*** (`ns_n`) (voir l'annexe C.4, p. 318), ***on*** (voir l'annexe C.4, p. 319–320), ***jenž*** (`ns_j`) (voir l'annexe C.4, p. 321–322), ***všechen*** (`ns_v`) (voir l'annexe C.4, p. 322–323).

Étant donné leur spécificité formelle, les paradigmes des pronoms *já, ty, my, vy, se, kdo, co, nic, on, jenž, všechen* sont inscrits dans le fichier *pdgm.xml* comme une liste de formes lexicales complètes (comme le paradigme des formes casuelles des pronoms en question), et non pas comme une liste de désinences (comme le paradigme de désinences casuelles). Techniquement, ces listes sont considérées comme des listes de désinences qui s'attachent à un radical de longueur nulle, ce qui permet un traitement unique pour les deux types de représentation des paradigmes.

### 3.4.7 Déclinaison numérale

La flexion numérale (la valeur de l'attribut `tagMorph` est N) contient un ensemble de paradigmes fermés des expressions du nombre qui ne peuvent pas être déclinées d'après aucun type nominal, adjectival et pronominal. Il s'agit des types fermés ***dva*** (`dv`) (voir l'annexe C.5, p. 324– 324), ***pět*** (`nu_i`) (voir l'annexe C.5, p. 325–325), ***kolik*** (`nu_a`) (voir l'annexe C.5, p. 326–326) et ***málo*** (`nu_o`) (voir l'annexe C.5, p. 327–327).

### 3.4.8 Fréquence des types paradigmatiques

Les substantifs dans le lexique A1A2, extrait de Hádková *et al.* (2005) et Bidlas *et al.* (2005), ont été examinés par rapport à leur appartenance aux types paradigmatiques. La connaissance de leur répartition dans le lexique recommandé aux apprenants débutants permet d'établir une correspondance entre ce lexique – qui est un ensemble de lexèmes choisis selon des critères communicationnels et pragmatiques – et les moyens morphologiques qui sont mis en exercice dans l'emploi des lexèmes qu'il contient.

D'après les résultats de la couverture du lexique A1A2 par rapport au lexique des mots tchèques les plus fréquents[16], nous pouvons supposer que la répartition des types paradigmatiques devrait correspondre à leur répartition pour la langue tchèque en général.

Pour obtenir l'information sur la fréquence des types, il est nécessaire d'abord d'extraire les substantifs du lexique et de leur attribuer leur type paradigmatique. Cette tâche a été effectuée par des procédés semi-automatiques utilisés pour le traitement des données textuelles. Les lexèmes dans le lexique A1A2 ont été annotés par l'analyseur morphologique de Hajič[17]. La sortie est un texte avec une annotation morphologique non désambiguïsée, voir un extrait :

```
<f>bratr<MMl>bratr<MMt>NNMS1-----A----
<f>budík<MMl>budík<MMt>NNIS1-----A----<MMt>NNIS4-----A----
<f>bufet<MMl>bufet<MMt>NNIS1-----A----<MMt>NNIS4-----A----
<f>bunda<MMl>bunda<MMt>NNFS1-----A----
```

L'indication de la catégorie lexicale et du genre grammatical est fiable dans la plupart des cas, les erreurs éventuelles ont été enlevées à la main pendant la vérification manuelle du fichier. Le lexique annoté, transformé automatiquement à l'aide d'un script Perl en un fichier au format XML, contient le lemme et les informations grammaticales associées. Il est

[16] Voir la section 3.1.3.2, pp. 38–43.

[17] L'analyseur est disponible sur `http://ufal.mff.cuni.cz/pdt/Morphology_and_Tagging/Morphology/index.html`. Pour l'explication des étiquettes voir `http://ufal.mff.cuni.cz/pdt/Morphology_and_Tagging/Doc/hmptagqr.html`.

donc désormais possible d'extraire du lexique A1A2 (2775 unité lexicales) un ensemble de 1402 substantifs. Leur catégorie lexicale correspond au type morphologique dans la plupart des cas à part 6 lexèmes de déclinaison adjectivale (*dovolená, spropitné, vstupné, poštovné, nájemné, průvodčí*).

Pour chaque substantif, son type paradigmatique lui a été attribué manuellement, souvent à l'aide des expressions régulières qui permettent de choisir un ensemble de lexèmes conformes à un patron, typique pour un certain type. Le résultat de cette recherche a été vérifié et trié manuellement pour l'élimination des erreurs et pour l'attribution éventuelle de différents sous-types. Le format d'une entrée du lexique contenu dans le fichier XML a la forme suivante :

```
<U niv="A1" op="pov" tag="NN" pdgm="mt_r" rankARF="3835">břicho</U>
<U niv="A1" op="nep" tag="NM" pdgm="pn_b" rankARF="1050">bůh</U>
<U niv="A1" op="nep" tag="NI" pdgm="hd_v" rankARF="23920">celer</U>
```

La fréquence est calculée avec le nombre de substantifs appartenant à un certain type paradigmatique par rapport au nombre total de substantifs dans le lexique. Ce calcul est utilisé dans les travaux sur la fréquence de types paradigmatiques que nous avons pu consulter, voir Jelínek *et al.* (1961) et Osolsobě *et al.* (1998) qui traitent des listes de mots extraits de différents corpus textuels. Cependant, les informations sur la fréquence de l'emploi effectif des types paradigmatiques dans le discours ne sont pas disponibles. Un tel calcul devrait être effectué avec les substantifs les plus employés, ce qui permettrait d'obtenir une idée sur la fréquence de l'occurrence des différents types dans le discours. Cette information permettrait d'établir la probabilité de l'occurrence de types paradigmatiques dans des textes réels et non pas dans une liste de mots comme dans notre cas. Ainsi, l'importance des types « irréguliers » qui n'englobent qu'un nombre restreint de lexèmes mais qui sont fréquents dans le discours (par exemple *člověk / lidé, ruka*) serait mise en relief.

Les types paradigmatiques les plus représentés sont *žena* (435 / 31,03%) et *hrad* (406 / 28,96%) qui sont assignés approximativement aux deux tiers de tous les substantifs. La portion restante contient les types qui représentent chacun à leur tour moins de 8% du lexique A1A2 – les types *růže* (109 / 7,77%), *město* (109 / 7,77%), *pán* (83 / 5,92%), *stavení* (73 / 5,21%), *muž* (43 / 5,21%), *stroj* (42 / 3,00 %), *kost* (43 / 2,43%), *píseň* (27 / 1,93 %), *moře* (15 / 1,07%), *předseda* (15 / 1,07%), *kuře* (7 / 0,50%), *soudce* (3 / 0,21%), *klima* (2 / 0,14%).

Parmi les sous-types établis dans notre classification des types paradigmatiques nominaux, les suivants ne sont pas représentés dans le lexique A1A2 : *džigolo* (*pn_d*), *kámen* (*hd_k*), *Hradčany* (*hd_h*), *kůň* (*mz_k*), *paňáca* (*pd_p*), *skica* (*zn_s*), *Stáňa* (*zn_t*), *idea* (*zn_i*), *Lidice* (*rz_l*), *lečo* (*mt_l*), *dvířka* (*mt_d*). Les autres, ouverts ou fermés, sont représentés par au moins un lexème pour chaque sous-type.

Les fréquences des différents types paradigmatiques dans le lexique A1A2 ont été comparés avec des données équivalentes (voir tab. 3.4.8, les chiffres représentent des pourcentage), publiées dans Jelínek *et al.* (1961), Šmilauer (1972) et Osolsobě *et al.* (1998).

| | Jelínek *et al.* (1961) | Šmilauer (1972) | Osolsobě *et al.* (1998) | **A1A2** |
|---|---|---|---|---|
| *pán* | 8,9 | 8,9 | 7,3 | 5,9 |
| *hrad* | 25,1 | 25,1 | 27,7 | 29,0 |
| *muž* | 4,0 | 3,8 | 3,5 | 3,1 |
| *stroj* | 2,7 | 3,0 | 2,6 | 3,0 |
| *předseda* | 1,1 | 1,1 | 0,7 | 1,1 |
| *soudce* | 0,4 | 0,4 | 0,5 | 0,2 |
| *žena* | 22,0 | 21,9 | 22,4 | 31,0 |
| *růže* | 6,2 | 6,2 | 8,2 | 7,8 |
| *píseň* | 2,2 | 2,2 | 1,7 | 1,9 |
| *kost* | 7,0 | 7,2 | 5,0 | 2,4 |
| *město* | 5,6 | 5,6 | 6,2 | 6,9 |
| *moře* | 0,7 | 0,7 | 0,4 | 1,1 |
| *kuře* | 0,4 | 0,4 | 0,3 | 0,5 |
| *stavení* | 11,8 | 11,6 | 9,0 | 5,21 |

TAB. 3.2 – Fréquence relative de types paradigmatiques

Une correspondance relativement forte existe pour les fréquences de la plupart des types (différences de 4% maximum), sauf pour le type *žena*. L'explication de l'excédent dans ce cas devrait être cherchée dans la présence des formes féminines dérivées systématiquement des substantifs masculins, indiquées dans le lexique pour des raisons didactiques, par exemple *cizinec* × *cizinka*.

Le résultat de cette comparaison confirme notre hypothèse que le vocabulaire, défini d'après des critères communicationnels et pragmatiques pour les niveaux A1 et A2, reflète l'usage général de la déclinaison du tchèque. Il ne s'agit pas d'un ensemble de moyens restreints, mais d'un échantillon lexical qui représente le système de la déclinaison du tchèque dans toute sa complexité. Il est possible de conclure, que dans les matériaux didactiques qui sont ou seraient construits à partir des spécifications lexicales et grammaticales dans Hádková *et al.* (2005) et Bidlas *et al.* (2005), les apprenants rencontrent la quasi-totalité des types paradigmatiques représentés dans des proportions proches de la situation générale en tchèque.

## 3.5 Alternances

Les alternances dans la déclinaison du tchèque sont des changements du radical dans certaines formes casuelles, par exemple *kluk–#* > *kluc–i, ves–#* > *vs–i, mýdl–o* > *mýdel–#, kráv–a* > *krav–#*.

### 3.5.1 Sens de l'alternance

Dans notre approche, le phonème participant à l'alternance et figurant dans la forme de base d'un lexème – son lemme – est considéré comme le point de départ de cette alternance ; le phonème correspondant, figurant dans la forme lexicale alternée, est considéré comme le point d'arrivée.

C'est ainsi que nous parlons du sens de l'alternance à partir de la forme de base vers la forme alternée et que nous marquons cette alternances par un schéma au format *élément de départ* > *élément d'arrivée*, par exemple l'alternance dans *kluk–#* > *kluc–i* est marquée *k* > *c*, l'alternance dans *mýdl–o* > *mýdel–#* est marquée comme # > *e*, etc. Cette représentation est en accord avec notre conception de la génération des formes lexicales qui est produite par la concaténation du radical et de la désinence et par l'application des règles d'alternance.

Dans un nombre de cas limité, un radical peut subir une double alternance, par exemple *vejce* > *vajec* avec la réalisation simultanée des alternances *e* > *a* et # > *e*. Dans cette situation, l'ordre de la réalisation des alternances est calculé à partir de la gauche.

### 3.5.2 Vérification de la réalisation des alternances

La présentation des alternances est assez éparse dans les grammaires disponibles. Le plus souvent, elles font partie des exposés sur la dérivation ou les types de déclinaison mais il existe rarement des exposés systématiques visant une présentation complète. Lorsqu'une partie dédiée spécifiquement aux alternances est présentée, voir par exemple Petr *et al.* (1986a), pp. 185–187, Čechová *et al.* (2000), pp. 176–177, les alternances effectuées dans les processus de dérivation et celles réalisées pendant la flexion ne sont pas distinguées explicitement, mais elles sont classées ensemble. Si cette distinction est faite, voir par exemple Šmilauer (1972), pp. 233–236, les alternances effectuées uniquement pendant la déclinaison et celle qui sont réalisées pendant la conjugaison ne sont pas dissociées.

Les insuffisances des grammaires traditionnelles pour un traitement automatique sont soulevées dans Bémová et Králíková (1988). En effet, certaines alternances sont réalisées

régulièrement dans un contexte donné mais d'autres dépendent souvent de critères plus complexes qui ne sont pas explicitement décrits ou dont la réalisation ou l'absence est complètement arbitraire du point de vue synchronique.

Il a donc été nécessaire de faire une étude de la réalisation des alternances produites pendant la déclinaison pour pouvoir définir des règles de leur exécution automatique dans le diagnostic des erreurs. La réalisation ou l'absence des différents types d'alternances a été vérifiée sur l'ensemble des 50 000 lexèmes les plus fréquents, contenus dans le *Frekvenční slovník češtiny*, Čermák et Křen (2004).

Les lexèmes susceptibles de subir une alternance ont d'abord été sélectionnés dans la liste à l'aide des expressions régulières correspondantes aux propriétés formelles définies en fonction d'une alternance spécifique, par exemple un lexème terminé par la suite *consonne + no* pour l'alternance de -e- intercalaire du type *okno* > *oken*. Dans les listes obtenues, les formes non fléchies ont été éliminées manuellement. L'ensemble des lexèmes restants a été examiné selon la réalisation de l'alternance. Pour des cas où l'introspection (ou la connaissance du tchèque d'un locuteur natif), ne pouvait pas suffire pour établir immédiatement la possibilité de réalisation d'une alternance pour un lexème donné, nous avons cherché à nous appuyer sur les recommandations éventuelles dans *Slovník spisovné češtiny pro školu a veřejnost*, Filipec et Daneš (1978), *Pravidla českého pravopisu*, Martincová et kol. (1993) et *Příruční mluvnice češtiny*, Karlík *et al.* (1995). Pour vérifier ces informations en cas de doute, nous avons également consulté le corpus SYN au sein du ČNK pour rechercher des attestations des formes alternées.

Cette technique de travail avec des données langagières est équivalente à une recherche dans les dictionnaires rétrogrades comme ceux de Slavíčková (1975) et Těšitelová *et al.* (1986). L'avantage de cette approche est l'actualité des données soumises à l'analyse ainsi que la flexibilité et la vitesse des recherches à l'aide d'outils offrant la possibilité de travailler avec des expressions régulières. Les listes de lexèmes ainsi établies sont un composant important de la procédure de génération des formes casuelles.

### Exemple de la démarche

Pour illustration, nous allons montrer la procédure de vérification de la réalisation de l'alternance de *-e-* intercalaire dans les substantifs terminé par *–𝒞ež*. L'expression régulière pour rechercher les lexèmes conformes à cette restriction formelle est :

```
^.*[^aáeéěyýuúůiíoó]ež$
```

Cette expression est vérifiée par toutes les suites dont le troisième caractère à partir de la fin n'est pas une voyelle (c'est donc une consonne) et les deux caractères restants sont *e* et *ž*. Avec l'outil `egrep`[18] et la commande :

```
egrep ^.*[^aáeéěyýuúůiíoó]ež$ 50000alphabetic.txt
```

où *50000alphabetic.txt* est le nom du fichier contenant le lexique (un mot par ligne, classés alphabétiquement), nous obtenons une liste de lexèmes correspondante à cette expression : *drůbež, fermež, krádež, lež, loupež, mládež, načež, než, papež, pročež, svatokrádež, verbež, řež*. Nous éliminons les lexèmes *načež* et *než*, car ce ne sont pas des substantifs. Parmi le reste, nous identifions uniquement le lexème *lež* qui peut subir l'alternance (*lež* > *lži, lží, lžím, lžích, lžemi*). Les *Pravidla českého pravopisu* attestent cette alternance, le *Slovník spisovné češtiny pro školu a veřejnost* également. Le corpus SYN indique 4417 occurrences du génitif singulier *lži*, la forme hypothétique *leži* qui serait produite sans la réalisation de l'alternance *e* > # est attestée 8 fois, mais il s'agit uniquement d'une forme courte dialectale de la 3e personne du pluriel du verbe *ležet – leži*.

### 3.5.3 Alternances consonantiques

Dans le cadre de la déclinaison du tchèque, nous distinguons quatre types d'alternances consonantiques, toutes liées au phénomène historique de la palatalisation : la mouillure (ou palatalisation 0), la palatalisation A, la palatalisation B et la palatalisation de groupe consonantique.

- La **mouillure** (ou *palatalisation* 0) est la variation des consonnes alvéolaires *d, t, n, r* avec leur variantes palatalisées respectives *d', t', ň, ř*, par exemple *doktor* > *doktoři, zahrada* > *zahradě*.
- La **palatalisation A** est la variation des consonnes vélaires *k, g, h, ch* avec leur variantes palatalisées respectives *c, z, z, š*, par exemple *vlk* > *vlci, moucha* > *mouše*.
- La **palatalisation B** est l'alternance des consonnes vélaires *k, c, h* avec leur variantes palatalisées respectives *č, č, ž*, par exemple *člověk* > *člověče, bůh* > *bože*.
- La **palatalisation de groupe** est l'alternance des groupes de consonnes *sk, ck* avec leur variantes palatalisées respectives *št', čt'*, par exemple *českú* > *češt'í, anglický* > *angličt'í*.

[18] Voir `http://www.ss64.com/bash/egrep.html`.

#### 3.5.3.1 Mouillure

Dans le cadre de la déclinaison, la mouillure est effectuée au contact du radical terminé par *d, t, n, r* avec une désinence casuelle dont la périphérie gauche commence par une voyelle molle (*i, í, ě*). La direction de cette alternance (au sens *lemme* > *forme fléchie*) suit généralement celle de la palatalisation, bien qu'il existe des cas où elle est effectuée dans le sens inverse (par exemple *paměť–#* > *paměť–ech*).

En forme écrite, un marquage explicite du graphème exprimant la consonne alternée est effectué uniquement pour la consonne *r* (par exemple *doktor–#* > *doktoř–i*) à l'aide du signe diacritique ˇ . Pour les graphèmes exprimant les consonnes *d, t, n*, l'alternance reste sans marquage (par exemple *nálad–a* > *nálad–ě, žen–a* > *žen–ě, vět–a* > *vět–ě*), car la mouillure de la consonne est signalée, dans le système orthographique du tchèque, par la voyelle qui suit cette consonne (*ě, i, í*).

De ce fait, l'ajout ou la suppression d'un signe diacritique ne signifie pas forcément (à part le couple *r* > *ř*) une alternance au plan morpho-phonologique (par exemple *mlád–ě* > *mláď–ata, jabloň–#* > *jablon–ě, odpověď–#* > *odpověd–i*) car la valeur phonologique de la consonne en question ne change pas : elle reste palatalisée. C'est un constatation importante pour l'élaboration des procédures automatiques travaillant avec des symboles écrits car il va être nécessaire de dissocier les deux types de variation en fonction du niveau de traitement (morpho-phonologique et graphique).

Sur la base des critères cités plus haut, nous distinguons (1) la mouillure marquée graphiquement, (2) la mouillure non marquée graphiquement, (3) l'alternance graphique et (4) la dépalatalisation.

**(1) Mouillure marquée graphiquement** Le seul cas où l'alternance phonologique de la mouillure coïncide avec l'alternance graphique de la consonne en question est l'alternance **r > ř** . Elle est effectuée devant la désinence *–e* dans le dat. et loc. sg. du type *žena* (par exemple *pár–a* > *pář–e*), dans le loc. sg. du type *hrad* (par exemple *pivovar–#* > *pivovař–e*) et *město* (par exemple *pouzdr–o* > *pouzdř–e*), dans le voc. sg. des substantifs du type *pán* terminés par *–r* précédé d'une consonne (par exemple *bratr–#* > *bratř–e, Petr–#* > *Petř–e)*[19] ; devant la désinence *–i* dans le nom. pl. du type *pán* (par exemple *doktor–#* > *doktoř–i*) ; devant la désinence *–í* dans le nom. pl. du type *mladý* (par exemple *star–ý* > *stař–í*). L'alternance analogue *z* > *ž* se produit dans les formes archaïques de pluriel du substantif *kněz* (*kněž–í*, *kněž–ím*, *kněž–ích*, *kněž–ími*), il s'agit dans ce cas de la seule occurrence de l'alternance *z* > *ž* dans la flexion nominale.

[19]Cette alternance se produit également dans les variantes archaïques de pluriel du substantif *bratr* : *bratř–í*, *bratř–ím*, *bratř–ích*, *bratř–ími*.

**(2) Mouillure non marquée graphiquement** Il s'agit d'une alternance réalisée au plan phonologique mais qui n'est pas reflétée graphiquement, elle n'est pas donc pertinente pour une procédure de génération des formes fléchies travaillant avec des symboles écrits. Elle est effectuée sur les radicaux terminés par *d, t, n* devant la désinence *–ě* dans le dat. et loc. sg. du type *žena* (par exemple *zahrad–a* > *zahrad–ě*, *tet–a* > *tet–ě*, *žen–a* > *žen–ě*), dans le loc. sg. du type *hrad* (par exemple *hrad–#* > *hrad–ě*, *svět–#* > *svět–ě*, *mlýn–#* > *mlýn–ě*) et *město* (par exemple *hnízd–o* > *hnízd–ě*, *měst–o* > *měst–ě*, *vlákn–o* > *vlákn–ě*) ; devant la désinence *–i* dans le nom. pl. du type *pán* (par exemple *had–#* > *had–i*, *kmet–#* > *kmet–i*, *pán–#* > *pán–i*) ; devant la désinence *–í* dans le nom. pl. du type *mladý* (par exemple *mlad–ý* > *mlad–í*, *ušat–ý* > *ušat–í*, *jemn–ý* > *jemn–í*).

**(3) Alternance graphique** Les alternances **d > d'** , **t > t'** , **n > ň** et **d' > d** , **t' > t** , **ň > n** ne reflètent pas une alternance morpho-phonologique[20] – elles sont uniquement un effet des règles orthographiques qui ne permettent pas la cooccurrence de deux marqueurs de la mouillure (de la consonne diacritée et de la voyelle molle, par exemple **odpověd'–i*). L'alternance graphique dans le sens *sans diacritique* > *avec diacritique* est effectuée dans le gen. pl. des types *růže* et *moře* avec la désinence zéro –# (par exemple *žákyn–ě* > *žákyň–#*, *letišt–ě* > *letišt'–#*) et devant les désinences *–ata, –at, –aty, –atech, –atům* du type *kuře* (par exemple *mlád–ě* > *mlád'–ata*). L'alternance graphique dans le sens inverse *avec diacritique* > *sans diacritique* est effectuée devant les désinences molles (*–i, –ě, -ěmi, –í, –ím, –ích*) pour tous les substantifs terminés par une consonne molle (par exemple *lod'–#* > *lod–i*, *kůň–#* > *kon–ě*, *větroň–#* > *větron–ěm*, *chot'–#* > *chot–ěm*).

**(4) Dépalatalisation** L'alternance inverse (durcissement) est rare – seulement quelques lexèmes appartenant au type *kost* terminés par une consonne molle sont sujets à la dépalatalisation. L'alternance **d' > d** est effectuée dans le dat., loc. et inst. pl. du lexème *zed'* (*zd–em*, *zd–ech*, *zd–mi* avec des variantes dans les désinences qui ne provoquent pas la dépalatalisation du radical *zd–ím*, *zd–ích*) ; l'alternance **t' > t** est effectuée dans le même cas pour les lexèmes *obět'* et *pamět'* (*obětem*, *pamětem*), possédant également des variantes avec des désinences molles (*obětím*, *pamětím*).

#### 3.5.3.2 Palatalisation A

La palatalisation A est effectuée au contact du radical terminé par une des consonnes *k / g / h / ch* avec une des désinences suivantes : *-i / -e / -ích / -í* . La direction de cette alternance

[20] A part de quelques cas rares de dépalatalisation pour l'alternance dans le sens *avec diacritique* > *sans diacritique*, voir plus loin.

suit sans exception celle de la palatalisation. La forme écrite reflète sans ambiguïté la réalité phonologique.

Les alternances **k > c** , **g > z** , **h > z** et **ch > š** sont effectuées dans le dat. et loc. sg. du type *žena* (par exemple *holk–a* > *holc–e*, *Olg–a* > *Olz–e*, *touh–a* > *touze–e*, *střech–a* > *střeš–e*), dans le loc. sg. du type *hrad* (par exemple *rybník–#* > *rybníc–e*), du type *město* (par exemple *mlék–o* > *mléc–e*, *rouch–o* > *rouš–e*), dans le nom. pl. du type *pán* (par exemple *vlk–#* > *vlc-i*, *hoch–#* > *hoš-i*) et dans le loc. pl. du type *pán* (par exemple *vlk–#* > *vlc-ích*, *hoch–#* > *hoš-ích*), du type *hrad* (par exemple *rybník–#* > *rybníc–ích*) et du type *město* (par exemple *středisk–o* > *středisc–ích*).

#### 3.5.3.3 Palatalisation B

Dans le cadre de la déclinaison, la palatalisation B est limitée uniquement au vocatif singulier de quelques lexèmes du type *pán*. Pour l'alternance **k > č** , il s'agit du lexème *člověk – #* > *člověč–e*, pour l'alternance **c > č** , *otec–#* > *otč–e*) et pour l'alternance **h > ž** (*bůh–#* > *bož–e*). L'alternance *k* > *č* est effectuée également dans les formes spécifiques du pluriel du substantif *oko* (*oči, očí, očí, očích, očima*) qui sont d'anciennes formes du duel. L'alternance *c* > *č* apparaît également dans le paradigme du pronom *co* / *nic* (*čeho / ničeho, čím / ničím, ...*). La direction de cette alternance suit sans exception celle de la palatalisation. La forme écrite reflète sans ambiguïté la réalité phonologique.

#### 3.5.3.4 Palatalisation des groupes consonantiques

Dans le cadre de la déclinaison, la palatalisation des groupes consonantiques **sk > št** et **ck > čt** est liée exclusivement à des adjectifs de type *mladý* dont le radical est terminé par un des deux groupes consonantique, attaché à la désinence *–í* exprimant le nom. et voc. pl. du masculin animé (par exemple *francouzský* > *francouzští, řecký* > *řečtí*).

### 3.5.4 Alternances vocaliques

Dans le cadre de la déclinaison du tchèque, il existe trois types d'alternances vocaliques : les alternances qualitatives, quantitatives et les alternances avec le *-e-* intercalaire.

- Les **alternances qualitatives** sont des variations exceptionnelles de deux voyelles qui ne sont pas dans l'opposition phonologique de quantité (la longueur vocalique) : *a* > *e, e* > *a, í* > *á*, par exemple *přítel* > *přátelé*.

- Les **alternances quantitatives** sont des variations de deux voyelles qui sont en opposition de quantité. Du point de vue de la situation synchronique du système phonologique du tchèque, cette opposition peut être directe (á > a, é > e, í > i) ou indirecte (í > ě, ou > u, ů > o)[21], par exemple *kráva > krav, kůň > koně*.
- Les **alternances de *-e-* intercalaire** sont définies soit par une suppression (*e* > #), soit par une insertion (# > *e*) de *-e-* mobile (intercalaire) dans la racine ou dans le suffixe d'un substantif, par exemple *chlap-e-c > chlap-#-ce, hes-#-lo > hes-e-l*.

Dans les contextes spécifiés, les alternances consonantiques se déroulent sans exception. Ce n'est pas le cas pour les alternances vocaliques. D'après la régularité de la réalisation de l'alternance, nous pouvons classer les substantifs conformes à la réalisation d'un patron donné (par exemple *𝒞ek, 𝒞ka*) dans trois groupes :

- l'alternance est réalisée pour la majorité des lexèmes conformes à ce patron. En d'autres mots, l'alternance est réalisée toujours sauf pour un petit ensemble d'exceptions : nous les appelons les **exceptions positives**.
- la réalisation et l'absence de l'alternance sont reparties de façon plus ou moins égale. Autrement dit, il existe deux ensembles plus ou moins volumineux dont l'un contient des substantifs qui alternent et l'autre ceux qui n'alternent pas.
- l'alternance n'est pas effectuée pour la majorité des lexèmes conformes au patron donné. En d'autres mots, l'alternance n'est réalisée jamais sauf pour un petit ensemble d'exceptions : nous les appelons les **exceptions négatives**.

Ce classement, basé sur un critère quantitatif, nous permet d'établir des listes d'exceptions positives (le premier et le deuxième groupe selon le cas) ou négatives (le deuxième et le troisième groupe), en respectant toujours le critère du nombre d'éléments le moins élevé. Ces exceptions sont inscrites dans le fichier *alt.xml*, nécessaires pour la génération des formes fléchies (voir la section 3.5.5, p. 98).

#### 3.5.4.1 Alternances qualitatives

Les alternances qualitatives sont effectuées dans certaines formes d'un ensemble fermé de lexèmes. Il s'agit de l'alternance optionnelle ***a > e*** effectuée uniquement dans le voc. sg. du nom propre *Jan > Jene* (une forme sans alternance *Jane* est possible) et des alternances obligatoires ***e > a*** , effectuée dans le gen. pl. de *vejce > vajec*, et ***í > á***, effectuée dans les formes de pluriel de *přítel / nepřítel > přátelé, přátel, přátelům, přátele, přátelích, přáteli*.

[21] Une telle alternance peut être interprétée comme une opposition de quantité changements historiques du système vocalique tchèque. Elle est parfois considérée comme une alternance simultanée de qualité et de quantité, voir Karlík *et al.* (2002), p. 35.

#### 3.5.4.2 Alternances quantitatives

Les substantifs susceptibles de subir une alternance quantitative peuvent être classés dans deux groupes en fonction du nombre de syllabes qu'ils contiennent et de leur type paradigmatique :

- l'ensemble **(A)** contient des substantifs à deux syllabes avec une voyelle longue dans la racine du lemme appartenant aux types *žena*, *město* ou *růže* (des paradigmes vocaliques), par exemple *tráva, dílo, plíce*.
- l'ensemble **(B)** contient des substantifs monosyllabiques avec une voyelle longue dans la racine du lemme appartenant aux types *pán, hrad, kost* (des paradigmes consonantiques), par exemple *vůl, sníh, sůl*.

Un cas spécifique d'une alternance quantitative est l'alternance *ů* > *o* dans le suffixe *–ův* des formes masculines des adjectifs possessifs du type *otcův*. Il s'agit de la seule alternance qualitative dont le distribution est entièrement régulière.

Si l'alternance a lieu, les substantifs dans la classe (A) alternent au gén. pl. (où la désinence zéro –# rend le lexème monosyllabique, par exemple *kráva* > *krav, chvíle* > *chvil, jméno* > *jmen*), moins souvent et uniquement pour les substantifs qui alternent dans le gen. pl., l'alternance peut être effectuée dans le dat., loc. et inst. pl. *kráva* > *kravami*) et exceptionnellement dans l'inst sg. (par exemple *práce* > *prací*)[22].

Pour l'ensemble (B), si l'alternance a lieu, elle est effectuée dans tous les formes casuelles sauf dans le nom. sg. et dans le acc. sg. des masculins inanimés et dans le acc. sg. des féminins (par exemple *kůň* > *koně, hrách* > *hrachu, líh* > *lihu*).

***á > a*** **(A)** Pour les substantifs bisyllabiques du type *žena*, l'alternance est effectuée approximativement dans un tiers des cas (16 sur 51) – elle est effectuée dans *blána, brána, bába, čára, dáma, dráha, chvála, jáma, kráva, pára, rána, skála, tráva, vrána, váha, záda, žába, práce*, elle n'est pas effectuée dans les mots d'origine étrangère (*čáka, fáma, gáza, káva, kára, láva, máta, páka, sága, šála, váza*) mais aussi dans les mots d'origine autochtone, dont une partie peut être considérée comme un ensemble de dérivés verbaux plus ou moins opaques (*hnáta, kláda, schrána, sláva, správa, spála, spára, spása, šťára, šťáva, škvára, vláda, vláha, zkáza, zpráva, ztráta*), la partie restante contient des substantifs divers (*krása, káča, máma, nána, sláma, škála, žláza*). Pour les substantifs bisyllabiques du type *růže*, l'alternance est effectuée uniquement pour le lexème *práce*, elle n'est pas effectuée

[22]Pour la plupart des lexèmes alternés dans le gen. pl., l'alternance dans le dat., loc. et inst. pl. est optionnelle, les formes alternées font souvent partie des locutions figées (voir Karlík *et al.* (1995), p.257).

dans les mots d'origine étrangère *báje, báze, fráze, fáze, gáže, láce* mais aussi dans *ráže, záře*. Il n'y aucun mot bisyllabique du type *město* qui subisse l'alternance *á* > *a* (*bláto, dláto, fáro, kápo, právo, ráno, stádo, ságo, tágo*). Cette alternance est également effectuée dans tous les cas sauf nom. et acc. du lexème *kámen* et dans la déclinaison des pronom *náš, váš*. **(B)** Pour les lexème monosyllabiques, ce ne sont que les substantifs *hrách, mráz, práh* et les pronoms *náš, váš* qui sont alternés, les autres susceptibles ne le sont pas (*báj, bál, báň, čáp, dráb, dráp, drát, flák, flám, fáč, hnát, háj, hák, háv, chrám, chán, král, krám, kád', křáp, lhář, lák, lán, mág, máj, mák, plán, plát, pláč, pláň, pláž, prám, pták, pád, pár, pás, páv, rváč, ráj, řád, rám, ráz, sklář, skráň, snář, spád, spár, spáč, sráz, sráč, stráň, stráž, stáj, stát, stáž, svár, sáh, sál, škvár, špás, šrám, štváč, štáb, šváb, šál, tchán, tkát, tkáň, tác, vpád, vál, zád', zář, žhář, žák, žár*). Une réalisation spécifique de l'alternance *á* > *a* est attestée pour le substantif *pán*. Au cas où il figure dans un groupe nominal en juxtaposition avec un nom (par exemple *pan Novák – monsieur Novák*), la voyelle est toujours courte. S'il n'est pas spécifié, la voyelle est longue et elle alterne uniquement dans le voc. sg. (*pán* > *pane*).

***é > e*** **(A)** Pour les substantifs bisyllabiques du type *město*, elle est effectuée dans *jméno, léto, péro*, elle n'est pas effectuée dans *krédo, léno, mléko, éro*. Pour les types *růže*(*léze, péče*) et *žena* (*céva, fréza, réva, schéma, scéna, sféra, stéla, tréma, éra*), l'alternance ne se produit pas. **(B)** Pour les lexèmes monosyllabiques, l'alternance est effectuée uniquement dans *chléb*, elle ne se produit pas dans *fén, lék, též, šéf, šém, drén, klér, krém, kvér, pléd*.

***í > i*** **(A)** Pour les substantifs bisyllabiques du type *žena*, l'alternance n'est effectuée nulle part (*bříza, cícha, číča, hlína, hlíva, hlíza, hříva, jíva, křída, míza, mícha, pípa, řípa, říza, slída, střída, škvíra, šmíra, špína, tíha, třída, víla*) à part les lexèmes *lípa, mísa, slíva, síla, síra, žíla* ; pour le type *růže*, elle est effectué dans *chvíle, lžíce, míle, plíce, svíce* et elle ne l'est pas dans *díže, píce, píle, příze, tíže, číše, říje, říše, šíje, šíře, žíně*. Pour les lexèmes appartenant au type *město* (*síto, víko, víno*), l'alternance *í* > *i* ne se produit pas. **(B)** Pour les lexèmes monosyllabiques, l'alternance n'est pas effectuée (*blín, cíl, cín, cíp, číš, díl, dříč, dřík, džíp, fík, hřích, jíl, klíč, klín, kmín, knír, kříž, kvíz, líc, míč, mír, mříž, píd', příd', psík, říz, síň, šíp, sít', smích, smír, špíz, spíž, stín, štír, vír, žír*) sauf dans le lexème *líh*.

***í > ě*** **(A)** Pour les substantifs bisyllabiques du type *žena*, l'alternance n'est effectuée nulle part (voir la liste plus haut) sauf dans les lexèmes *bída, díra, míra, víra* ; pour le type *město*, elle est effectuée uniquement dans *dílo* et pour le type *růže*, il existe une variante dans la forme alternée du gén. pl. du lexème *svíce* > *svic / svěc*. Cette alternance s'effectue également dans les formes de pluriel du substantif *dítě* (*děti, dětí, dětem, dětech, dětmi*), dans tous les cas

sauf le nom. et acc. du substantif *vítr* et dans toutes les formes du pluriel sauf le nominatif d'un seul substantif trisyllabique *peníze* > *peněz, penězům, penězích, penězi*. **(B)** Pour les substantifs monosyllabiques, l'alternance *í* > *ě* s'effectue uniquement pour le lexème *sníh*.

***ou* > *u*** **(A)** Pour les substantifs bisyllabiques du type *žena*, l'alternance est effectuée dans *bouda, houba, hrouda, louka, smlouva, strouha, trouba*, elle ne l'est pas dans *chlouba, coura, gouda, mouka, moula, ňouma, roura, routa, stoupa, šmouha, čmouha, touha, vzpoura, zhouba* ; pour le type *město*, elle ne se produit jamais (*pouto, rouno, kouzlo, moudro, poutko, roucho, sousto*) et il en est de même pour le type *růže* (*boule, bouře, koule, louže, nouze*). **(B)** Pour les substantifs monosyllabiques, l'alternance *ou* > *u* ne se produit jamais (*bloud, brouk, fous, houf, kloub, kňour, kouč, kouř, kout, louč, louh, louž, mour, pout', proud, roub, sloup, soud, souš, spoušť, šroub, šťouch, toust, troud, troup, vous, žrout*).

***ů* > *o*** **(A)** Pour les substantifs bisyllabiques du type *žena*, l'alternance est effectuée dans *hůra, fůra, smůla*, elle ne se produit pas dans *blůza, chůva, hrůza, krůta, kůra, lhůta, lůza, můra, půda, šňůra, stvůra, zrůda* ; pour le type *město*, l'alternance ne se produit jamais (*slůvko, lůno* ; pour le type *růže* l'alternance ne s'effectue pas (*chůze, nůše, půle, růže, schůze, vůle, zvůle*) sauf pour le lexème *kůže*. L'alternance *ů* > *o* se produit sans exception dans tous les cas sauf le nom. sg. et le acc. sg. i. des formes masculines des adjectifs possessifs du type *otcův* (par exemple *otcova, otcovi, etc.*). **(B)** Pour les substantifs monosyllabiques, l'alternance *ů* > *o* s'effectue approximativement dans deux tiers des cas – elle se produit dans *bůh, dvůr, důl, dům, hnůj, hůl, kůň, lůj, vůz, nůž, stůl, sůl, vůl, vůz* et elle ne se produit pas dans *kůl, kůr, půl, růž, trůn, tůň*.

#### 3.5.4.3 Alternance -*e*- intercalaire

Étant donnée l'origine de cette alternance (la disparition ou la vocalisation des voyelles courtes réduites dans le vieux tchèque d'après leur position dans le mot[23]), le -*e*- mobile peut être situé dans la racine du substantif alterné (par exemple *pes* > *psa*) ou dans un suffixe (par exemple *znalec* > *znalce, píseň* > *písně*).

Du point de vue synchronique, la distribution de cette alternance est déterminée d'abord par des contraintes de prononciation. Les alternances avec le -*e* mobile peuvent être divisées en deux groupes : dans le premier groupe, le -*e* est supprimé lors de l'alternance (l'alternance ***e* > #**, par exemple *věd-ec* > *věd-#-c–e*) ; dans le second groupe, le -*e* est inséré (l'alternance

[23] Voir la règle de Havlík dans Karlík *et al.* (2002), p. 163 et 203.

***# > e***, par exemple *hlás-#-ka* > *hlás-e-k*)[24]. Une alternance particulière est effectuée pour le lexème *tkadl–ec* > *tkal-#-c–e* où l'alternance # > *e* est accompagné par l'effacement de la consonne *d*.

Dans le cadre de la déclinaison, la suppression du *-e-* peut s'effectuer uniquement si celui-ci est situé à la seconde position à partir de la fin du lemme et suivi par une consonne (par exemple *pes* > *psa, broskev* > *broskve, kozel* > *kozla*). Il existe trois lexèmes subissant l'alternance *e* > # qui ne succombent pas à cette règle : *čest, lest, křest*. Dans ces cas, le -e- alterné (par exemple *cti, lsti, křtu*) est situé à la troisième position à partir de la fin, les autres lexèmes avec une périphérie droite identique n'alternent pas (*arest, atest, azbest, bolest, chřest, fest, incest, interest, manifest, motorest, neřest, pelest, počest, protest, psychotest, ratolest, rest, svěžest, test, trest, šelest, šest*). Les lexèmes subissant cette alternances régulièrement appartiennent aux types paradigmatiques consonantiques (*pán, hrad, muž, stroj, píseň, kost*). Si l'alternance a lieu, elle est effectuée dans tous les cas sauf dans le nom. sg. de tous les genres (la forme du lemme avec le *-e-* intercalaire) et dans la forme de l'acc. sg. des masculins inanimés et des types féminins *píseň, kost*, qui est identique avec la forme du nominatif.

L'alternance opposée, l'insertion de *-e-*, peut être effectuée uniquement dans le gén. pl. (exprimé par la désinence zéro *–#*) des substantifs appartenant aux types vocaliques *žena, město* et, pour une seule exception (lexème *vejce*), au type *moře*. Leur radical doit être terminé par deux consonnes consécutives – le *-e-* est inséré entre ces deux consonnes (par exemple *stovka* > *stovek, stehno* > *stehen, vejce* > *vajec*). La seule exception à cette règle semble être uniquement l'alternance du lexème *mzda* > *mezd* où le *-e-* est déplacé devant les deux consonnes. Une certaine difficulté pour évaluer la possibilité de la réalisation de l'alternance # > *e* se présente pour les lexèmes dont les formes de pluriel sont théoriquement possibles mais qui ne s'emploient pas - il s'agit souvent des noms abstraits comme *hudba* > *?hudeb, hanba* > *?haneb/hanb*. Le même problème pourrait exister pour les formes adverbiales en *–o* dérivées des adjectifs (les prédicatifs, par exemple *horko, vlhko, těžko*, voir Petr *et al.* (1986b), p. 15-16). Dans certains contextes, ces formes peuvent fonctionner comme un substantif (par exemple *Nemám rád vlhko*), elles peuvent donc être théoriquement déclinées et subir une alternance dans leurs formes casuelles appropriées (*vlhko* > *?vlhek/vlchek/vlhk*). Il s'agit cependant d'un phénomène plutôt hypothétique et nous n'allons pas examiner la capacité de ces formes à être alternées. Dans certains cas, la réalisation ou l'absence de l'alternance peut être déterminée par des raisons sémantiques – par exemple *karta* > *karet* pour

[24]D'après nos recherches, ils n'existent que trois lexèmes qui peuvent réaliser les deux alternances à la fois : *švec* > *ševce, pestřec* > *pesterce*. Pour les deux substantifs, il existe des variantes avec une simple alternance (*švec* > *švece, pestřec* > *pestrce*) qui sont des variantes acceptables, pour *pestřec*, il s'agit même de la variante recommandée (voir Prošek (2007), p. 178).

*karta* au sens de carte de jeu mais *karta* > *kart* pour *karta* au sens de feuille dans un classeur (voir Karlík *et al.* (1995), p. 256). Ces exemples sont toutefois marginaux, les résultats de l'évolution historique et les critères de prononciation qui gouvernent cette alternance étant assez puissants pour que la variation dans l'alternance reste faible.

Comme pour les alternances quantitatives et qualitatives, les lexèmes avec des propriétés formelles favorables à la réalisation d'une alternance ont été recherchés dans la liste de 50 000 mots à l'aide des expressions régulières et la réalisation de l'alternance a été vérifiée cas par cas. Pour l'alternance *e* > #, le patron est $[\textit{consonne } \mathscr{C} + e + \textit{consonne } \mathscr{C}]$ à la fin du lemme ; pour l'alternance # > *e*, le patron est $[\textit{consonne } \mathscr{C} + \textit{consonne } \mathscr{C} + a/o/e]$ à la fin du lemme.

***e* > #** L'alternance est réalisée plus ou moins régulièrement pour les lemmes terminés par *−ek, −ec, −eň*.

Les substantifs avec la périphérie ***–ek***, appartenant donc aux types *pán* (par exemple *ptáček*) et *hrad* (par exemple *vzorek*) alternent sans exception si la périphérie *–ek* a la fonction d'un suffixe dérivationnel (diminutif ou autre). Pour les substantifs qui n'alternent pas (23 sur 777 possibles), deux critères pour identifier leur insensibilité à l'alternance peuvent être établis : 1) ils sont monosyllabiques (*bek, flek, jek, klek, rek, skřek, šnek, špek, šek, vlek, vztek*) et / ou 2) le *-e-* se trouve dans la racine d'un dérivé verbal (*dotek, oblek, poklek, průsek, úlek, úsek, výsek, zásek*). Nous avons trouvé une exception à cette règle : le lexème *nářek*, qui alterne régulièrement. Les substantifs suivants qui n'alternent pas, sont des lexèmes d'origine étrangère *bek, krosček, bodyček* et le substantif *česnek*.

La situation est semblable pour la périphérie ***–ec***. Ce sont des substantifs appartenant principalement au type *muž* (par exemple *herec*) et *stroj* (par exemple *hrnec*) qui sont terminés par cette suite. Il existe quelques substantifs terminés par *–ec* appartenant au type *píseň* (par exemple *obec, klec, pec*). Parmi les 381 substantifs susceptibles d'alterner, ce ne sont que 5 lexèmes, tous monosyllabiques, qui n'alternent pas (*hec, kec, klec, pec, plec*).

La périphérie ***–eň*** se trouve dans les substantifs appartenant au type *muž* (par exemple *učeň*), *stroj* (par exemple *stupeň*), *píseň* (par exemple *báseň*). L'alternance s'effectue régulièrement (43 lexèmes sur 49 susceptibles) sauf pour les lexèmes monosyllabiques (*dřeň, žeň*) et les lexèmes où le groupe *–eň* est précédé par *l* (*holeň, zeleň, tuleň*) ou par un groupe consonantique (*červeň*).

La réalisation ou l'absence de l'alternance *e* > # sont reparties de façon équivoque pour les substantifs terminés par *–em*, *–ev*, *–en* et *–eš*.

Pour la périphérie ***–em***, l'alternance est effectuée dans *dojem, nezájem, nájem, pojem, pronájem, podnájem, příjem, zájem, průjem*, elle ne se produit pas dans les mots d'origine

étrangère (*džem, faxmodem, gem, golem, modem, rekviem, tandem, totem*), dans les lexèmes monosyllabiques (*lem, vjem, zem*) et leurs composés (*černozem*), mais aussi dans *plynojem, vodojem*. Les mots étrangers à part, les motifs de la distribution de cette alternance pourrait être vues dans la différence entre les lexèmes déverbatifs (*dojem* < *dojímat, průjem* < *projmout,* etc.) qui alternent régulièrement et les lexèmes *plynojem, vodojem*, créés par la composition (*plyn-o-jem*) qui n'alternent pas.

Pour la périphérie **–ev** , l'alternance est effectuée dans les lexèmes multisyllabiques *broskev, brukev, církev, konev, koroptev, korouhev, lahev, láhev, mrkev, pánev, plástev, ploutev, podešev, rakev, ředkev, tykev, větev* et monosyllabiques *krev, lev, šev*, elle n'est pas effectuée lorsque la périphérie *–ev* fait partie de la racine d'un dérivé verbal *jev, nálev, název, objev, ohřev, projev, rozsev, řev, výjev, výlev, výsev, zjev* mais aussi dans *tetřev*.

La périphérie **–en** (77 substantifs du type *pán* ou *hrad*) fait souvent partie des mots d'origine étrangère (par exemple *eben, karoten, antigen*) mais aussi des mots autochtones (par exemple *jelen, jícen, křen*). Après l'élimination des mots d'origine étrangère, où l'alternance ne se produit jamais, nous obtenons un groupe de substantifs qui alternent (*blázen, březen, buben, červen, den, duben, hrozen, jícen, květen, leden, osten, půlden, říjen, sen, srpen, svícen, týden* et le pronom *jeden*) et un groupe de substantifs qui n'alternent pas (*čeřen, člen, hlen, hřeben, ječmen, jelen, kámen, kmen, kořen, křemen, křen, len, lumen, lupen, plamen, pramen, prsten, pupen, řemen, rodokmen, sten, třmen, ženšen*).

Pour les quatre substantifs avec la périphérie **–eš** , il y en a deux (*veš, faleš*) qui alternent tandis que *pleš, veteš* n'alternent pas.

La réalisation exceptionnelle de l'alternance *e* > # est effectuée dans des substantifs dont le lemme est terminé par *–el, –et, –es, –er, –eb, –ed', –ež*.

La périphérie **–el** a été trouvée dans 269 substantifs dont la plupart sont des masculins animés terminés par le suffixe *–tel* (type *muž*, par exemple *obyvatel*) mais aussi des lexèmes appartenant au type stroj (par exemple *chmel*), hrad (par exemple *účel, kabel*) ou *píseň* (par exemple *ocel, sardel*). L'alternance *e* > # n'est effectuée nulle part sauf dans les substantifs *d'ábel, chumel, kachel, kašel, kel, kotel, kozel, kyčel, obratel, orel, osel, posel, pytel, uhel, uzel, úhel, živel*.

La périphérie **–et** a été trouvée dans 139 substantifs du type *pán* (par exemple *atlet, kmet, brunet*) ou *hrad* (par exemple *hřbet, výlet*). Ce sont uniquement les dérivées du verbe *počítat* qui alternent (*bezpočet, kmitočet, letopočet, počet, přepočet, odpočet, součet, výpočet, účet, zápočet, výčet*) avec les substantifs *dehet, loket, nehet, ocet, ret, tucet, půltucet*.

La périphérie **–es** , retrouvée dans 39 substantifs, fait souvent partie des mots étrangers (par exemple *dres, blues, exces, stres*). Après leur élimination, nous obtenons un ensemble

de quatre substantifs où l'alternance est effectué (*pes, kočkopes, ves, náves*) et un ensemble où elle ne se produit pas (*les, okres, otřes, ples, pokles, prales, přednes, třes, výkres, vřes, útes, účes*).

Parmi les 116 substantifs terminés par **–er** , à part *celer, džber, houser, kačer, klášter, kráter, večer, podvečer, úder, uher, protiúder, průser, sever*, tous les autres sont d'origine étrangère. C'est seulement *uher* qui alterne.

Quatre mots (*hřeb, pohřeb, škleb, web*) sont terminés par **–eb** dont le seul qui alterne est *pohřeb*. Par **–ed'** sont terminés quatre mots (*čeled', sled', zed'*) dont le seul qui alterne est *zed'*.

Pour **–ež**, uniquement *lež* alterne à la différence de *drůbež, fermež, krádež, loupež, mládež, papež, svatokrádež, verbež, řež*. Parmi les substantifs dont le lemme est terminé par *–eč, –ed, –eh, –ech, –ej, –ep, –eř, –et', –ez*, nous n'en avons trouvé aucun qui puisse alterner. Cette affirmation devrait être vraie, du moins pour les noms communs. Cependant, ils peut exister des noms propres qui sont terminés par certains de ces groupes et qui subissent l'alternance (par exemple les toponymes *Ledeč, Skuteč, Bubeneč*).

**# > e** Concernant les substantifs du type *žena*, l'alternance # > *e* est effectuée régulièrement ou avec un petit nombre d'exceptions pour les lemmes terminés par *–𝒞la*, *–𝒞ma*, *–𝒞ra*, *–𝒞ka*, *–𝒞na*, *–𝒞ba*, *–𝒞va*.

Tous les substantifs terminés par **–𝒞la** (*bedla, berla, cihla, jehla, kudla, kukla, metla, modla, perla, piksla, šmudla, truhla*) alternent sans exception. La situation est la même pour la périphérie **–𝒞ma** (*drachma, farma, firma, forma, helma, karma, krčma, kulma, parma, platforma, reforma, rozedma, sedma, šelma, újma, uniforma, vědma*) et pour les substantifs terminées par **–𝒞ra** (54 substantifs, par exemple *hra, sestra, šifra*). Seulement quelques mots étrangers avec des géminées *mulla, bulla, gamma* pourraient être considérées, grâce à leur orthographe, comme conformes au patron donné et insensibles à l'alternance en même temps.

L'alternance # > *e* est effectuée quasiment régulièrement pour les substantifs avec la périphérie **–𝒞ka** (par exemple *židlička, dívka, kočka*) sauf pour les lexèmes *odmlka, pomlka, zámlka* et *banka* (4 exceptions sur 1871 substantifs susceptibles).

Les substantifs terminés par **–𝒞na** (254 substantifs, par exemple *továrna, slečna, panna, flétna*) alternent dans la plupart des cas sauf dans *bavlna, vlna, lajna, obrna, skvrna, srna* et dans les mots étrangers *moderna, postmoderna*.

Les substantifs terminés par **–𝒞ba** (92 substantifs, par exemple *tržba, hudba, léčba*) alternent dans la plupart des cas sauf dans les substantifs d'origine étrangère avec la périphérie *–mba* (*bomba, sexbomba, gamba, plomba, rumba, samba*).

Pour la périphérie **–𝒞va** , il existe 25 substantifs dont *antikva, barva, bezva, bitva, bulva, břitva, jizva, kotva, kurva, larva, lichva, odezva, pastva, pitva, pochva, průrva, vrstva, výzva, žatva* alternent et *brva, konzerva, mrva, rezerva, salva, želva* n'alternent pas.

Un ensemble de périphéries pour lesquels il existe une répartition approximativement égale de la présence et de l'absence d'une alternance, comme c'était le cas pour l'alternance *e* > #, n'a pas pu être établi pour l'alternance # > *e*.

La réalisation exceptionnelle de l'alternance # > *e* est effectuée dans les substantifs dont le lemme est terminé par *–𝒞ta*, *–𝒞pa*, *–𝒞sa*, *–𝒞da*.

La périphérie **–𝒞ta** figure dans 58 substantifs (par exemple *fakulta, pošta, úcta*) dont ce ne sont que le substantif *karta* et tous ceux terminés par *–chta* (*buchta, jachta, kruchta, plachta, plichta, rychta, šachta, šichta, kuchta*) qui alternent.

La périphérie **–𝒞pa** est présente dans 11 substantifs dont *výspa, zácpa, škarpa* alternent et *chrpa, krempa, lampa, pompa, pumpa, rampa, šerpa, žumpa* n'alternent pas. Pour la périphérie **–𝒞sa** (*apokalypsa, elipsa, kapsa, římsa*) ce n'est que le substantif *kapsa* qui alterne.

Pour la périphérie **–𝒞da** (62 substantifs, par exemple *pravda, bunda, uzda*) ce n'est que le substantif *mzda* qui alterne (*mzda* > *mezd*).

Parmi les substantifs dont le lemme est terminé par *–𝒞ca, –𝒞ča, –𝒞d'a, –𝒞fa, –𝒞ga, –𝒞ha, –𝒞cha, –𝒞ja, –𝒞ňa, –𝒞řa, –𝒞ša, –𝒞t'a, –𝒞za, –𝒞ža*, nous n'en avons trouvé aucun qui puissent alterner.

Concernant le type *město*, l'alternance est réalisée régulièrement dans les substantifs terminés par *–𝒞lo*, *–𝒞ko*, *–𝒞mo*, *–𝒞no*, *–𝒞ro* et *–𝒞vo*.

Pour la périphérie **–𝒞lo** (après l'élimination du prédicatif *mdlo*, il reste 124 substantifs, par exemple *clo, máslo, umyvadlo*), l'alternance est effectuée partout sauf dans deux mots d'origine étrangère avec un orthographe spécifique – *cello, cemballocello, cemballo*.

Tous les substantifs terminés par **–𝒞ko** (après l'élimination des prédicatifs *horko, vlhko, úzko, těžko, lehko, nízko, nizoučko*, ils reste 209 substantifs, par exemple *brčko, zrnko, očko*) alternent régulièrement, sauf *brko, vojsko* et les mots d'origine étrangère *disko, manko*.

Pour la périphérie **–𝒞mo** , la plupart des lexèmes sont des prédicatifs ou des adverbes (*dvojmo, trojmo, kolmo, koňmo, kradmo, letmo, nadarmo, obkročmo, potažmo, rozkročmo, zadarmo, zkusmo, zřejmo, šikmo*), ceux qui restent (*bělmo, jařmo, pižmo, pásmo, písmo*) alternent régulièrement.

Les lexèmes terminés par la suite $\boxed{\textbf{–}\mathscr{C}\textbf{no}}$ peuvent être divisés en deux groupes. Dans le premier groupe sont les prédicatifs (par exemple *plno, možno, smutno*) ou les adverbes agglutinés (par exemple *nalačno, nadrobno*) qui ne nous intéressent pas du point de vue des alternances. Dans le second groupe, nous pouvons distinguer entre les substantifs abstraits déadjectifs du type *jsoucno, absolutno, nekonečno*, qui peuvent tous alterner[25] et les autres *bahno, břevno, dno, hovno, hejno, lejno, mračno, okno, plátno, porno, prkno, ráhno, stehno, storno, sukno, vápno, techno, terno, vlákno, mikrovlákno, zrno* parmi lesquels ce sont les mots étrangers terminés par *–rno* (*porno, storno, terno*)[26], le substantif *zrno* et les substantifs *hejno, lejno* qui n'alternent pas.

Pour la périphérie $\boxed{\textbf{–}\mathscr{C}\textbf{ro}}$ . Après l'élimination de deux prédicatifs (*modro, mokro*), ils restent les substantifs *bistro, dobro, futro, jitro, jádro, kvádro, lejstro, makro, metro, mezipatro, moudro, nitro, ňadro, patro, pouzdro, stříbro, vedro, vnitro, vědro, žebro* qui alternent régulièrement.

Après l'élimination du prédicatif *mrtvo*, les 58 substantifs avec la périphérie $\boxed{\textbf{–}\mathscr{C}\textbf{vo}}$ sont tous des noms collectifs dérivés par le suffixe *–stvo / –ctvo* (par exemple *ptactvo, svalstvo, panstvo*) qui forment rarement le pluriel. Il existe néanmoins des substantifs sans la restriction sémantique sur le pluriel (par exemple *mužstvo* > *mužstev, včelstvo* > *včelstev*) qui montrent que la réalisation potentielle de l'alternance est régulière pour cette périphérie.

Aucun substantif du type *město* qui puisse subir l'alternance # > *e* n'a pas été trouvé pour les périphéries *–$\mathscr{C}$bo*, *–$\mathscr{C}$co*, *–$\mathscr{C}$do*, *–$\mathscr{C}$fo*, *–$\mathscr{C}$ho*, *–$\mathscr{C}$cho*, *–$\mathscr{C}$jo*, *–$\mathscr{C}$so*, *–$\mathscr{C}$to*. Les périphéries suivantes ne font partie d'aucun substantif dans la liste de 50 000 mots avec laquelle nous avons travaillé : *–$\mathscr{C}$čo*, *–$\mathscr{C}$d'o*, *–$\mathscr{C}$ňo*, *–$\mathscr{C}$po*, *–$\mathscr{C}$řo*, *–$\mathscr{C}$šo*, *–$\mathscr{C}$t'o*, *–$\mathscr{C}$zo*, *–$\mathscr{C}$žo*.

Le seul substantif appartenant au type *moře* dans lequel nous pouvons observer la réalisation de l'alternance # > *e* est le lexème *vejce* – la périphérie correspondante est $\boxed{\textbf{–}\mathscr{C}\textbf{ce}}$ .

### 3.5.5 Fichier *alt.xml*

Le fichier *alt.xml* (voir un extrait sur la fig. 3.7, p. 100) contient les informations concernant les alternances dans un format lisible par ordinateur. Ces informations sont nécessaires pour la procédure de génération automatique des formes casuelles utilisée pendant le diagnostic des erreurs. La structure de ce fichier (voir fig. 3.6) est déterminée par les propriétés formelles des alternances qui viennent d'être exposées.

[25] Bien qu'il s'agisse des substantifs abstraits avec des contraintes sémantiques sur la création du pluriel, l'alternance est possible et dans certains cas, sa réalisation est attestée – voir par exemple les 69 occurrences de la forme *jsoucen* dans le corpus SYN du ČNK.

[26] SYN donne une seule occurrence de la forme *storen*.

```
<!ELEMENT root (type+)>
<!ELEMENT type (soustype+)>
<!ATTLIST type phono ( vocalique | consonantique ) #REQUIRED>
<!ELEMENT soustype (alt+)>
<!ATTLIST soustype proces CDATA #REQUIRED >
<!ELEMENT alt (except*)>
<!ATTLIST alt  mnemo CDATA #REQUIRED
               dep   CDATA #REQUIRED
               arr   CDATA #REQUIRED >
<!ELEMENT except (u+)>
<!ATTLIST except neg  (oui | non) #REQUIRED
                 att_spec CDATA   #IMPLIED
                 val_spec CDATA   #IMPLIED >
<!ELEMENT u (#PCDATA)>
```

FIG. 3.6 – Fichier *alt.dtd*

L'élément racine `<root>` contient deux éléments `<type>`, définis par la valeur de l'attribut `@phono` : il s'agit soit des alternances *vocaliques*, soit des alternances *consonantiques*. Chaque élément `<type>` contient plusieurs éléments `<soustype>` qui sont spécifiés par la valeur de l'attribut `@proces`. Pour les alternances consonantiques, cet attribut prend les valeurs *mouillure, dépalatalisation, alternance graphique, palatalisation A, palatalisation B, alternance de groupe* ; pour les alternances vocaliques, l'attribut `@proces` prend les valeurs *qualitative, quantitative, -e- intercalaire*.

Chaque élément `<soustype>` contient plusieurs éléments `<alt>` : ce sont les différentes alternances qui se réalisent dans le cadre d'un certain sous-type, par exemple les alternance *á* > *a* et *í* > *ě* dans le cadre des alternances vocaliques quantitatives, les alternances *k* > *c* et *h* > *z* dans le cadre des alternances consonantiques du type palatalisation A, etc. Chaque élément `<alt>` est défini par l'attribut `@mnemo` qui contient l'inscription symbolique de l'alternance, par exemple `a > á`, `í > ě`, et les attributs `@dep`, `@arr` dont les valeurs sont la consonne ou la voyelle de « départ » de l'alternance (`a` pour `a > á`) et la consonne ou la voyelle d'« arrivée » de l'alternance (`á` pour `a > á`). Cette information est utilisée pour certaines étapes dans les traitements de génération de formes fléchies.

S'il est nécessaire de spécifier la réalisation d'une certaine alternance pour un ensemble d'exceptions positives ou négatives, les éléments `<alt>` peuvent contenir un ou plusieurs éléments `<except>`. Un élément `<except>` peut contenir à son tour plusieurs éléments `<u>` dont le contenu sont les exceptions. L'attribut `@neg` de l'élément `<alt>` spécifie la nature des exceptions :

```
  <SOUSTYPE proces="palatalisation B">
      <ALT mnemo="k > č" dep="k" arr="č"/>
      <ALT mnemo="c > č" dep="c" arr="č"/>
      <ALT mnemo="h > ž" dep="h" arr="ž"/>
   </SOUSTYPE >
   <SOUSTYPE proces="palatalisation de groupe">
      <ALT mnemo="sk > št" dep="sk" arr="št"/>
      <ALT mnemo="ck > čt" dep="ck" arr="čt"/>
   </SOUSTYPE >
</TYPE >
<TYPE phono="vocalique">
   <SOUSTYPE proces="qualitative">
      <ALT mnemo="a > e" dep="a" arr="e">
         <EXCEPT neg="oui">
            <U>Jan</U>
         </EXCEPT>
      </ALT>
      <ALT mnemo="e > a" dep="e" arr="a">
         <EXCEPT neg="oui">
            <U>vejce</U>
         </EXCEPT>
      </ALT>
      <ALT mnemo="í > á" dep="í" arr="á">
         <EXCEPT neg="oui">
            <U>přítel</U>
            <U>nepřítel</U>
         </EXCEPT>
      </ALT>
   </SOUSTYPE >
   <SOUSTYPE proces="quantitative">
      <ALT mnemo="á > a" dep="á" arr="a">
         <EXCEPT neg="oui" att_spec="NbSyllabes" val_spec="2">
            <U>blána</U>
            <U>brána</U>
            <U>bába</U>
```

FIG. 3.7 – Extrait du fichier *alt.xml*

- la valeur *oui* signifie que les lexèmes listés sont des exceptions négatives, c'est à dire que l'alternance spécifiée dans l'élément `<alt>` est effectuée uniquement dans les lexèmes listés.
- la valeur *non* signifie que les lexèmes listés sont des exceptions positives, c'est à dire que l'alternance spécifiée dans l'élément `<alt>` est effectuée partout sauf dans les lexèmes listés.

Un élément `<except>` peut contenir les attributs `@att_spec` et `@val_spec` dont la fonction est de poser des restrictions sur la forme des exceptions. Il est possible de spécifier la nature de cette restriction comme un attribut dans `@att_spec` et fournissant la valeur de cet attribut dans `@val_spec`. Pour donner un exemple, comme nous l'avons vu plus haut, l'alternance de quantité *í* > *ě* peut être effectuée soit dans des substantifs dont le lemme est bisyllabique (par exemple *dílo* > *děl*), soit dans des substantifs dont le lemme est monosyllabique (par exemple *sníh* > *sněhu*). Pour la distribution de cette alternance, le trait pertinent de la forme à alterner est donc le nombre de syllabes. Ainsi, l'attribut `@att_spec` prend la valeurs *NbSyllabes* et la valeur de l'attribut `@val_spec` peut être respectivement 1 et 2. La partie correspondante du fichier *alt.xml* est affichée ci-dessous :

```
<ALT mnemo="í > ě" dep="í" arr="ě">
   <EXCEPT neg="oui" att_spec="NbSyllabes" val_spec="2">
      <U>bída</U>
      <U>díra</U>
      <U>víra</U>
      <U>míra</U>
      <U>dílo</U>
      <U>peníze</U>
      <U>dítě</U>
      <U>svíce</U>
   </EXCEPT>
   <EXCEPT neg="oui" att_spec="NbSyllabes" val_spec="1">
      <U>sníh</U>
      <U>vítr</U>
   </EXCEPT>
</ALT>
```

Le nombre d'éléments `<except>` est déterminé par le nombre de valeurs différentes que peut prendre une restriction posée sur la forme du lexème alterné. Dans l'exemple présenté ci-dessus, il y deux éléments `<except>` pour les alternances vocaliques quantitatives – 1 syllabe ou 2 syllabes du lemme. Pour l'alternance # > *e* dont la distribution est spécifiée

par la forme de la périphérie droite du lemme (`@att_spec` prend la valeur *PeriphLemme*), le nombre d'éléments `<except>` est 17, car l'attribut `@val_spec` peut prendre les valeurs *la, ma, ra, ka, na, ba, va, ta, pa, sa, da, lo, ko, mo, no, ro, vo, ce*. Pour l'alternance *e > #* dont la distribution est spécifiée également par la forme de la périphérie droite du lemme (la valeur de l'attribut `@att_spec` est *PeriphLemme*), le nombre d'éléments `<except>` est 15, car l'attribut `@val_spec` peut prendre les valeurs *k, c, ň, m, v, n, š, l, t, s, r, b, d', ž, ž*.

L'intégralité des informations contenues dans le fichier *alt.xml*, peut être consultée dans l'annexe D (p. 329). Cette annexe a été générée automatiquement à partir du fichier *alt.xml*.

## 3.6 Procédure *AlterneRadical*

La procédure *AlterneRadical* est cruciale pour la génération des formes fléchies dans les traitements automatiques sur CETLEF. Elle assure la réalisation d'une éventuelle alternance sur la concaténation du radical d'un lexème et d'une désinence, appartenant à son paradigme de désinences casuelles. La réalisation d'une alternance est décidée en fonction des propriétés formelles du radical et des propriétés formelles de la désinence, spécifiée par les valeurs des catégories grammaticales, voir la section 3.5 *Alternances*, pp. 83–115. Cette procédure est utilisée pendant l'assistance à l'annotation (la procédure *Annote*) et dans le diagnostic des erreurs (la procédure *Diagnostic*).

En bref, cette procédure décide si la concaténation d'un radical avec une désinence doit rester intacte (par exemple *hrad* + *u* > *hradu*), ou s'il doit y avoir une alternance (par exemple *kočk* + *e* > *kočce*).

### 3.6.1 Description

Les données à l'ENTRÉE de la procédure[27] sont : le **lemme** du mot à décliner, le **radical** du mot à décliner[28], une **désinence casuelle**, les valeurs du **cas**, du **genre** et du **nombre**.

Les données à la SORTIE de la fonction sont le **radical alterné** et l'**identifiant de l'alternance** effectuée.

Par exemple, pour la vérification de l'alternance du radical dans le locatif singulier du mot *kočka*, l'entrée de la fonction sera *kočka, kočk, e, loc, f, sg*, la sortie va contenir le

[27] Techniquement, il s'agit d'une fonction implémentée en PHP contenant une série de tests conditionnels. Les données à l'entrée de la procédure sont les arguments de cette fonction.

[28] A priori, le radical pourrait être directement déterminé au sein de la fonction, sans figurer comme son argument. La solution adoptée a été choisie pour des raisons d'économie de temps dans l'exécution de la fonction. Pour un lemme, le radical est déterminé une seule fois, avant le lancement des procédures qui appellent cette fonction à plusieurs reprises.

radical alterné et l'identifiant de l'alternance *kočc, k > c*. Si la concaténation du radical et de la terminaison ne provoque pas d'alternance (par exemple pour l'entrée *kočka, kočk, y, gen, f, sg*, la forme du radical à la sortie est égale à celle de l'entrée et l'identifiant de l'alternance a la valeur *sans* : *kočk, sans*.

L'alternance est réalisée dans tous les mots qui remplissent les conditions nécessaires pour être alternées. Par exemple, pour l'alternance du radicale dans le datif singulier d'un mot fictif, comme *tropka*, qui a les propriétés formelles d'un substantif féminin, avec l'entrée de la fonction *tropka, tropk, e, dat, sg*, la sortie sera *tropc, k > c* ; pour la génération de la forme du génitif pluriel, avec l'entrée *tropka, tropk, #, gen, pl*, la sortie sera *tropek, # > e*.

La fonction est composée d'une série de tests qui vérifient la réalisation d'un certain type d'alternance. Les conditions nécessaires pour la réalisation d'une alternance sont inscrites en dur dans l'algorithme du test en question ; des informations supplémentaires sur les exceptions sont lues à partir du fichier *alt.xml*. Les tests pour les différents types d'alternances sont branchés en série, le succès d'un test (manifesté par la réalisation d'une alternance) n'interrompt pas cette série car il est nécessaire de vérifier si une autre alternance ne doit pas être réalisée. Par exemple, pour l'entrée *vejce, vejc, #, gen, pl*, la sortie sera *vajec, e > a | # > e* car le test vérifiant la réalisation de l'alternance vocalique qualitative *e > a* sera passé avec succès ainsi que le test vérifiant la réalisation de l'alternance *# > e*. Les doubles alternances sont en effet assez fréquentes à cause des alternances consonantiques graphiques, par exemple *píseň > písně*, où l'alternance *e > #* est accompagnée par l'alternance graphique *ň > n*.

Avant d'exposer la suite des tests dans la fonction *AlterneRadical*, nous allons décrire, en guise d'exemple, le déroulement d'un seul test – il s'agit du test pour la réalisation de l'alternance *e > #* avec la suppression de *-e-* intercalaire. Nous allons tester la réalisation de l'alternance dans le génitif singulier du mot *celek* qui est formé par la désinence *–u* (*celku*). L'entrée de la fonction sera la suivante *celek, celek, u, gen, i, sg*. Le test procédera par les étapes énumérées ci-dessous :

(1) Si le deuxième graphème à partir de la fin du lemme est la voyelle *e*, le test se poursuit, sinon il s'arrête et le test sur l'alternance suivante est commencé. Pour le mot *celek*, le test continue car il remplit cette condition.

(2) Pour les mots remplissant la condition précédente, l'alternance *# > e* peut être réalisée partout, sauf dans le nominatif singulier pour tous les genres, et dans l'accusatif singulier des féminins et des masculins inanimés[29]. Les valeurs de catégories grammaticales dans l'entrée ne correspondent pas à cette spécification, le test peut donc continuer.

[29]Car la désinence est ici la même qu'au nominatif (la désinence zéro –#).

(3) Dans le fichier *alt.xml*, la réalisation de l'alternance # > *e* est spécifiée par le graphème périphérique du lemme : la valeur de l'attribut `att_spec` est *PeriphLemme*, la valeur de l'attribut `val_spec` correspond à la consonne qui suit le *-e-* intercalaire dans le lemme, par exemple *k, c, ň*, etc. Le dernier graphème du mot *celek* est la consonne *k* qui appartient à la liste des valeurs de l'attribut `val_spec` pour l'alternance # > *e*, le test peut donc continuer.

(4) La dernière étape consiste à vérifier si l'alternance # > *e* peut être réalisée dans le mot analysé. Le fichier *alt.xml* contient une liste d'exceptions positives qui contient tous les mots qui passeraient avec succès les étapes précédentes, mais qui ne permettent pas la réalisation de l'alternance (*bek, biftek, dotek, flek, jek, klek, oblek, poklek, rek, skřek, vlek, úsek, výsek, průsek, zásek, česnek, šek, šnek, špek*). Le mot *celek* n'appartient pas à cette liste, son radical est donc alterné d'après l'alternance *e* > # : *celk+u*. Le test est fini et réussi.

La sortie de la fonction sera donc *celk, e* > # et la concaténation du radical résultant avec la désinence *–u* donnera la forme lexicale *celku* qui est la forme correcte du génitif singulier.

Si le mot dans l'exemple ci-dessus était *česnek* (entrée de la fonction : *česnek, česnek, u, gen, i, sg*), le test pour la réalisation de l'alternance # > *e* serait terminé par un échec dans l'étape (4) car le mot *česnek* appartient à la liste des exceptions qui ne permettent pas l'alternance. Le même test pour le mot *soupeř* (entrée de la fonction : *soupeř, soupeř, e, gen, m, sg*) se serait arrêté à l'étape (3) car la consonne *ř* n'est pas spécifiée dans l'attribut `val_spec` comme une des périphéries possibles pour la réalisation de l'alternance. De la même façon, l'étape (2) ne pourrait pas être passée avec succès pour les mots avec les valeurs correspondantes de la catégorie du cas et du genre et l'étape (1) ne pourrait pas être passée par un mot comme *člověk*.

Pour l'illustration, le fragment du code PHP, avec le test pour l'alternance *e* > #, est le suivant :

```
if ($lemme[count($lemme)-2] == 'e')
{
   if (!($cas =='nom' && $num =='sg') &&
       !($cas == 'acc' && $num =='sg' && ($gen == 'i' || $gen == 'f')))
   {
      $mnemo_alt = 'e > #';

      if(Appartient(PeriphDroite($lemme,1), TrouveSpec($mnemo_alt)))
      {
         if (DecideAlt($mnemo_alt, PeriphDroite($lemme,1), $lemme))
         {
            $radical = Enleve_e_mobile($radical);
            $alt_sortie .= '|'. $mnemo_alt;
         }
      }

      $mnemo_alt = array();
```

```
    }
}
```

### 3.6.2 Tests algorithmiques des alternances

Dans cette partie, nous allons décrire les conditions posées dans les différents tests dans la fonction *AlterneRadical*. L'ordre des alternances testées, a priori arbitraire, est le suivant : (1) vocaliques quantitatives, (2) vocaliques qualitatives, (3) suppression de *-e-* intercalaire, (4) insertion de *-e-* intercalaire, (5) mouillure marquée graphiquement, (6) palatalisation A, B et groupes consonantiques, (7) alternance graphique, (8) dépalatalisation.

Si un test est passé avec succès, le radical modifié d'après l'alternance correspondante et l'identifiant de l'alternance sont enregistrés pour être présentés à la sortie. Le radical alterné continue à passer les tests pour vérifier la réalisation des doubles alternances. Si une seconde alternance a lieu, le radical est modifié d'après cette alternance et son identifiant est ajouté à l'identifiant de l'alternance précédente.

Si aucun test n'a pu être passé avec succès, le radical alterné à la sortie sera égal au radical à l'entrée de la fonction ; l'identifiant de l'alternance sera *sans*.

#### 3.6.2.1 Gestion des exceptions

Il est possible que les spécifications génériques dans les tests ne suffisent pas pour traiter les réalisations des alternances qui sont rares, liées uniquement à quelques unités lexicales, qui n'ont pas été intégrées dans le fichier *alt.xml* ou qui ne nécessitent pas sa consultation (les alternances vocaliques qualitatives et la dépalatalisation). Dans ces cas, la réalisation de l'alternance est gérée par un test spécifique qui pose la première condition directement sur le lemme et non pas sur des propriétés morphologiques comme dans les autres tests. Il s'agit des substantif suivants : *dítě, peníze* pour l'alternance vocalique quantitative *í* > *ě* ; *Jan, vejce, přítel* pour les alternances vocaliques qualitatives ; *čest, křest, lest* pour l'alternance vocalique avec la suppression de *-e-* intercalaire ; *mzda* pour l'alternance vocalique avec l'insertion de *-e-* intercalaire ; *oko* pour la réalisation de la palatalisation B *k* > *č* ; *zeď, paměť, oběť* pour la dépalatalisation.

#### 3.6.2.2 Test 1 : Vocaliques quantitatives

Ce test vérifie la réalisation des alternances vocaliques quantitatives dans le radical des mots monosyllabiques ou bisyllabiques : *á* > *a* (par exemple *skála* > *skal ; náš* > *našeho*), *é* > *e* (par exemple *jméno* > *jmen ; chléb* > *chleba*), *í* > *i* (par exemple *lípa* > *lip ; líh* > *lihu*),

*í* > *ě* (par exemple *díra* > *děr ; sníh* > *sněhu*), *ou* > *u* (par exemple *smlouva* > *smluv*), *ů* > *o* (par exemple *fůra* > *for ; dům* > *domu*).

1. **si** la dernière voyelle du radical est longue
   **alors** passe à l'étape (2)
   **sinon** passe au test suivant **fin**
2. **si** le nombre de syllabes dans le lemme est égal à 1 OU le lemme est terminé par *ův* (adjectifs possessifs)
   **alors** passe à l'étape (3)
   **sinon** passe à l'étape (6)
3. **si** les valeurs des catégories grammaticales sont autres que le nom. sg. et le acc. sg. i. et f.
   **alors** passe à l'étape (4)
   **sinon** passe au test suivant **fin**
4. **si** la dernière voyelle du radical appartient à la liste des valeurs de l'attribut `dep` dans le fichier *alt.xml* (les voyelles longues au départ d'une alternance)
   **alors** passe à l'étape (5)
   **sinon** passe au test suivant **fin**
5. **si** le lemme peut être alterné d'après la liste des exceptions dans le fichier *alt.xml* OU le lemme est terminé par *ův*
   **alors** ALTERNE le radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
   **sinon** passe au test suivant **fin**
6. **si** le nombre de syllabes dans le lemme est égal à 2
   **alors** passe à l'étape (7)
   **sinon** passe au test suivant **fin**
7. **si** le lemme est terminé par les voyelles *a, o, e*
   **alors** passe à l'étape (8)
   **sinon** passe au test suivant **fin**
8. **si** les valeurs de la catégorie du cas et du nombre sont *gen* et *pl*
   **alors** passe à l'étape (9)
   **sinon** passe au test suivant **fin**
9. **si** la dernière voyelle du radical appartient à la liste des valeurs de l'attribut `dep` dans le fichier *alt.xml* (les voyelles longues au départ d'une alternance)
   **alors** passe à l'étape (10)
   **sinon** passe au test suivant **fin**

10. **si** le lemme peut alterner d'après la liste des exceptions pour la dernière voyelle du radical dans le fichier *alt.xml*
    **alors** ALTERNE le radical d'après une des alternances *á* > *a*, *é* > *e*, *í* > *i*, *í* > *ě*, *ou* > *u*, *ů* > *o* en fonction de la dernière voyelle du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
    **sinon** passe au test suivant **fin**

**Remarque** Dans l'étape (8), la réalisation de l'alternance des substantifs avec deux syllabes est limitée seulement au génitif pluriel, où elle est régulière et obligatoire. Dans les autres formes casuelles possibles (dat., loc. et inst. pl. et inst. sg.), l'alternance peut être réalisée uniquement pour quelques substantifs et elle est optionnelle. C'est pourquoi elle n'est pas intégrée dans la procédure de génération. La réalisation de l'alternance *í* > *ě* dans le seul substantif trisyllabique *peníze* qui peut alterner (*peníze* > *peněz, penězům, penězích, penězmi/penězi*), est gérée par une règle spécifique, liée à cette unité lexicale. La situation est la même pour les formes plurielles alternées du substantif *dítě* > *děti, dětí, dětem, dětech*.

#### 3.6.2.3 Test 2 : Vocaliques qualitatives

Ce test vérifie la réalisation des alternances vocaliques qualitatives, liées uniquement à des lexèmes spécifiques : *a* > *e* (*Jan* > *Jene*), *e* > *a* (*vejce* > *vajec*), *í* > *á* (*přítel* > *přátelé*). La condition est posée sur le lemme (*Jan, vejce, přítel*) et les valeurs des catégories grammaticales pour la réalisation de l'alternance. Il y a donc trois test successifs, liés à ces unité lexicales spécifiques, qui vérifient les données fournies et qui, dans le cas du succès, permettent la réalisation de l'alternance comme pour les substantifs avec les alternances irrégulières.

#### 3.6.2.4 Test 3 : Suppression de *-e-* intercalaire

Ce test vérifie la réalisation des alternances vocaliques avec la suppression de *-e-* intercalaire dans les substantifs terminés par la suite *e* + *consonne*, par exemple *vrcholek* > *vrcholku, konec* > *konce, píseň* > *písně*, etc. (voir aussi l'exemple ci-dessus, p. 103).

1. **si** le deuxième graphème à partir de la fin du lemme est un *e*
   **alors** passe à l'étape (2)
   **sinon** passe au test suivant **fin**
2. **si** les valeurs des catégories grammaticales sont autres que le nom. sg. et le acc. sg. i. et f.

**alors** passe à l'étape (3)
**sinon** passe au test suivant **fin**

3. **si** la dernière consonne (précédée par *e*) du lemme appartient à la liste des valeurs de l'attribut `val_spec` dans le fichier *alt.xml*
   **alors** passe à l'étape (4)
   **sinon** passe au test suivant **fin**
4. **si** le lemme peut alterner d'après la liste des exceptions pour la dernière consonne du lemme dans le fichier *alt.xml*
   **alors** ALTERNE le radical d'après l'alternance *e* > # ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
   **sinon** passe au test suivant **fin**

**Remarque** La réalisation de l'alternance *e* > # dans les mots *čest, křest, lest*, avec le *-e-* situé à la troisième position à partir de la fin du lemme, est gérée par des règles spécifiques, liées à ces unités lexicales.

#### 3.6.2.5 Test 4 : Insertion de *-e-* intercalaire

Ce test vérifie la réalisation des alternances vocaliques avec l'insertion de *-e-* intercalaire dans le génitif pluriel exprimé par la désinence zéro –#), par exemple *cihla* > *cihel, okno* > *oken, vejce* > *vajec*.

1. **si** le troisième graphème à partir de la fin du lemme est une consonne ET le lemme est terminé par les voyelles *a, o, e*
   **alors** passe à l'étape (2)
   **sinon** passe au test suivant **fin**
2. **si** les valeurs des catégories grammaticales sont le génitif pluriel féminin ou neutre ET la désinence est #
   **alors** passe à l'étape (3)
   **sinon** passe au test suivant **fin**
3. **si** la périphérie droite du lemme appartient à la liste des valeurs de l'attribut `val_spec` dans le fichier *alt.xml*
   **alors** passe à l'étape (4)
   **sinon** passe au test suivant **fin**
4. **si** le lemme peut alterner d'après la liste des exceptions pour la périphérie droite dans le fichier *alt.xml*

**alors** ALTERNE le radical d'après l'alternance # > *e* ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
**sinon** passe au test suivant **fin**

**Remarque** La réalisation de l'alternance # > *e* dans le mot *mzda*, avec le -*e*- inséré devant les deux consonnes finales (*mzda* > *mezd*) est gérée par une règle spécifique, liée à cette unité lexicale.

#### 3.6.2.6 Test 5 : Mouillure marquée graphiquement

Ce test vérifie la réalisation de l'alternance consonantique de la mouillure marquée graphiquement *ř* > *r*, par exemple *doktor* > *doktoři, bratr* > *bratře, hora* > *hoře, starý* > *staří*, etc.

1. **si** la désinence appartient à l'ensemble *i, e, í*
   **alors** passe à l'étape (2)
   **sinon** passe au test suivant **fin**
2. **si** le radical est terminé par *r*
   **alors** passe à l'étape (3)
   **sinon** passe au test suivant **fin**
3. **si** le graphème qui précède le *r* final du radical n'est pas une voyelle ET la valeur de la catégorie du cas n'est pas vocatif ET la désinence n'est pas *e*
   **alors** ALTERNE le radical d'après l'alternance *r* > *ř* ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
   **sinon** passe au test suivant **fin**

**Remarque** La condition dans l'étape (2) permet d'éliminer l'alternance dans le vocatif singulier des masculins terminés par la suite *voyelle* + *r* (*doktor* > *doktore, vzor* > *vzore*) qui n'alternent pas dans ce cas, contrairement à ceux qui sont terminés par la suite *consonne* + *r* (*bratr* > *bratře, machr* > *machře*, etc.).

#### 3.6.2.7 Test 6 : Palatalisation A, B et groupes consonantiques

Ce test vérifie la réalisation de trois types de palatalisation. La palatalisation A : *k* > *c*, *g* > *z*, *h* > *z*, *ch* > *š*) ; la palatalisation B : *k* > *č*, *c* > *č*, *h* > *ž* ; la palatalisation de groupes consonantiques : *sk* > *št*, *ck* > *čt*.

1. **si** la désinence est une des désinences molles *i, imi, ima, í, ím, ích, e, ě, ěm, ěmi, ěma*
   **alors** passe à l'étape (2)
   **sinon** passe au test suivant **fin**
2. **si** le radical est terminé par une des consonnes suivantes : *k, c, g, h, ch*
   **alors** passe à l'étape (3)
   **sinon** passe au test suivant **fin**
3. **si** le lemme est terminé par *cký* ou *ský*
   **alors** passe à l'étape (4)
   **sinon** passe à l'étape (5)
4. **si** la désinence est *í*
   **alors** ALTERNE le radical d'après une des alternances *sk* > *št, ck* > *ct* en fonction des deux consonnes finales du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
5. **si** le cas et le nombre ne sont pas vocatif singulier
   **alors** ALTERNE le radical d'après une des alternances *k* > *c, g* > *z, h* > *z, ch* > *š* en fonction de la consonne finale du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
   **sinon** ALTERNE le radical d'après une des alternances *k* > *č, h* > *ž, c* > *č* en fonction de la consonne finale du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**

**Remarque** La réalisation de l'alternance *k* > *č* dans les formes du pluriel du substantif *oko* (*oko* > *oči, očím, očích, očima*) est gérée par une règle spécifique, liée à cette unité lexicale.

#### 3.6.2.8 Test 7 : Alternances graphiques

Ce test vérifie la réalisation des alternances graphiques : *t'* > *t* (par exemple *plet'* > *pleti, chot'* > *chotěm*), *d'* > *d* (par exemple *lod'* > *lodi, výpověd'* > *výpovědi*), *ň* > *n* (par exemple *jabloň* > *jabloně, úroveň* > *úrovně*), *t* > *t'* (par exemple *kotě* > *kot'at*), *d* > *d'* (par exemple *mládě* > *mlád'at*), *n* > *ň* (par exemple *štěně* > *štěňat*).

1. **si** le radical est terminé par une des consonnes suivantes : *d', t', ň*
   **alors** passe à l'étape (2)
   **sinon** passe à l'étape (3)
2. **si** la désinence est une des désinences suivantes *i, imi, ima, í, ím, ích, ě, ěm, ěmi, ěma*
   **alors** ALTERNE le radical d'après une des alternances *d'* > *d, t'* > *t, ň* > *n* en fonction

de la consonne finale du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**

3. **si** le lemme est terminé par une des suites suivantes : *dě, tě, ně*
   **alors** passe à l'étape (4)
   **sinon** passe au test suivant **fin**
4. **si** la désinence est une des désinences suivantes *#, ata, at, aty, atům, atum, atech, atama*
   **alors** ALTERNE le radical d'après une des alternances *d > d', t > t', n > ň* en fonction de la consonne finale du radical ; inscrit l'identifiant de l'alternance ; passe au test suivant **fin**
   **sinon** passe au test suivant **fin**

#### 3.6.2.9 Test 8 : Dépalatalisation

La réalisation de la dépalatalisation (*d' > d, t' > t*) pour les substantifs *zed'*, *paměť*, *oběť* dans le datif, locatif et instrumental pluriel (*zed' > zdem, zdech, zdmi ; paměť > pamětem, pamětech, pamětmi ; oběť > obětem, obětech, obětmi*) est gérée par des règles spécifiques, liées à ces unité lexicales.

## 3.7 Procédure *Annote*

La procédure *Annote* est enclenchée pendant le création d'une tâche dans un exercice sur la plateforme auteur (voir la section 6.3.1.2, p. 226). Son rôle est de proposer une annotation morphologique de la forme requise ce qui permet de faciliter le processus de la création des tâches et d'augmenter la cohérence de l'annotation, car les erreurs liées au traitement manuel sont ainsi minimisées. L'annotation à la sortie de cette procédure doit être vérifiée et désambiguïsée manuellement par l'auteur avant qu'elle puisse être inscrite dans la base de données.

### 3.7.1 Description

Pendant la création d'une tâche, l'auteur saisit les informations nécessaires exigées par sa structure et par le schéma de l'annotation. Prenons par exemple la tâche : *Vedle našeho (dům, sg.) ... je hezká zahrada.* Les informations concernant sa structure[30] sont les suivantes :

- le contexte gauche, valeur de l'attribut `cntx_gauche` : *Vedle našeho*

[30] Voir la section 6.2.5, p. 209.

- la forme requise, valeur de l'attribut `requis` : *domu*
- le lemme de la forme requise, valeur de l'attribut `lemme` : *dům*
- la traduction française de la forme requise, valeur de l'attribut `trad_fr` : *maison*
- la consigne grammaticale, valeur de l'attribut `consigne` : *sg.*
- le contexte droit, valeur de l'attribut `cntx_droit` : *je hezká zahrada*

L'annotation morphologique de la forme requise[31], saisie à la main ou générée automatiquement par la procédure *Annote*, devrait être inscrite dans la base de données avec les informations suivantes :

- catégorie lexicale, valeur de l'attribut `tagLex` : *subst*
- type morphologique, valeur de l'attribut `tagMorph` : *N*
- type paradigmatique, valeur de l'attribut `pdgm` : *hd*
- cas, valeur de l'attribut `cas` : *gen*
- nombre, valeur de l'attribut `num` : *sg*
- genre, valeur de l'attribut `gen` : *i*
- alternance, valeur de l'attribut `alt` : *ů* > *o*

La procédure *Annote* lit les données structurelles, saisies à la main par l'auteur, et calcule toutes les valeurs possibles des attributs de l'annotation. Pour faire ceci, elle utilise la définition des types paradigmatiques dans le fichier *pdgm.xml* et la procédure *AlterneRadical*, qui se sert de la spécification des alternances dans le fichier *alt.xml*.

Les données à l'ENTRÉE de la procédure sont : le **lemme** du mot à décliner, la **forme requise**, la **consigne grammaticale**.

Les données à la SORTIE de la procédure sont toutes les valeurs possibles des attributs `tagLex`, `tagMorph`, `pdgm`, `cas`, `num`, `gen`, `alt`.

Le principe de cette procédure est de trouver une correspondance entre la forme requise et des formes qui sont le résultat de la concaténation du radical (extrait à partir du lemme, et modifié éventuellement par la procédure *AlterneRadical*) et d'une désinence, lue dans les listes spécifiant les types paradigmatiques dans le fichier *pdgm.xml*. Si une correspondance est trouvée, les informations grammaticales liées à la désinence (les valeurs des attributs `tagLex`, `tagMorph`, `pdgm`, `cas`, `num`, `gen`, `alt`) sont considérées comme une annotation possible de la forme requise.

Le calcul de l'annotation peut se dérouler de deux manières différentes. Si le lemme figure parmi les exemples spécifiant les différents sous-types paradigmatiques dans le fichier *pdgm.xml*, le lemme est traité comme *connu* ; dans le cas contraire, il est traité comme

[31] Voir la section 6.2.6, p. 211.

*inconnu*. Les exemples ont été choisis parmi les mots les plus fréquents appartenant au sous-type donné, avec la préférence de ceux qui figurent dans le lexique A1A2, voir la section 3.5, p. 69. Ainsi, les sous-types fermés, représentant les particularités dans la déclinaison, peuvent être annotés confortablement, car tous les mot déclinés d'après ces types sont énumérés dans les exemples.

### 3.7.2 Annotation des lemmes connus

Si le lemme a été trouvé dans les exemples spécifiant les sous-types dans le fichier *pdgm.xml*, les traitements de l'annotation se déroulent de la façon suivante :

(1) Les valeurs des attributs `tagMorph` et `pdgm` sont inscrites comme les valeurs des attributs correspondants de l'annotation et la valeur de l'attribut `tagLex` est calculée à partir de la valeur `tagMorph` d'après la relation non marquée entre ces deux attributs (voir la section 3.2 p. 50) : *N = subst, A = adj, M = adj, P = pron, C = num, X = subst*. La valeur de l'attribut `consigne` (restriction sur le nombre grammatical de la forme requise) est inscrite comme la valeur de l'attribut `num` de l'annotation.

(2) Le radical est extrait à partir du lemme.

(3) Les désinences (le contenu de l'élément `term`) ; appartenant au sous-type spécifié par `pdgm` et dont la valeur de l'attribut `num` correspond à l'attribut `consigne` ; sont lues une par une, et sont soumises comme argument à la fonction *AlterneRadical*, avec le lemme, le radical et les valeurs des attributs `cas`, `num` et `gen` correspondants à la désinence en cours.

(4) À l'intérieur de la boucle de lecture des désinences, le test suivant est effectué :
**si** il y a une correspondance entre la forme requise et la forme créée par la concaténation du radical alterné (sortie de la fonction *AlterneRadical*) et de la désinence en cours
**alors** les valeurs des attributs *cas*, *gen* et *alt* (sortie de la fonction *AlterneRadical*) sont inscrits comme les valeurs des attributs correspondants de l'annotation.

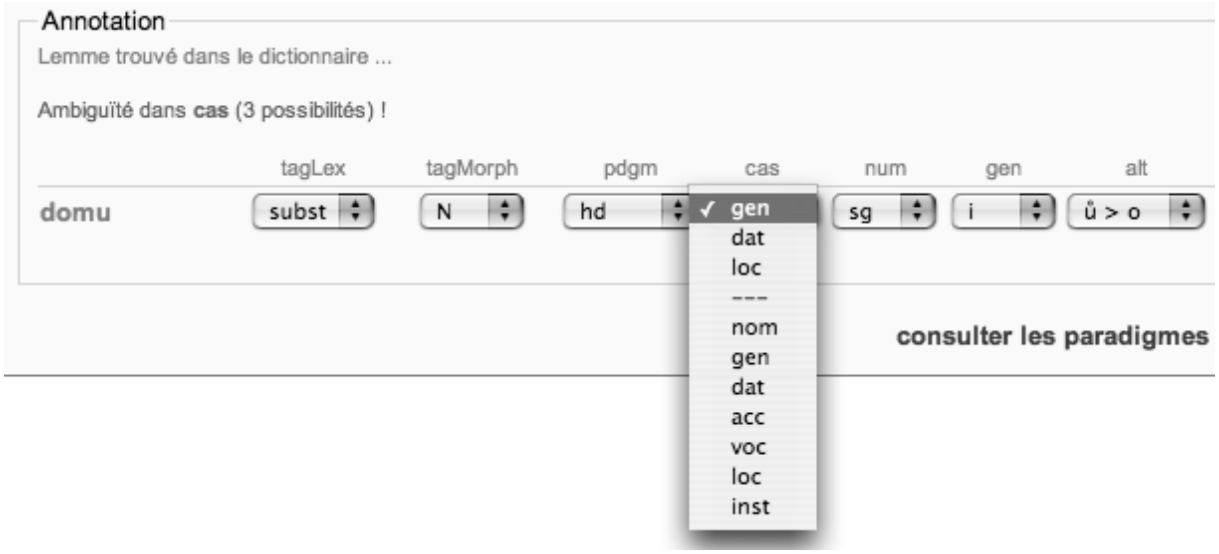
Ainsi, pour l'exemple de la tâche ci-dessus avec le lemme *dům*, qui appartient à la liste des exemples du sous-type *hrad* (la valeur de l'attribut `pdgm` est égale à *hd*) et qui est un type morphologique nominal (la valeur de l'attribut `tagMorph` est égale à *N*), la lecture des désinences et le test dans (4) sont effectués de la façon suivante :

| n° lecture | cas | num | gen | term | radical alterné | alt | requis | succès |
|---|---|---|---|---|---|---|---|---|
| 1 | nom | sg | i | # | dům | sans | domu | 0 |
| 2 | gen | sg | i | u | dom | ů > o | domu | 1 |
| 3 | dat | sg | i | u | dom | ů > o | domu | 1 |
| 4 | acc | sg | i | # | dům | sans | domu | 0 |
| 5 | voc | sg | i | e | dom | ů > o | domu | 0 |
| 6 | loc | sg | i | u | dom | ů > o | domu | 1 |
| 7 | loc | sg | i | ě | dom | ů > o | domu | 0 |
| 8 | inst | sg | i | em | dom | ů > o | domu | 0 |

Comme nous pouvons le voir, le test sur la correspondance de la forme requise avec la forme générée à l'aide de la fonction *AlterneRadical* a réussi à l'étape 2, 3 et 6. L'annotation proposée à l'auteur est donc :

| tagLex | tagMorph | pdgm | cas | num | gen | alt |
|---|---|---|---|---|---|---|
| subst | N | hd | gen | sg | i | ů > o |
| | | | dat | | | |
| | | | loc | | | |

Sur la plateforme auteur, la sortie de l'annotation est présentée de la manière suivante[32] :




Les ambiguïtés, dont le nombre dépend de l'homonymie des désinences au sein d'un certain paradigme, sont expressément soulignées et l'auteur peut modifier la sortie de l'annotation manuellement, en choisissant la valeur appropriée. Il s'agit le plus souvent des ambiguïtés dans la valeur du cas, mais également du genre pour les adjectifs (par exemple

[32] Voir également l'annexe *A Plateforme auteur*, fig. A.5, p. 246.

*mladých* peut être le locatif pluriel de tous les genres) et les pronoms (par exemple *těm* peut être le datif pluriel de tous les genres).

### 3.7.3 Annotation des lemmes inconnus

Si le lemme n'a pas été trouvé dans les exemples spécifiant les sous-types dans le fichier *pdgm.xml*, la situation est plus compliquée car l'homonymie des désinences ainsi que le caractère arbitraire de l'attribution du genre à des unités lexicales de certains types paradigmatiques accroît le nombre d'ambiguïtés. Par exemple, pour le lemme *vkus*, la forme requise *vkuse* et la consigne *sg*, l'annotation peut théoriquement contenir trois valeurs de la catégorie du genre (masculin animé, inanimé, féminin) ; pour chaque genre, il peut y avoir plusieurs types paradigmatiques consonantiques qui peuvent avoir une forme lexicale au singulier terminée par la désinence *–e* ; cette désinence peut être à son tour homonyme au sein du paradigme (type *pán* : voc. sg. ; type *hrad* : voc. et loc. sg. ; type *muž* : gen., acc. sg. ; type *muž* : gen. sg. ; type *píseň* : gen. sg.).

L'annotation des lemmes inconnus est assez imparfaite et l'intervention de l'auteur dans le choix des bonnes étiquettes doit être plus importante. Les traitements se déroulent de cette façon :

(1) Le radical et la désinence sont extraits à partir du lemme. Étant donné, que les lemmes inconnus peuvent appartenir uniquement aux types morphologiques nominaux, adjectivaux ou adjectivaux mixtes[33], la désinence du lemme peut être uniquement l'une des suivantes : *–#, –a, –e, –ě, –í, –ý, –o*[34].

(2) Le début d'une boucle de lecture effectuée sur l'ensemble des sous-types paradigmatiques (spécifiés par la valeur de l'attribut `pdgm`) ayant la même désinence au nominatif singulier que le lemme recherché. Pour chacun de ces sous-types, la même procédure que pour l'annotation de lemmes connus à l'étape (3) et (4) est réalisée.

L'algorithme pourrait être perfectionné surtout pour une restriction de types paradigmatiques possibles plus efficace que celle basée uniquement sur la désinence du nominatif singulier. La condition qui devrait être prise en compte prioritairement serait la nature de la dernière consonne du radical pour pouvoir distinguer les types paradigmatiques durs et mous.

[33]Le type pronominal et numéral sont des classes fermées, gérées par la procédure de l'annotation de lemmes connus.

[34]La désinence *–us* pour les sous-types *virus, Celsius* n'est pas prise en compte par ce traitement.

# Chapitre 4

# Diagnostic des erreurs

Ce chapitre présente le module de diagnostic utilisé sur la plateforme CETLEF pour l'identification des types d'erreurs commises par les apprenants au sein des exercices de déclinaison. Une classification des erreurs d'après des critères formels morphologiques est établie et argumentée : la production erronée et la forme requise sont représentées par une structure de traits morphologiques et la nature de l'erreur est interprétée en fonction de la différence dans les valeurs des attributs de ces structures. Les sections suivantes détaillent la description des procédures automatiques générant un message d'erreur caractérisant une production erronée, utilisé pour la génération du message de diagnostic de cette production, publié sur la plateforme apprenant.

## 4.1 Hypothèses sur les erreurs de déclinaison

Les hypothèses sur les erreurs possibles, commises dans les exercices de déclinaison par des apprenants francophones, peuvent être établies sur la base de la comparaison globale du système nominal tchèque et français. Les fonctions syntaxiques des substantifs, des adjectifs et des pronoms dans la phrase tchèque sont exprimées par des formes casuelles, accompagnées éventuellement par des prépositions. Dans le système français, les fonctions syntaxiques sont marquées uniquement par l'ordre des mots ou à l'aide des prépositions, la flexion nominale n'existe pas sauf pour exprimer l'opposition du singulier et du pluriel. Ainsi, la déclinaison du tchèque présente pour un apprenant français, ou pour tout autre apprenant dont la langue maternelle ne dispose pas de la déclinaison, une sorte d'**idiosyncrasie** par rapport au système de sa langue maternelle. La variation des formes fléchies rajoute de la complexité dans la production langagière et peut être naturellement une source d'erreurs.

Un cadre intéressant pour la prédiction des erreurs possibles sur la base du système morphologique d'une langue est présenté dans les travaux sur la *morphologie naturelle*, voir Dressler (1985), Dressler *et al.* (1987), Kilani-Schoch (1988). Cette théorie définit la « naturalité » dans la morphologie sur la base de la notion de *diagrammaticité*, issue de la sémiotique de Pierce, et sur la base de l'opposition *marqué* × *non marqué*, introduite dans la linguistique générale par le fonctionnalisme pragois. La diagrammaticité est définie comme « iconicité constructionnelle » du signe linguistique : la lisibilité des différents composants (des morphèmes) d'un mot peut être plus ou moins prononcée en fonction du nombre d'opérations nécessaires pour extraire les différents composants. Dans cette perspective, le génitif pluriel *pánů* du substantif *pán* serait plus diagrammatique que le génitif pluriel *vajec* du substantif *vejce*, car il est plus transparent par rapport à ses composants.

La régularité dans la formation des formes fléchies est considérée comme plus naturelle que l'irrégularité, qui peut être la source d'erreurs pour un apprenant étranger mais également pour un enfant qui est en train d'acquérir sa langue maternelle et qui transpose les procédés utilisés pour la création des formes régulières dans le cas des formations idionsyncrasiques. Dans l'acquisition de la langue maternelle ou étrangère, on parle souvent dans ce cas de la surgénéralisation des règles, voir par exemple Clark (2001), Porquier (1977), Gaonac'h (1991). Par rapport à l'opposition *marqué* × *non marqué*, les éléments non marqués sur un certain niveau linguistique devraient être mieux accessibles pour l'apprentissage que les éléments marqués.

La déclinaison tchèque représente pour un francophone une sorte d'« irrégularité », qui doit être maîtrisée par l'assimilation progressive des formes casuelles, avec leur réalisation morphologique ainsi qu'avec la fonction syntaxique qui leur est attribuée dans la phrase. Dans la période préalable à cet maîtrise, des erreurs, présentes nécessairement dans les productions des apprenants, devraient à priori refléter les défaillances dans la manipulation des deux parties de la forme casuelle.

Du point de vue de l'activité de production langagière de l'apprenant[1], il est possible de distinguer plusieurs étapes, qui sont nécessaires pour produire une forme casuelle correcte au sein d'une tâche de déclinaison : (1) le choix des valeurs de la catégorie du cas, du nombre et du genre ; (2) le classement du lexème dans le paradigme approprié et le choix de la désinence correspondante aux valeurs des catégories grammaticales ; (3) la réalisation, si cela est nécessaire, des alternances vocaliques ou consonantiques.

L'étape (1) est effectuée en fonction des critères purement syntaxiques pour l'attribution du cas. Les valeurs du genre et du nombre sont attribuées en fonction de l'accord entre la

[1]La notion de *production* est issue des recherches en psycholinguistique, où la réalisation des énoncés est compris comme une suite organisée de différentes activités d'ordre cognitif qui traduisent un message préverbal vers sa réalisation linguistique, prononcée ou écrite, voir par exemple Ferrand (2001), Fayol (2002).

forme requise et son régisseur où en fonction des critères d'ordre sémantiques, relatifs au contenu cognitif exprimé par la phrase. Un dysfonctionnement dans cette opération serait reflété dans la production par le choix de désinences exprimant les valeurs inappropriées des catégories respectives, par exemple *pána* ou lieu de *pánovi*.

L'attribution d'un certain type paradigmatique au lemme de la forme requise pendant l'étape (2) délimite le répertoire des désinences permettant d'exprimer les valeurs des catégories grammaticales choisies à l'étape précédente. Le classement du lemme dans un paradigme peut être dans certains cas effectué uniquement sur la base de ses propriétés formelles (la périphérie droite du lemme). Dans cette situation, la forme du mot à décliner peut être considérée comme une sorte spécifique d'un signe linguistique au sens de la triade piercienne *symbole*, *icone*, *index*[2]. En effet, la forme en elle même peut fonctionner comme un index de son genre et son classement dans un certain paradigme. Plus le degré de cette indexicalité est élevé, moins il y a d'ambiguïtés de classement dans des groupes différents. Par exemple un mot tchèque dont le lemme est terminé par *–ý* ne peut être classé que dans un seul paradigme de déclinaison, celui des adjectifs durs du type *mladý*[3]. Un mot terminé par la désinence *–a* serait avec le plus de probabilité un substantif féminin décliné d'après le modèle *žena*, mais il peut être également un substantif masculin, décliné d'après le type *předseda*.

Cependant, étant donné la complexité de la déclinaison du tchèque qui prend en compte non seulement la forme des lexèmes, mais également des facteurs d'ordre lexical et sémantique, le critère formel ne suffit pas toujours pour l'attribution d'un type sans qu'il puisse y avoir de l'ambiguïté. Ainsi, nous pouvons supposer l'existence de formes produites par les apprenants qui peuvent être correctes, en ce qui concerne les valeurs des catégories grammaticales liées à la désinence choisie pour la génération de cette forme, mais qui ne sont pas appropriées pour exprimer les significations grammaticales au sein du paradigme propre à la forme requise. Par exemple, le génitif pluriel *turist* du lexème *turista* n'est pas généré d'après le paradigme approprié (type *předseda*), demandant la désinence *–ů* dans le génitif pluriel, mais d'après le type *žena* qui emploie effectivement la désinence zéro *–#* pour exprimer le cas et le nombre correspondants. Dans cet exemple, l'apprenant aurait classé le lexème *turista* dans le type paradigmatique *žena* justement sur la base des ressemblances formelles entre ce type et le type approprié *předseda* (la désinence *–a* dans le nominatif singulier).

La dernière étape (3) représente une opération sur la forme composée du radical et la désinence. Au contact avec certaines désinence, le radical, extrait du lemme, peut changer à cause des alternances consonantiques ou vocaliques. La réalisation des alternances est condi-

[2]Voir par exemple Peirce (1987). La problématique de l'indexicalité des composants morphologiques est abordée également dans Čermák (2000), p. 119

[3]Dans notre modèle, nous n'acceptons pas la définition des types de déclinaison substantivale comme *hajný*, *vstupné*, etc.

tionnée par des facteurs phonologiques, morphologiques et lexicaux. La connaissance des règles pour cette réalisation est nécessaire pour la création adéquate des formes casuelles. Nous pouvons supposer, qu'un apprenant fera des erreurs dans les alternances, car c'est une opération nécessitant des connaissances supplémentaires.

Le diagnostic automatique des erreurs dans CETLEF est construit sur la base de ces hypothèses : une production erronée est considérée comme une combinaison du radical de la forme requise et d'une désinence. Grâce à l'annotation d'une forme requise au sein d'une tâche, il est possible de générer automatiquement différentes formes hypothétiques qui pourraient être produites par un apprenant : (1) des formes casuelles exprimant d'autres valeurs des catégories grammaticales que celles demandées dans la forme requise, (2) des formes qui sont le résultat de la combinaison du radical avec une désinence qui n'appartient pas à son paradigme et (3) les formes sans la réalisation des alternances vocaliques ou consonantiques obligatoires.

## 4.2 Définition des types d'erreurs

Dans cette section, les différents types d'erreurs sont définis sur la base des propriétés morphologiques des formes hypothétiques, générés à partir du radical de la production requise et des désinences employées pour la génération des formes casuelles du tchèque. Ces formes hypothétiques sont comparées à la production erronée et s'il y une correspondance, la production erronée est interprétée par les différentes propriétés de la forme hypothétique correspondante.

### 4.2.1 Définition de l'univers U

L'univers dans lequel les écarts formels entre la forme requise et la production erronée sont interprétés, est défini ici à l'aide des notions issues de la théorie des ensembles et des langages formels[4]. Ces définitions sont établies directement sur le niveau graphique car c'est cette représentation qui est manipulée dans le cadre du diagnostic automatique sur CETLEF.

- Soit un alphabet $L$, ensemble fini de caractères ; l'espace, les chiffres et les signes de ponctuation inclus $L = \{$A, À, Á, Â, ..., 2, ...$\}$[5] ;

[4]Voir par exemple Brugère et Mollard (2003). Dans ce qui suit, le terme langage est utilisé pour désigner l'ensemble des mots sur un alphabet ; le terme alphabet est utilisé pour désigner un ensemble de symboles.

[5]Du point de vue technique, il s'agit dans le cadre de ce travail de tous les caractères supportés par la norme UNICODE (`http://www.unicode.org`) qui peuvent être saisis dans les formulaires dans le cadre d'une application Web.

- Soit un langage $L^*$, ensemble infini de toutes les chaînes possibles sur l'alphabet $L$ : $L^*$ = {*A, AÁ, Aa, Aá, ..., BcD, ..., commençer, ... , slovo, ...*} ;
- Soit un alphabet $L_{CZ} \subset L$, ensemble fini contenant tous les caractères du tchèque à part l'espace, les chiffres et les signes de ponctuation, et qui est une union des ensembles de caractères minuscules, majuscules, minuscules diacritées et majuscules diacritées : $L_{CZ}$ = {*A, Á, a, á, B, b, C, c, Č, č, D, d, Ď, d', E, e, ě, É, é, F, f, G, g, H, h, I, i, J, j, K, k, L, l, M, m, N, n, Ň, ň, O, o, Ó, ó, P, p, R, r, Ř, ř, S, s, Š, š, T, t, Ť, t', U, u, Ú, ú, ů, V, v, Y, y, Z, z, Ž, ž*} ;
- Soit un langage $L_{CZ}^* \subset L^*$, ensemble infini de toutes les chaînes possibles sur l'alphabet $L_{CZ}$ : $L_{CZ}^*$ = {*A, AÁ, Aa, ... , mládovi, ..., ův, ..., zžŽŽŽ*} ;
- Soit une fonction *Min*, opération de minusculisation qui attribue à chaque mot $m \in L_{CZ}^*$ sa forme correspondante uniquement en minuscules : par exemple, pour un mot $m$ = *ČžčRs*, le résultat de l'opération de minusculisation est *Min*(*ČžčRs*) = *čžčrs*.
- Soit une fonction *Dia*, opération d'enlèvement du diacritique qui attribue à chaque mot $m \in L_{CZ}^*$ sa forme correspondante en caractères sans diacritique : par exemple, pour un mot $m$ = *ČžčRs*, le résultat de l'opération d'enlèvement du diacritique est *Dia*(*ČžčRs*) = *CzcRs*.
- Soit une fonction *St*, opération de standardisation telle que $St(m) = Dia(Min(m))$ : par exemple, pour un mot $m$ = *ČžčRs*, le résultat de l'opération de standardisation est *St*(*ČžčRs*) = *czcrs*.
- Soit un langage $N \subset L_{CZ}^*$, ensemble fini de toutes les formes lexicales des mots tchèques appartenant aux types morphologiques nominal, adjectival, adjectival mixte, pronominal ou numéral : $N$ = {*abatyše, ..., mládě, mláděte, ..., žížalami*}.
- Soit un ensemble $R \subset L_{CZ}^*$, ensemble fini de tous les radicaux extraits par l'enlèvement de la désinence casuelle de la forme du lemme des mots tchèques appartenant au type morphologique nominal, adjectival, adjectival mixte, pronominal ou numéral : $R$ = {*abatyš, ..., mlád, ..., pán, ..., žížal*}.
- Soit un ensemble $D \subset L_{CZ}^*$, ensemble fini de toutes les désinences des types paradigmatiques des mots tchèques appartenant au type morphologique nominal, adjectival, adjectival mixte, pronominal[6] ou numéral : $N$ = {*#, a, á, ách, ám, ama, ami, at, ata, atama, atami, atech, atem, atu, atum, atům, aty, e, é, ě, ech, ěch, eho, ého, ej, ejch, ejm, ejma, ejo, em, ém, ěm, ema, ěma, emi, ěmi, emu, ému, en, ete, ěte, etem, ětem, eti,*

[6]Les paradigmes des pronoms spécifiques *já, ty, my, vy, se, kdo, co, nic, on, jenž, všechen*, qui sont inscrits dans le fichier *pdgm.xml* comme des listes de formes lexicales complètes (voir la section 3.4.6, p. 79), font partie de l'ensemble $D$ au niveau technique (pour les procédures automatiques). Pour des raisons de place, elles sont représentées comme un ensemble S qui est l'ensemble des formes lexicales des pronoms spécifiques.

*ěti, ho, i, í, ich, ích, ího, im, ím, ima, íma, imi, ími, ímu, ma, o, oho, oje, oji, ojí, om, omu, on, ona, oni, ono, ony, ou, oum, ové, ovi, u, ů, ůj, um, ům, us, y, ý, ých, ýho, ym, ým, ýma, ými, ýmu, S*}.

- Soit un langage $H \subset L_{CZ}*$, ensemble fini de toutes les formes lexicales hypothétiques des mots tchèques $h$ appartenant au type morphologique nominal, adjectival, adjectival mixte, pronominal ou numéral ; ainsi que leurs formes minusculisées $Min(h)$, sans diacritique $Dia(h)$ et standardisées $St(h)$. Ces formes sont le résultat de la concaténation des couples contenus dans le produit des ensembles $R \times D$, avec ou sans la réalisation de l'alternance sur la chaîne résultante : $H$ = {*abatyša, ... , kupec, kupece, ..., Kočke, Kocke, kočke, kočce, kocce, ..., mládě, ..., mládovi, ..., páněCH, ...., páněte, ...,* }

### 4.2.2 Forme requise et production erronée dans U

Dans l'univers défini précedemment, nous allons définir la forme requise et la production erronée de la manière suivante :

- La forme requise $r$ dans une tâche $x$ est un mot tel que $r \in N$. Chaque $r$ est caractérisé par son annotation morphologique.
- La production erronée $p$ dans une tâche $x$ est un mot tel que $p \in L*$ et $p \neq r$.
- Une production erronée $p$ peut être interprétée morphologiquement si $p$ correspond à une des formes lexicales hypothétiques $h \in H$, générées à partir du radical de la forme requise $r$.
- Une production erronée $p$ ne peut pas être interprétée morphologiquement si $p$ ne correspond à aucune des formes hypothétiques $h \in H$, générées à partir du radical de la forme requise $r$.

### 4.2.3 Interprétation morphologique

Chaque forme requise $r$ dans le cadre d'une tâche $x$ est caractérisée par son annotation morphologique. Cette annotation spécifie son type morphologique, le cas, le nombre, le genre et une éventuelle alternance qui doit être effectuée pour créer cette forme à partir de son lemme. Pour les besoins de la description formelle des erreurs, cette annotation peut être représentée à l'aide d'une **structure de traits**[7]. Pour une forme requise $r$, la structure de traits est la suivante :

[7]Cette façon de représenter les propriétés d'un objet par une structure contenant des couples *trait : valeur* est utilisée dans les formalismes grammaticaux tels que HPSG, LFG, TAG, etc. dites également grammaires d'unification. Pour une revue synthétique, voir par exemple Abeillé (1993).

$$\begin{bmatrix} cas : cas \\ num : nombre \\ gen : genre \\ alt : identifiant\ de\ l'alternace \\ pdgm : soustype\ paradigmatique \\ tagMorph : type\ morphologique \end{bmatrix}$$

Les différentes valeurs des attributs dans cette structure sont instanciées en fonction des propriétés morphologiques de *r* inscrites dans l'annotation. Les noms des attributs sont identiques à ceux utilisés dans l'annotation morphologique sur CETLEF.

Par exemple, pour une forme requise *r* = *matce* qui est le datif singulier du substantif *matka*, l'instanciation de la structure de traits est la suivante :

$$\begin{bmatrix} cas : dat \\ num : sg \\ gen : f \\ alt : k > c \\ pdgm : zn_Re \\ tagMorph : N \end{bmatrix}$$

L'ensemble des formes hypothétiques *h*, générées à partir du radical *matka*, est créé par toutes les combinaisons possibles du radical *matk* avec toutes les désinences dans *D*, avec ou sans la réalisation des alternances, avec ou sans diacritique et avec toutes les possibilités dans la casse des caractères : *matka, matky, ..., matek, matk, ..., matkám, matkam, ..., Matkách, matkach, ..., matkovi, matkem, ..., matkému, ...*.

À chacune de ces formes *h* peut être assignée au moins une structure de traits. Les valeurs des attributs dans la structure de *h* sont déterminées uniquement sur la base de propriétés formelles de ses composants en fonction (1) des différentes valeurs des catégories morphologiques qui peuvent être exprimées par la désinence employée, (2) de la réalisation d'une alternance sur le radical ou (3) de sa forme graphique. Une structure de traits assignée à une forme *h* est appelée son **interprétation**.

### 4.2.4 Attributs de graphie

Les informations sur la graphie d'une forme hypothétiques sont inscrites dans deux attributs supplémentaires qui spécifient la nature de la forme par rapport à la forme graphique canonique :

$$\begin{bmatrix} dia : schéma\ du\ diacritique \\ casse : schéma\ de\ la\ casse \end{bmatrix}$$

En effet, les composantes d'une forme hypothétique, telle qu'elle est définie dans l'univers U, peuvent avoir des variantes dans leur forme graphique en ce qui concerne l'opposition *minuscule* × *majuscule* ou *diacrité* × *non diacrité*. Par exemple, une forme *h* = *stAveni* peut être interprétée sur la base des mêmes valeurs des catégories grammaticales qui sont assignées à la forme *stavení* (la forme canonique).

Pour traiter ce phénomène, les valeurs des attributs *dia* et *casse* sont des structures de traits qui contiennent un attribut pour chaque caractère dans la forme canonique, nommé par ce caractère. La valeur de chacun de ces attributs est une valeur binaire qui indique si les propriétés de la forme canonique ont été observée (la valeur 1) ou non (la valeur 0). Pour l'exemple-ci dessus (*stavení* × *stAveni*), les valeurs des attributs *dia* et *casse* dans la structure de traits caractérisant la forme *stAveni* sont les suivants :

$$\boxed{dia} \begin{bmatrix} s : 1 \\ t : 1 \\ a : 1 \\ v : 1 \\ e : 1 \\ n : 1 \\ \acute{\imath} : 0 \end{bmatrix} \quad \boxed{casse} \begin{bmatrix} s : 1 \\ t : 1 \\ a : 0 \\ v : 1 \\ e : 1 \\ n : 1 \\ \acute{\imath} : 1 \end{bmatrix}$$

Cette spécification est nécessaire car les propriétés graphiques entrent en jeu dans l'interprétation des productions erronées. Par exemple, la production erronée *akci* au lieu de *akcí* peut être interprétée soit comme un erreur de cas, soit comme une erreur de diacritique. Théoriquement, il n'y a pas de moyen de décider quelle interprétation est correcte, cependant les deux doivent être considérées comme possibles. Pour la structure de traits d'une forme requise, les valeurs de ces attributs sont toujours positifs.

### 4.2.5 Exemple d'interprétations

Le nombre des interprétations des formes hypothétiques n'est pas déterminé uniquement par l'homonymie des formes casuelles existantes pour un certain lemme, mais également par toutes les combinaisons possibles du radical et des désinences appartenant aux autres paradigmes. Pour donner un exemple de la richesse des possibilités dans l'attribution des interprétations à une forme hypothétique, voici quatre interprétations pour la forme *h* = *růži* :

$$\boxed{1}\begin{bmatrix} cas : acc \\ num : sg \\ gen : f \\ alt : sans \\ pdgm : rz \\ tagMorph : N \\ dia : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} casse : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} \end{bmatrix} \quad \boxed{2}\begin{bmatrix} cas : inst \\ num : sg \\ gen : f \\ alt : sans \\ pdgm : rz \\ tagMorph : N \\ dia : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 0 \end{bmatrix} casse : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} \end{bmatrix}$$

$$\boxed{3}\begin{bmatrix} cas : inst \\ num : pl \\ gen : i \\ alt : sans \\ pdgm : sj \\ tagMorph : N \\ dia : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} casse : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} \end{bmatrix} \quad \boxed{4}\begin{bmatrix} cas : nom \\ num : pl \\ gen : m \\ alt : sans \\ pdgm : jn \\ tagMorph : A \\ dia : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 0 \end{bmatrix} casse : \begin{bmatrix} r : 1 \\ \mathring{u} : 1 \\ \check{z} : 1 \\ i : 1 \end{bmatrix} \end{bmatrix}$$

L'ambiguïté d'une forme hypothétique n'est pas déterminée seulement sur la base de l'homonymie de la désinence au sein du paradigme de la forme requise, mais également par ses fonctions dans tous les autres paradigmes possibles. En plus, les variantes de diacritique multiplient les interprétations possibles des désinences ayant des signes diacritiques opposés (par exemple $i \times \acute{i}$, $e \times \acute{e}$).

L'interprétation de *h* dans la structure $\boxed{1}$ est qu'il s'agit d'une forme régulière appartenant au paradigme casuel du substantif *růži* qui exprime l'accusatif singulier. En fonction de l'homonymie de la désinence *–i* au sein de ce paradigme, trois autres interprétations sont possibles avec des valeurs différentes dans l'attribut cas : le datif et le locatif.

L'interprétation dans la structure $\boxed{2}$ est qu'il s'agit de la forme régulière de l'instrumental pluriel au sein du paradigme *růže* avec un écart dans le diacritique. Une autre interprétation avec une erreur de diacritique est possible pour le génitif pluriel.

Dans $\boxed{3}$ , la forme *růži* est interprétée comme une forme exprimant l'instrumental singulier au sein du paradigme *stroj*, car dans ce paradigme, ce sont ces valeurs des catégories du cas et du nombre qui sont exprimées par la désinence *i*. Une autre interprétation au sein du même paradigme serait qu'il s'agit du datif, du vocatif ou du locatif singulier.

L'interprétation $\boxed{4}$ est déterminée par les valeurs des catégories morphologiques exprimés par la désinence *–i* au sein du type paradigmatique *jarní* avec un écart dans le diacritique. D'autres interprétations seraient possibles d'après la fonction de la désinence *í* ou *i* au sein de ce paradigme.

### 4.2.6 Erreur d'après l'attribut atteint

Comme nous l'avons défini plus haut, une production erronée *p* peut être interprétée morphologiquement si elle correspond à une des formes lexicales hypothétiques *h* générées à partir du radical de la forme requise *r*. S'il y a une correspondance entre *p* et *h*, *p* peut être interprétée par l'ensemble des structures assignée à *h*.

Ainsi nous pouvons établir que **les attributs communs portant des valeurs différentes dans les structures de *r* et de *p* peuvent servir pour la définition des types d'erreurs** car chaque structure assignée à *p* est nécessairement différente de la structure de *r* dans la valeur d'un attribut minimum.

En fonction des attributs qui diffèrent dans les structures de *r* et de *p* (qui sont atteints par l'erreur), les erreurs sont appelées de la manière suivante :

| attribut | type d'erreur |
|---|---|
| cas | **erreur de cas** |
| num | **erreur de nombre** |
| gen | **erreur de genre** |
| alt | **erreur d'alternance** |
| pdgm | **erreur de type paradigmatique** |
| pdgm | **erreur de sous-type paradigmatique** |
| tagMorph | **erreur de type morphologique** |
| dia | **erreur de diacritique** |
| casse | **erreur de casse** |

Ces erreurs peuvent se combiner librement entre elles en fonction des attributs atteints par l'erreur dans une interprétation donnée. Il peut exister par exemple une erreur de cas et de nombre, une erreur de cas et d'alternance, une erreur de nombre et de graphie, etc. Dans ces appellations, chaque attribut qui contient une valeur différente par rapport à la forme requise doit être spécifié.

### 4.2.7 Erreur par rapport au paradigme de la forme requise

Sur la base des observations des erreurs authentiques produites par les apprenants dans les exercices de déclinaison[8], nous avons établi quatre groupes dans lesquels l'ensemble des interprétations des formes hypothétiques *h*, employées pour la recherche d'une correspondance avec une production erronée *p* pour une forme requise *r*, est limité aux formes qui sont générées par les désinences ayant les propriétés suivantes :

- la désinence appartient au sous-type paradigmatique de la forme requise *r* avec la même valeur de cas, de genre et de nombre. S'il y a une correspondance entre une telle forme hypothétique et la production erronée, nous allons parler d'une **erreur locale**.
- la désinence appartient au sous-type paradigmatique de la forme requise *r* avec la valeur de cas autre que celle de la forme requise. S'il y a une correspondance entre une telle forme hypothétique et la production erronée, nous allons parler d'une **erreur verticale**.
- la désinence appartient aux autres sous-types dans le type paradigmatique de la forme requise *r* avec les mêmes valeurs de cas, de nombre et de genre. S'il y a une correspondance entre une telle forme hypothétique et la production erronée, nous allons parler d'une **erreur horizontale interne**.
- la désinence appartient aux autres types paradigmatiques dans le cadre du même type morphologique avec les mêmes valeurs de cas et de nombre et qui ont la même désinence dans le nominatif singulier comme le lemme de *r*. S'il y a une correspondance entre une telle forme hypothétique et la production erronée, nous allons parler d'une **erreur horizontale externe**.

#### 4.2.7.1 Erreur locale

**Définition** Une *erreur locale* est une interprétation d'une production erronée *p* telle que *p* appartient à l'ensemble des formes hypothétiques dans *H* générées à partir du radical de la production requise *r* et qui a obligatoirement les mêmes valeurs de traits que *r* à part celles des attributs *alt*, *dia* et *casse*.

Dans les deux structures de traits pour *r* et *p*, où *p* est une erreur locale, les traits qui peuvent contenir des valeurs différentes sont mis en gras :

[8] Voir l'enquête préliminaire dans le chapitre *6 Évaluation*, p. 186.

$$\begin{array}{ccc} r & \neq & p \\ \begin{bmatrix} cas : a \\ num : b \\ gen : c \\ \textbf{alt : X} \\ pdgm : e \\ tagMorph : f \\ \textbf{dia : Z} \\ \textbf{casse : P} \end{bmatrix} & & \begin{bmatrix} cas : a \\ num : b \\ gen : c \\ \textbf{alt : Y} \\ pdgm : e \\ tagMorph : f \\ \textbf{dia : W} \\ \textbf{casse : Q} \end{bmatrix} \end{array}$$

Par exemple, une partie de l'ensemble de toutes les productions erronées possibles, qui sont interprétées comme une erreur locale par rapport à la forme requise *kočce*, est le suivant : *kočke, kóčke, kočké, kočkě, kockě, kóckě, kocce, kócce, kóccě, koccě, kocče, kócče, kóccě, koccě, kočče, kóčče, kóčče, kočče, ..., KOCKE, ....*

**Exemples**

(1) *Petr vzal* ***Olge*** *všechny peníze.*
Olze dat.sg.f.
Pierre a pris à Olga tout argent
'Pierre a pris à Olga tout l'argent' (0.88.11)

$$\begin{array}{ccc} Olze & \neq & Olge \\ \begin{bmatrix} cas : dat \\ num : sg \\ gen : f \\ \textbf{alt : g > z} \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : dat \\ num : sg \\ gen : f \\ \textbf{alt : sans} \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *Olge* est une erreur locale d'alternance[9].

(2) *Za* ***vesi*** *bylo obilné pole.*
vsí inst.sg.f.
Derrière village était de blé champ
'Derrière le village était un champ de blé' (0.80.3)

[9]Pour simplifier la notation pour les valeurs des attributs *dia* et *casse*, toute la structure qu'ils contiennent est caractérisée par la valeur *1* s'il n'y a pas de différence par rapport à la structure de la forme canonique, et par la valeur *0*, s'il y existe une différence.

$$\begin{array}{ccc} \textit{vsí} & \neq & \textit{vesi} \\ \begin{bmatrix} cas : inst \\ num : sg \\ gen : f \\ \textbf{alt : e > \#} \\ pdgm : kt_n \\ tagMorph : N \\ \textbf{dia : 1} \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : inst \\ num : sg \\ gen : f \\ \textbf{alt : sans} \\ pdgm : kt_n \\ tagMorph : N \\ \textbf{dia : 0} \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *vesi* est une erreur locale d'alternance et de diacritique.

(3) *Řekněte několik typicky českých* ***jmén**.*
jmen gen.pl.n.
Dites quelques typiquement tchèques noms
'Dites quelques noms typiquement tchèques' (1.41.30)

$$\begin{array}{ccc} \textit{jmen} & \neq & \textit{jmén} \\ \begin{bmatrix} cas : gen \\ num : pl \\ gen : n \\ \textbf{alt : é > e} \\ pdgm : mt \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : gen \\ num : pl \\ gen : n \\ \textbf{alt : sans} \\ pdgm : mt \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée est une erreur locale d'alternance.

#### 4.2.7.2 Erreur verticale

**Définition** Une *erreur verticale* est une interprétation d'une production erronée $p$ tel que $p$ appartient à l'ensemble des formes hypothétiques dans $H$ générées à partir du radical de la production requise $r$, qui a obligatoirement une valeur différente de l'attribut *cas* et obligatoirement une valeur identique de l'attribut *pdgm* et *tagMorph*.

Dans les deux structures de traits pour $r$ et $p$ ou $p$ est une erreur verticale, les traits qui peuvent contenir des valeurs différentes sont mis en gras, ceux qui doivent être obligatoirement différents sont encadrés :

$$\begin{array}{ccc} r & \neq & p \\ \begin{bmatrix} \boxed{\textbf{cas}} : \textbf{X} \\ \textbf{num} : \textbf{Z} \\ \textbf{gen} : \textbf{E} \\ \textbf{alt} : \textbf{P} \\ pdgm : e \\ tagMorph : f \\ \textbf{dia} : \textbf{R} \\ \textbf{casse} : \textbf{T} \end{bmatrix} & & \begin{bmatrix} \boxed{\textbf{cas}} : \textbf{Y} \\ \textbf{num} : \textbf{W} \\ \textbf{gen} : \textbf{F} \\ \textbf{alt} : \textbf{Q} \\ pdgm : e \\ tagMorph : f \\ \textbf{dia} : \textbf{S} \\ \textbf{casse} : \textbf{U} \end{bmatrix} \end{array}$$

Une erreur verticale est donc toute forme hypothétique générée à partir du radical de la production requise avec les désinences de son sous-type paradigmatique, et qui ne peut pas être interprétée comme une erreur locale. La valeur de l'attribut de genre peut être différente uniquement pour les paradigmes adjectivaux et pronominaux. Par exemple, par rapport à la forme requise *kočce*, quelques productions erronées interprétées comme une erreur verticale, sont les suivantes : *kočka, kocka, kočky, ..., kočko, ..., kočkou, koček, kočk, kock, Kock, ....*

**Exemples**

(1) *Výklad našeho* ****průvodci*** *byl velice zajímavý.*
průvodce gen.sg.m.
Exposé notre guide était très intéressant
'L'exposé de notre guide a été très intéressant' (1.32.43)

$$\begin{array}{ccc} \textit{průvodce} & \neq & \textit{průvodci} \\ \begin{bmatrix} \textbf{cas : gen} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : sc \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} \textbf{cas : dat | loc} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : sc \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *průvodci* est une erreur verticale de cas.

(2) *Líbí se mi* ****mori*** *v Bretani.*
moře gen.sg.m.
me plaît mer en Bretagne
'La mer en Bretagne me plaît' (1.12.39)

$$\begin{array}{ccc} \textit{moře} & \neq & \textit{mori} \\ \begin{bmatrix} \textbf{cas : nom} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : sc \\ tagMorph : N \\ \textbf{dia : 1} \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} \textbf{cas : dat | loc} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : sc \\ tagMorph : N \\ \textbf{dia : 0} \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *mori* est une erreur verticale de cas et de diacritique

(3) *Jezdit na* ****koně*** *je* *umění.*
koni loc.sg.m.
Chevaucher cheval est art
'Faire du cheval est un art' (1.86.16)

$$\begin{array}{ccc} \textit{koni} & \neq & \textit{koně} \\ \begin{bmatrix} \textbf{cas : loc} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : mz_k \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} \textbf{cas : gen | acc} \\ num : sg \\ gen : m \\ alt : sans \\ pdgm : mz_k \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *koně* est une erreur verticale de cas.

#### 4.2.7.3 Erreur horizontale interne

**Définition** Une *erreur horizontale interne* est une interprétation d'une production erronée $p$ tel que $p$ appartient à l'ensemble des formes hypothétiques dans $H$ générées à partir du radical de la production requise $r$ et qui a obligatoirement les mêmes valeurs pour les attributs *cas*, *num*, *gen* et *tagMorph* et obligatoirement une valeur différente dans l'attribut *pdgm*. Il y a une restriction sur les valeurs possibles de l'attribut *pdgm* : il doit s'agir uniquement des sous-types du même type paradigmatique que la forme requise.

Dans les deux structures de traits pour $r$ et $p$ ou $p$ est une erreur horizontale interne, les traits qui peuvent contenir des valeurs différentes sont mis en gras, celui qui doit être obligatoirement différent est encadré :

$$\begin{array}{ccc} r & \neq & p \\ \begin{bmatrix} cas : a \\ num : b \\ gen : c \\ \textbf{alt : X} \\ \boxed{\textbf{pdgm}} \textbf{ : Z} \\ tagMorph : d \\ \textbf{dia : P} \\ \textbf{casse : T} \end{bmatrix} & & \begin{bmatrix} cas : a \\ num : b \\ gen : c \\ \textbf{alt : Y} \\ \boxed{\textbf{pdgm}} \textbf{ : W} \\ tagMorph : d \\ \textbf{dia : Q} \\ \textbf{casse : U} \end{bmatrix} \end{array}$$

Il s'agit des formes qui prennent une désinence correspondante aux catégories grammaticales exigées dans la forme requise mais qui appartient à un autre sous-type paradigmatique dans le cadre d'un type paradigmatique.

**Exemples**

(1) *Petr přijel z* ****výleta*** *v Bretani.*
moře gen.sg.i.
Petr est rentré de voyage en Bretagne
'Pierre est rentré du voyage en Bretagne' (1.38.27)

$$\begin{array}{ccc} výletu & \neq & výleta \\ \begin{bmatrix} cas : gen \\ num : sg \\ gen : i \\ alt : sans \\ \textbf{pdgm : hd} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : gen \\ num : sg \\ gen : i \\ alt : sans \\ \textbf{pdgm : hd_l} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *výleta* est une erreur horizontale interne de sous-type paradigmatique.

(2) *Jdeme s kamarády na výlet do* ****lesu****.*
lesa gen.sg.i.
Allons avec camarades à ballade dans forêt
'Nous allons avec les camarades nous promener à la forêt' (1.66.38)

$$\begin{array}{ccc} \textit{lesa} & \neq & \textit{lesu} \\ \begin{bmatrix} cas : gen \\ num : sg \\ gen : i \\ alt : sans \\ \textbf{pdgm : hd_l} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : gen \\ num : sg \\ gen : i \\ alt : sans \\ \textbf{pdgm : hd} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *lesa* est une erreur horizontale interne de sous-type paradigmatique.

(3) *Čteme v novinách o* ****následkech*** *přírodní katastrofy*
následcích loc.pl.i.
Lison dans journal de conséquences naturelle catastrophe
'Nous lisons dans le journal à propos des conséquences d'une catastrophe naturelle' (0.89.24)

$$\begin{array}{ccc} \textit{následcích} & \neq & \textit{následkech} \\ \begin{bmatrix} cas : loc \\ num : pl \\ gen : i \\ \textbf{alt : e > \#, k > c} \\ \textbf{pdgm : hd_z} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : loc \\ num : pl \\ gen : i \\ \textbf{alt : e > \#} \\ \textbf{pdgm : hd} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *následkech* est une erreur horizontale interne de sous-type paradigmatique et d'alternance.

#### 4.2.7.4 Erreur horizontale externe

**Définition** Une *erreur horizontale externe* est une interprétation d'une production erronée $p$ tel que $p$ appartient à l'ensemble des formes hypothétiques dans l'ensemble $H$, restreint uniquement sur les types ayant la même désinence dans le lemme que le lemme de la production requise $r$. Elle doit avoir en plus obligatoirement les mêmes valeurs pour les attributs *cas*, *num* et *tagMorph* et une valeur obligatoirement différentes dans l'attribut *pdgm*.

Dans les deux structures de traits pour $r$ et $p$ ou $p$ est une erreur horizontale externe, les traits qui peuvent contenir des valeurs différentes sont mis en gras, celui qui doit être obligatoirement différent est encadré :

$$\begin{array}{ccc} r & \neq & p \\ \begin{bmatrix} cas : a \\ num : b \\ \textbf{gen : X} \\ \textbf{alt : Z} \\ \boxed{\textbf{pdgm}} \textbf{ : P} \\ tagMorph : c \\ \textbf{dia : T} \\ \textbf{casse : E} \end{bmatrix} & & \begin{bmatrix} cas : a \\ num : b \\ \textbf{gen : Y} \\ \textbf{alt : W} \\ \boxed{\textbf{pdgm}} \textbf{ : Q} \\ tagMorph : c \\ \textbf{dia : U} \\ \textbf{casse : F} \end{bmatrix} \end{array}$$

Il s'agit des erreurs qui sont causées par une confusion de type paradigmatique, liée éventuellement à une confusion de genre. Pour la forme requise *kočce*, qui appartient au type *žena (zn)* et qui est un féminin, une erreur horizontale externe est théoriquement possible par la confusion avec les deux types ayant la même désinence dans le lemme : *předseda (pd)* ou *klima (kl)*. La production erronée d'après les deux types est respectivement *kočkovi, kočkatu*.

**Exemples**

(1) *Maso bez* ***sole*** *není většinou příliš chutné*
soli gen.sg.f
viande sans sel n'est pas d'habitude très appétissant
'La viande sans sel n'est pas d'habitude très appétissante' (1.164.16)

$$\begin{array}{ccc} soli & \neq & sole \\ \begin{bmatrix} cas : gen \\ num : sg \\ gen : f \\ alt : \mathring{u} > o \\ \textbf{pdgm : kt_n} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} & & \begin{bmatrix} cas : gen \\ num : sg \\ gen : f \\ alt : \mathring{u} > o \\ \textbf{pdgm : ps_Re} \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *sole* est une erreur horizontale externe de type paradigmatique.

(2) *K obědu si dáme* ***kuři*** *s rýží*
kuře acc.sg.f
pour déjeuner nous prenons poulet avec riz
'Pour le déjeuner, nous prenons du poulet avec du riz' (1.15.22)

$$\begin{array}{ccc} \textit{kuře} & \neq & \textit{kuři} \\ \begin{bmatrix} \textit{cas} : \textit{acc} \\ \textit{num} : \textit{sg} \\ \textbf{gen : n} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : kr} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix} & & \begin{bmatrix} \textit{cas} : \textit{acc} \\ \textit{num} : \textit{sg} \\ \textbf{gen : f} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : rz} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *kuři* est une erreur horizontale externe de type paradigmatique et de genre.

(3) *Bez* ***túrist*** *je v Praze klid*
turistů gen.pl.m.
sans touristes est à Prague calme
'Sans touristes, Prague est calme' (1.37.24)

$$\begin{array}{ccc} \textit{turistů} & \neq & \textit{túrist} \\ \begin{bmatrix} \textit{cas} : \textit{gen} \\ \textit{num} : \textit{pl} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pd} \\ \textit{tagMorph} : N \\ \textbf{dia : 1} \\ \textit{casse} : 1 \end{bmatrix} & & \begin{bmatrix} \textit{cas} : \textit{gen} \\ \textit{num} : \textit{pl} \\ \textbf{gen : f} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : zn} \\ \textit{tagMorph} : N \\ \textbf{dia : 0} \\ \textit{casse} : 1 \end{bmatrix} \end{array}$$

D'après cette interprétation, la production erronée *turist* est une erreur horizontale externe de type paradigmatique, de genre et de diacritique.

### 4.2.8 Diagnostic morphologique d'une production erronée

Le diagnostic morphologique d'une production erronée $p$ dans une tâche $x$ est l'ensemble de ses interprétations qui sont le plus plausibles du point de vue de l'activité langagière de l'apprenant. Pour illustrer ceci, prenons une tâche avec une forme requise $r$ = *matce* et une production erronée $p$ = *matky* :

*Řekl* ***matky*** *o svém rozhodnutí.*
matce dat.sg.f.
Dit mère de son décision

'Il a parlé de sa décision à sa mère'

La structure de traits correspondant à cette à la forme requise *matce* est la suivante :

$$\begin{bmatrix} cas : dat \\ num : sg \\ gen : f \\ alt : k > c \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix}$$

Pour la production erronée *matky* (décomposés en radical *matk* + la désinence *y*), une interprétation qui vient immédiatement à l'idée est qu'il s'agit d'une erreur de cas : l'apprenant chosit la forme de génitif singulier au lieu de datif singulier. Ainsi nous avons fait un diagnostic qui peut être satisfaisant. Cependant, ce n'est qu'une seule interprétation parmi d'autres qui sont formellement possibles dans l'univers U défini plus haut. À part les différentes combinaisons possibles des valeurs des attributs morphologiques, le nombre d'interprétations d'une forme hypothétique monte exponentiellement en fonction des critères graphiques. Pour un diagnostic automatique, il sera nécessaire de choisir des critères qui puissent permettre de retenir parmi toutes les interprétations possibles uniquement celles, qui sont les plus plausibles.

Pour la production erronée *matky*, quelques interprétations possibles sont présentés ci-dessous. Les couples *trait : valeur* mis en valeur dans les structures sont ceux qui diffèrent par rapport à la structure de la forme requise.

$$\boxed{1}\begin{bmatrix} \mathbf{cas : gen} \\ num : sg \\ gen : f \\ alt : sans \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \quad \boxed{2}\begin{bmatrix} \mathbf{cas : nom} \\ \mathbf{num : pl} \\ gen : f \\ alt : sans \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \quad \boxed{3}\begin{bmatrix} \mathbf{cas : acc} \\ \mathbf{num : pl} \\ gen : f \\ alt : sans \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix} \quad \boxed{4}\begin{bmatrix} \mathbf{cas : voc} \\ \mathbf{num : pl} \\ gen : f \\ alt : sans \\ pdgm : zn_Re \\ tagMorph : N \\ dia : 1 \\ casse : 1 \end{bmatrix}$$

$\boxed{5}\begin{bmatrix} \textbf{cas : acc} \\ \textbf{num : pl} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pn} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{6}\begin{bmatrix} \textbf{cas : inst} \\ \textbf{num : pl} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pn} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{7}\begin{bmatrix} \textbf{cas : gen} \\ \textbf{num : sg} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{8}\begin{bmatrix} \textbf{cas : acc} \\ \textbf{num : pl} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$

$\boxed{9}\begin{bmatrix} \textbf{cas : inst} \\ \textbf{num : pl} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : pd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{10}\begin{bmatrix} \textbf{cas : nom} \\ \textbf{num : pl} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : hd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{11}\begin{bmatrix} \textbf{cas : acc} \\ \textbf{num : pl} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : hd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{12}\begin{bmatrix} \textbf{cas : voc} \\ \textbf{num : pl} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : hd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$

$\boxed{13}\begin{bmatrix} \textbf{cas : inst} \\ \textbf{num : pl} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : hd} \\ \textit{tagMorph} : N \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{14}\begin{bmatrix} \textbf{cas : inst} \\ \textbf{num : pl} \\ \textbf{gen : n} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : mt} \\ \textbf{tagMorph : N} \\ \textit{dia} : 1 \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{15}\begin{bmatrix} \textbf{cas : nom} \\ \textit{num} : \textit{sg} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : md} \\ \textbf{tagMorph : A} \\ \textbf{dia : 0} \\ \textit{casse} : 1 \end{bmatrix}$

$\boxed{16}\begin{bmatrix} \textbf{cas : voc} \\ \textit{num} : \textit{sg} \\ \textbf{gen : m} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : md} \\ \textbf{tagMorph : A} \\ \textbf{dia : 0} \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{17}\begin{bmatrix} \textbf{cas : nom} \\ \textit{num} : \textit{sg} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : md} \\ \textbf{tagMorph : A} \\ \textbf{dia : 0} \\ \textit{casse} : 1 \end{bmatrix}$ $\boxed{18}\begin{bmatrix} \textbf{cas : acc} \\ \textit{num} : \textit{sg} \\ \textbf{gen : i} \\ \textit{alt} : \textit{sans} \\ \textbf{pdgm : md} \\ \textbf{tagMorph : A} \\ \textbf{dia : 0} \\ \textit{casse} : 1 \end{bmatrix}$ ...

Il est évident que pour un diagnostic adéquat, seulement quelques interprétations sont acceptables. Dans l'exemple ci-dessus, l'interprétation la plus probable est la structure $\boxed{1}$ qui est une erreur verticale de cas. Cette erreur pourrait être une manifestation d'une identification erronée de la fonction syntaxique de la forme requise ou de la rection du verbe, ou tout simplement une erreur dans le choix de la désinence sans une motivation spécifique.

Les interprétations $\boxed{2}$, $\boxed{3}$, $\boxed{4}$, qui sont des erreurs verticales de cas et de nombre, sont moins probables car ce sont les valeurs de deux attributs qui diffèrent : l'apprenant se trompe et dans le cas, et dans le nombre.

Les interprétations suivantes sont formellement possibles mais elle ne sont pas vraisemblables pour servir comme un diagnostic adéquat de la production erronée. Les structures $\boxed{5}$ – $\boxed{14}$ donnent une interprétation qui voudrait dire que l'apprenant se trompe également de genre et de types paradigmatiques. Les structures $\boxed{15}$ – $\boxed{18}$ interprètent la forme *matky* comme si c'était une forme adjectivale avec une erreur de diacritique dans la désinence. Aucune de ces interprétations ne rentre dans le cadre des types d'erreurs définis par rapport au paradigme de la forme requise (locale, verticale, horizontale interne et externe) et par définition, elles ne peuvent pas être prises en compte pour un diagnostic.

### 4.2.9 Plausibilité d'une interprétation morphologique

La plausibilité d'une interprétation peut être établie sur la base des critères morphologiques pour les erreurs locales et horizontales. Par contre, pour les erreurs verticales, où la valeur de la catégorie du cas est une variable, des facteurs syntaxiques entrent nécessairement en jeu.

Dans la procédure de diagnostic, telle qu'elle est proposée dans ce travail, des informations d'ordre syntaxique ne sont pas disponibles. Il serait envisageable d'accompagner les phrases dans les tâches par leur structures syntaxiques, d'assigner à la forme requise sa fonction dans la phrase et inclure ces données dans la procédure du diagnostic. Il serait également utile de disposer de l'information sur la valence du régisseur de la forme requise, pour pouvoir identifier les erreurs de cas causées par le choix erroné parmi plusieurs cadres syntaxiques possibles qui déterminant les propriétés morphologiques de la forme requise.

Le diagnostic automatique sur CETLEF ne peut prendre en compte que les informations morphologiques et son ambition n'est que de proposer la meilleure solution dans le cadre donnée. Cette solution peut être par la suite confirmée ou rejetée à l'aide d'une étude « manuelle », effectuée par un humain qui prend en compte des critères divers qui lui permettent de choisir l'interprétation la plus probable.

Dans le cadre morphologique, nous définissons que **la plausibilité d'une interprétation est déterminée par le nombre d'attributs atteints par l'erreur**. Moins il y a d'attributs qui diffèrent dans la structure de $r$ et dans une certaine interprétation de $p$, plus cette interprétation est plausible. Si le nombre d'attributs atteints dans deux interprétations est égal, soit elles sont retenues toutes les deux, soit c'est le classement des attributs dans une échelle spécifiant son pertinence pour le diagnostic, qui décide. Par exemple, la production erronée

*jmén* au lieu de *jmen* (lemme *jméno*) est diagnostiquée prioritairement comme une erreur dans l'alternance *é* > *e* que d'une erreur de diacritique, car ce diagnostic est plus pertinent pour l'apprenant.

Cette solution devrait être acceptable du point de vue de l'apprenant pour lequel le diagnostic est destiné : l'explication de l'écart entre la forme requise et la production erronée sur la base du nombre le plus petit des attributs dont il est nécessaires de changer les valeurs pour rendre la production correcte, peut être considérée comme adéquate au niveau didactique.

## 4.3 Procédure *Diagnostic*

La procédure *Diagnostic* est employée sur la plateforme CETLEF pour diagnostiquer les productions qui ne correspondent à aucune des formes requises dans le cadre d'une tâche. Ce diagnostic est effectué par la recherche de différentes interprétations de la production erronée et par le choix de celles qui sont les plus plausibles. À la sortie de la procédure, un message qui spécifie l'erreur est généré. Ce message sert comme critère pour les recherches des productions dans la base de données en fonction des différents types d'erreurs et pour la génération du message de diagnostic qui spécifie la production erronée sur la plateforme apprenant de CETLEF.

### 4.3.1 Description

La procédure *Diagnostic* est composée d'une série de tests qui tentent d'interpréter l'écart entre la production erronée et la forme requise. Les tests sons rangés dans des blocs en fonction de leur complexité au niveau du calcul. Ainsi, si une interprétation est suffisante pour un diagnostic adéquat, la procédure est terminée, ce qui permet d'économiser les ressources au niveau informatique.

D'abord, l'interprétation de l'erreur est effectuée à l'aide des techniques simples qui n'impliquent pas l'utilisation des données morphologiques. Le but est d'identifier, à l'aide des calculs sur les caractères et les chaînes de caractères qui représentent la forme requise et la production erronée, une différence trop importante entre ces deux éléments, pour qu'il puisse y avoir une interprétation morphologique. Si cette étape n'a pas été suffisante pour déterminer le bon diagnostic, une série de tests morphologiques est commencée pour interpréter l'erreur comme une forme hypothétique générée à partir du radical de la forme requise.

Les tests morphologiques utilisent les informations linguistiques dans l'annotation de la forme requise, le modèle de la déclinaison, structuré dans les fichiers *pdgm.xml* et *alt.xml*,

et la procédure *AlterneRadical*. Pendant ce traitement, des formes hypothétiques sont générées à partir du radical de la forme requise. Ces formes doivent nécessairement observer les restrictions posées sur les erreurs locales, verticales, horizontales internes et horizontales externes. Chaque forme hypothétique est ensuite systématiquement comparée à la production erronée. S'il y a une correspondance, l'erreur est interprétée sur la base des propriétés morphologiques de cette forme et cette interprétation est inscrite parmi les autres possibles.

Les différentes interprétations possibles pour une production erronée sont ensuite triées et filtrées pour ne retenir que les interprétations plausibles. Cette procédure classe les différentes interprétations d'après une échelle, établie sur la nature de l'erreur, et ne retient que celles qui sont placées le plus haut. Prenons les différentes interprétations de l'erreur dans la tâche suivante :

| | | | |
|---|---|---|---|
| *David* | *nemůže* | *jíst* | ****rajče**.* |
| | | | rajčata acc.pl.n. |
| David | ne peut pas | manger | les tomates |

'David ne peut pas manger des tomates'(1.18.43)

La comparaison de la production erronée avec les formes hypothétiques détermine qu'il peut s'agir des erreurs suivantes : (a) une erreur verticale de nombre d'après l'accusatif singulier ; (b) une erreur verticale de cas et de nombre d'après le nominatif singulier ; (c) une erreur verticale de cas et de nombre d'après le vocatif singulier ; (d) une erreur horizontale externe de type paradigmatique d'après l'accusatif pluriel du type *moře* ; (e) une erreur horizontale externe de type paradigmatique et de genre d'après l'accusatif pluriel du type *růže* ; (f) une erreur horizontale externe de type paradigmatique et de genre d'après l'accusatif pluriel du type *soudce*.

Pendant l'étape de filtrage des interprétations possibles, les interprétations retenues comme les plus probables sont les interprétations (a) et (d) : l'apprenant se trompe soit dans le nombre et met la forme du singulier au lieu de la forme du pluriel ; soit il confond le type paradigmatique *kuře* avec le type *moře*, qui a le même genre. Cette décision est prise sur la base du nombre de traits morphologiques atteintes par l'erreur : il s'agit d'un seul trait pour les interprétations (a) et (d) ; et de deux traits pour les interprétation (b), (c), (e) et (f). Sur la base de ce critère, les interprétations (b), (c), (e), (f) peuvent être rejetées, car deux interprétations, classées plus haut sur l'échelle de la plausibilité, ont été trouvées.

La dernière étape de la procédure *Diagnostic* consiste en une traduction des interprétations retenues par le filtrage au format qui sera publié sur la plateforme apprenant comme un retour sur sa production. Pour l'exemple ci-dessus, sa forme est la suivante : *« Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le*

*mot d'après le type moře. 2) Le cas et le genre sont corrects mais vous avez mis une forme de singulier au lieu de pluriel. »*

### 4.3.2 Structure de la procédure

Les données à l'ENTRÉE de la procédure sont : la **production erronée**, la **forme requise**, le **lemme** de la forme requise et l'**annotation** de la forme requise (catégorie lexicale, type morphologique, sous-type paradigmatique, cas, nombre, genre, alternance)[10].

Les données à la SORTIE de la procédure sont : le **message d'erreur**, contenant l'ensemble des interprétations retenues par le diagnostic et le **diagnostic**, qui contient le message destiné pour la plateforme apprenant.

Pour la génération des formes hypothétiques, la procédure emploie les informations morphologiques contenues dans le fichier *pdgm.xml* et la procédure *AlterneRadical*, qui utilise les informations sur les alternances, structurées dans le fichier *alt.xml*.

Il y a quatre étapes principales dans cette procédure : (1) **traitement non morphologique**, (2) **traitement morphologique**, (3) **filtrage des interprétations** et (4) **formatage du diagnostic**. Si une interprétation satisfaisante est trouvée pendant les tests non morphologiques, les étapes (2) et (3) ne sont pas effectuées et l'étape suivante est l'étape (4). Si aucune interprétation n'est trouvée dans l'étape (1), les autres se suivent successivement.

## 4.4 Message d'erreur

Le **message d'erreur** sert pour stocker les différentes interprétations d'une production erronée qui ont été trouvées au cours de la procédure pendant les étapes (1) et (2). Ces interprétations sont soit des interprétations morphologiques d'une forme hypothétique qui correspond à la production erronée ; soit des interprétations non morphologiques, établies sur la base des calculs avec les caractères et les chaînes de caractères dans la production erronée et la forme requise.

Au niveau technique, le message d'erreur est représenté par une table à deux dimensions dans laquelle les lignes contiennent les différentes interprétations d'une production erronée et les colonnes les valeurs des différents attributs de cette interprétation.

[10]Le traitement des erreurs de casse n'est pas inclus dans cette procédure. À l'entrée, toute la production erronée, la forme requise et le lemme de la forme requises sont minusculisées pour éviter l'influence de ce facteur.

### 4.4.1 Attribut `erreur`

Chaque interprétation est obligatoirement spécifiée par la valeur de l'attribut `erreur`.

| attribut obligatoire pour toutes les interprétations | | |
|---|---|---|
| *attribut* | *description* | *exemple de la valeur* |
| `erreur` | type global de l'erreur | `HOR EXT` |

Pour les productions qui n'ont pas d'interprétation morphologique, cet attribut spécifie le type de l'écart identifié entre la production erronée et la forme requise, calculé à l'aide des techniques non morphologiques.

| interprétations non morphologiques sur les caractères | |
|---|---|
| *valeur de* `erreur` | *description* |
| `ETRANG` | production erronée contient des caractères étrangers |
| `NUMERIC` | production erronée contient des caractères numériques |
| `PUNCT` | production erronée contient des signes de ponctuations |

| interprétations non morphologiques sur les chaînes | |
|---|---|
| *valeur de* `erreur` | *description* |
| `SIGNE UNIQUE` | production erronée est une chaîne composé d'un seul caractère |
| `COURT` | production erronée est trop court |
| `LONG` | production erronée est trop long |
| `LEV` | distance de Levenshtein trop grande entre production erronée et la production requise |

Pour les productions ayant une ou plusieurs interprétations morphologiques, cet attribut spécifie le type de l'erreur en fonction de son positionnement par rapport à la forme requise :

| interprétations morphologiques | |
|---|---|
| *valeur de* `erreur` | *description* |
| `LOC` | erreur locale |
| `VERT` | erreur verticale |
| `HOR INT` | erreur horizontale interne |
| `HOR EXT` | erreur horizontale externe |

Au cas où, à la fin du *Diagnostic*, il n'y aurai aucune interprétation de la production erronée, l'attribut `erreur` contient la valeur `INCONNU`.

| aucune interprétation | |
|---|---|
| *valeur de* `erreur` | *description* |
| `INCONNU` | aucune interprétation n'a été trouvée |

### 4.4.2 Attributs morphologiques

Les productions avec une interprétation morphologique sont obligatoirement spécifiées par les valeurs des attributs `pdgm` (identifiant du sous-type paradigmatique), `term` (la désinence casuelle), `cas` (cas), `num` (nombre), `gen` (genre) dont les valeurs sont lues dans le fichier *pdgm.xml* pour la désinence employée dans la génération de la forme hypothétique.

| attributs obligatoires pour les interprétations morphologiques | | |
|---|---|---|
| *attribut* | *description* | *exemple de la valeur* |
| `pdgm` | sous-type paradigmatique | `zn_Re` |
| `term` | la désinence casuelle | `a` |
| `cas` | le cas | `dat` |
| `num` | le nombre | `pl` |
| `gen` | le genre | `f` |

### 4.4.3 Attribut `spec`

L'attribut `spec` spécifie les interprétations des erreurs verticales et horizontales par rapport au catégories morphologiques de cas, de nombre et de genre atteintes par l'erreur.

| attribut `spec` | | |
|---|---|---|
| *attribut* | *description* | *exemple de la valeur* |
| `spec` | spécification de l'erreur | `erreur_num` |

Les valeurs possibles de l'attribut `spec` qui spécifient les erreurs verticales sont les suivantes :

| valeurs possibles de `spec` pour les erreurs verticales | | | | |
|---|---|---|---|---|
| *valeur de* `spec` | *description* | `cas` | `num` | `gen` |
| `erreur_cas` | erreur de cas | 0 | 1 | 1 |
| `erreur_num` | erreur de nombre | 1 | 0 | 1 |
| `erreur_gen` | erreur de genre | 1 | 1 | 0 |
| `erreur_cas_num` | erreur de cas et de nombre | 1 | 0 | 0 |
| `erreur_cas_gen` | erreur de cas et de genre | 0 | 1 | 1 |
| `erreur_num_gen` | erreur de nombre et de genre | 1 | 0 | 0 |
| `erreur_cas_num_gen` | erreur de cas, de nombre et de genre | 0 | 0 | 0 |

La seule valeur possible de l'attribut `spec` pour les erreurs horizontales internes est la valeur `ident` car par définition, le cas, le nombre et le genre sont identiques à la forme requise :

| valeurs possibles de `spec` pour les horizontales internes | | | | |
|---|---|---|---|---|
| *valeur de* `spec` | *description* | `cas` | `num` | `gen` |
| `ident` | erreur de sous-type paradigmatique | 1 | 1 | 1 |

Les valeurs possibles pour spécifier les erreurs horizontales externes sont les suivantes :

| valeurs possibles de `spec` pour les erreurs horizontales externes | | | | |
|---|---|---|---|---|
| *valeur de* `spec` | *description* | `cas` | `num` | `gen` |
| `erreur_gen` | erreur de type paradigmatique et de genre | 1 | 1 | 0 |
| `ident` | erreur de type paradigmatique | 1 | 1 | 1 |

Chaque valeur de l'attribut `spec` spécifie nécessairement tous les attributs parmi *cas*, *num* et *gen* atteints par l'erreur. Ceux qui ne sont pas mentionnés, ont la même valeur que la forme requise. Pour chaque interprétation, il y a uniquement une valeur de l'attribut `spec`.

A priori, la spécification dans `spec` apporte une information redondante, car la nature de l'erreur peut être calculée par la comparaison de l'annotation de la forme requise et les valeurs des catégories morphologiques dans le message d'erreur. Cependant, cette spécification rend le message plus autonome dans les traitements de filtrage des interprétations et elle permet de formuler plus confortablement les requêtes de recherche, une fois le message d'erreur inscrit dans la base de données pour accompagner une production erronée.

### 4.4.4 Attributs concernant les alternances

Les attributs `alt` (identifiant de l'alternance), `valid_alt` (informations sur la réalisation de l'alternance), sont ajoutés à l'interprétation d'une production erronée à l'aide de la procédure *AlterneRadical*. Pour les alternances doubles (par exemple les alternance *ů* > *o* et *r* > *ř* dans *dvůr* > *dvoře*), les attributs ajoutées sont `alt_1` (identifiant de l'alternance 1), `valid_alt_1` (informations sur la réalisation de l'alternance 1), `alt_2` (identifiant de l'alternance 1), `valid_alt_2` (informations sur la réalisation de l'alternance 2).

| attributs spécifiant les alternances | | |
|---|---|---|
| *attribut* | *description* | *exemple de la valeur* |
| `alt` | alternance | `e > #` |
| `valid_alt` | réalisation de l'alternace | `OK` |
| `alt_1` | alternance 1 | `ů > o` |
| `valid_alt_1` | réalisation de l'alternace 1 | `ERR` |
| `alt_2` | alternance 2 | `r > ř` |
| `valid_alt_2` | réalisation de l'alternace 2 | `OK` |

Si dans une forme hypothétique une alternance doit être réalisée, les attributs `valid_alt`, `valid_alt_1`, `valid_alt_2` spécifient si elle a été réalisée (valeur `'OK'`) ou non (valeur `'ERR'`).

### 4.4.5 Attribut `dia`

L'attribut `dia` avec une seule valeur possible `dia` figure dans l'interprétation de la production erronée avec une erreur de diacritique.

| Valeur possible de l'attribut `dia` | |
|---|---|
| *valeur de* `dia` | *description* |
| `dia` | erreur de diacritique |

### 4.4.6 Attribut `var`

L'attribut `var` avec deux valeurs possibles `reg` et `fnct` sert pour marquer les interprétations des productions erronées qui contiennent une désinence inscrite dans le fichier *pdgm.xml* comme une variante de registre ou comme une variante fonctionnelle qui n'est pas acceptable au sein d'une certaine tâche (par exemple *v roku* au lieu de *v roce*)

| valeur possible de l'attribut var | |
|---|---|
| *valeur de* var | *description* |
| reg | variante de registre |
| fnct | variante fonctionnelle inacceptable |

### 4.4.7 Exemples

(1) *Od* ***muzeum*** *přichází Jana*
muzea acc.sg.i.
De musée vient Jana
'Jana vient du musée.' (1.57.63)

Le message d'erreur pour la production erronée dans la tâche ci-dessus contient, avant le début de la procédure de filtrage des interprétations, cinq lignes avec cinq interprétations différentes :

| message_erreur | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| erreur | pdgm | term | cas | num | gen | alt | spec_erreur | var | dia |
| VERT | mt_m | um | nom | sg | n | sans | | | |
| VERT | mt_m | um | acc | sg | n | sans | | | |
| VERT | mt_m | um | voc | sg | n | sans | | | |
| VERT | mt_m | ům | dat | pl | n | sans | erreur_num_cas | reg | dia |
| VERT | mt_m | um | dat | pl | n | sans | erreur_num_cas | reg | |

(2) Exemple d'une ligne contenant l'interprétation de la production *turist* au lieu *turistů*$_{gen.pl.m.}$ qui est une erreur horizontale externe de genre :

| erreur | pdgm | term | cas | num | gen | alt | spec |
|---|---|---|---|---|---|---|---|
| HOR EXT | zn | # | gen | pl | f | sans | erreur_gen |

(3) Exemple d'une ligne contenant l'interprétation de la production *koček*$_{gen.pl.f.}$ au lieu de *kočkám*$_{dat.pl.f.}$, qui est une erreur verticale de cas :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt |
|---|---|---|---|---|---|---|---|
| VERT | zn_Re | # | gen | pl | f | # > e | OK |

(4) Exemple d'une ligne contenant l'interprétation de la production *koček*$_{gen.pl.f.}$ au lieu de *kočkám*$_{dat.pl.f.}$, qui est une erreur verticale de cas et d'alternance :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt |
|---|---|---|---|---|---|---|---|
| VERT | zn_Re | # | gen | pl | f | # > e | ERR |

(5) Au cas où il y deux alternances qui doivent être réalisées, les productions sont interprétées en fonction de la réalisation des deux alternances. Pour la production erronée *vajc* ou lieu de *vajec*, l'interprétation correspondante est :

| erreur | pdgm | term | cas | num | gen | alt_1 | valid_alt_1 | alt_2 | valid_alt_2 |
|---|---|---|---|---|---|---|---|---|---|
| LOC | mr_v | # | gen | pl | n | e > a | OK | # > e | ERR |

(6) Pour la production erronée *vejc* ou lieu de *vajec*, l'interprétation est :

| erreur | pdgm | term | cas | num | gen | alt_1 | valid_alt_1 | alt_2 | valid_alt_2 |
|---|---|---|---|---|---|---|---|---|---|
| LOC | mr_v | # | gen | pl | n | e > a | ERR | # > e | ERR |

(7) Pour la production erronée *vejec* ou lieu de *vajec*, l'interprétation est :

| erreur | pdgm | term | cas | num | gen | alt_1 | valid_alt_1 | alt_2 | valid_alt_2 |
|---|---|---|---|---|---|---|---|---|---|
| LOC | mr_v | # | gen | pl | n | e > a | ERR | # > e | OK |

## 4.5 Traitement non morphologique

La première étape de la procédure *Diagnostic* consiste dans l'interprétation des productions erronées à l'aide des méthodes qui n'impliquent pas l'emploi de données morphologiques dans le fichier *pdgm.xml* et *alt.xml*. L'objectif de cette étape est d'éviter le traitement morphologique pour les productions erronées qui ne pourraient pas avoir une interprétation dans les tests morphologiques de la procédure *Diagnostic*. Elle doivent avoir au moins une des propriétés suivantes : (a) elles contiennent des caractères qui ne sont pas des signes alphabétiques du tchèque ; (b) elles sont trop différentes par rapport à la forme requise pour qu'elles puissent être interprétées comme des formes hypothétiques générées à partir de son lemme.

En pratique, les productions avec ces caractéristiques sont des fautes de frappe, des confusions de tout genre, des productions qui sont saisies par les apprenant comme volontairement erronées, etc. Formellement, en nous appuyant sur les définitions à la page 122, une production erronée qui n'a pas d'interprétation morphologique, est un mot appartenant

à la différence des langages $L^*-H$ où $L^*$ contient tous les mots possibles créés par tous les caractères possibles et $H$ est l'ensemble des formes hypothétiques pour une forme requise.

Les productions remplissant les caractéristiques établies dans le type (a) sont repérées pendant les tests sur les caractères, les productions de type (b) sont repérées pendant les tests sur les chaînes.

### 4.5.1 Tests sur les caractères

1. **si** $p$ contient des caractères étrangers
   **alors** inscrit dans le message d'erreur `erreur <= ETRANG`
2. **si** $p$ contient des caractères numériques
   **alors** inscrit dans le message d'erreur `erreur <= NUMERIC`
3. **si** $p$ contient des signes de ponctuation
   **alors** inscrit dans le message d'erreur `erreur <= PUNCT`
4. **si** au moins un des tests (1), (2), (3), (4) est réussi
   **alors** passe à l'étape *Formatage du diagnostic* **fin**
   **sinon** passe au tests sur les chaînes

**Remarque** L'élimination des productions erronées qui sont interprétées par ces tests pourrait être considérée comme trop stricte dans certains cas, notamment pour celles qui contiennent des signes de ponctuation ou l'espace. En effet, s'il y a uniquement une virgule ou un point attachée à la production, le diagnostic pourrait se poursuivre après l'enlèvement de ce symbole. Il s'agit d'une des améliorations de la procédure *Diagnostic* à intégrer dans le futur ce qui permettra d'augmenter la robustesse de dispositif. Concernant le traitement des caractères « parasites » dans les entrées textuelles des dispositifs informatiques, rajoutant du bruit à une donnée autrement valide, Antoniadis *et al.* (2004) parlent du « spacebar syndrom » comme d'une déficience majeure des outils actuels d'enseignement assisté par ordinateur.

### 4.5.2 Tests sur les chaînes

1. **si** la production erronée contient un seul caractère répété une ou plusieurs fois
   **alors** inscrit dans le message d'erreur `erreur <= SIGNE UNIQUE`
2. **si** la différence entre la longueur de $p$ et la longueur du lemme de $r$ est supérieure à 4
   **alors** inscrit dans le message d'erreur `erreur <= LONG`
   **sinon si** la différence entre la longueur du lemme de $r$ et la longueur de $p$ est supérieure

à 1
**alors** inscrit dans le message d'erreur `erreur <= COURT`

3. **si** la distance de Levenshtein entre *p* et *r* est supérieure à 4
   **alors** inscrit dans le message d'erreur `erreur <= LEV`
4. **si** au moins un des tests (1), (2), (3) est réussi
   **alors** passe à l'étape *Formatage du diagnostic* **fin**
   **sinon** passe au tests morphologiques

**Remarque 1** Les limitations sur la différence entre la longueur de *p* et la longueur du lemme de *r* dans le test (2) sont motivées linguistiquement. En effet, dans la déclinaison du tchèque, le nombre maximum de caractères qui peuvent être ajoutés au lemme pour créer une forme casuelle est 4 : c'est le cas du locatif pluriel du type *kuře* et *klima* où la forme casuelle est créée par la chaîne *atech* (*ch* est considéré ici comme deux caractères) ou avec la variante non standard *atama*. Une production erronée ne peut pas être interprétée comme une forme hypothétique si sa longueur est supérieure de 4 points par rapport à celle du lemme de la forme requise.

**Remarque 2** Dans le test (3), la distance de Levenshtein[11] entre *p* et *r* est et le nombre minimal d'opérations de substitution, d'ajout ou de suppression d'un caractère dans *p* pour le rendre égale à *r*. Il s'agit d'une technique qui est couramment utilisée dans les correcteurs orthographiques basés sur des lexiques : les mots proposés pour la correction d'un mot qui n'est pas inscrit dans le lexique sont ordonnés en fonction de la distance de Levehnstein, celui avec la valeur la plus petite est en tête de la liste. Par exemple, la distance entre *žena* et *ženu* est égale à 1, entre *žena* et *ženou* à 2, entre *žena* et *ženami* à 2, entre *žena* et *ženách* à 3, etc. L'intérêt d'utiliser cette mesure ici est de pouvoir éliminer des *p* qui sont trop différents par rapport à *r*. La valeur maximale utilisée dans le test (3) a été établie sur la base des mesures effectuées par la comparaison des différentes formes casuelles au sein des types paradigmatiques. À part les différences causées par les désinences, il est également nécessaire de prendre en compte les modifications provoquées par les alternances. La distance maximale trouvée est égale à 4. Il s'agit des différences entre les couples de formes casuelles comme *pes* × *psovi, blan* × *blánami, kuře* × *kuřatům, idea* × *idejích, voják* × *vojácích, hub* × *houbami*, etc.

[11] Voir Levenshtein (1966). Dans la procédure *Diagnostic*, la distance de Levenshtein est calculée à l'aide de la fonction *levenshtein*, qui fait partie de la bibliothèque standard du langage PHP à partir de la version 4.0.1.

**Remarque 3** Des productions erronées authentiques, interprétées comme trop différentes à l'aide de la distance de Levenshtein, sont par exemple *houbovych* ou lieu de *hub*, *vejcatach* ou lieu de *vajec*, *rad* ou lieu de *radost*. Une certaine motivation morphologique de ces productions est évidente, mais elles ne pourraient pas être interprétées morphologiquement par la procédure automatique telle qu'elle est conçue et présentée dans ce travail.

## 4.6 Traitement morphologique

La seconde étape de la procédure *Diagnostic* consiste dans la génération des formes hypothétiques à partir du radical de la forme requise et des désinences casuelles lues dans le fichier *pdgm.xml*. Cette génération est effectuée à l'aide de la procédure *AlterneHypothèse*[12] qui génère toutes les combinaisons possibles pour un radical et une désinence (la réalisation ou l'absence des alternances simples et doubles) ainsi que les variantes sans diacritique. Ces formes sont systématiquement comparées à la production erronée. S'il y a une correspondance, les informations morphologiques liées à la désinence dans la forme hypothétique et sur les alternances éventuelles du radical sont inscrites comme une interprétation possible dans le message d'erreur.

Le traitement morphologique est divisé en quatre étapes en fonction de l'axe de la lecture des désinences dans le fichier *pdgm.xml* : lecture locale, verticale, horizontale interne et externe. Les interprétations trouvées pendant chacune de ces lectures correspondent aux types d'erreurs établies sur la base du même critère.

Il y a une condition dans le déroulement de ce traitement : l'exécution des deux lectures horizontales est conditionnée par l'échec de la lecture locale. Si une interprétation locale est trouvée, il est inutile de chercher dans les autres sous-types car les interprétations horizontales seraient forcément moins probables. Par contre, la lecture verticale est exécutée même après une interprétation locale retrouvée car une erreur locale de diacritique peut également avoir une interprétation probable comme une erreur verticale de cas (par exemple *akci* au lieu de *akcí*).

À la fin de la lecture verticale et des lectures horizontales, la valeur de l'attribut `spec` est calculée d'après les différences dans les attributs `cas`, `num`, `gen` entre la forme hypothétique est la forme requise.

[12] Voir plus bas, pp. 152–154.

### 4.6.1 Lecture locale

La lecture locale est exercée uniquement sur les désinences dont toutes les valeurs des attributs `cas`, `num`, `gen`, `pdgm` sont identiques à celle de la forme requise : cette désinence est utilisée pour la génération de la forme hypothétique *h*. Si une correspondance entre la production erronée *p* et *h* est trouvée au cour de ce traitement, les informations lues dans le fichier *pdgm.xml* pour la désinence courante sont inscrites dans le message d'erreur :

| | | |
|---|---|---|
| `classe <= LOC` | `pdgm <=` *pdgm courant* | `des <=` *désinence courante* |
| `cas <=` *cas courant* | `num <=` *num courrant* | `gen <=` *gen courant* |

### 4.6.2 Lecture verticale

La lecture verticale est exercée sur toutes les désinences au sein du sous-types paradigmatique de la forme requise à part celle de la forme requise : ces désinences sont utilisées pour la génération de la forme hypothétique *h*. Si une correspondance entre la production erronée *p* et *h* est trouvée au cour de ce traitement, les informations lues dans le fichier *pdgm.xml* pour la désinence courante sont inscrites dans le message d'erreur :

| | | |
|---|---|---|
| `classe <= VERT` | `pdgm <=` *pdgm courant* | `des <=` *désinence courante* |
| `cas <=` *cas courant* | `num <=` *num courant* | `gen <=` *gen courant* |

### 4.6.3 Lecture horizontale interne

La lecture horizontale interne est exercée uniquement sur les désinences en dehors du sous-type de la forme requise mais dans le cadre du même type. Uniquement les désinences avec le même cas, nombre et genre sont utilisées pour la génération de la forme hypothétique *h*. Si une correspondance entre la production erronée *p* et *h* est trouvée au cour de ce traitement, les informations lues dans le fichier *pdgm.xml* pour la désinence courante seront inscrites dans le message d'erreur :

| | | |
|---|---|---|
| `classe <= HOR INT` | `pdgm <=` *pdgm courant* | `des <=` *désinence courant* |
| `cas <=` *cas courant* | `num <=` *num courant* | `gen <=` *gen courant* |

### 4.6.4 Lecture horizontale externe

La lecture horizontale externe est exercée uniquement sur les désinences en dehors du type de la forme requise et uniquement dans le cadre de ces types qui ont la même désinence

dans le lemme que celle dans le lemme de la forme requise. Comme pour le traitement horizontal interne, uniquement les désinences avec le même cas, nombre et genre sont utilisées pour la génération de la forme hypothétique. S'il y a une correspondance entre $p$ et $h$ trouvée au cour de ce traitement, les informations lues dans le fichier *pdgm.xml* pour la désinence courante seront inscrites dans le message d'erreur :

| | | |
|---|---|---|
| `classe <= HOR EXT` | `pdgm` <= *pdgm courrant* | `des` <= *désinence courrante* |
| `cas` <= *cas courrant* | `num` <= *num courrant* | `gen` <= *gen courrant* |

### 4.6.5 Procédure *AlterneHypothèse*

Pendant chacune des quatre lectures présentées ci-dessus, une procédure de génération des formes hypothétiques *AlterneHypothèse* est lancée pour chaque désinence. Cette procédure prend le radical du lemme de la forme requise ainsi que la désinence courante, lue dans le fichier *pdgm.xml*, et elle génère à l'aide de la procédure *AlterneRadical* les formes hypothétiques possibles pour chacune de ces combinaisons. Ces formes sont comparées à la production erronée $p$ et s'il y a une correspondance, les informations sur l'alternance sont ajoutée au message d'erreur.

Dans les formes hypothétiques, un radical $R$ concaténée à une désinence $D$ qui provoque dans cette combinaison une alternance $a$ ($D_a$) (par exemple *kočk* + *e*), peut être soit alterné d'après les règles de la réalisation des alternances ($alt(R,a)$), soit il peut rester intact ($R$). Par rapport à la réalisation des alternances, il y a au maximum 7 combinaisons possibles :

| *radical* | *désinence* | *exemple* |
|---|---|---|
| $R$ | $D$ | *kočk* + *ou* |
| $R$ | $D_a$ | *kočk* + *e* |
| $alt(R,a)$ | $D_a$ | *kočc* + *e* |
| $R$ | $D_{ab}$ | *vejc* + # |
| $alt(R,a)$ | $D_{ab}$ | *vajc* + # |
| $alt(R,b)$ | $D_{ab}$ | *vejec* + # |
| $alt(R,ab)$ | $D_{ab}$ | *vajec* + # |

La procédure *AlterneHypothèse* choisi le cas échéant pour une combinaison $R + D$ (aucune alternance à réaliser ; une alternance à réaliser ; deux alternances à réaliser) et elle génère les formes hypothétiques en les comparant à $p$. Le déroulement de la procédure, où les variables $a$ et $b$ contiennent l'identifiant de l'alternance (par exemple $e > a$, $á > a$, $k > c$, etc.), est le suivant :

1. **si** $h = R + D$ (aucune alternance à réaliser)
   **alors** passe à (2)
   **sinon** passe à (3)
2. **si** $R + D = p$
   **alors** succès, inscrit dans le message d'erreur : `alt <= sans`
3. **si** $h = R + D_{ab}$ (deux alternances à réaliser)
   **alors** passe à (4)
   **sinon** passe à (5)
4. **si** $R + D_{ab} = p$
   **alors** succès, inscrit dans le message d'erreur :

   `alt_1 <= a | valid_alt_1 <= ERR | alt_2 <= b | valid_alt_2 <= ERR`

   **sinon si** $alt(R,a) + D_{ab} = p$
   **alors** succès, inscrit dans le message d'erreur :

   `alt_1 <= a | valid_alt_1 <= OK | alt_2 <= b | valid_alt_2 <= ERR`

   **sinon si** $alt(R,b) + D_{ab} = p$
   **alors** succès, inscrit dans le message d'erreur :

   `alt_1 <= a | valid_alt_1 <= ERR | alt_2 <= b | valid_alt_2 <= OK`

   **sinon si** $alt(R,ab) + D_{ab} = p$
   **alors** succès, inscrit dans le message d'erreur :

   `alt_1 <= a | valid_alt_1 <= OK | alt_2 <= b | valid_alt_2 <= OK`

   **fin**
5. si $R + D_a = p$ (une alternance à réaliser)
   **alors** succès, inscrit dans le message d'erreur :

   `alt <= a | valid_alt <= ERR`

   **sinon si** $alt(R,a) + D_a = p$
   **alors** succès, inscrit dans le message d'erreur :

   `alt <= a | valid_alt <= OK`

   **fin**

Pour chaque forme hypothétique $h$ dans cette procédure, sa variante sans diacritique $dia(h)$ est comparée à la production erronée $dia(p)$, sans diacritique également. S'il y une correspondance et la production erronée n'exigeait pas la réalisation d'une alternance qui se

manifeste par l'ajout d'un signe diacritique (par exemple *jméno*×*jmen*, *dvůr*×*dvoře*), cette production est interprétée comme une erreur de diacritique, et l'attribut `dia` est spécifié par sa valeur dans le message d'erreur : `dia <= dia`.

Si la désinence employée pour la génération d'une forme hypothétique correspondante à la production erronée est spécifiée dans le fichier *pdgm.xml* comme une variante de registre (par exemple *ama* dans *pánama*), l'attribut `var` est spécifié par sa valeur dans le message d'erreur : `var <= reg`.

Si la désinence employée pour la génération d'une forme hypothétique correspondante à la production erronée est spécifiée dans le fichier *pdgm.xml* comme une variante de fonctionnelle (par exemple *ama* dans *pánama*), l'attribut `var` est spécifié par sa valeur dans le message d'erreur : `var <= fnct`.

### 4.6.6 Exemple

Prenons le déroulement du test morphologique pour une production erronée *sole* dans une tâche avec la forme requise $soli_{gen.sg.f.}$, dont le lemme ainsi que le radical est *sůl*. La génération de cette forme requise nécessite la réalisation de l'alternance *ů* > *o*. L'ensemble des formes hypothétiques *h* différentes, générées pendant la lecture locale est le suivant :

| radical | désinence | *h* | *p* | succes |
|---|---|---|---|---|
| *sůl* | *i* | ***sůli*** | *sole* | 0 |
| *sůl* | *i* | ***soli*** | *sole* | 0 |

Aucune correspondance entre *p* et *h* n'a pas trouvée. Pendant la lecture verticale, les formes hypothétiques suivantes sont générées :

| radical | désinence | *h* | *p* | succes |
|---|---|---|---|---|
| *sůl* | *#* | ***sůl*** | *sole* | 0 |
| *sůl* | *i* | ***sůli*** | *sole* | 0 |
| *sůl* | *i* | ***soli*** | *sole* | 0 |
| *sůl* | *í* | ***sůlí*** | *sole* | 0 |
| *sůl* | *í* | ***solí*** | *sole* | 0 |
| *sůl* | *ím* | ***sůlím*** | *sole* | 0 |
| *sůl* | *ím* | ***solím*** | *sole* | 0 |
| *sůl* | *ích* | ***sůlích*** | *sole* | 0 |
| *sůl* | *ích* | ***solích*** | *sole* | 0 |
| *sůl* | *emi* | ***sůlemi*** | *sole* | 0 |
| *sůl* | *emi* | ***solemi*** | *sole* | 0 |
| *sůl* | *emi* | ***sůlema*** | *sole* | 0 |
| *sůl* | *emi* | ***solema*** | *sole* | 0 |

Aucune correspondance entre *p* et *h* n'a été trouvée. À ce moment, la lecture horizontale interne peut commencer car aucune correspondance n'a pas été déterminée pendant la lecture locale. Ce traitement lit les désinences de même cas, genre et nombre au sein du même type paradigmatique. L'ensemble des formes hypothétiques générées avec ce désinences est égal à celui généré pendant le traitement local, car pour le type *kost* auquel appartient le mot *sůl*, il n'y a aucune autre désinence possible au génitif singulier que la désinence *–i* :

| radical | désinence | *h* | *p* | succes |
|---|---|---|---|---|
| *sůl* | *i* | ***sůli*** | *sole* | 0 |
| *sůl* | *i* | ***soli*** | *sole* | 0 |

Étant donné, qu'aucune correspondance entre *p* et *h* n'a été trouvée, la lecture horizontale externe peut commencer. Ce traitement consulte toutes les désinences dans le génitif singulier des types *píseň (ps) et tvář (ps_Re), pán, hrad (hd) et čas (hd_Re), muž (mz) et učeň (mz_Re), stroj (sj) et stupeň (sj_Re)* et il génère les formes hypothétiques suivantes :

| radical | désinence | *h* | *p* | succes |
|---|---|---|---|---|
| *sůl* | *ě* | ***sůlě*** | *sole* | 0 |
| *sůl* | *ě* | ***solě*** | *sole* | 0 |
| *sůl* | *e* | ***sůle*** | *sole* | 0 |
| *sůl* | *e* | ***sole*** | ***sole*** | 1 |
| *sůl* | *u* | ***sůlu*** | *sole* | 0 |
| *sůl* | *u* | ***solu*** | *sole* | 0 |
| *sůl* | *a* | ***sůla*** | *sole* | 0 |
| *sůl* | *a* | ***sola*** | *sole* | 0 |

Une correspondance a été trouvée pour la désinence *–e* avec la réalisation de l'alternance *ů* > *o*. Cette forme hypothétique a les interprétations suivantes :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
|---|---|---|---|---|---|---|---|---|---|
| HOR EXT | mz | e | gen | sg | m | ů > o | OK | erreur_gen | |
| HOR EXT | mz_Re | ě | gen | sg | m | ů > o | OK | erreur_gen | dia |
| HOR EXT | sj | e | gen | sg | i | ů > o | OK | erreur_gen | |
| HOR EXT | sj_Re | ě | gen | sg | i | ů > o | OK | erreur_gen | dia |
| HOR EXT | ps | ě | gen | sg | f | ů > o | OK | ident | dia |
| HOR EXT | ps_Re | e | gen | sg | f | ů > o | OK | ident | |

## 4.7 Filtrage des interprétations

Pendant cette étape, le message d'erreur est filtré pour réduire au minimum le nombre des interprétations possibles afin d'en retenir uniquement celles qui sont les plus plausibles. Cette réduction est effectuée en fonction du nombre d'attributs atteints par l'erreur qui détermine leur classement dans l'échelle de plausibilité pour un diagnostic.

Les différentes interprétations dans le message d'erreur sont d'abord classées dans des groupes *v* d'après la valeur de l'attribut `erreur`, qui spécifie les interprétations **locales**, **verticales**, **horizontales internes**, **horizontales externes** et les interprétations issues des traitements non morphologiques. À l'intérieur des groupes *v*, les différentes interprétations sont classées dans des groupes *g* en fonctions des valeurs des attributs `spec` et `dia`. Deux règles générales sont appliquées pendant le filtrage : (1) les interprétations avec une erreur de diacritique sont toujours placées derrières les interprétations équivalentes sans erreur de diacritique et (2) les interprétations avec moins d'attributs atteints par l'erreur sont placées devant celles qui en ont plus.

Pour chaque groupe *v*, ce n'est qu'un seul groupe *g*, celui placé le plus haut dans l'échelle de plausibilité, qui peut être inscrit dans le message filtré.

### 4.7.1 Exemple du groupement des interprétations

Dans l'exemple ci-dessus (*sůl* × *sole*), il n'y a dans le message d'erreur qu'un seul groupe *v*, celui des erreurs horizontales externes. Les différentes groupes *g* (g1, g2, g3, g4), qui sont établis sur la base des valeurs des attributs `spec` et `dia`, sont les suivants :

g1

| erreurs de genre | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
| HOR EXT | mz | e | gen | sg | m | ů > o | OK | erreur_gen | |
| HOR EXT | sj | e | gen | sg | i | ů > o | OK | erreur_gen | |

g2

| erreurs de genre et de diacritique | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
| HOR EXT | mz_Re | ě | gen | sg | m | ů > o | OK | erreur_gen | dia |
| HOR EXT | sj_Re | ě | gen | sg | i | ů > o | OK | erreur_gen | dia |

g3

| erreur de type paradigmatique | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
| HOR EXT | ps_Re | e | gen | sg | f | ů > o | OK | ident | |

g4

| erreur de type paradigmatique et de diacritique | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
| HOR EXT | ps | ě | gen | sg | f | ů > o | OK | ident | dia |

### 4.7.2 Filtrage des interprétations locales

Tous les groupes *g* dans le groupe *v* contenant toutes les interprétations locales sont parcourus un par un pour chaque point de l'échelle ci-dessous, qui établie la priorité des différentes interprétations.

| n° | spec | dia | description | cas | num | gen |
|---|---|---|---|---|---|---|
| **1** | *aucune valeur* | 1 | pas d'erreur de diacritique | 1 | 1 | 1 |
| **2** | *aucune valeur* | 0 | erreur de diacritique | 1 | 1 | 1 |

Pour une production erronée, il peut exister uniquement deux interprétations locales concurrentes. En effet, il y a des production erronées comme *dvorum* × $dvorům_{dat.pl.i}$ qui peuvent être interprétées soit comme une erreur dans la variante de registre, soit comme erreur de diacritique. La situation est la même pour les erreurs d'alternance et leur contrepartie avec une erreur de diacritique, par exemple $jmen_{gen.pl.n.}$ × *jmén*.

Si cette éventualité existe, ce n'est que l'interprétation sans l'erreur de diacritique qui est inscrite dans le message filtré, car elle est plus pertinente au niveau linguistique. S'il y a uniquement une interprétation, elle est inscrite automatiquement dans le message filtré. Si dans cette interprétation, il n'y a pas d'erreur de diacritique, le filtrage du message d'erreur est arrêté car l'interprétation locale est considérée comme la plus plausible. S'il y a une erreur de diacritique, le filtrage continue pour rechercher une éventuelle erreur verticale.

### 4.7.3 Filtrage des interprétations verticales

Tous les groupes *g* dans le groupe *v* contenant toutes les interprétations verticales sont parcourus un par un pour chaque point de l'échelle ci-dessous, qui établie la priorité des différentes interprétations.

| n° | spec | dia | description | cas | num | gen |
|---|---|---|---|---|---|---|
| **1** | erreur_num | 1 | erreur de nombre | 1 | 0 | 1 |
| **2** | erreur_gen | 1 | erreur de genre | 1 | 1 | 0 |
| **3** | erreur_cas | 1 | erreur de cas | 0 | 1 | 1 |
| **4** | erreur_num | 0 | erreur de nombre et de diacritique | 1 | 0 | 1 |
| **5** | erreur_gen | 0 | erreur de genre et de diacritique | 1 | 1 | 0 |
| **6** | erreur_cas | 0 | erreur de cas et de diacritique | 0 | 1 | 1 |
| **7** | erreur_cas_num | 1 | erreur de cas et de nombre | 0 | 0 | 1 |
| **8** | erreur_cas_gen | 1 | erreur de cas et de genre | 0 | 1 | 0 |
| **9** | erreur_cas_num | 0 | erreur de cas, de nombre et de diacritique | 0 | 0 | 1 |
| **10** | erreur_cas_gen | 0 | erreur de cas, de genre et de diacritique | 0 | 1 | 0 |
| **11** | erreur_cas_num_gen | 1 | erreur de cas, de num et de genre | 0 | 0 | 0 |
| **12** | erreur_cas_num_gen | 0 | erreur de cas, de num, de genre et de diacritique | 0 | 0 | 0 |

S'il existe un élément qui correspond à la spécification de l'erreur dans un point donné (la valeur correspondante de l'attribut `spec` et `dia`), il est inscrit dans le message filtré et le parcours des interprétations verticales et terminé ; sinon le point suivant est considéré.

Par exemple, si dans le groupe *g*, il y a trois interprétations verticales dont l'une est l'erreur de genre, l'autre erreur de cas et la troisième une erreur de genre et de diacritique, c'est l'erreur de genre, classée le plus haut dans l'échelle de pertinence des interprétations qui est inscrit dans le message filtré et les autres sont rejetées. Les valeurs dans les attributs caractérisant les alternances n'ont pas de rôle pour ce filtrage.

Des exemples d'erreurs définies à chaque point de l'échelle sont donnés ci-dessous :

| erreur de nombre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **1** | *městu* | dat | sg | n | 1 | *městům* | dat | **pl** | n | 1 |

| erreur de genre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **2** | *mladého* | acc | sg | m | 1 | *mladou* | acc | sg | **f** | 1 |

| erreur de cas | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **3** | *akci* | acc | sg | f | 1 | *akcí* | **inst** | sg | f | 1 |

| erreur de nombre et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **4** | *kočce* | dat | sg | f | 1 | *kockam* | dat | **pl** | f | **0** |

| erreur de genre et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **5** | *mladému* | dat | sg | m | 1 | *mlade* | dat | sg | **f** | **0** |

| erreur de cas et de genre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **6** | *kočce* | dat | sg | f | 1 | *koček* | **gen** | **pl** | f | 1 |

| erreur de cas et de nombre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **7** | *mladé* | dat | sg | f | 1 | *mladého* | **gen** | sg | **n** | 1 |

| erreur de cas, de genre et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **8** | *kočce* | dat | sg | f | 1 | *kocek* | **gen** | **pl** | f | **0** |

| erreur de cas, de nombre et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **9** | *mladé* | dat | sg | f | 1 | *mladeho* | **gen** | sg | **n** | **0** |

| erreur de cas, de genre et de nombre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **10** | *mladé* | dat | sg | f | 1 | *mladých* | **loc** | **pl** | **n** | 1 |

| erreur de cas, de genre, de nombre et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **11** | *mladé* | dat | sg | f | 1 | *mladych* | **loc** | **pl** | **n** | **0** |

### 4.7.4 Filtrage des interprétations horizontales internes

Les éléments dans le groupe *g* contenant toutes les interprétations horizontales internes sont parcourus un par un pour chaque point de l'échelle ci-dessous, qui établit la priorité des différentes interprétations.

| n° | spec | dia | description | cas | num | gen |
|---|---|---|---|---|---|---|
| **1** | ident | 1 | erreur de sous-type paradigmatique | 1 | 1 | 1 |
| **2** | ident | 0 | erreur de sous-type paradigmatique et de diacritique | 1 | 1 | 1 |

S'il existe un élément qui correspond à la spécification de l'erreur dans un point donné (la valeur correspondante de l'attribut spec et dia), il est inscrit dans le message filtré et le parcours des interprétations horizontales internes et terminé ; sinon le point suivant est considéré.

Des exemples des erreurs définies à chaque point de l'échelle sont donnés ci-dessous :

| erreur de sous-type paradigmatique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **1** | *letišt'* | gen | pl | n | 1 | *letiští* | gen | pl | n | 1 |

| erreur de sous-type paradigmatique et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **1** | *letišt'* | gen | pl | n | 1 | *letišti* | gen | pl | n | **0** |

### 4.7.5 Filtrage des interprétations horizontales externes

Les éléments dans le groupe *g* contenant toutes les interprétations horizontales externes sont parcourus un par un pour chaque point de l'échelle ci-dessous, qui établie la priorité des différentes interprétations.

| n° | spec | dia | description | cas | num | gen |
|---|---|---|---|---|---|---|
| **1** | ident | 1 | erreur de type paradigmatique | 1 | 1 | 1 |
| **2** | ident | 0 | erreur de type paradigmatique et de diacritique | 1 | 1 | 1 |
| **3** | erreur_gen | 1 | erreur de type paradigmatique et de genre | 1 | 1 | 0 |
| **4** | erreur_gen | 0 | erreur de type paradigmatique, de genre et de diacritique | 1 | 1 | 0 |

S'il existe un élément qui correspond à la spécification de l'erreur dans un point donné (la valeur correspondante de l'attribut `spec` et `dia`), il est inscrit dans le message filtré et le parcours des interprétations horizontales externes et terminé ; sinon le point suivant est considéré.

Des exemples des erreurs définies à chaque point de l'échelle sont donnés ci-dessous :

| erreur de type paradigmatique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **1** | *mláděti* | dat | sg | n | 1 | *mládi* | dat | sg | n | 1 |

| erreur de type paradigmatique et de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **2** | *mláděti* | dat | sg | n | 1 | *mladi* | dat | sg | n | **0** |

| erreur de type paradigmatique et de genre | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **3** | *mláděti* | dat | sg | n | 1 | *mládovi* | dat | sg | **m** | 1 |

| erreur de type paradigmatique, de genre ou de diacritique | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n° | *r* | cas | num | gen | dia | *p* | cas | num | gen | dia |
| **4** | *mláděti* | dat | sg | n | 1 | *mladovi* | dat | sg | **m** | **0** |

### 4.7.6 Exemple

Pour le message d'erreur spécifiant la production erronée *sole* (voir p. 156), il n'y aucune interprétation locale, verticale ni horizontale interne, le filtrage correspondant à ces groupes

n'est pas donc effectué. Sur la base de l'échelle des différentes interprétations au sein des erreurs horizontales, ce n'est que l'interprétation ci-dessous qui est retenue, car elle est classée devant les autres :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
|---|---|---|---|---|---|---|---|---|---|
| HOR EXT | ps_Re | e | gen | sg | f | ů > o | OK | ident | |

## 4.8 Formatage du diagnostic

La dernière étape de la procédure *Diagnostic* consiste en une « traduction » des interprétations retenues dans le message filtré en langue naturelle pour qu'elles puissent être publiées sur la plateforme apprenant afin de servir comme une explications des production erronées. Cette procédure est basée sur un principe simple de transfert des valeurs contenues dans le message d'erreur filtré dans des phrases préformatées avec des variables à instancier.

S'il y plusieurs interprétations de type différent (erreur locale, verticale, horizontale interne et externe) retenues dans le message filtré, le diagnostic est préfixé par le préambule ci-dessous et les différentes interprétations sont numérotées :

*Il y a plusieurs interprétations possibles de votre erreur :*

L'ordre de classement des différentes types d'erreurs est le suivant : les erreurs non morphologiques, les erreurs locales, les erreurs horizontales internes, les erreurs horizontales externes et les erreurs verticales. La motivation de ce classement est la préférence des interprétations qui ne varient pas dans la valeur de la catégorie de cas.

La traduction de chaque interprétation dans le message de diagnostic peut avoir au maximum trois composantes, classées dans cet ordre :

1. spécification des erreurs en fonction des valeurs des attributs `erreur` et `spec` portant les informations sur les écarts dans le **cas**, le **nombre**, le **genre** ou le **type paradigmatique** pour les erreurs avec une interprétation morphologique ; pour les erreurs sans interprétation morphologique, cette composante explique la raison pour laquelle la production a été exclue du traitement morphologique ;
2. spécification des erreurs dans l'**alternance** (simple ou double) en fonction des valeurs des attributs `alt`, `valid_alt`, `alt_1`, `valid_alt_1`, `alt_2`, `valid_alt_2` ;
3. spécification des erreurs de **diacritique** ou dans le choix d'une **variante** inappropriée d'une désinence en fonction des attributs `dia` et `var`.

La composante 1 est obligatoire pour toutes les productions erronées. Au moins une des composantes 2 ou 3 doit être présente dans la spécification des erreurs locales, car c'est dans cette partie que l'erreur est décrite (erreur d'alternance, de diacritique et de variante). Pour les erreurs verticales et horizontales, la composante 1 est obligatoire, car elle porte l'information principale sur cette erreur, les composantes 2 et 3 figurent dans le message de diagnostic si la production erronée contient des erreurs d'alternance, de diacritique ou de variante.

Pour les erreurs verticales et horizontales internes et externes, la première des composantes 2 et 3, qui figure dans le message de diagnostic, est préfixée par la glose suivante :

| |
|---|
| *Concernant votre forme,* |

### 4.8.1 Composante 1 : Traduction des attributs `erreur` et `spec`

Pour les interprétations non morphologiques, les traductions des valeurs de l'attribut `erreur` sont les suivantes :

| `erreur` | traduction |
|---|---|
| `ETRANG` | *Votre production contient des caractères qui n'appartiennent pas à l'alphabet tchèque.* |
| `NUMERIC` | *Votre production contient des chiffres. Vous aimez les maths ?* |
| `PUNCT` | *Votre production contient des signes non alphabétiques.* |

| `erreur` | traduction |
|---|---|
| `SIGNE UNIQUE` | *Vous devez plaisanter ...* |
| `COURT` | *Votre production est trop courte pour que nous puissions vous prendre au sérieux...* |
| `LONG` | *Votre production est trop longue ...* |
| `LEV` | *Je pense que vous avez saisi nímporte quoi ..* |
| `INCONNU` | *Nous n'avons pas réussi à interpréter votre erreur.* |

La composante 1 pour les **erreurs locales** est la suivante :

| `erreur` | traduction |
|---|---|
| `LOC` | *Aucune erreur de cas, de genre ni de nombre mais* |

La phrase n'est pas terminé car les autres composantes du message apportent les informations spécifiant cette erreur en fonction des autres attributs atteints par l'erreur.

La composante 1 pour les **erreurs horizontales internes** est ci-dessous :

| erreur | traduction |
|---|---|
| HOR INT | *Le cas, le genre et le nombre sont corrects mais la désinence* X *ne peut pas être utilisée avec ce mot.* |

La valeur de la variable X est instanciée en fonction de la désinence inscrite dans l'interprétation de cette erreur (la valeur de l'attribut des dans le message d'erreur filtré).

Les différentes formes de la composante 1 pour les **erreurs horizontales externes** sont présentées ci-dessous :

| erreur | spec | traduction |
|---|---|---|
| HOR EXT | ident | *Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le mot d'après le type* X. |
| HOR EXT | erreur_gen | *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* X. |

La valeur de la variable X est instanciée en fonction de l'identifiant du type paradigmatique dans l'interprétation de cette erreur. La valeur de l'attribut pdgm dans le message d'erreur est converti en la valeur correspondant de l'identifiant mnémotechnique du type modèle, par exemple zn_Re est converti en *žena*.

Les différentes formes de la composante 1 pour les **erreurs verticales** sont ci-dessous :

| erreur | spec | traduction |
|---|---|---|
| VERT | erreur_cas | *Le nombre et le genre sont corrects mais il y a une erreur de cas.* |
| VERT | erreur_num | *Le cas et le genre sont corrects mais vous avez mis une forme de* X *au lieu de* Y. |
| VERT | erreur_gen | *Le cas et le nombre sont corrects mais vous avez mis une forme de* X *au lieu de* Y. |
| VERT | erreur_cas_num | *Le cas et le nombre sont incorrects.* |
| VERT | erreur_cas_gen | *Le cas et le genre sont incorrects.* |
| VERT | erreur_cas_num_gen | *Le cas, le genre et le nombre sont incorrects.* |

Pour les erreurs de nombre et de genre, les valeurs des variables `X` et `Y` sont instanciées par les valeurs des attributs correspondants. Pour une erreur de cas, il n'y a qu'un seul message, même s'il y a plusieurs interprétations possibles pour les désinences homonymes au sein d'un paradigme et d'un nombre donné.

Dans chaque type d'erreur qui ont une interprétation morphologique et qui ont été traduites dans la première phase du formatage du message d'erreur, il peut y avoir des erreurs supplémentaires liées à la réalisation des alternances, au diacritique et au choix des variantes inappropriées des désinences. Pour les erreurs locales, cette information fait partie intégrante de la composante principale du message de diagnostic. Dans les autres types d'erreurs, cette information est adjointe à cette composante comme une spécification de l'erreur principale.

### 4.8.2 Composante 2 : Traduction des attributs concernant les alternances

Les erreurs d'alternance sont marquées dans une interprétation par les valeurs des attributs `alt`, `valid_alt`, `alt_1`, `valid_alt_1`, `alt_2`, `valid_alt_2`.

Les différentes formes de la composante 2 sont les suivantes :

| `alt` | `valid_alt` | traduction |
|---|---|---|
| `X` | `ERR` | *vous n'avez pas effectué l'alternance* `X` |

| `alt_1` | `valid_alt_1` | `alt_2` | `valid_alt_2` | traduction |
|---|---|---|---|---|
| `X` | `ERR` | `Y` | `OK` | *bien que l'alternace* `X` *soit correcte, l'alternance* `Y` *n'a pas été effectuée* |
| `X` | `OK` | `Y` | `ERR` | *l'alternace* `X` *est correcte, l'alternance* `Y` *n'a pas été effectuée* |
| `X` | `ERR` | `Y` | `ERR` | *les deux alternances nécessaires* `X` *et* `Y` *n'ont pas été effectuées* |

Les variables `X` et `Y` contiennent les identifiants des alternances tels qu'ils sont inscrits dans l'interprétation de l'erreur, par exemple *k > č, h > ž, ou > u*, etc.

### 4.8.3 Composante 3 : Traduction des attributs `dia` et `var`

La troisième composante du message de diagnostic contient les informations sur les éventuelles erreurs de diacritique ou de variantes de la désinence

Les erreurs de diacritique sont signalées dans une interprétation par la valeur de l'attribut `dia` :

| dia | traduction |
|---|---|
| dia | *vous avez fait une erreur de diacritique* |

L'attribut `var` porte l'information sur le caractère de la désinence au cas où la désinence employée appartient au registre non standard ou qu'elle est une variante fonctionnelle qui n'est pas acceptable au sein d'une certaine tâche. La valeur de cet attribut est traduite de cette façon :

| var | traduction |
|---|---|
| reg | *la désinence* X *est une variante parlée* |
| fnct | *la désinence* X *est une variante qui ne peut pas être employée ici* |

La valeur de la variable `X` est instanciée en fonction de la désinence inscrite dans l'interprétation de l'erreur.

### 4.8.4 Exemples

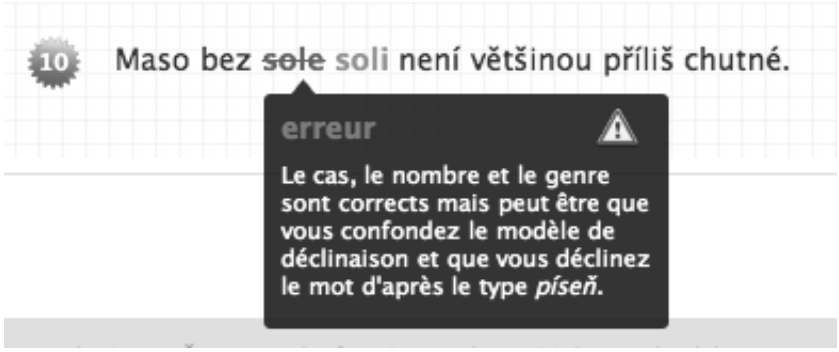
(1) *Maso bez* ****sole*** *není většinou příliš chutné*
soli gen.sg.f
viande sans sel n'est pas d'habitude très bonne
'Habituellement, la viande sans sel n'est pas très bonne' (1.164.16)

Le message filtré contenant l'interprétation de la production erronée dans la tâche (1), contient les attributs et les valeurs suivantes :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec | dia |
|---|---|---|---|---|---|---|---|---|---|
| HOR EXT | ps_Re | e | gen | sg | f | ů > o | OK | ident | |

La traduction de cette interprétation est la suivante : *« Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le mot d'après le type* píseň. *»*

Le message de diagnostic, tel qu'il est affiché sur la plateforme apprenant de CETLEF, a cette forme :



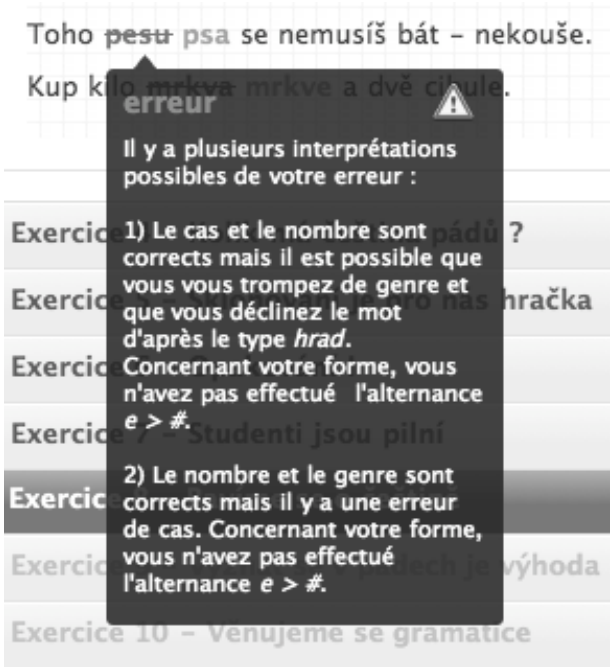

(2) *Toho* ***pesu*** *se* *nemusíš* *bát* – *nekouše*
psa gen.sg.m
De ce chien *refl* ne dois pas avoir peur ne mord pas
'Tu ne dois pas avoir peur de ce chien – il ne mord pas' (1.30.38)

Le message filtré contenant l'interprétation de la production erronée dans la tâche (2) contient les attributs et les valeurs suivantes :

| erreur | pdgm | term | cas | num | gen | alt | valid_alt | spec |
|---|---|---|---|---|---|---|---|---|
| VERT | pn | u | dat | sg | m | e > # | ERR | erreur_cas |
| VERT | pn | u | loc | sg | m | e > # | ERR | erreur_cas |
| HOR EXT | hd | u | gen | sg | i | e > # | ERR | erreur_gen |

La traduction de ces interprétations est la suivante : *« Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas et le nombre sont corrects mais il est possible que vous vous trompez de genre et que vous déclinez le mot d'après le type* hrad. *Concernant votre forme, vous n'avez pas effectué l'alternance* e > #. *2) Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous n'avez pas effectué l'alternance* e > #. *»*

Le message de diagnostic, tel qu'il est affiché sur la plateforme apprenant de CETLEF, a cette forme :


Toho pesu psa se nemusíš bát – nekouše.
Kup kilo mrkva mrkve a dvě cibule.
erreur
Il y a plusieurs interprétations possibles de votre erreur :
1) Le cas et le nombre sont corrects mais il est possible que vous vous trompez de genre et que vous déclinez le mot d'après le type *hrad*.
Concernant votre forme, vous n'avez pas effectué l'alternance *e > #*.
2) Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous n'avez pas effectué l'alternance *e > #*.
Exercice 10 – Věnujeme se gramatice

# Chapitre 5

# Évaluation

Dans ce chapitre, nous présentons la mise en service expérimentale de la plateforme CETLEF et les données authentiques recueillies par deux enquêtes différentes. Ces enquêtes donnent un aperçu général des erreurs qui sont commises par les apprenants et créent un point de départ pour des recherches ultérieures. La représentativité de ces données doit être cependant considérée avec prudence, notamment du point de vue du nombre d'apprenants qui ont participé à ces enquêtes (30 pour l'enquête publique et 13 pour l'enquête préliminaire).

Dans ce qui suit, nous allons présenter les données collectées et nous allons examiner la couverture et l'adéquation du diagnostic automatique. Nous allons également nous arrêter sur les productions erronées pour lesquelles le diagnostic n'a trouvé aucune interprétation et nous essaierons d'identifier les causes de cet échec.

## 5.1 Enquête publique

L'enquête publique est composée de 16 exercices publiés sur la plateforme apprenant de CETLEF. L'identifiant de cette enquête, tel qu'il est marqué dans les identifiants des tâches dans ce travail est 1.

### 5.1.1 Présentation de l'enquête

Les exercices contiennent 162 tâches (10 tâches par exercice en moyenne). Ils ont été ordonnées d'après l'introduction des différents cas dans la définition des niveaux A1 et A2 du Cadre européen pour le tchèque : accusatif singulier (25 tâches) et pluriel (15 tâches), génitif singulier (22 tâches) et pluriel (19 tâches), nominatif pluriel (11 tâches), locatif singulier (18

tâches) et pluriel (9 tâches), datif singulier (22 tâches) et pluriel (7 tâches), vocatif singulier (9 tâches) et pluriel (2 tâches), instrumental singulier (17 tâches) et pluriel (5 tâches).

La couverture des types paradigmatiques dans les tâches, respectant dans les grandes lignes la fréquence des types paradigmatique en général, est la suivante : *žena* (26 tâches), *hrad* (24 tâches), *pán* (19 tâches), *město* (14 tâches), *muž* (12 tâches), *stroj* (10 tâches), *mladý* (8 tâches), *růže* (6 tâches), *předseda* (5 tâches), *stavení* (5 tâches), *moře* (5 tâches), *kost* (4 tâches), *kuře* (4 tâches), *soudce* (3 tâches), *píseň* (3 tâches), *jarní* (2 tâches), *matčin* (1 tâche), *otcův* (1 tâche) et les différents types de la déclinaison pronominale (31 tâches).

Concernant les alternances, l'enquête contient 27 tâches avec les alternances vocaliques ($3 \times \# > e$, $11 \times e > \#$, $2 \times ou > u$, $1 \times á > a$, $1 \times é > e$, $1 \times í > ě$, $2 \times í > á$, $6 \times ů > o$,) et 13 tâches avec les alternances consonantiques ($1 \times c > č$, $1 \times g > z$, $1 \times h > z$, $7 \times k > c$, $1 \times k > č$, $2 \times r > ř$). Dans 9 tâches, il était nécessaire d'effectuer une double alternance ($1 \times e > \#$ et $c > č$, $1 \times e > \#$ et $k > c$, $3 \times e > \#$ et $ň > n$, $1 \times e > a$ et $\# > e$, $1 \times ů > o$ et $h > ž$ et 2 tâches $ů > o$ et $n > ň$).

### 5.1.2 Caractéristique des apprenants et leurs scores

Sur les quatre mois de mise en ligne de CETLEF[1], 43 apprenants ont créé leur compte utilisateur et 30 parmi eux ont effectué au moins un exercice. Par définition, la plateforme apprenant est accessible à tous les apprenants de tous les niveaux, il est donc évident que se sont que quelques uns parmi eux qui ont terminé tous les exercices au sein de cette enquête. Pour des enquêtes effectuées à l'avenir, il sera nécessaire d'atteindre un certain nombre d'apprenants qui auraient terminé toute l'enquête et de restreindre l'analyse uniquement à cet ensemble pour pouvoir disposer de données plus équilibrées.

Le nombre d'apprenants par exercice effectué au sein de l'enquête publique, est présenté dans la table ci-dessous :

| n° exercice | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| apprenants | 30 | 23 | 18 | 13 | 10 | 10 | 8 | 7 | 5 | 4 | 4 | 4 | 4 | 4 | 3 | 3 |

La progression dans les exercices reflète fidèlement l'auto-évaluation de la maîtrise du tchèque par les apprenants qu'ils indiquent dans le formulaire d'enregistrement. Parmi ceux qui ont terminé le premier exercice, il y en a 8 qui estiment leur niveau du tchèque inférieur au niveau A1, 5 apprenants s'estiment capable de maîtriser le tchèque au niveau A1, 5 au niveau A2, 7 au niveau B1, 2 au niveau B2, 2 au niveau C1 et 1 au niveau C2. Les deux derniers

[1] La mise en ligne de la plateforme apprenant CETLEF a été faite le 11 juin 2008 et la collecte des données utilisées dans cette évaluation a été réalisée le 10 septembre 2008.

exercices ont été effectués uniquement par trois apprenants qui estiment leur maîtrise du tchèque correspondante aux niveaux C1 et C2.

Le score moyen[2] dans tous les exercices est égal à 70, c'est à dire qu'en moyenne, il y a 3 erreurs dans un exercice de 10 tâches. Les scores moyens des apprenants par exercice sont les suivants :

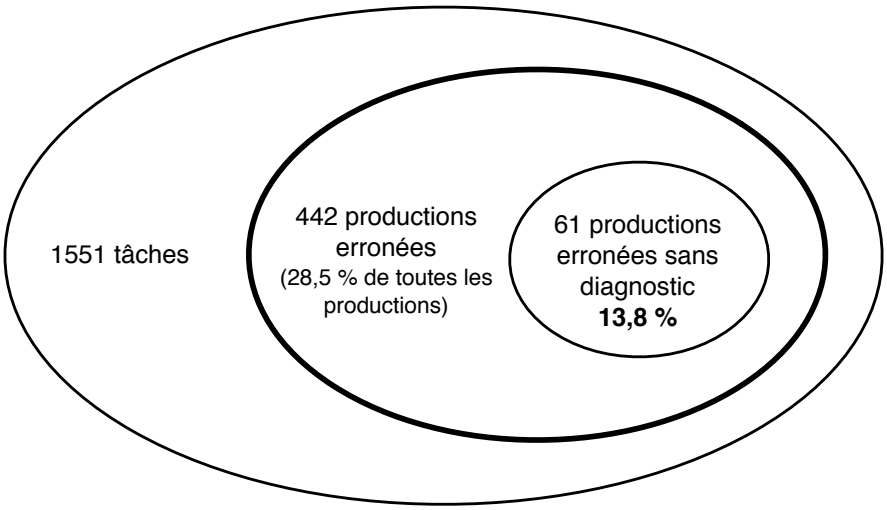
| n° exercice | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| score | 63,5 | 68,0 | 63,9 | 54,8 | 70,0 | 74,0 | 56,3 | 71,4 | 62 | 80 | 85 | 80 |

| n° exercice | 13 | 14 | 15 | 16 |
|---|---|---|---|---|
| score | 65 | 82,5 | 76,7 | 66,7 |

Le score dans les différents exercices doit être considéré par rapport au nombre décroissant des apprenants qui les effectuent. Les scores faibles se font de plus en plus rares avec le nombre décroissant des apprenants qui effectuent les exercices, car les apprenants qui n'ont pas de bons résultats abandonnent l'enquête.

### 5.1.3 Présentation des erreurs recueillies

Dans l'enquête publique, 159 exercices ont été envoyés à la correction. Au sein de ces exercices, 1551 tâches ont été effectuées. Le nombre de productions erronées dans ces tâches est égal à 442 (28,5 % de toutes les productions). Pour 61 productions erronées (13,8 % de toutes les productions erronées), le diagnostic automatique n'a pas réussi d'identifier la nature de l'erreur. Une représentation schématique de cette situation est proposée ci-dessous :




[2]Le score est le pourcentage des tâches effectuées sans erreur.

Parmi toutes les productions erronées, il y en a 373 (84,4 %) qui ont été diagnostiquées comme une erreur avec une ou plusieurs interprétations morphologiques. Le nombre de productions erronées, diagnostiquées comme un des types d'erreurs établies d'après les restrictions sur les formes hypothétiques (ou une combinaison possible de deux types différents), est représentée dans la table ci-dessous :

| *attribut* `erreur` | *type de l'erreur* | *productions* | *%* |
|---|---|---|---|
| LOC | erreur locale | 123 | 33,0 |
| VERT | erreur verticale | 145 | 38,9 |
| HOR INT | erreur horizontale interne | 2 | 0,5 |
| HOR EXT | erreur horizontale externe | 19 | 5,1 |
| LOC ou VERT | erreur locale ou verticale | 39 | 18,5 |
| VERT ou HOR INT | erreur verticale ou horizontale interne | 1 | 0,2 |
| VERT ou HOR EXT | erreur verticale ou horizontale externe | 40 | 10,7 |
| HOR INT ou HOR EXT | erreur horizontale interne ou externe | 4 | 1,1 |
| INCONNU | sans diagnostic | 61 | |

Le nombre d'erreurs qui ont été diagnostiquées comme ayant une interprétation non morphologique (8 en total, 1,8 % de toutes les productions erronées) est présenté dans le tableau ci-dessous :

| attribut `erreur` | type de l'erreur | productions |
|---|---|---|
| LONG | production trop longue | 1 |
| COURT | production trop courte | 5 |
| LEV | distance de Levehnstein | 2 |

#### 5.1.3.1 Erreurs locales

Dans l'enquête publique, les erreurs les plus fréquentes parmi les erreurs locales sont les **erreurs de diacritique** (93 productions erronées, 75,6 %). Pour les productions erronées qui ne contiennent aucun des signes diacritiques nécessaires (par exemple *kolac* au lieu de *koláč*, *prekvapeni* au lieu de *překvapení*), ceci doit être nécessairement imputé à l'interface web de l'application CETLEF et les restrictions techniques du côté des apprenants (manque de clavier tchèque). Il est fréquent que les apprenants oublient de marquer la diacritique ou, au cas où ils n'ont pas la possibilité de la saisir à l'aide de leur clavier, ils n'utilisent pas le clavier virtuel qui leur est proposé à cet effet.

Il y a cependant des erreurs de diacritique qui contiennent des signes inadéquats : les apprenants ont la possibilité de saisir les caractères diacrités, mais ils font des erreurs dans

leur emploi, par exemple *véže* au lieu de *věže*, *profesoří* au lieu de *profesoři*, *počitatčich* au lieu de *počitatčích*, etc.

Les **erreurs d'alternance** sont le second type d'erreurs locales le plus fréquent (20 productions erronées, 16,3 %). Il est intéressant de constater que ce sont uniquement les formes requises avec les alternances vocaliques quantitatives qui étaient à l'origine de ces erreurs (*kráv* au lieu de *krav*, *jmén* au lieu de *jmen*, *díl* au lieu de *děl*, *houb* au lieu de *hub*, *nůžem* au lieu de *nožem*, *penízmi* au lieu de *penězmi*, *smlouv* au lieu de *smluv*, etc.). Le diagnostic généré pour une telle erreur, par exemple pour la production *smlouv* au lieu de *smluv*, est le suivant : *« Aucune erreur de cas, de genre ni de nombre mais vous n'avez pas effectué l'alternance* ou > u. *»*

Aucune des tâches avec des alternances consonantiques n'a été effectuée avec erreur, sauf la production erronée *profesori* au lieu de *profesoři*, qui a été diagnostiquée comme une erreur d'alternance, mais qui pourrait aussi bien être considérée comme une erreur de diacritique.

Parmi les autres occurrences des erreurs locales, les productions dans les tâches (1) et (2)[3] ont été diagnostiquées comme une **erreur de variante de registre**.

(1) *Bez* ***vejcí** *bábovku neuděláš.*
vajec gen.pl.f.
Sans oeufs kouglof ne fais pas
'Tu ne peux pas faire un kouglof sans oeufs' (1.42.18)
**Diagnostic** : *Aucune erreur de cas, de genre ni de nombre mais la désinence* -í *est une variante parlée.*

(2) *Tehdy byl člověk ještě divoký a nedalo se s* ***nim** *žít.*
ním inst.sg.m.
Jadis était homme encore sauvage et ne pouvait pas avec lui vivre
'Jadis, l'homme était encore sauvage et on ne pouvait pas vivre avec lui' (1.143.18)
**Diagnostic** : *Aucune erreur de cas, de genre ni de nombre mais la forme* nim *est une variante parlée.*

Dans la tâche (3), la production erronée a été diagnostiquée comme une **erreur de variante fonctionnelle** qui n'est pas acceptable dans le contexte donné car ce n'est que la variante fonctionnelle *–e* du locatif singulier du substantif *rok* qui peut être employée pour une expression adverbiale de temps.

(3) *Narodila jsem se v* ***roku** *devatenáct set osmdesát pět.*
roce loc.sg.i.
Je suis née en an mille neuf cents quatre-vingt cinq

[3]La numérotation des tâches dans cette section est utilisée uniquement pour faciliter la lecture. L'identifiant de chaque tâche est affichée à côté de sa traduction française. Nous rappelons son format : n° enquête.identifiant de la tâche.identifiant de ou des apprenants ayant fourni la production erronée affichée dans la tâche, par exemple 1.83.24|27|65. Les identifiants des apprenants ayant produit la même erreur sont séparés par le signe |.

'Je suis née en mille neuf cents quatre-vingt cinq' (1.83.24|27|65)

**Diagnostic** : *Aucune erreur de cas, de genre ni de nombre mais la désinence* -u *est une variante qui ne peut pas être employée ici.*

#### 5.1.3.2 Erreurs verticales

Parmi les erreurs verticales, les productions les plus fréquentes sont les **erreurs de cas** (71 productions, 49,0 %) qui reflètent probablement des dysfonctionnements au niveau syntaxique. Les erreurs les plus courantes sont celles où la production erronée a été laissée au nominatif singulier, par exemple dans les tâches (4) ou (5).

(4) *Kup kilo* ***mrkev*** *a dvě cibule.*
mrkve gen.sg.f
Achète kilo de carrotte et deux oignons
'Achète un kilo de carottes et deux oignons' (1.31.51|60|63)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas.*

(5) *Do* ***místnost*** *vstoupil nějaký člověk.*
místnosti gen.sg.f
Dans pièce est entré un homme
'Un homme est entré dans la pièce' (1.29.24|43|63)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas.*

Parmi les erreurs verticales de cas où la production a une autre forme que le nominatif singulier, nous pouvons citer par exemple les productions dans les tâches (6), (7) ou (8).

(6) *Po semináři jdeme do* ***kavárnu****.*
kavárny gen.sg.f
Après seminaire allons au café
'Nous allons au café après le séminaire' (1.29.23|63)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas.*

(7) *Bez* ***svačinu*** *mám večer hlad.*
svačiny gen.sg.f.
Sans goûter j'ai soir faim
'Sans avoir goûté, j'ai faim le soir' (1.24.21|43|51)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas.*

(8) *Fotbalový klub Slavia Praha má mnoho* ***fanoušky****.*
fanoušků gen.pl.m.
Football club Slavia Praha a beaucoup de fans
'Le club de football Slavia Praha a beaucoup de fans' (1.39.63)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas.*

Parmi les erreurs verticales de cas combinées avec un autre type d'erreur, la production *fanoušeky* au lieu de *fanoušků* dans la tâche (8) a été diagnostiquée comme une **erreur de cas** (l'accusatif pluriel au lieu de génitif pluriel) **et d'alternance** *e* > # (la forme avec l'alternance réalisée correctement serait *fanoušky*).

La production erronée dans la tâche (9) a été diagnostiquée comme une **erreur de cas** (le nominatif pluriel au lieu de l'accusatif pluriel) **et de diacritique**, de la même façon que la production dans la tâche (10) où il s'agit d'une erreur de cas (la forme du génitif pluriel au lieu du nominatif pluriel) avec une erreur de diacritique.

(9) *Dobře znám své* ****spolužací***.
spolužáky acc.pl.m.
Bien connais mes collègues
'Je connais bien mes collègues de classe' (1.16.24)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous avez fait une erreur de diacritique.*

(10) *Petr a Pavel jsou nerozluční* ****přatel***.
přátelé nom.pl.m.
Pierre et Paul sont inséparables amis
'Pierre et Paul sont des amis inséparables' (1.52.24)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous avez fait une erreur de diacritique.*

Le second groupe d'erreurs verticales les plus fréquentes sont les **erreurs de nombre** (19 productions avec un remplacement d'une forme de pluriel par une forme de singulier ; 8 productions avec un remplacement d'une forme de singulier par une forme de pluriel), par exemple les productions dans les tâches (11) ou (12).

(11) *Večer Honza nepracuje, ale poslouchá* ****písničku***.
písničky acc.pl.f.
Soir Jean ne travaille pas mais écoute chansons
'Le soir, Jean ne travaille pas mais il écoute des chansons' (1.17.21|27|53|63)
**Diagnostic** : *Le cas et le genre sont corrects mais vous avez mis une forme de singulier au lieu de pluriel.*

(12) *Mařenka nese plný koš* ****houby***.
hub gen.pl.f.
Marie porte plein panier de champignons
'Marie porte un panier plein de champignons' (1.36.21)
**Diagnostic** : *Le cas et le genre sont corrects mais vous avez mis une forme de singulier au lieu de pluriel.*

Les limites du diagnostic automatique actuel se voient dans l'interprétation de la production erronée dans la tâche (13).

(13) *Vážený pane* ***soudci**, jsem nevinný.*
soudce voc.sg.m.
Respecté monsieur juge suis innocent
'Monsieur le juge, je suis innocent' (1.129.18|65)
**Diagnostic** : *Le cas et le genre sont corrects mais vous avez mis une forme de pluriel au lieu de singulier.*

La production *soudci* est interprétée automatiquement comme une **erreur de nombre** (le vocatif pluriel de *soudce* est effectivement *soudci*), il y a cependant une interprétation morphologique plus probable qui n'a pas été trouvée par le diagnostic : la forme *soudci* pourrait être interprétée comme une erreur de type paradigmatique (*muž* au lieu de *soudce*) au sein des paradigmes masculins animés car le vocatif singulier du type mou *muž* est exprimé par la désinence *–i*.

Le diagnostic ne génère pas cette forme hypothétique car il y a une restriction sur les erreurs horizontales externes qui définit que les types paradigmatiques parcourus pour cette génération doivent avoir la forme du nominatif singulier correspondante au nominatif singulier de la forme requise (pour les types consonantiques, la forme du nominatif peut être terminée par une consonne mixte, dure ou molle ; pour des types vocaliques, elle peut être terminé par *–a*, *–e/–ě*, *–í* ou *–o*). Avec cette restriction, ce sont donc uniquement les types paradigmatiques avec la désinence *–e* au nominatif singulier qui sont parcourus pour la génération des formes hypothétiques et aucun de ces types n'a la désinence *–i* dans le vocatif singulier. Pour prendre compte de l'interprétation plus probable de l'emploi d'une désinence d'un type paradigmatique du même genre mais sans une correspondance dans le nominatif singulier, il serait intéressant dans les versions futures du diagnostic d'enlever cette restriction et de parcourir tous les types paradigmatiques au sein des types du même genre grammatical.

#### 5.1.3.3 Erreurs horizontales internes

Uniquement deux occurrences d'une production erronée ont été diagnostiquées comme des erreurs horizontales internes, voir la tâche (14) :

(14) *Náš dům má sedm* ***poschody**.*
poschodí voc.sg.m.
Notre immeuble a sept étages
'Notre immeuble a sept étages' (1.20.33|52)
**Diagnostic** : *Le cas, le genre et le nombre sont corrects mais la désinence -ý ne peut pas être utilisée avec ce mot. Concernant votre forme, vous avez fait une erreur de diacritique.*

En effet, la forme *poschody* est interprétée comme une confusion des sous-types paradigmatiques *stavení* et *úterý* avec une erreur de diacritique dans la forme de l'accusatif pluriel.

Nous pouvons voir aisément qu'un diagnostic de ce type, bien qu'adéquat par rapport aux règles établies dans la procédure, n'est pas trop pertinent pour l'apprenant car le sous-type *úterý* est un sous-type fermé, d'après lequel est décliné uniquement un seul lexème.

#### 5.1.3.4 Erreurs horizontales externes

Contrairement aux erreurs horizontales internes, le diagnostic des erreurs horizontales externes se montre plutôt satisfaisant. Les plus nombreuses sont les **erreurs de type paradigmatique et de genre**, causées par la confusion du type paradigmatique en fonction des ambiguïtés qui peuvent exister dans l'attribution de ce type à un mot en fonction de sa forme.

Dans les tâches (15) et (16), il s'agit d'une confusion entre les types *předseda* et *žena*.

(15) *Bez* ***turist*** *je v Praze klid*
turistů gen.sg.m
sans touristes est à Prague calme
'Sans touristes, Prague est calme' (1.37.24|27)
**Diagnostic** : *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* žena.

(16) *Hanička se líbí* ***Jirce***.
Jirkovi gen.sg.f
Hanička *refl* plaît à Jirka
'Hanička plaît à Jirka' (1.107.18|65)
**Diagnostic** : *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* žena.

Dans la tâche (17), il s'agit d'une confusion entre les types *hrad* et *kost*.

(17) *Jdeme s kamarády na výlet do* ***lesi***.
lesa gen.sg.i.
Allons avec camarades à ballades à forêt
'Nous allons nous balader dans la forêt avec les camarades' (1.66.22)
**Diagnostic** : *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* kost.

Dans les tâches (18) et (19), la production erronée (produite d'ailleurs par un seul apprenant) reflète une confusion entre les types *hrad* et *píseň*.

(18) *Šli jsme* ***lesí*** *a hledali houby*
lesem inst.sg.i.
Marchions forêt et cherchions champignons
'Nous marchions dans la forêt et nous cherchions des champignons' (1.135.65)
**Diagnostic** : *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* píseň.

(19) *Pod* ***stolí*** *leží* *náš* *pejsek* *Dášenka*
stolem inst.sg.i.
Sous table est allongé notre chien Dášenka
'Sous la table est allongé notre chien Dášenka' (1.134.65)
**Diagnostic** : *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* píseň.

Pour les productions dans les tâches (18) et (19), il pourrait également s'agir d'une confusion entre le type *hrad* et *kost*, comme dans la tâche (17), car les deux types paradigmatiques (tous les deux consonantiques, ce qui imposé par la restriction sur les erreurs horizontales externes) ont la désinence *–í* dans l'instrumental singulier.

Cette interprétation est effectivement retrouvée par le diagnostic et elle est inscrite dans le message d'erreur avant son formatage, mais elle n'est pas retenue car une seule interprétation peut être affichée. Il serait dans ce cas peut être plus adéquat d'annoncer uniquement une erreur de genre, sans spécifier le type paradigmatique auquel appartient la désinence identifiée dans la production erronée.

Une production diagnostiquée comme une **erreur de type paradigmatique et de genre** (confusion du type *píseň* avec le type masculin *hrad*) ainsi qu'une erreur **d'alternance** *e* > # a été produite par deux apprenants dans la tâche (20).

(20) *Kup* *kilo* ***mrkevu*** *a* *dvě* *cibule.*
mrkve gen.sg.f
Achète kilo de carrotte et deux oignons
'Achète un kilo de carottes et deux oignons' (1.31.21|52)
**Diagnostic** : *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* hrad. *Concernant votre forme, vous n'avez pas effectué l'alternance* e > #.

Deux exemples de productions qui ont été diagnostiquées comme des **erreurs de type paradigmatique** au sein du même genre, sont donnés dans les tâches (21) et (22).

(21) *Maso* *bez* ***sole*** *není* *většinou* *příliš* *chutné*
soli gen.sg.f.
viande sans sel n'est pas habituellement très bonne
'Habituellement, la viande sans sel n'est pas très bonne' (1.164.16|18)
**Diagnostic** : *Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le mot d'après le type* píseň.

(22) *V* *zoologické* *zahradě* *je* *mnoho* *zvláštních* ***zvíří.***
zvířat gen.pl.n.
Dans zoologique jardin est beaucoup particuliers animaux
'Au jardin zoologique, il y a beaucoup d'animaux particuliers' (1.34.19|30)
**Diagnostic** : *Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le mot d'après le type* moře.

Dans la tâche (21), il s'agit d'une confusion entre les types *kost* et *píseň* (la désinence dans le génitif singulier du type *píseň* est –e). Dans la tâche (22), il s'agit d'une confusion entre les types *kuře* et *moře* (le génitif pluriel du type moře est exprimé par la désinence *–í*).

#### 5.1.3.5 Erreurs locales ou verticales

La majorité des productions erronées pour lesquelles le diagnostic propose soit une interprétation locale, soit une interprétation verticale, sont des **erreurs de diacritique** en ce qui concerne l'interprétation locale. Le diagnostic peut être jugé comme adéquat si la seconde interprétation – l'interprétation verticale – est **une erreur de cas ou de nombre** sans contenir une erreur de diacritique, par exemple dans la tâche (23) ou (24).

(23) *Každou neděli dávám ženě* ***růží.**
růži acc.sg.f.
Chaque dimanche donne à femme rose
'Je donne une rose à ma femme chaque dimanche' (1.4.24)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Aucune erreur de cas, de genre ni de nombre mais vous avez fait une erreur de diacritique. 2) Le nombre et le genre sont corrects mais il y a une erreur de cas.*

(24) *Petr a Pavel jsou nerozluční* ***přátele.**
přátelé nom.pl.m.
Pierre et Paul sont inséparables amis
'Pierre et Paul sont des amis inséparables' (1.52.16|18|65)
**Diagnostic** : *Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous avez fait une erreur de diacritique.*

Dans la tâche (23), l'interprétation verticale est une erreur de cas causée par l'emploi de l'instrumental singulier au lieu de l'accusatif singulier. Dans la tâche (24), l'interprétation verticale est une erreur de cas causée par l'emploi de l'accusatif pluriel au lieu du nominatif pluriel.

Le diagnostic est par contre plutôt redondant quand l'interprétation verticale contient une erreur de diacritique comme pour la production *ruzi* au lieu de *růži* dans la tâche (25).

(25) *Každou neděli dávám ženě* ***ruzi.**
růži acc.sg.f.
Chaque dimanche donne à femme rose
'Je donne une rose à ma femme chaque dimanche' (1.4.24)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Aucune erreur de cas, de genre ni de nombre mais vous avez fait une erreur de diacritique. 2) Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous avez fait une erreur de diacritique.*

L'interprétation verticale de la forme *ruzi* dans la tâche (25) est qu'il s'agit d'une **erreur de cas et de diacritique** : il peut ainsi s'agir du datif, de l'accusatif, du locatif ou de l'instrumental singulier. Il est évident que cette précision dans le diagnostic n'apporte pas une information trop pertinente pour l'apprenant et elle peut même avoir un effet négatif car elle paraît confuse.

Pour donner un autre exemple d'une production erronée avec deux interprétations concurrentes locale ou verticale, nous pouvons mentionner la production erronée *přítele* au lieu de *přátelé* dans la même tâche (26).

(26) *Petr a Pavel jsou nerozluční* ***přítele**.
přátelé nom.pl.m.
Pierre et Paul sont inséparables amis
'Pierre et Paul sont des amis inséparables' (1.52.38|27|66)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Aucune erreur de cas, de genre ni de nombre mais vous n'avez pas effectué l'alternance* í > á *et vous avez fait une erreur de diacritique. 2) Le nombre et le genre sont corrects mais il y a une erreur de cas. Concernant votre forme, vous n'avez pas effectué l'alternance* í > á.

La production erronée est diagnostiquée soit comme une **erreur locale d'alternance** *í > á* **et de diacritique** (à cause de la désinence *–e* au lieu de *–é*), soit comme **une erreur verticale de cas** (l'accusatif pluriel au lieu du nominatif pluriel) **et d'alternance** *í > á*. Les autres interprétations possibles, par exemple une erreur verticale de cas et de nombre pour le génitif singulier, n'ont pas été retenues par le diagnostic pendant l'étape de filtrage des interprétations possibles à cause du nombre d'attributs atteints par l'erreur.

#### 5.1.3.6 Erreurs verticales ou horizontales internes

Uniquement une seule production erronée a été interprétée soit comme une erreur verticale, soit comme une erreur horizontale interne. Il s'agit de la production *pane* au lieu de *pánové* ou *páni* dans la tâche (27).

(27) *Vážené dámy, vážení* ***pane**, *vítám Vás na našem*
pánové/páni voc.pl.m.
Chères mesdames, chers messieurs, bienvenus à notre
*představení.*

spectacle
'Mesdames, messieurs, soyez les bienvenus à notre spectacle' (1.125.65)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas, le genre et le nombre sont corrects mais la désinence* -é *ne peut pas être utilisée avec ce mot. Concernant votre forme, vous avez fait une erreur de diacritique. 2) Le cas et le genre sont corrects mais vous avez mis une forme de singulier au lieu de pluriel.*

Le diagnostic automatique interprète cette production soit comme une **erreur verticale de nombre** (le vocatif singulier au lieu du vocatif pluriel), soit comme une **erreur horizontale interne de sous-type paradigmatique**, causée par la confusion des sous-type *pán* et *občan* (la désinence *–é* dans le vocatif pluriel) et accompagnée par une **erreur de diacritique** sur cette désinence. Nous pouvons voir encore une fois, que l'interprétation horizontale interne de cette production, bien que théoriquement possible, n'a pas de justification réelle pour servir comme un retour à l'apprenant.

#### 5.1.3.7 Erreurs verticales ou horizontales externes

Les productions diagnostiquées soit comme une erreur verticale, soit comme une erreur horizontale externe, sont par exemple celles dans les tâches (28), (29) et (30). Dans la tâche (28), il s'agit soit d'une **erreur verticale de nombre** (l'accusatif singulier au lieu de l'accusatif pluriel), soit d'une **erreur horizontale externe de type paradigmatique** à cause de la confusion des types *kuře* et *moře*. Il est intéressant de noter que cette forme a été produite par 7 apprenants.

(28) *David nemůže jíst* ****rajče**.*
rajčata acc.pl.n.
David ne peut pas manger tomates
'David ne peut pas manger les tomates' (1.18.33|27|43|52|53|60|63)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas, le nombre et le genre sont corrects mais peut être que vous confondez le modèle de déclinaison et que vous déclinez le mot d'après le type* moře*. 2) Le cas et le genre sont corrects mais vous avez mis une forme de singulier au lieu de pluriel.*

Dans la tâche (29), la production erronée est diagnostiquée soit comme une **erreur verticale de nombre**, soit comme une **erreur horizontale externe de type paradigmatique et de genre** à cause de la confusion des types *stroj* et *muž*. Effectivement, l'interprétation verticale semble plus adéquate.

(29) *Tatínek upekl výborný* ****koláče*** *a dal nám ho.*
koláč acc.sg.i.
Papa a cuisiné excellent gâteau et donné à nous le.
'Papa a cuisiné un excellant gâteau et il nous l'a donné' (1.10.54|53)
**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) 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* muž*. 2) Le cas et le genre sont corrects mais vous avez mis une forme de pluriel au lieu de singulier.*

Dans la tâche (30), la production erronée *moří* est diagnostiquée soit comme une **erreur verticale de nombre** (l'instrumental pluriel *moři* au lieu de l'instrumental singulier) **et de diacritique**, soit comme **une erreur horizontale externe de type paradigmatique et de genre** à cause de la confusion du type *moře* et *růže*.

(30) *Concorde může letět nadzvukovou rychlostí pouze nad*

Concorde peut voler supersonique vitesse seulement au dessus
***moří.***
mořem acc.pl.n.
mer
'Concorde peut voler à la vitesse supersonique seulement au dessus de la mer' (1.133.16|18|65)

**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) 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* růže. *2) Le cas et le genre sont corrects mais vous avez mis une forme de pluriel au lieu de singulier. Concernant votre forme, vous avez fait une erreur de diacritique.*

L'interprétation horizontale externe est plus probable, notamment grâce au fait qu'il peut s'agir d'une interférence avec le système lexical du français dans lequel le substantif *la mer* est féminin.

#### 5.1.3.8 Erreurs horizontales internes ou externes

Dans l'enquête publique, il y a eu uniquement deux tâches dans lesquelles la production erronée a été diagnostiquée soit comme une erreur horizontale interne, soit comme une erreur horizontale externe.

(31) *Vedle našeho ***doma*** *je hezká zahrada.*
domu gen.sg.i.
à côté notre maison est joli jardin
'A côté de notre maison, il y a un joli jardin' (1.27.38|63|65)

**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas, le genre et le nombre sont corrects mais la désinence* -a *ne peut pas être utilisée avec ce mot. 2) 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* pán.

(32) *Petr přijel z ***výleta** úplně nadšený.*
výletu gen.sg.i.
Pierre est rentré de voyage complètement enthousiasmé.
'Pierre est rentré de voyage complètement enthousiasmé' (1.38.27)

**Diagnostic** : *Il y a plusieurs interprétations possibles de votre erreur : 1) Le cas, le genre et le nombre sont corrects mais la désinence* -a *ne peut pas être utilisée avec ce mot. 2) 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* pán.

Pour les tâches (31) et (32), l'interprétation horizontale interne est basée sur la présence de la désinence *–a* dans le génitif singulier qui est employée dans certains sous-types du type *hrad* (*ostrov, rybník* etc.). Ainsi, la production erronée est considérée comme une **erreur de sous-type paradigmatique**.

Pour l'interprétation horizontale externe, il s'agit d'une confusion entre le type *hrad* et *pán* et donc d'une **erreur de genre et de type paradigmatique**. Le diagnostic généré

pour ces productions est redondant, il est cependant difficile de décider la quelle des deux interprétations est la plus adéquate.

#### 5.1.3.9 Erreurs non morphologiques

Les productions qui ont été diagnostiquées comme n'ayant aucune interprétation morphologique mais qui remplissent cependant une des conditions posées dans les tests non morphologiques[4] sont les suivantes : la production *pesmrkev* au lieu de *pes* a été diagnostiquée comme trop longue. Il s'agit évidemment d'une inattention, les productions de deux tâches distinctes ont été saisies dans un seul champ de formulaire sur la plateforme apprenant. La production *rad* au lieu de *radost* dans la tâche (33) ont été diagnostiquées comme trop courtes.

(32) *Mám* ***rad,** *že se učíte česky.*
radost acc.sg.f.
Ai joie que *refl* apprenez tchèque
'Pierre est rentré du voyage complètement enthousiasmé' (1.5.39)
**Diagnostic** : *Votre production est trop courte pour que nous puissions vous prendre au sérieux...*

Il s'agit ici probablement d'une confusion avec l'adjectif nominal *rád*, figurant dans la locution *Jsem rád, že ...* (*Je suis content que ...*), et qui remplace dans cette tâche la forme casuelle appropriée du substantif *radost*.

Deux productions ont été éliminées du traitement morphologique grâce au test sur la distance de Levehnstein entre la forme requise et la production erronée. Il s'agit des productions *houbovych* au lieu de *hub* et *vejcatach* au lieu de *vajec*.

#### 5.1.3.10 Erreurs sans diagnostic

Les productions qui n'ont pas été diagnostiquées comme aucun des types d'erreurs présentés ci-dessus représentent 13,8 % de toutes les productions erronées. La plus grande partie de ces productions contient des fautes de frappe, manifestées le plus souvent par un ajout, un remplacement ou un effacement d'un graphème dans le radical de la production erronée. Cette modification rend le radical distinct par rapport au radical de la forme requise mais cette différence n'est pas assez prononcée pour que la production erronée puisse être diagnostiquée dans les traitements non morphologiques. Il s'agit par exemple de la production *spolužci* au lieu de *spolužáci*, *mínosti* au lieu de *místnosti*, *pkoje* au lieu de *pokoje*, *studenky* au lieu de *studentky*, *písnchí* au lieu de *písní*, etc.

[4]Voir le chapitre *4 Diagnostic des erreurs*, section *4.5*, pp. 147–150.

D'autres productions sans diagnostic sont celles dans lesquelles l'apprenant a réalisé une alternance vocalique ou consonantique inappropriée, par exemple dans la tâche (33).

(33) *Indiáni přivázali zajatce ke* ***kolu**.
kůlu dat.sg.i.
Indiens attachèrent prisonnier à poteau
'Les indiens ont attachée le prisonnier à un poteau' (1.169.16|17|18)
**Diagnostic** : *Aucun*

En effet, le substantif *kůl* fait partie des substantifs monosyllabiques avec la voyelle *ů* dans le radical (*kůl, kůr, půl, růž, trůn, tůň*) qui ne réalisent pas l'alternance *ů* > *o* dans la flexion, contrairement à l'ensemble des substantifs avec les mêmes propriétés formelles (*bůh, dvůr, důl, dům, hnůj, hůl, kůň, lůj, vůz, nůž, stůl, sůl, vůl, vůz*) qui alternent régulièrement.

Le diagnostic automatique actuel est capable de reconnaître uniquement le non respect des alternances appropriées, par exemple la production *důmu* au lieu de *domu* est interprétée automatiquement comme une erreur d'alternance. Pour pouvoir diagnostiquer la forme *kolu* comme une telle erreur, il serait nécessaire de modifier la procédure pour qu'elle effectue systématiquement les alternances dans une situation spécifique et non seulement, comme c'est le cas actuellement, dans les mots qui réalisent régulièrement l'alternance donnée.

## 5.2 Enquête préliminaire

L'enquête préliminaire contient un seul exercice qui a été effectué par les apprenants du tchèque à l'INALCO sur un support papier. Un lexique contenant tous les mots dans l'exercice a été distribué aux apprenants afin qu'ils puissent se concentrer uniquement sur les aspects grammaticaux des tâches. Les données ont été rentrées dans la base de données CETLEF manuellement à partir des copies des apprenants.

L'objectif de cette enquête, entreprise avant la conception même du dispositif CETLEF, était de collecter un matériau de base pour recenser les différentes erreurs afin de pouvoir établir les hypothèses pour un diagnostic automatique. La saisie de ces productions dans la base de données à posteriori devrait d'abord servir pour élargir le volume de données soumis au diagnostic automatique en guise de son test et pour pouvoir faire une comparaison basique avec les productions collectées à l'aide de l'enquête publique.

### 5.2.1 Présentation de l'enquête

Le seul exercice dans cette enquête contient 100 tâches. Les cas des formes requises dans les tâches de cette enquête sont les suivants : nominatif pluriel (3 tâches), génitif singulier

(13 tâches) et pluriel (22 tâches), datif singulier (10 tâches) et pluriel (3 tâches), accusatif singulier (16 tâches) et pluriel (15 tâches), locatif singulier (6 tâches) et pluriel (4 tâches), instrumental singulier (6 tâches) et pluriel (8 tâches).

Les types paradigmatiques des formes requises sont les suivants : *žena* (15 tâches), *město* (11 tâches), *hrad* (8 tâches), *pán* (8 tâches), *muž* (4 tâches), *stroj* (4 tâches), *mladý* (16 tâches), *růže* (6 tâches), *kost* (5 tâches), *kuře* (5 tâches), *předseda* (3 tâches), *stavení* (2 tâches), *moře* (2 tâches), *soudce* (2 tâches), *píseň* (5 tâches), *jarní* (3 tâches), *matčin* (1 tâche), et les différents types de déclinaison pronominal (12 tâches).

Les alternances à effectuer dans les tâches sont les suivantes : 19 tâches avec les alternances vocaliques ($4 \times \# > e$, $7 \times e > \#$, $4 \times \text{ů} > o$, $1 \times \text{í} > \text{ě}$, $1 \times \text{í} > \text{á}$, $1 \times \text{é} > e$, $1 \times \text{á} > a$,) et 9 tâches avec les alternances consonantiques ($5 \times k > c$, $2 \times k > \text{č}$, $1 \times h > z$, $1 \times g > z$). Dans 6 tâches, il était nécessaire d'effectuer une double alternance ($3 \times e > \#$ et $\text{ň} > n$, $1 \times e > \#$ et $k > c$, $1 \times e > a$ et $\# > e$).

### 5.2.2 Caractéristique des apprenants et leurs scores

L'enquête préliminaire a été effectuée par 13 apprenants assistant à des cours à l'INALCO dans le cadre de la deuxième (4 apprenants) et de la troisième année (9 apprenants) de la licence de tchèque. Le niveau correspondant de la maîtrise de ces apprenants devrait à priori varier entre le niveaux B1 et C1 d'après l'échelle du Cadre européen.

Contrairement à l'enquête publique, toutes le tâches au sein de l'enquête ont été effectuées par tous les apprenants ce qui peut donner une image plus complète en ce qui concerne l'évaluation générale de leurs résultats. Le score moyen dans l'unique l'exercice de l'enquête est égal à 56,9. Les scores obtenus par les apprenants, classés en ordre décroissant, sont les suivants[5] :

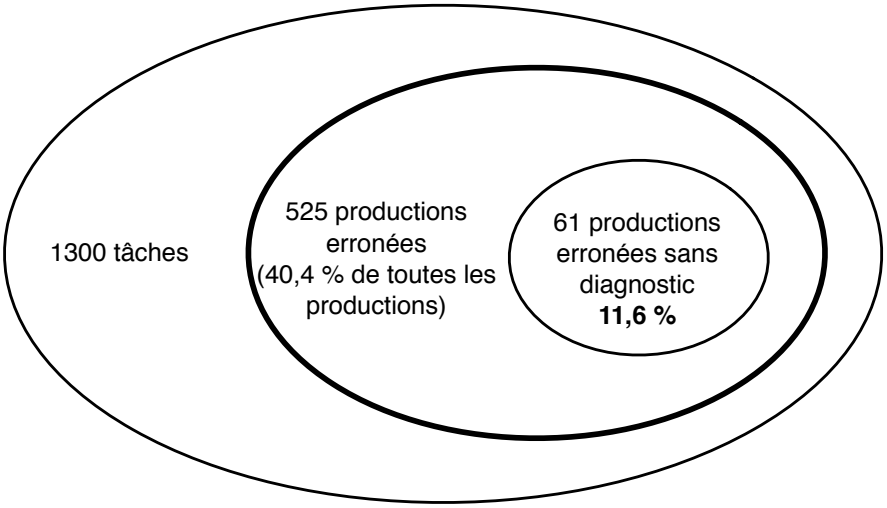
| `id_app` | 15 | 2 | 5 | 13 | 4 | 7 | 6 | 10 | 9 | 14 | 12 | 11 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| score | 87 | 77 | 66 | 62 | 59 | 57 | 53 | 52 | 50 | 48 | 44 | 42 | 42 |

### 5.2.3 Présentation des erreurs recueillies

Dans l'enquête préliminaire, le nombre total de tâches effectuées au sein de l'unique exercice de l'enquête est égal à 1300. Ce nombre est à peu près correspondant au nombre de tâches effectuées dans l'enquête publique (1551 tâches). Le nombre de productions erronées

[5]La valeur de l'attribut `id_app` représente l'identifiant de l'apprenant dans la base de données.

dans ces tâches est égal à 525 (40,4 % de toutes les productions). Pour 61 productions erronées (11,6 % de toutes les productions erronées), le diagnostic automatique n'a pas réussi d'identifier la nature de l'erreur. Une représentation schématique de cette situation est proposée ci-dessous :




Parmi toutes les productions erronées, il y en a 451 (85,9 %) qui ont été diagnostiquées comme une erreur avec une ou plusieurs interprétations morphologiques. Le nombre de productions erronées diagnostiquées comme un des types d'erreurs établis d'après les restrictions sur les formes hypothétiques (ou une combinaison possible de deux types différents), est représentés dans la table ci-dessous :

| *attribut* `erreur` | *type de l'erreur* | *productions* | *%* |
|---|---|---|---|
| LOC | erreur locale | 67 | 14,9 |
| VERT | erreur verticale | 283 | 62,7 |
| HOR INT | erreur horizontale interne | 1 | 0,2 |
| HOR EXT | erreur horizontale externe | 28 | 6,2 |
| LOC ou VERT | erreur locale ou verticale | 24 | 5,3 |
| VERT ou HOR INT | erreur verticale ou horizontale interne | 0 | 0 |
| VERT ou HOR EXT | erreur verticale ou horizontale externe | 42 | 9,3 |
| HOR INT ou HOR EXT | erreur horizontale interne ou externe | 6 | 1,3 |
| INCONNU | sans diagnostic | 61 | |

Le nombre d'erreurs qui ont été diagnostiquées comme n'ayant pas d'interprétation morphologique (13 en total, 2,5 % de toutes les productions erronés) est présenté dans la table ci-dessous :

| attribut `erreur` | type de l'erreur | productions |
|---|---|---|
| LONG | production trop longue | 3 |
| PUNCT | ponctuation dans la production | 4 |
| LEV | distance de Levehnstein[6] | 6 |

### 5.2.4 Comparaison avec l'enquête publique

En comparant les résultats obtenus dans l'enquête préliminaire avec ceux obtenus dans l'enquête publique, nous pouvons d'abord constater une correspondance assez forte dans le pourcentage des productions pour lesquelles le diagnostic n'a pas réussi de trouver une interprétation morphologique ou non morphologique (11,6 % pour l'enquête préliminaire et 13,8 % pour l'enquête publique).

Dans les deux enquêtes, le pourcentage des erreurs locales additionné au pourcentage des erreurs verticale représente à peu près 3/4 de toutes les interprétations (77,6 % pour l'enquête préliminaire et 71,7 % pour l'enquête publique) bien que le rapport entre les deux types dans les deux enquêtes est très différent (14,9 % des erreurs locales et 62,7 % des erreurs verticales pour l'enquête préliminaire contre et 32,9 % des erreurs locales et 38,8 % des erreurs verticales pour l'enquête publique). Ceci devrait être probablement imputé au nombre élevé des erreurs locales de diacritique commises par les apprenants sur la plateforme Web pour des raisons d'ordre technique.

Les autres interprétations le plus représentées dans les deux enquêtes (approximativement entre 5 et 20 % pour chacune) sont les erreurs horizontales externes et les doubles interprétations des productions avec des erreurs locales ou verticales et verticales ou horizontales externes. Les erreurs horizontales internes et leur combinaison avec d'autres types d'erreurs se montrent comme marginales dans les deux enquêtes.

### 5.2.5 Illustration d'une analyse d'erreurs par tâche

Contrairement à la présentation des productions dans l'enquête publique (classement en fonction des différents types d'erreurs), nous allons proposer ici une illustration de l'analyse d'erreurs sur le matériau de trois tâches seulement. Nous essaieront de sortir du cadre strict de la morphologie pour chercher des causes probables des erreurs au niveau syntaxique et nous allons également considérer les différences entre le tchèque et le français.

Cette approche est plus centrée sur la nature linguistique de la tâche à effectuer car elle permet la prise en compte et la comparaison des facteurs qui peuvent être à l'origine des différentes erreurs. L'étude des productions dans l'enquête préliminaire, où toutes les tâches

ont été effectuées par tous les apprenants, permet de comparer les erreurs dans une tâche et de dégager des tendances générales en fonction du nombre d'apprenants ayant commis une erreur identique. Ainsi, cette approche est plus adéquate du point de vue d'une recherche sur l'acquisition de la déclinaison du tchèque contrairement à un simple recensement des différents types d'erreurs comme nous l'avons fait pour l'enquête publique.

#### 5.2.5.1 Exemple 1

La première tâche que nous allons examiner est relativement simple. Il s'agit de produire une forme de génitif singulier du substantif féminin *vůně* après la préposition *bez*.

(1) *Ta káva je bez* **...** *(vůně, sg.).*
vůně gen.sg.f.
Ce café est sans odeur
'Ce café est sans odeur' (0.30)

| forme requise | tagLex | tagMorph | pdgm | cas | num | gen | alt |
|---|---|---|---|---|---|---|---|
| vůně | subst | N | rz_Re | gen | sg | f | sans |

À priori, les 13 apprenants entre les niveaux B1 et C1 impliqués dans l'enquête préliminaire ne devraient pas avoir beaucoup de difficulté à effectuer cette tâche, car le génitif, recommandé dans l'enseignement à partir du niveau A2, est le seul cas lié à la préposition *bez*. De plus, la forme du génitif singulier de *vůně* est égale à la forme du nominatif singulier.

Effectivement, 8 apprenants ont produit la forme correcte. Les 5 apprenants restants ont produit ces quatre formes[7] : **vonu* (11), **voně* (8 | 7), **vůni* (3), **vůní* (9).

Les productions *vonu* et *voně* n'ont pas pu être diagnostiquées automatiquement à cause de la réalisation d'une alternance inappropriée *ů* > *o*[8]. Dans la forme *voně*, il s'agit du seul écart par rapport à la forme requise, dans la forme *vonu*, une désinence qui peut exprimer le génitif singulier des substantifs du type *hrad* a été employée, il pourrait donc éventuellement s'agir d'une erreur de type paradigmatique et de genre.

La production *vůni* a été diagnostiquée automatiquement comme une erreur de cas (le datif, l'accusatif ou le locatif singulier au lieu du génitif singulier) et il est effectivement possible que l'apprenant attribue à la préposition *bez* une rection demandant un de ces cas. La désinence *–i* est également employée pour exprimer le génitif singulier au sein du type féminin *kost*, il est cependant peut probable que l'apprenant attribue ce type au substantif *vůně* car, formellement, il n'y a pas d'ambiguïté possible dans cette attribution. Enfin, la

[7] Le chiffre entre parenthèse représente l'identifiant de l'apprenant.
[8] Nous avons parlé de ce problème du diagnostic automatique sur la page 186.

production *vůní* a été diagnostiquée comme une erreur de nombre (le génitif pluriel au lieu du génitif singulier).

#### 5.2.5.2 Exemple 2

Le cas de la forme requise dans l'exemple (2) est également imposé par une préposition, il s'agit ici de produire l'instrumental singulier du substantif féminin *ves* avec une alternance vocalique de -*e*- intercalaire *e* > #.

(2) *Za* **...** *(ves, sg.)* *bylo* *obilné* *pole.*
vsí inst.sg.f.
Derrière village était de blé champ
'Derrière le village, il y avait un champ de blé' (0.80)

| forme requise | tagLex | tagMorph | pdgm | cas | num | gen | alt |
|---|---|---|---|---|---|---|---|
| vsí | subst | N | kt_n | inst | sg | f | e > # |

Uniquement deux apprenants ont produit la forme correcte *vsí*. Nous allons donc examiner les productions erronées des onze apprenants restants (**ves* (9), **vsem* (10), **vesem* (5 | 6 | 12), **vesi* (3 | 7), **vesu* (11), **vse* (8), **vsy* (4), **vese* (14)) et nous essaieront de les interpréter.

La préposition *za* peut être employée avec l'instrumental, comme dans cette tâche, mais également avec l'accusatif ou avec le génitif. Ce ne sont cependant que l'accusatif et l'instrumental qui peuvent avoir une signification spatiale, déductible du contenu lexical de la tâche[9]. Nous pouvons donc prévoir qu'il y aura probablement des erreurs de cas avec l'accusatif au lieu de l'instrumental. Ceci est effectivement le cas pour la production *ves* qui a été diagnostiquée comme une erreur verticale de cas et qui peut signaler justement une confusion dans la rection de la préposition *za*.

Avant d'examiner les erreurs suivantes, il est nécessaire de noter que le substantif *ves* ne figure pas dans le lexique pour les niveaux A1 et A2 et même dans le tchèque actuel, il a une valeur stylistique plutôt archaïsante[10]. Il est donc probable que les apprenants ne connaissaient pas ce lexème, ni ses propriétés morphologiques (le genre féminin et le besoin de la réalisation de l'alternance *e* > #) et que ce manque était la source probable des erreurs.

Ceci est sûrement le cas pour les productions *vsem* et *vesem*. La désinence –*em* est typique pour l'instrumental des types paradigmatiques masculins ou neutres. Nous pouvons

[9]La préposition *za* avec le génitif et l'accusatif peut en effet avoir également une signification temporelle, par exemple *za starých časů* (*dans les vieux temps*) ou *za hodinu* (*dans une heure*).

[10]D'après le *Frekvenční slovník češtiny*, il s'agit cependant d'un mot qui appartient au 5000 mots les plus fréquents (la valeur de rank ARF est 4766).

supposer, que les apprenants font une confusion de genre à cause de la ressemblance formelle du substantif *ves* avec des substantifs masculins (par exemple *pes, les*) et qu'ils le considère comme un masculin. Ceci peut être également justifié par une interférence provenant du français où les mots correspondants (*le village, le bourg*) sont effectivement des masculins. Dans la production *vesem*, il y a également une erreur d'alternance *e* > #. Cette alternance a été réalisée régulièrement dans la production *vsem*.

Deux apprenants ont produit la forme *vesi* qui a été diagnostiquée automatiquement comme ayant deux interprétations possibles : il s'agit soit d'une erreur locale d'alternance et de diacritique, soit d'une erreur verticale de cas (le génitif, le datif, le vocatif ou le locatif singulier au lieu de l'instrumental singulier) avec une erreur d'alternance. Pour la première interprétation, la motivation de l'erreur serait donc uniquement morphologique. Dans la seconde, elle serait motivée par des facteurs syntaxiques : le seul cas qui pourrait être imposé par une confusion des rections de la préposition *za* serait le génitif, il n'a pas cependant de signification de l'expression de lieu, donc cette interprétation est peu probable.

Les productions *vesu, vse, vsy, vese* n'ont pas été interprétées par le diagnostic automatique. Aucune des désinences dans ces formes (*–u, –y, –e*) ne peut exprimer l'instrumental singulier au sein d'un type paradigmatique, leur interprétation pourrait être donc cherchée au plan syntaxique comme une erreur de cas mais avec de telles hypothèses, nous nous livrons déjà à des spéculations qui peuvent être difficilement vérifiées.

#### 5.2.5.3 Exemple 3

La tâche dans l'exemple 3 est la plus complexe du point de vue d'un apprenant francophone car la forme requise figure dans une construction qui n'a pas d'équivalent en français et qui doit être traduite par une construction différente.

(3) **...** *(mládě, sg.)* *se* *nechtělo* *z* *nory.*
mláděti dat.sg.n.
jeune animal *refl* ne voulait pas de terrier
'Le jeune animal ne voulait pas sortir du terrier' (0.51)

| forme requise | tagLex | tagMorph | pdgm | cas | num | gen | alt |
|---|---|---|---|---|---|---|---|
| Mláděti | subst | N | kr_Re | dat | sg | n | sans |

Uniquement deux apprenants ont réussi à produire la forme correcte. Les autres ont produit les formes suivantes : **Mládovi* (3 | 7), **Mládě* (14 | 12 | 11 | 9 | 8 | 6 | 5 | 4), **Mládětí* (10).

La difficulté dans cette tâche réside dans le fait que le cas de la forme requise (le datif) est imposé par la rection du verbe *chtít* (*vouloir*) employé avec le pronom réfléchi dans une

construction impersonnelle. L'élément de la phrase tchèque qui est un complément d'objet indirect exprimé par un datif (*mláděti*), figure dans la traduction française équivalente dans la position du sujet (*l'animal ne voulais pas*). Ceci est effectivement reflété dans les productions des apprenants, dont 8 ont produit une forme du nominatif (*mládě*) qui est le cas type pour exprimer le sujet en tchèque. La motivation de cette erreur de cas est donc facilement justifiable par des facteurs syntaxiques.

Les deux productions restantes (*mládovi, mládětí*) reflètent des erreurs d'un type différent. La production *mládovi* a été diagnostiquée comme une erreur horizontale externe de genre et de type paradigmatique à cause de la confusion entre les types *kuře* et *soudce*. La production *mládětí* est une erreur de diacritique. Dans les deux cas, ce n'est pas donc le manque dans la connaissance d'une construction syntaxique spécifique mais les facteurs morphologique pour *mládovi* et graphique pour *mládětí* qui ont causé l'erreur dans cette tâche.

# Chapitre 6

# Implémentation

CETLEF est un outil qui permet de créer des exercices de déclinaison annotés morphologiquement, de les mettre en ligne pour un public d'apprenants non restreint et d'offrir, à côté des fonctionnalités supplémentaires, la correction et le diagnostic des erreurs commises dans ces exercices. La collecte de données langagières à l'aide de cet outil et leur stockage dans une base de données sont destinés à servir comme un matériau pour une recherche sur l'acquisition de la déclinaison tchèque. Dans ce chapitre, nous présentons les solutions techniques adoptées pour le développement de cet outil.

## 6.1 Application Web dynamique

Le choix des moyens techniques pour la réalisation de CETLEF est déterminé par les exigences suivantes :

- besoin d'acquérir les données linguistiques directement au format numérique ;
- stockage et une structuration des données acquises dans une base de données assurant un accès rapide et efficace et permettant des recherches basées sur l'annotation linguistique et d'autres informations contenues dans la base ;
- possibilité d'utiliser les moyens de la programmation procédurale pour le traitement de ces données dans les procédures automatiques ;
- interface utilisateur intuitive, ergonomique et esthétique ;
- large disponibilité et accessibilité des moyens technologiques utilisés pour la conception de cette interface.

La solution de développer une application Web dynamique s'est imposée naturellement. En effet, un navigateur Web (parmi les plus répandus, nous pouvons noter MS Explorer, Safari, Firefox), nécessaire pour la consultation d'une telle application, fait partie des logiciels de base sur les ordinateurs personnels. Ainsi, l'utilisateur peut accéder à cette plateforme par le réseau Internet sans installation de programmes supplémentaires sur son ordinateur. Les technologies de développement des pages Web universellement utilisées (HTML, CSS, Javascript) permet de créer une interface dynamique et riche au niveau graphique au sein d'une unique fenêtre de navigateur.

### 6.1.1 Techniques de programmation employées

L'application CETLEF a été développée sous l'environnement MAMP – acronyme pour $MacOSX + Apache + MySQL + PHP$[1]. Mac OS X est un système d'exploitation basé sur un noyau Unix – en ce qui concerne l'intégration des différents outils et leur communication mutuelle, il est donc équivalent à Linux, qui est le système d'exploitation standard pour la majorité des installations des serveurs sur le Web. L'ensemble des outils intégrés dans cet environnement permet un développement confortable des applications Web dynamiques et leur fonctionnement dans la mise en service quotidienne.

Chacun des quatre composants de cet environnement a son rôle spécifique. Le système d'exploitation est l'élément de base qui assure l'attribution des ressources aux autres composants. Le logiciel *Apache*[2] est un logiciel libre de serveur HTTP qui assure la communication entre le serveur (installation du dispositif sur un ordinateur distant) et le client (le navigateur internet de l'utilisateur sur son ordinateur personnel). *MySQL*[3] est un système de gestion de bases de données relationnelles (SGBD) qui permet de stocker, d'organiser et de manipuler des données. Le langage de script *PHP*[4] est un langage interprété de programmation procédurale. Il permet la communication avec le serveur MySQL et la génération dynamique du code HTML[5] qui est interprété par le navigateur du côté client. Le fonctionnement de cet environnement est schématiquement représenté sur la figure 6.1.

Pour la consultation de CETLEF, les moyens techniques requis sont un ordinateur personnel avec un système d'exploitation doté d'une interface graphique (Windows, Linux, Mac OS etc.), une connexion sur Internet et un navigateur Web supportant le Javascript. Le Javascript[6] est un langage interprété par le navigateur qui permet l'intégration des éléments

[1] Disponible sur `http://www.mamp.info`.
[2] Voir `http://www.apache.org`.
[3] Voir `http://www-fr.mysql.com`.
[4] Voir `http://www.php.net`.
[5] *Hypertext Markup Language* – format de données conçu pour les pages web.
[6] Voir par exemple `http://developer.mozilla.org/fr/docs/JavaScript`.

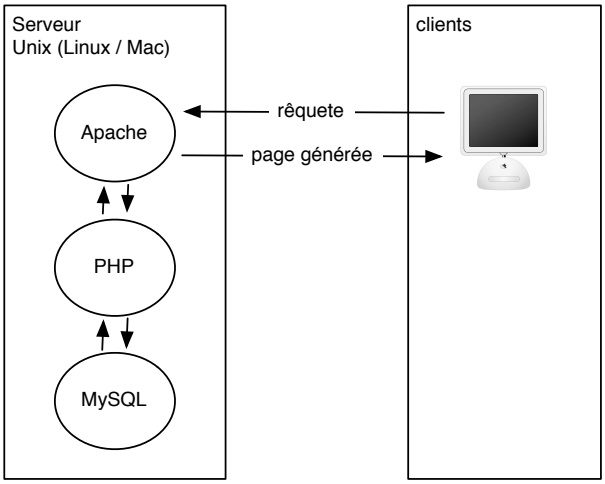



FIG. 6.1 – Architecture client serveur

interactifs et dynamiques dans les pages Web. Les cookies sont une technologie permettant de conserver des informations concernant un client (utilisateur consultant un site Internet) pendant sa connexion sur un serveur Web et lors de son retour sur ce serveur à l'aide d'un fichier (appelé *cookie*) stocké dans l'ordinateur du client[7].

### 6.1.2 Architecture de CETLEF

Du point de vue technique, l'application CETLEF, logée le serveur web, contient trois composants :

- la base de données MySQL ;
- les scripts PHP destinés à la génération des pages HTML avec du code Javascript ;
- les scripts PHP qui exécutent les procédures linguistiques lors de l'annotation des tâches des exercices et lors du diagnostic des erreurs (procédures *AlterneRadical*, *Annote*, *Diagnostic*).

Du point de vue de l'interface entre le système et l'utilisateur, nous distinguons deux plateformes distinctes :

- la **plateforme auteur** est destinée à la création des exercices, elle est utilisée par un enseignant ou un expérimentateur.

[7]Pour plus d'informations, voir par exemple `http://www.cookiecentral.com/`.

- la **plateforme apprenant** sert à rendre les exercices accessibles au public, elle est utilisée par l'apprenant.

Dans la partie suivante, nous allons d'abord exposer l'organisation de la base de données. Une idée de sa structure et des informations qui y sont stockées est nécessaire pour la compréhension de l'architecture des deux plateformes (auteur et apprenant) et de l'intégration des traitements linguistiques décrits dans le chapitre *4 Diagnostic des erreurs*.

## 6.2 Base de données

La gestion des données dans le cadre de CETLEF est effectuée par MySQL. Il s'agit d'un système de gestion de base de données (SGBD), basé sur le modèle relationnel proposé par Codd (1970). D'après ce modèle, élaboré avec la théorie des ensembles et des relations, toutes les données sont structurées au sein de tables (relations). Chaque table possède un nom et contient des colonnes (attributs). Chaque ligne dans ce tableau (un enregistrement / un tuple / un uplet) contient une valeur par attribut.

### 6.2.1 Rappels formels et définitions

Avant de décrire la structure de la base de données employée au sein de CETLEF, nous faisons un rappel des fondements formels du modèle relationnel et nous définissons les termes qui seront utilisés par la suite. Dans ce rappel, nous nous servons des ouvrages de Contensin (2004), Connoly et Begg (2005) et DelVigna (2005).

Le modèle relationnel s'établit sur le concept mathématique de *relation*. Supposons que :

- $\mathscr{R}$ est un ensemble fini ou infini de *symboles relationnels* ($r$).
- $S$ est un sous-ensemble fini de $\mathscr{R}$ appelé *schéma de base de données*.
- un *domaine D* est un ensemble fini ou infini de valeurs appelé *domaine des valeurs*.
- un *attribut* est une propriété définie sur un sous-ensemble du *domaine D*.
- l'*arité* de $r$ est une application de $r$ dans l'ensemble des entiers naturels $\mathbb{N}$.
- un *tuple / uplet* d'arité $n > 0$ est une application de l'intervalle $[1,n]$ dans $D$.

Maintenant, nous pouvons définir une *base de données* comme un couple $(S,\Delta)$ où $\Delta$ est une application qui associe à tout symbole $r$ du schéma $S$ un ensemble fini d'*uplets* dont l'*arité* est égale à l'*arité* du symbole $r$. Le symbole $r$ est le nom de la relation $R$ qui est un sous-ensemble du produit cartésien des ensembles des valeurs admissibles pour les attributs

de cette relation – c'est un ensemble de $n$-uplets contenant des valeurs sur le domaine $D$. De façon moins formelle, nous pouvons établir les définitions suivantes :

- une *relation* est une table avec des colonnes et des lignes.
- un *attribut* est le nom d'une colonne de la relation.
- un *domaine* est un ensemble de valeurs admissibles pour un ou plusieurs attributs.
- un *tuple / uplet* est une ligne distincte (unique) dans une table. Il ne peut y avoir de tuples en double dans une relation.
- l'*arité / degré* d'une relation est le nombre d'attributs qu'elle contient.
- une *base de données relationnelle* est une collection de relations portant des noms distincts.

Pour assurer les liens entre les différentes relations, nous avons besoin des moyens spécifiques qui permettent la connexion entre deux uplets dans deux relations différentes qui concernent le même objet :

- une *clé primaire* est un attribut ou un ensemble d'attributs choisis pour identifier de façon unique les tuples au sein d'une relation. Deux valeurs de cet attribut au sein d'une relation doivent être toujours distinctes.
- une *clé étrangère / secondaire* est un attribut ou un ensemble d'attributs d'une relation qui contiennent des valeurs des attributs ayant la fonction de la clé primaire d'une autre relation. La clé étrangère permet de lier deux lignes de deux tables différentes qui stockent les informations sur un objet identique.

D'après la manière dont les clés primaires et étrangères sont utilisées dans la base de données, il existe deux types différents de liens entre deux tables :

- lien *1 à 1* – la clé primaire d'un uplet d'une relation correspond à la valeur d'un seul uplet d'une autre relation. Dans ce cas, l'existence de deux relations séparées peut être mise en doute, car les informations nécessaires pourraient être stockées au sein d'une table unique.
- lien *1 à plusieurs* – la clé primaire d'un uplet d'une relation peut correspondre à la valeur de la clé étrangère de plusieurs uplets d'une autre relation.

Dans la présentation suivante, nous allons utiliser le terme ***table*** pour désigner une *relation* et le terme ***attribut*** pour désigner le nom d'une colonne dans une table.

### 6.2.2 Tables dans la base de données `cetlef`

Les données sont insérées dans la base de données par quatre agents différents :

- les données saisies par l'auteur lors de la création des exercices : les énoncés des exercices, l'annotation des tâches, etc.
- les données saisies par l'apprenant lors de son inscription sur CETLEF et au cours des exercices : les informations socio-linguistiques, les productions, etc.
- les données générées automatiquement par le SGBD : les identifiants uniques, les dates d'inscriptions, etc.
- les données générées automatiquement par des scripts PHP : les résultats des calculs sur la note moyenne, le diagnostic des erreurs etc.

Le choix des symboles relationnels et la segmentation du domaine en attributs sont arbitraires par principe. Ces deux opérations doivent mener à une modélisation adéquate de la structure des informations censées être stockées dans la base de données. La situation que nous modélisons est la suivante : *un enseignant crée des exercices auxquels il ajoute l'annotation grammaticale ; les apprenants fournissent des informations personnelles qui peuvent être pertinentes pour la compréhension de leurs productions linguistiques et ils effectuent ensuite les exercices. Leurs productions sont conservées avec la sortie du diagnostic des erreurs*.

Notre choix des tables dans la base de données `cetlef` est présenté dans le tab. 6.2.2, les attributs des différentes tables ainsi que la motivation de leur établissement seront présentés en détail plus bas.

| tables_in_cetlef |
|---|
| apprenants |
| exercices |
| taches |
| requis |
| lexique |
| prod_exercices |
| prod_taches |

TAB. 6.1 – Tables dans la base de données MySQL `cetlef`

L'organisation des liens entre les différentes tables de la base de données est exposée sur la fig. 6.2, qui présente aussi les différents attributs. Le nom des attributs avec la fonction de clé primaire commencent par la chaîne `id_` ; les attributs avec la fonction des clés étrangères reprennent le nom de la clé primaire dans la table correspondante, précédé par la chaîne [`LIEN_id_...`].

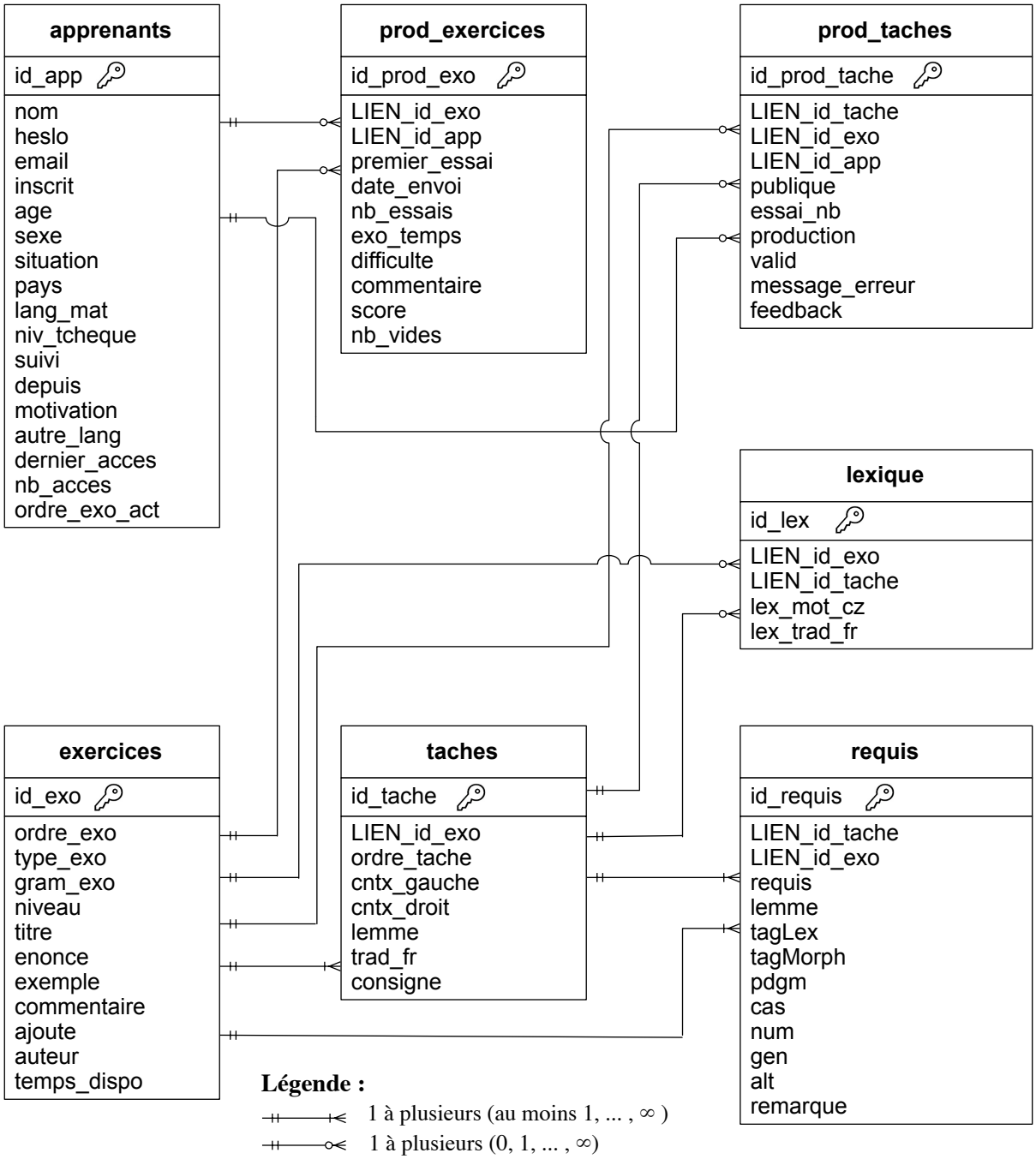



FIG. 6.2 – Schéma de la base de donnée `cetlef`

### 6.2.3 Table apprenants

La table apprenants stocke les informations concernant les apprenants observés dans le cadre expérimental d'une enquête donnée. Dans cette table, il y a d'une part les informations que les apprenants fournissent au moment de l'inscription sur http://www.cetlef.fr/inscription.php, d'autre part les informations liées aux activités de chaque apprenant, qui sont ajoutées dans la table automatiquement. Les données dans cette table permettent de spécifier les productions des apprenants par des informations supplémentaires d'ordre sociolinguistique comme l'âge de l'apprenant, son suivi pédagogique dans l'apprentissage du tchèque, les autres langues maîtrisées, etc. qui peuvent être pertinentes pour l'interprétation de ses erreurs.

| apprenants | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_app | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| nom | varchar(40) | NO | UNI | | |
| heslo | char(40) | NO | | | |
| email | varchar(40) | NO | | | |
| inscrit | datetime | NO | | | |
| age | varchar(40) | NO | | | |
| sexe | char(5) | NO | | | |
| situation | varchar(40) | NO | | | |
| pays | varchar(40) | NO | | | |
| lang_mat | varchar(40) | NO | | | |
| niv_tcheque | char(2) | NO | | | |
| suivi | varchar(40) | NO | | | |
| depuis | year(4) | NO | | | |
| motivation | text | YES | | NULL | |
| autre_lang | varchar(128) | YES | | NULL | |
| dernier_acces | datetime | YES | | NULL | |
| nb_acces | mediumint(8) unsigned | NO | | 0 | |
| ordre_exo_act | mediumint(8) unsigned | NO | | 1 | |

TAB. 6.2 – Table apprenants

#### Attributs de la table apprenants

id_app *« identifiant de l'apprenant »*

obligatoire, sans doublon, généré automatiquement par le SGBD

un nombre entier entre 0 et 16 777 215, clé primaire de la table

`nom` *« nom d'utilisateur »*

obligatoire, sans doublon, saisi par l'apprenant

une chaîne de caractères de longueur variable (40 caractères maximum)

*Remarque* : Les apprenants peuvent fournir un nom fictif pour que leur inscription reste anonyme. Nous estimons que ceci peut encourager les apprenants qui pourraient être hésitants s'ils étaient obligés de fournir leur véritable identité.

`heslo` *« mot de passe »*

obligatoire, saisi par l'apprenant

une chaîne de longueur fixe de 40 caractères.

*Remarque* : Le mot de passe permet un accès sécurisé à la section privée. Il peut être changé à tout moment par l'administrateur de la base de donnée mais il ne peut pas être lu par celui-ci, car il est chiffré au moment de l'inscription de l'apprenant.

`email` *« adresse électronique »*

obligatoire, saisi par l'apprenant

une chaîne de caractères de longueur variable (40 caractères maximum)

*Remarque* : L'adresse électronique est vérifiée lors de l'inscription par une procédure éliminant des entrées qui ne sont pas conformes à la syntaxe d'une adresse e-mail bien formée. Après l'inscription, les apprenants reçoivent un message avec leur nom d'utilisateur et leur mot de passe. Lors de la publication des messages sur le forum de CETLEF, l'adresse peut être attachée à leur message, pour favoriser des échanges entre les apprenants en dehors du forum de CETLEF.

`inscrit` *« date de l'inscription de l'apprenant »*

obligatoire, généré automatiquement par le SGBD

une chaîne en format `datetime`, par ex. : `2008-06-03 13:21:31`

`age` *« âge de l'apprenant »*

obligatoire, saisi par l'apprenant

une chaîne de caractères de longueur variable (40 caractères maximum)

valeurs possibles : `moins de 18 ans | 18 - 30 ans | 31 - 40 ans | 41 - 50 ans | 51 - 60 ans | 61 ans et plus`

`sexe` *« sexe de l'apprenant »*

obligatoire, saisi par l'apprenant

une chaîne de longueur fixe de 5 caractères.

valeurs possibles : `homme | femme`

`situation` *« statut de l'apprenant »*

obligatoire, saisi par l'apprenant
une chaîne de caractères de longueur variable (40 caractères maximum)
valeurs possibles : `oui | non`

*Remarque* : Pour le moment, cet attribut stocke la réponse à la question *Êtes-vous étudiant(e) ?* Des renseignements supplémentaires peuvent être intégrés dans le futur mais nous estimons qu'une curiosité trop importante de la part du formulaire à remplir peut avoir un effet dissuasif.

`pays` *« pays de l'apprenant »*

obligatoire, saisi par l'apprenant
une chaîne de caractères de longueur variable (40 caractères maximum)
valeurs possibles : une des valeurs dans une liste de 190 pays du monde

`lang_mat` *« langue maternelle de l'apprenant »*

obligatoire, saisi par l'apprenant
une chaîne de caractères de longueur variable (40 caractères maximum)
valeurs possibles : une des valeurs dans une liste de 120 langues du monde

*Remarque* : La liste avec des langues à choisir est une adaptation de la liste publiée sur `http://fr.wikipedia.org/wiki/Liste_des_langues_officielles`. Des précisions peuvent être toujours ajoutées à l'aide de l'attribut `motivation`.

`niv_tcheque` *« niveau de l'apprenant en tchèque »*

obligatoire, saisi par l'apprenant
une chaîne de longueur fixe de 2 caractères
valeurs possibles : `nc | A1 | A2 | B1 | B2 | C1 | C2`

*Remarque* : Les apprenants sont invités à auto-évaluer leur maîtrise du tchèque d'après l'échelle du Cadre européen commun de référence pour les langues, CECRL (2001). La valeur `nc` signifie que l'apprenant juge sa maîtrise du tchèque inférieure au niveau A1.

`suivi` *« suivi pédagogique de l'apprenant »*

obligatoire, saisi par l'apprenant
une chaîne de caractères de longueur variable (40 caractères maximum)
valeurs possibles : `autodidacte | cours individuel |`
`cours collectifs | cours universitaire | diplôme universitaire`

*Remarque* : Les apprenants sont invités à choisir la valeur qui correspond au degré le

plus haut de leur formation. Des précisions peuvent être toujours ajoutées à l'aide de l'attribut `motivation`.

`depuis` *« l'année du début de l'étude du tchèque »*

obligatoire, saisi par l'apprenant

valeurs possibles : une valeur dans une liste de chaîne en format `year`, par exemple `2008`, à partir de l'année 1988 ou la valeur *avant*.

*Remarque* : Les apprenants répondent à la question *Depuis quand apprenez-vous le tchèque ?*

`motivation` *« motivation pour apprendre le tchèque »*

optionnel, saisi par l'apprenant

un texte de longueur maximale 65 535 caractères

*Remarque* : Les apprenants peuvent s'exprimer librement sur leurs motivations pour apprendre le tchèque et/ou pour s'inscrire sur CETLEF. Les apprenants fictifs – les personnes qui souhaitent uniquement tester le dispositif et ne sont pas des apprenants du tchèque – sont repérés « manuellement » par l'administrateur qui remplit pour cet attribut la valeur `__SMAZAT__` (*effacer* en tchèque).

`autre_langue` *« autres langues maîtrisées »*

optionnel, saisi par l'apprenant

une chaîne de longueur variable de 128 caractères maximum.

valeurs possibles : ... au maximum 4 couples (`langue, niveau`), par exemple `english C2, italiano C2`.

*Remarque* : Cette information peut être utile pour l'identification des interférences inter-langue dans les productions de l'apprenant.

`dernier_acces` *« dernière connexion de l'apprenant dans la section privée »*

obligatoire, généré par le script approprié

une chaîne en format `datetime`, par exemple `2008-06-15 13:21:31`

`nb_acces` *« nombre d'accès de l'apprenant dans la section privée »*

obligatoire, généré par le script approprié

un nombre entier entre 0 et 16 777 215

*Remarque* : La valeur de cet attribut donne un aperçu rapide de la fréquentation du site par un apprenant donné car elle est augmenté à chaque connexion de l'apprenant sur CETLEF.

`ordre_exo_act` *« ordre de l'exercice actif pour l'apprenant »*

obligatoire, généré par le script approprié

un nombre entier entre 0 et 16 777 215

*Remarque* : Cette valeur est augmentée d'un point à chaque fois qu'un apprenant termine un exercice. Elle doit être stockée dans la base de données pour que l'apprenant retrouve son exercice actif à chaque nouvelle connexion sur CETLEF, car un parcours unique à travers les exercices, hiérarchisés d'après leur difficulté, a été établi au sein d'une enquête.

### 6.2.4 Table `exercices`

La table `exercices` stocke les informations caractérisant les exercices, il s'agit de l'énoncé et des informations complémentaires, et non pas du contenu de l'exercice lui-même. Les informations nécessaires sont saisies par l'auteur pendant la création de l'exercice à l'aide de la plateforme auteur. Ces informations peuvent être modifiées à tout moment à l'aide des formulaires disponibles sur cette plateforme.

| exercices | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_exo | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| ordre_exo | mediumint(8) unsigned | NO | | | |
| type_exo | varchar(40) | NO | | | |
| gram_exo | varchar(40) | NO | | | |
| niveau | char(2) | NO | | | |
| titre | varchar(40) | NO | | | |
| enonce | mediumtext | NO | | | |
| exemple | mediumtext | NO | | | |
| commentaire | mediumtext | YES | | NULL | |
| ajoute | datetime | NO | | | |
| auteur | varchar(40) | NO | | | |
| temps_dispo | tinyint(4) | NO | | | |

TAB. 6.3 – Table `exercices`

#### Attributs de la table `exercices`

`id_exo` *« identifiant de l'exercice »*

obligatoire, sans doublon, généré automatiquement par le SGBD

un nombre entier entre 0 et 16 777 215, clé primaire de la table

`ordre_exo` *« ordre de l'exercice »*

obligatoire, saisi par l'auteur

un nombre entier entre 0 et 16 777 215

*Remarque* : Cette valeur – entre 1 et *n* où *n* est le nombre d'exercices dans une enquête donnée – désigne l'ordre de l'exercice dans cette enquête. Cette valeur peut être modifiée par l'auteur en fonction de ses besoins – par exemple, pour changer l'ordre des exercices d'après leur difficulté. Si un exercice est inséré entre deux exercices déjà existants, un mécanisme de mise à jour assure la modification de cet attribut pour les exercices concernés.

`type_exo` *« type de l'exercice »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (40 caractères maximum)

valeurs possibles : `trous`

*Remarque* : Pour le moment, le seul type d'exercice supporté par CETLEF est l'exercice « à trous ». Cet attribut a été créé pour permettre un élargissement du nombre de types d'exercices possibles, par exemple des exercices avec une modification des structures grammaticales, des exercices de traductions, etc.

`gram_exo` *« grammaire abordée dans l'exercice »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (40 caractères maximum)

valeurs possibles : `déclinaison`

*Remarque* : Comme pour l'attribut `type_exo`, une seule valeur est admissible pour le moment, étant donné l'orientation de CETLEF sur la déclinaison dans le cadre de cette thèse. Un élargissement sera possible dans le futur.

`niveau` *« niveau des apprenants auxquels l'exercice est destiné »*

obligatoire, saisi par l'auteur

une chaîne de longueur fixe de 2 caractères.

valeurs possibles : `A1 | A2 | B1 | B2 | C1 | C2`

*Remarque* : La difficulté de l'exercice peut être mesurée d'après la définition des compétences dans le Cadre européen commun de référence pour les langues CECRL (2001), ou d'après la hiérarchisation de la matière exposée dans l'application du Cadre européen pour le tchèque, voir Hádková *et al.* (2005), Bidlas *et al.* (2005), Adamovičová *et al.* (2005), qui attribuent des éléments de grammaire à des niveaux différents en fonction de leur difficulté.

`titre` *« titre de l'exercice »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (40 caractères maximum)

*Remarque* : Le titre de l'exercice sert surtout comme une aide mnémotechnique. Dans l'enquête publique n° 1, le titre donne un indice sur le cas dont l'emploi est examiné au sein de cet exercice, par exemple `Máme rádi češtinu` pour l'accusatif singulier.

`enonce` *« énoncé de l'exercice »*

obligatoire, saisi par l'auteur

un texte de longueur maximale 16 777 215 caractères

*Remarque* : La valeur de cet attribut est l'instruction, destinée aux apprenants, qui spécifie les actions à effectuer dans le cadre des tâches de l'exercice, par exemple `Déclinez les substantifs en respectant le nombre indiqué.`

`exemple` *« exemple d'une tâche dans l'exercice »*

obligatoire, saisi par l'auteur

un texte de longueur maximale 16 777 215 caractères

*Remarque* : Une illustration des tâches à effectuer sert à l'apprenant comme un aperçu des activités au sein de l'exercice (par exemple `Jdu na návštěvu k ..... (<b>sestra, sg.</b>).` ⇒ `Jdu na návštěvu k <b>sestře</b>`). Les balises HTML sont intégrées directement dans le message pour faciliter la mise en page sur la plateforme apprenant.

`commentaire` *« commentaire concernant l'exercice »*

optionnel, saisi par l'auteur

un texte de longueur maximale 16 777 215 caractères

*Remarque* : Cet attribut peut contenir des consignes de nature diverse. Pour donner un exemple, il s'est avéré nécessaire d'attirer l'attention des apprenants sur la possibilité d'utiliser le clavier virtuel pour la saisie des lettres diacritées.

`ajoute` *« date de l'ajout de l'exercice »*

obligatoire, généré automatiquement par le SGBD

une chaîne en format `datetime`, par ex. : `2008-06-03 13:21:31`

`auteur` *« nom de l'auteur de l'exercice »*

obligatoire, généré automatiquement par le script approprié

une chaîne de caractères de longueur variable (40 caractères maximum)

*Remarque* : Cet attribut est prévu pour une mise en service de la plateforme auteur pour plusieurs personnes, notamment pour des enseignants qui souhaiteraient créer des exercices pour leurs apprenants.

`temps_dispo` *« temps disponible pour l'exercice »*

obligatoire, saisi par l'auteur

un nombre entier entre 0 et 255

*Remarque* : Le temps disponible pour un exercice donné peut varier en fonction du nombre de tâches ou de leur complexité. Nous supposons qu'un temps limité par exercice peut avoir un effet motivant. Un temps limité peut également rendre plus difficile la consultation des matériaux extérieurs (grammaires, tables de déclinaison, dictionnaires) dont l'utilisation pourrait fausser le résultat de l'enquête.

### 6.2.5 Table `taches`

La table `tâches` stocke les éléments d'une tâche dans le cadre d'un exercice donné. Les informations nécessaires pour chaque tâche sont saisies par l'auteur pendant leur création à l'aide de la plateforme auteur. Elles peuvent être modifiées à tout moment également à l'aide de cet outil.

| taches | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_tache | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| LIEN_id_exo | mediumint(8) unsigned | NO | | | |
| ordre_tache | mediumint(8) unsigned | NO | | | |
| cntx_gauche | varchar(256) | YES | | NULL | |
| cntx_droit | varchar(256) | YES | | NULL | |
| lemme | varbinary(64) | NO | | | |
| trad_fr | varchar(64) | YES | | NULL | |
| consigne | varchar(40) | YES | | NULL | |

TAB. 6.4 – Table `taches`

#### Attributs de la table `taches`

`id_tache` *« identifiant de la tâche »*

obligatoire, sans doublon, généré automatiquement par le SGBD

un nombre entier entre 0 et 16 777 215, clé primaire de la table

`LIEN_id_exo` *« lien vers l'identifiant de l'exercice »*

obligatoire, inséré automatiquement par le script approprié

un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `exercices`, relation 1 à plusieurs – un exercice peut contenir plusieurs tâches.

`ordre_tache` *« ordre de la tâche »*

obligatoire, saisi par l'auteur

un nombre entier entre 0 et 16 777 215

*Remarque* : Cette valeur (entre 1 et $n$ où $n$ est le nombre de tâches dans un exercice donné) désigne l'ordre de la tâche dans un exercice, cette valeur peut être modifiée librement par l'auteur comme c'est le cas pour l'attribut `ordre_exo`.

`cntx_gauche` *« contexte gauche »*

optionnel, saisi par l'auteur

une chaîne de caractères de longueur variable (256 caractères maximum)

*Remarque* : La valeur de cet attribut est la partie de la phrase située à gauche de la forme requise dans une tâche. Cet attribut peut rester vide si la forme requise est située au début de la phrase.

`cntx_droit` *« contexte droit »*

optionnel, saisi par l'auteur

une chaîne de caractères de longueur variable (256 caractères maximum)

*Remarque* : La valeur de cet attribut est la partie de la phrase située à droite de la forme requise dans une tâche. Cet attribut peut rester vide si la forme requise est située à la fin de la phrase.

`lemme` *« lemme de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne binaire de longueur variable (64 octets maximum)

*Remarque* : La valeur de cet attribut est le lemme de la forme requise. Le format binaire de cet attribut est exigé par l'encodage adéquat des caractères diacrités au sein de la base MySQL. Sans l'attribution du format binaire, des valeurs qui se distingueraient uniquement par la diacritique (par exemple `řeč` et `rec`) seraient considérées dans les requêtes SQL comme égales. Ce n'est pas souhaitable car la valeur de cet attribut est pertinente du point de vue linguistique.

`trad_fr` *« traduction française du lemme de la forme requise »*

optionnel, saisi par l'auteur

une chaîne de caractères de longueur variable (64 caractères maximum)

*Remarque* : La valeur de cet attribut est la traduction française du lemme de la forme requise. Éventuellement, il peut y avoir plusieurs traductions possibles séparées par une virgule.

`consigne` *« consigne spécifiant la tâche »*

optionnel, saisi par l'auteur

une chaîne de caractères de longueur variable (40 caractères maximum)

*Remarque* : La valeur de cet attribut ajoute l'information qui est nécessaire pour que l'apprenant puisse effectuer correctement la tâche spécifiée. Dans le cas des exercices de déclinaison, la consigne spécifie le nombre grammatical de la forme requise (par exemple `sg.` dans la tâche `Vedle našeho ... (dům, sg.) je hezká zahrada`).

### 6.2.6 Table `requis`

La table `requis` stocke la forme requise dans une tâche et l'annotation morphologique de cette forme. Si une tâche ne demandait à chaque fois qu'une seule forme requise, cette table serait inutile car toutes les informations nécessaires pourraient être stockés dans la table `taches`. L'existence d'une table spécifique est justifiée par le fait, que plusieurs formes requises (3 maximum[8]) peuvent figurer dans le cadre d'une tâche. Les informations nécessaires sont saisies par l'auteur à l'aide de la plateforme auteur. Ces valeurs peuvent également être modifiées à tout moment à l'aide de cet outil.

| requis | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_requis | mediumint(8) | NO | PRI | NULL | auto_increment |
| LIEN_id_tache | mediumint(8) | NO | | | |
| LIEN_id_exo | mediumint(8) | NO | | | |
| requis | varbinary(64) | NO | | | |
| lemme | varbinary(64) | NO | | | |
| tagLex | varchar(8) | NO | | | |
| tagMorph | varchar(8) | NO | | | |
| pdgm | varchar(8) | NO | | | |
| cas | varchar(8) | NO | | | |
| num | varchar(8) | NO | | | |
| gen | varchar(8) | NO | | | |
| alt | varbinary(64) | YES | | NULL | |
| remarque | varchar(256) | YES | | NULL | |

TAB. 6.5 – Table `requis`

[8]Pour la motivation de cette limitation, voir le chapitre *2 Cadre méthodologique*, section *2.4.2.4*, p. 26.

### Attributs de la table `requis`

`id_tache` *« identifiant de la tâche »*

obligatoire, sans doublon, généré automatiquement par le SGBD
un nombre entier entre 0 et 16 777 215, clé primaire de la table

`LIEN_id_tache` *« lien vers l'identifiant de la tâche »*

obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `taches`, relation 1 à plusieurs – une tâche peut avoir plusieurs formes requises.

`LIEN_id_exo` *« lien vers l'identifiant de l'exercice »*

obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `exercices`, relation 1 à plusieurs (dans un exercice, il peut y avoir plusieurs formes requises).

`requis` *« forme requise »*

obligatoire, saisi par l'auteur
une chaîne binaire de longueur variable (64 octets maximum)

*Remarque* : La valeur de cet attribut est la forme requise dans une tâches donnée (par exemple `domu` dans la tâche `Vedle našeho ... (dům, sg.) je hezká zahrada`).

`lemme` *« lemme de la forme requise »*

obligatoire, saisi par l'auteur
une chaîne binaire de longueur variable (64 octets maximum)

`tagLex` *« catégorie lexicale de la forme requise »*

obligatoire, saisi par l'auteur
une chaîne de caractères de longueur variable (8 caractères maximum)
valeurs possibles : `subst | adj | pron | num`

*Remarque* : Voir la section 3.2, p. 47.

`tagMorph` *« type morphologique de la forme requise »*

obligatoire, saisi par l'auteur
une chaîne de caractères de longueur variable (8 caractères maximum)
valeurs possibles : `N | A | M | P | C | X`

*Remarque* : Voir la section 3.2, p. 47.

`pdgm` *« type paradigmatique de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (8 caractères maximum)

valeurs possibles : un des identifiants mnémotechniques d'un type paradigmatique, par exemple `ph_h`.

*Remarque* : Voir la section 3.4, p. 56.

`cas` *« cas de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (8 caractères maximum)

valeurs possibles : `nom | gen | dat | acc | voc | loc | inst`

*Remarque* : Voir la section 3.3, p. 52.

`num` *« nombre grammatical de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (8 caractères maximum)

valeurs possibles : `sg | pl`

*Remarque* : Voir la section 3.3, p. 52.

`gen` *« genre grammatical de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne de caractères de longueur variable (8 caractères maximum)

valeurs possibles : `m | i | f | n | s`

*Remarque* : Voir la section 3.3, p. 52.

`alt` *« alternance pendant la génération de la forme requise »*

obligatoire, saisi par l'auteur

une chaîne binaire de longueur variable (64 octets maximum)

valeurs possibles : si une alternance a lieu, son identifiant mnémotechnique est saisi, par exemple `k > c`; si aucune alternance ne se produit, la valeur de cet attribut est `sans`.

*Remarque* : Voir la section 3.5, p. 83.

`commentaire` *« commentaire sur la forme requise »*

optionnel, saisi par l'auteur

une chaîne de caractères de longueur variable (256 caractères maximum)

*Remarque* : Un commentaire peut être utile par exemple pour indiquer la préférence pour une certaine variante si plusieurs formes requises sont possibles au sein d'une

tâche (par exemple `Avec une expression de lieu, préférez la terminaison -ě`).

### 6.2.7 Table `lexique`

La table `lexique` stocke les équivalents français des unités lexicales tchèques figurant dans les tâches au sein d'un exercice donné. Le lexique est saisi par l'auteur à l'aide de la plateforme auteur. Les entrées du lexique peuvent être ajoutées ou effacées à tout moment également à l'aide de cet outil.

| lexique | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_lex | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| LIEN_id_exo | mediumint(8) unsigned | NO | | | |
| LIEN_id_tache | mediumint(4) unsigned | NO | | | |
| lex_mot_cz | varchar(64) | NO | | | |
| lex_trad_fr | varchar(64) | NO | | | |

TAB. 6.6 – Table `lexique`

#### Attributs de la table `lexique`

`id_tache` *« identifiant de l'entrée du lexique »*

obligatoire, sans doublon, généré automatiquement par le SGBD

un nombre entier entre 0 et 16 777 215, clé primaire de la table

`LIEN_id_exo` *« lien vers l'identifiant de l'exercice »*

obligatoire, inséré automatiquement par le script approprié

un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `exercices`, relation 1 à plusieurs – pour un exercice, il peut y avoir plusieurs entrées du lexique.

`LIEN_id_tache` *« lien vers l'identifiant de la tâche »*

obligatoire, saisi par l'auteur

un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `taches`, relation 1 à plusieurs – dans une tâche, il peut y avoir plusieurs entrées du lexique.

`lex_mot_cz` *« mot en tchèque »*
obligatoire, saisi par l'auteur
une chaîne de caractères de longueur variable (40 caractères maximum)

`lex_mot_cz` *« mot en français »*
obligatoire, saisi par l'auteur
une chaîne de caractères de longueur variable (40 caractères maximum)

### 6.2.8 Table `prod_exercices`

La table `prod_exercices` stocke les informations concernant les exercices qui ont été effectuées par un apprenant donné. La plupart des valeurs dans cette table est générée automatiquement au moment de l'envoi de l'exercice à la correction, à l'exception de l'évaluation de la difficulté de l'exercice et du commentaire qui sont saisis directement par l'apprenant.

| prod_exercices | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_prod_exo | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| LIEN_id_exo | mediumint(8) unsigned | NO | | | |
| LIEN_id_app | mediumint(8) unsigned | NO | | | |
| premier_essai | datetime | NO | | | |
| date_envoi | datetime | YES | | NULL | |
| nb_essais | mediumint(8) unsigned | NO | | 1 | |
| exo_temps | mediumint(8) unsigned | YES | | NULL | |
| difficulte | tinyint(4) unsigned | YES | | NULL | |
| commentaire | mediumtext | YES | | NULL | |
| score | tinyint(3) unsigned | YES | | NULL | |
| nb_vides | tinyint(3) unsigned | YES | | NULL | |

TAB. 6.7 – Table `prod_exercices`

#### Attributs de la table `prod_exercices`

`id_exo` *« identifiant de l'exercice effectué »*
obligatoire, sans doublon, généré automatiquement par le SGBD
un nombre entier entre 0 et 16 777 215, clé primaire de la table

`LIEN_id_exo` *« lien vers l'identifiant de l'exercice »*
obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `exercices`, relation 1 à plusieurs – un exercice peut être effectué plusieurs fois.

`LIEN_id_app` *« lien vers l'identifiant de l'apprenant »*
obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `apprenants`, relation 1 à plusieurs – un apprenant peut effectuer plusieurs exercices.

`premier_essai` *« date du premier essai de l'exercice par l'apprenant »*
obligatoire, généré automatiquement par le SGBD
une chaîne en format `datetime`, par ex. : `2008-06-03 13 : 21 : 31`

*Remarque* : Cette valeur est inscrite lors de la première connexion de l'apprenant à l'exercice.

`date_envoi` *« date de l'envoi de l'exercice par l'apprenant »*
optionnel, généré automatiquement par le SGBD
une chaîne en format `datetime`, par ex. : `2008-06-03 13 : 21 : 31`

*Remarque* : Cette valeur est inscrite lors de l'envoi de l'exercice à la correction. L'existence des deux attributs `premier_essai` et `date_envoi` est justifié par le fait qu'une valeur inscrite uniquement dans la colone `premier_essai` (et aucune valeur de l'attribut `date_envoi`) permet d'identifier les exercices qui ont été abandonnés par l'apprenant entre les deux événements.

`nb_essais` *« nombre d'essais »*
obligatoire, généré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215

*Remarque* : Le nombre d'essais est le nombre de connexions sur un exercice donné sans que celui-ci soit envoyé à la correction.

`exo_temps` *« temps pris par l'apprenant pour effectuer l'exercice »*
optionnel, généré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215

*Remarque* : La valeur de cet attribut est la différence en secondes entre la dernière connexion sur un exercice et son envoi à la correction.

`difficulte` *« évaluation de la difficulté de l'exercice par l'apprenant »*
optionnel, saisi par l'apprenant
un nombre entier entre 0 et 255

valeurs possibles : `1 | 2 | 3 | 4 | 5`

*Remarque* : Après avoir envoyé l'exercice à la correction, l'apprenant peut évaluer la difficulté de l'exercice sur une échelle de 5 points : 1 – très facile ... 5 – très dificile, la valeur par défaut est 3.

`commentaire` *« évaluation de la difficulté de l'exercice par l'apprenant »*

optionnel, saisi par l'apprenant

un texte de longueur maximale 16 777 215 caractères

*Remarque* : L'apprenant peut s'exprimer librement sur l'exercice effectué.

`score` *« pourcentage des productions correctes dans l'exercice effectué »*

optionnel, généré automatiquement par le script approprié

un nombre entier entre 0 et 255

*Remarque* : La valeur de cet attribut est un nombre entre 0 et 100 qui est le pourcentage des formes correctes dans l'exercice. Cette valeur sert pour le calcul de la note (une note sur 20 d'après le système français) pour un exercice et un apprenant donnée ; pour le calcul de la note moyenne d'un apprenant et pour le calcul de la note moyenne de tous les apprenants. Ces notes sont publiées sur la plateforme apprenant.

`nb_vides` *« nombre de productions vides dans un exercice effectué »*

optionnel, généré automatiquement par le script approprié

un nombre entier entre 0 et 255

*Remarque* : La valeur de cette attribut est le nombre de tâches dans l'exercice qui ont été envoyées à la correction sans qu'aucune production soit remplie par l'apprenant.

### 6.2.9 Table `prod_taches`

La table `prod_taches` stocke les productions saisies par un apprenant dans le cadre d'une tâche au sein d'un exercice et les informations qui les concernent. C'est la table qui stocke les données les plus précieuses – les productions des apprenants.

| prod_taches | | | | | |
|---|---|---|---|---|---|
| Field | Type | Null | Key | Default | Extra |
| id_prod_tache | mediumint(8) unsigned | NO | PRI | NULL | auto_increment |
| LIEN_id_tache | mediumint(8) unsigned | NO | | | |
| LIEN_id_exo | mediumint(8) unsigned | NO | | | |
| LIEN_id_app | mediumint(8) unsigned | NO | | | |
| publique | tinyint(1) unsigned | NO | | 1 | |
| production | varbinary(64) | YES | | NULL | |
| valid | tinyint(1) unsigned | NO | | 0 | |
| message_erreur | text | YES | | NULL | |
| feedback | text | YES | | NULL | |

TAB. 6.8 – Table `prod_taches`

### Attributs de la table `prod_taches`

`id_prod_tache` *« identifiant de la production »*

obligatoire, sans doublon, généré automatiquement par le SGBD
un nombre entier entre 0 et 16 777 215, clé primaire de la table

`LIEN_id_tache` *« lien vers l'identifiant de la tâche »*

obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `taches`, relation 1 à plusieurs – pour une tâche, il peut y avoir plusieurs productions différentes.

`LIEN_id_exo` *« lien vers l'identifiant de l'exercice »*

obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `exercices`, relation 1 à plusieurs – pour un exercice, il peut y avoir plusieurs productions différentes.

`LIEN_id_app` *« lien vers l'identifiant de l'apprenant »*

obligatoire, inséré automatiquement par le script approprié
un nombre entier entre 0 et 16 777 215, clé étrangère

*Remarque* : Cet attribut assure le lien avec la table `apprenants`, relation 1 à plusieurs – un apprenant peut produire plusieurs productions.

`publique` *« consigne pour la publication de la production »*

obligatoire, généré automatiquement par le script approprié

une valeur booléenne
valeurs possibles : `0 | 1`

*Remarque* : La valeur de cet attribut indique si la production doit être affichée dans la correction de l'exercice sur l'interface apprenant. Si la valeur est 0, la production a été enregistrée dans la table après que l'apprenant ait abandonné l'exercice d'une manière illicite ou qu'il ait choisi l'option de le refaire. La production enregistrée ne sera pas affichée pour l'apprenant, mais elle reste intéressante pour l'analyse, c'est pour cette raison qu'elle est stockée dans la base.

`production` *« production de l'apprenant »*
optionnel, saisi par l'apprenant
une chaîne binaire de longueur variable (64 octets maximum)

*Remarque* : La valeur de cet attribut est la production saisie par l'apprenant au sein de la tâche. Si l'apprenant envoie l'exercice à la correction avec des tâches non remplies (sans saisir une ou plusieurs productions), la valeur `__VIDE __` est insérée automatiquement.

`valid` *« validité de la production »*
obligatoire, généré automatiquement par le script approprié
une valeur booléenne
valeurs possibles : `0 | 1`

*Remarque* : La valeur de cet attribut est générée automatiquement immédiatement après l'inscription de la production dans la base. Elle est calculée par une simple comparaison des valeurs correspondantes des attributs `production` dans la table `prod_taches` et `requis` dans la table `taches`.

`message_erreur` *« message d'erreur avec les interprétations retenus par le diagnostic »*
optionnel, généré automatiquement par le script approprié
un texte de longueur maximale 65 535 caractères

*Remarque* : La valeur de cet attribut est le message généré par la procédure de diagnostic si la valeur de l'attribut `valid` est 0. Il est stocké au format destiné à des requêtes automatisées (par exemple `HOR EXT|pn|a|acc|sg|m|sans|erreur_gen`).

`feedback` *« diagnostic de l'erreur »*
optionnel, généré automatiquement par le script approprié
un texte de longueur maximale 65 535 caractères

*Remarque* : La valeur de cet attribut est le message généré par la procédure de diagnostic. Pour le `message_erreur` donné comme exemple plus haut, l'attribut `feedback`

contient la valeur : `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 <em>pán</em>`.

### 6.2.10 Illustration des requêtes

Le SGBD MySQL permet d'effectuer des recherches dans les données à l'aide du langage SQL (Structured Query Language) qui est le standard pour l'exploitation des bases de données relationnelles. Ce langage permet de spécifier les conditions que doit satisfaire un enregistrement (ou un ensemble d'enregistrements) pertinent pour la recherche. Le résultat d'une requête est une relation (une table) avec les colonnes nommées d'après les attributs dont les valeurs ont été recherchés ; les lignes contiennent les enregistrements (les tuples, les uplets) qui satisfont les conditions posées sur les valeurs de ces attributs.

Pour une recherche manuelle, une requête peut être effectuée à l'aide d'une interface en ligne de commande comme l'outil `mysql`, ou par un gestionnaire doté d'une interface graphique comme `phpMyAdmin`. Le langage SQL a une syntaxe proche du langage naturel (commandes `SELECT, FROM, WHERE, ORDER,` etc.), les objets manipulés sont les tables et les attributs, les conditions sont posées sur les valeurs des différents attributs.

Les possibilités de recherche sont très larges, il est possible d'exprimer une grande variété de questions limitées uniquement par les restrictions du modèle de la réalité choisi pour la structure de la base de données. Voici quelques exemples de requêtes qui peuvent être effectuées sur la base `cetlef` :

#### 6.2.10.1 Exemple 1

Rechercher le nom d'utilisateur (attribut `nom` dans la table `apprenants`), l'adresse e-mail (attribut `email` dans la table `apprenants`) et le suivi pédagogique (attribut `suivi` dans la table `apprenants`) de tous les apprenants qui apprennent le tchèque depuis l'année 2004 (attribut `depuis` dans la table `apprenants`).

```
SELECT nom,
       email,
       suivi
FROM apprenants
WHERE depuis = '2004' ;
```

Le résultat de cette requête est la table :

| nom | email | suivi |
|---|---|---|
| mmaria | xxx@gmail.com | cours universitaire |
| marinka | xxx@yahoo.fr | diplôme universitaire |

Interprétation : *Il y a deux apprenants qui apprennent le tchèque depuis l'année 2004, l'apprenant* mmaria *a suivi des cours universitaires sans obtenir de diplôme, l'apprenant* marinka *a suivi des cours universitaires et il a obtenu un diplôme.*

#### 6.2.10.2 Exemple 2

Rechercher les exercices effectués par les apprenants qui apprennent le tchèque depuis l'année 2004 et afficher le nom d'utilisateur (attribut `nom` dans la table `apprenants`), son score (attribut `score` dans la la table `prod_exercices`) et l'identifiant de chaque exercice (attribut `LIEN_id_exo` dans la la table `prod_exercices`). Le résultat devrait être ordonnée d'après la valeur de l'identifiant de l'exercice.

```
SELECT apprenants.nom,
       prod_exercices.score,
       prod_exercices.LIEN_id_exo
FROM   apprenants
       prod_exercices
WHERE  apprenants.depuis = '2004' AND
       apprenants.id_app = prod_exercices.LIEN_id_app
ORDER BY prod_exercices.LIEN_id_exo ;
```

Le résultat de cette requête est la table :

| nom | score | LIEN_id_exo |
|---|---|---|
| marinka | 100 | 1 |
| marinka | 100 | 2 |
| marinka | 90 | 3 |
| marinka | 72 | 4 |
| marinka | 70 | 5 |
| marinka | 80 | 6 |
| marinka | 90 | 7 |
| marinka | 60 | 8 |
| marinka | 80 | 9 |
| marinka | 80 | 10 |
| marinka | 80 | 11 |
| marinka | 60 | 12 |
| marinka | 90 | 13 |
| marinka | 70 | 14 |
| marinka | 90 | 15 |
| marinka | 70 | 16 |

Interprétation : *Il y a uniquement l'apprenant* marinka *qui a terminé des exercices (l'apprenant* mmaria *n'a terminé aucun exercice).* Marinka *a fait 16 exercices, le score obtenu dans chaque exercice est affiché dans la colonne du milieu.*

#### 6.2.10.3 Exemple 3

Rechercher toutes les productions erronées de l'utilisateur *marinka* dans l'exercice n° 8 et afficher le nom de l'apprenant (attribut `nom` dans la table `apprenants`), la production erronée (attribut `production` dans la table `prod_taches`), le type paradigmatique de cette production (attribut `pdgm` dans la table `requis`) et la forme requise (attribut `requis` dans la table `requis`).

```
SELECT apprenants.nom,
       prod_taches.production,
       requis.pdgm,
       requis.requis
FROM   apprenants,
       prod_taches,
       requis
WHERE  apprenants.nom = 'marinka' AND
```

```
       apprenants.id_app = prod_taches.LIEN_id_app AND
       prod_taches.LIEN_id_exo = '8' AND
       prod_taches.LIEN_id_tache = requis.LIEN_id_tache AND
       prod_taches.valid = '0' ;
```

Le résultat de cette requête est la table :

| nom | production | pdgm | requis |
|---|---|---|---|
| marinka | sestry | zn_Re | sestře |
| marinka | tebe | ps_ty | tobě |
| marinka | trávniku | hd_z | trávníku |
| marinka | neznámého | md | neznámém |

Interprétation : *Dans l'exercice n$^o$ 8, l'apprenant* marinka *a fait quatre erreurs :* sestry *ou lieu de* sestře*, type paradigmatique* `zn_Re` *etc.*

#### 6.2.10.4 Exemple 4

Rechercher toutes les productions erronées issues des tâches où il fallait effectuer une alternance *e* > # et dans lesquelles le diagnostic a identifié une erreur dans la réalisation de cette alternance. Imprimer l'identifiant de l'apprenant pour chaque production (attribut `LIEN_id_app` dans la table `prod_taches`), la production erronée (attribut `production`, table `prod_taches`) et la forme requise (attribut `requis` dans la table `requis`), ordonnées d'après l'ordre alphabétique de la production.

```
SELECT prod_taches.LIEN_id_app,
       prod_taches.production,
       requis.requis,
FROM   prod_taches,
       requis
WHERE  requis.alt = 'e > #' AND
       message_erreur LIKE '%e > #|ERR%' AND
       prod_taches.LIEN_id_tache = requis.LIEN_id_tache
ORDER BY production ;
```

Le résultat de cette requête est la table :

| LIEN_id_app | production | requis |
|---|---|---|
| 11 | chlapece | chlapce |
| 21 | fanoušeky | fanoušků |
| 11 | fanoušeků | fanoušků |
| 37 | mrkeve | mrkve |
| 35 | mrkeve | mrkve |
| 10 | mrkeve | mrkve |
| 36 | mrkeve | mrkve |
| 34 | mrkeve | mrkve |
| 52 | mrkevu | mrkve |
| 21 | mrkevu | mrkve |
| 37 | pesa | psa |
| 10 | pesa | psa |
| 36 | pesa | psa |
| 34 | pesa | psa |
| 35 | pese | psa |
| 11 | pese | psa |
| 38 | pesu | psa |
| 52 | pesu | psa |

Interprétation : *Il y quatre tâches différentes dans lesquelles il fallait effectuer l'alternance* e > # *et dans lesquelles la production de l'apprenant, dont l'identifient est dans la colonne de gauche, a été diagnostiquée comme une erreur d'alternance.*

#### 6.2.10.5 Exemple 5

Cette exemple est assez semblable à l'exemple 4, cependant, cette fois il faut rechercher toutes les productions erronées issues des tâches où il fallait effectuer une alternance *e* > # et dans lesquels a été commise une erreur autre qu'une erreur dans l'alternance. L'impression du résultat doit contenir l'identifiant de l'apprenant pour chaque production (attribut `LIEN_id_app` dans la table `prod_taches`), la production erronée (attribut `production` dans la table `prod_taches`) et la forme requise (attribut `requis` dans la table `requis`), ordonnées d'après la production.

```
SELECT prod_taches.LIEN_id_app,
       prod_taches.production,
       requis.requis,
FROM   prod_taches,
```

```
       requis
WHERE  requis.alt = 'e > #' AND
       message_erreur NOT LIKE '%e > #|ERR%' AND
       prod_taches.LIEN_id_tache = requis.LIEN_id_tache
ORDER BY production ;
```

Le résultat de cette requête est la table :

| LIEN_id_app | production | requis |
|---|---|---|
| 38 | chlap | chlapce |
| 18 | chlapci | chlapce |
| 38 | fanouškach | fanoušků |
| 16 | kopcu | kopci |
| 51 | mrkev | mrkve |
| 43 | mrkevů | mrkve |
| 38 | mrkva | mrkve |
| 19 | mrkvi | mrkve |
| 22 | mrkvi | mrkve |
| 51 | pes | psa |
| 21 | pesmrkev | psa |
| 30 | psi | psa |
| 16 | pánvích | pánvi |
| 16 | vsem | vsí |

Interprétation : *Il y sept tâches différentes dans lesquelles il fallait effectuer l'alternance* e > # *et dans lesquelles l'erreur commise n'a pas été diagnostiquée comme une erreur d'alternance.*

## 6.3 Interface utilisateur CETLEF

Nous distinguons ici deux plateformes de l'interface utilisateur de CETLEF : la plateforme auteur et la plateforme apprenant. Chacune de ces plateformes joue le rôle de l'interface entre les deux types d'utilisateurs : l'auteur crée des exercices directement au format numérique, ces exercices sont proposées aux apprenants et leurs productions sont récupérées par la suite également au format numérique.

### 6.3.1 Plateforme auteur

Le fonction de la plateforme auteur est la création et la modification des exercices qui sont publiés sur la plateforme apprenant. C'est un outil destiné à l'enseignant ou à l'expérimentateur. L'interface graphique est réalisée en HTML, sa prise en main est assez intuitive. Pour le moment, la plateforme n'est pas publique mais il sera possible de la rendre accessible à un nombre restreint de personnes dans le futur. Cela nécessiterait une adaptation du dispositif pour supporter plusieurs batteries parallèles d'exercices, destinés à des publics différents, par exemple aux différentes classes d'un enseignant donné, ce qui n'a pas été réalisé dans le cadre de cette thèse.

La possibilité d'ajouter, de modifier et d'effacer les exercices librement rend le dispositif CETLEF très flexible. En effet, la plateforme auteur offre un cadre unique pour la création d'exercices ainsi que pour l'annotation de tâches sans fixer « en dur » leur contenu. Ce contenu peut varier en fonction des besoins de l'expérimentateur ou de l'enseignant.

Nous allons décrire le fonctionnement de cette plateforme en nous référant aux impressions d'écran dans l'annexe *A. CETLEF Plateforme auteur*, pp. 242–248 et au schéma de la fig. 6.3 (les chiffres encadrés sont repris dans le texte pour faciliter l'orientation du lecteur).

#### 6.3.1.1 Création des exercices

La page d'entrée de la plateforme auteur (voir fig. A.1, p. 242) offre une vue d'ensemble sur les exercices enregistrés dans la base de données. Les exercices sont lus $\boxed{1}$ au moment de l'entrée sur la page. Il est possible d'ajouter un exercice, afficher les tâches dans un exercice ou modifier un exercice existant.

L'ajout d'un exercice $\boxed{2}$ se fait par un formulaire (voir fig. A.2, p. 243) destiné à cet effet. L'auteur saisit les valeurs des attributs `ordre_exo`, `type_exo`, `gram_exo`, `niveau_exo`, `temps_dispo`, `titre`, `enonce`, `exemple` et `commentaire` de la table `exercices`. Après l'envoi de ce formulaire $\boxed{3}$ , les informations fournies par l'auteur sont enregistrés dans la base de données $\boxed{4}$ et l'auteur est redirigé $\boxed{5}$ sur la page d'accueil. La modification d'un exercice s'effectue de la même façon sauf que le formulaire est déjà rempli par les informations enregistrées dans la base de données pour l'exercice que l'on souhaite modifier.

#### 6.3.1.2 Ajout des tâches

Pour chaque exercice, il est possible d'afficher les tâches $\boxed{6}$ $\boxed{7}$ qu'il contient (voir fig. A.3, p. 244). Ensuite, il est possible d'afficher et de modifier le lexique, ajouter ou modifier une tâche ou revenir sur la page d'accueil $\boxed{19}$ . La modification du lexique (voir fig. A.4,

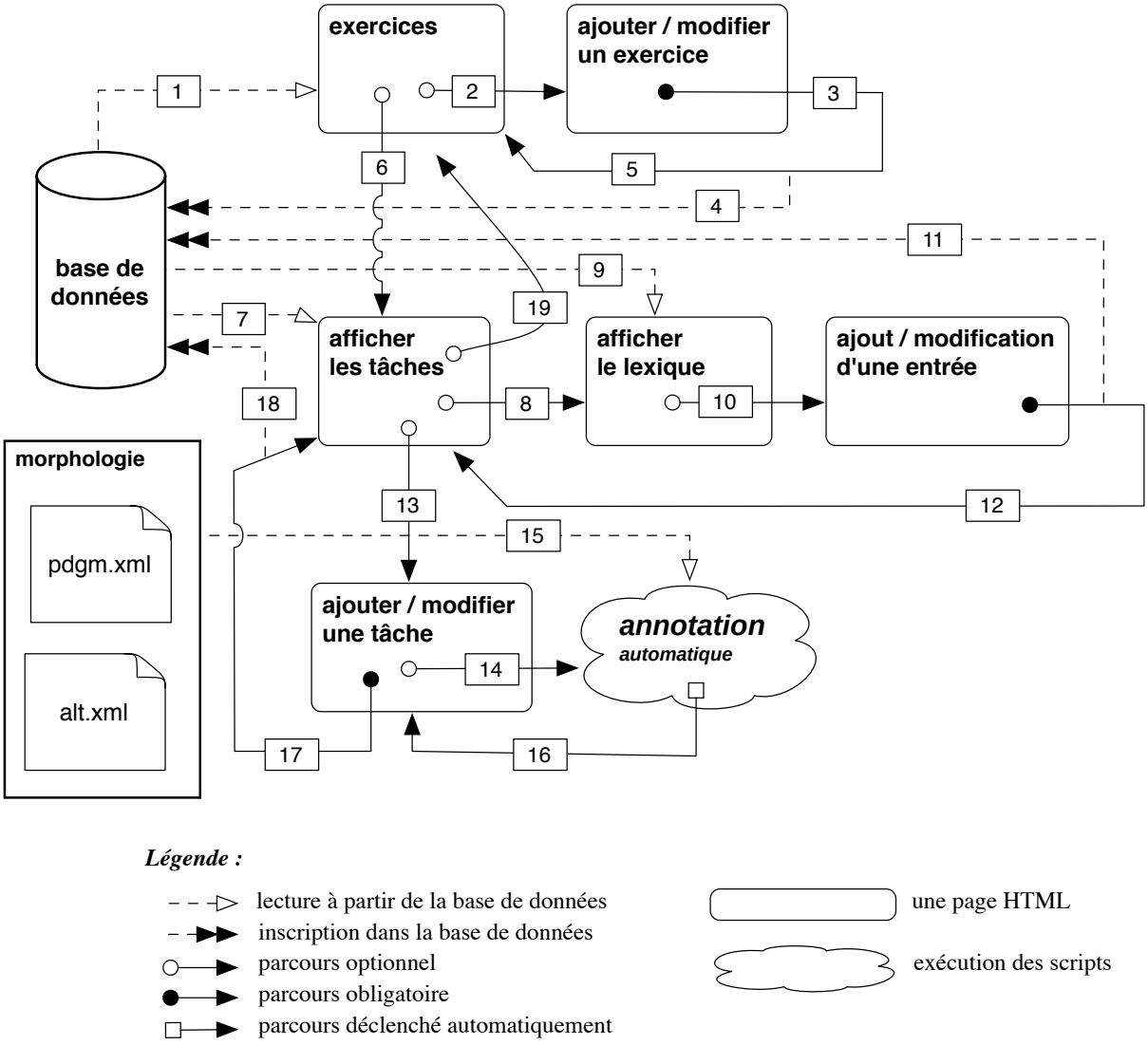



FIG. 6.3 – Architecture de la plateforme auteur

p. 245) s'effectue par les étapes 8 9 10 11 12 . L'auteur modifie les valeurs des attributs `lex_mot_cz`, `lex_trad_fr` et `LIEN_id_tache` de la table `lexique`.

Ajouter ou modifier une tâche 13 s'effectue à l'aide d'un formulaire (voir fig. A.5, p. 246). L'auteur saisit les valeurs des attributs `ordre_tache`, `cntx_gauche`, `lemme`, `trad_fr`, `consigne`, `cntx_droit` de la table `taches`. Concernant la forme requise au sein d'une tâche, il est possible d'en saisir trois au maximum. Pour chaque forme requise (attribut `requis` de la table `requis`), il est obligatoire de saisir les valeurs des attributs de l'annotation linguistique (les attributs `tagLex`, `tagMorph`, `pdgm`, `cas`, `num`, `gen`, `alt` et `remarque` de la table

`requis`). L'auteur peut annoter la forme requise en sélectionnant manuellement les valeurs appropriées dans les listes déroulantes contenant toutes les valeurs permises pour chaque attribut. Il peut également utiliser le module d'assistance de l'annotation (voir la procédure *Annote*, voir la section 3.7, p. 111) 14 15 16 qui propose les valeurs des attributs, les ambiguïtés éventuelles dans le choix pour certains attributs sont annoncées et doivent être levées manuellement.

Pour l'attribution de la valeur `pdgm`, l'auteur peut se servir d'un outil de consultation des paradigmes inscrits dans le fichier *pdgm.xml* (voir fig. A.6, p. 247, pour le fichier *pdgm.xml* voir la section 3.4.2, p. 65). Après que toutes les informations aient été remplies et l'annotation vérifiée (voir fig. A.7, p. 248), la tâche peut être ajoutée à l'exercice 17 18 .

Toute la procédure peut être répétée un nombre de fois illimité en fonction du nombre de tâches souhaité pour un exercice. Le retour sur la page d'accueil 19 permet de recommencer toute la procédure pour un exercice suivant.

## 6.3.2 Plateforme apprenant

La plateforme apprenant est la partie « visible » de CETLEF – elle est publiquement accessible sur `http://www.cetlef.fr`. Nous allons décrire le fonctionnement de cette plateforme en nous référant au schéma de la figure 6.4 et aux impressions de l'écran dans l'annexe *CETLEF Plateforme apprenant*, pp. 250–268.

### 6.3.2.1 Division section publique / section privée

En fonction de la disponibilité de ses différents composants, la plateforme apprenant est divisée en deux sections :

- la **section publique** est accessible directement par la saisi de l'URL[9] dans le navigateur et sa consultation ne nécessite aucune identification.
- la **section privée** est réservée aux apprenants qui se sont inscrits sur CETLEF et qui possèdent un nom d'utilisateur et un mot de passe nécessaires pour la connexion dans cette section.

La section publique est composée de quatre pages : *accueil*, *présentation*, *inscription* et *connexion*, le nom de chacune de ces pages indique clairement son contenu. La page *accueil* présente en grandes lignes l'objectif et les fonctionnalités de CETLEF. La page *présentation* offre une galerie de captures d'écran commentées montrant l'interface de l'apprenant dans la section privée. Sur la page *inscription*, l'apprenant peut s'inscrire sur CETLEF 1 à

[9] Uniform Ressource Locator - l'adresse d'une page Web.

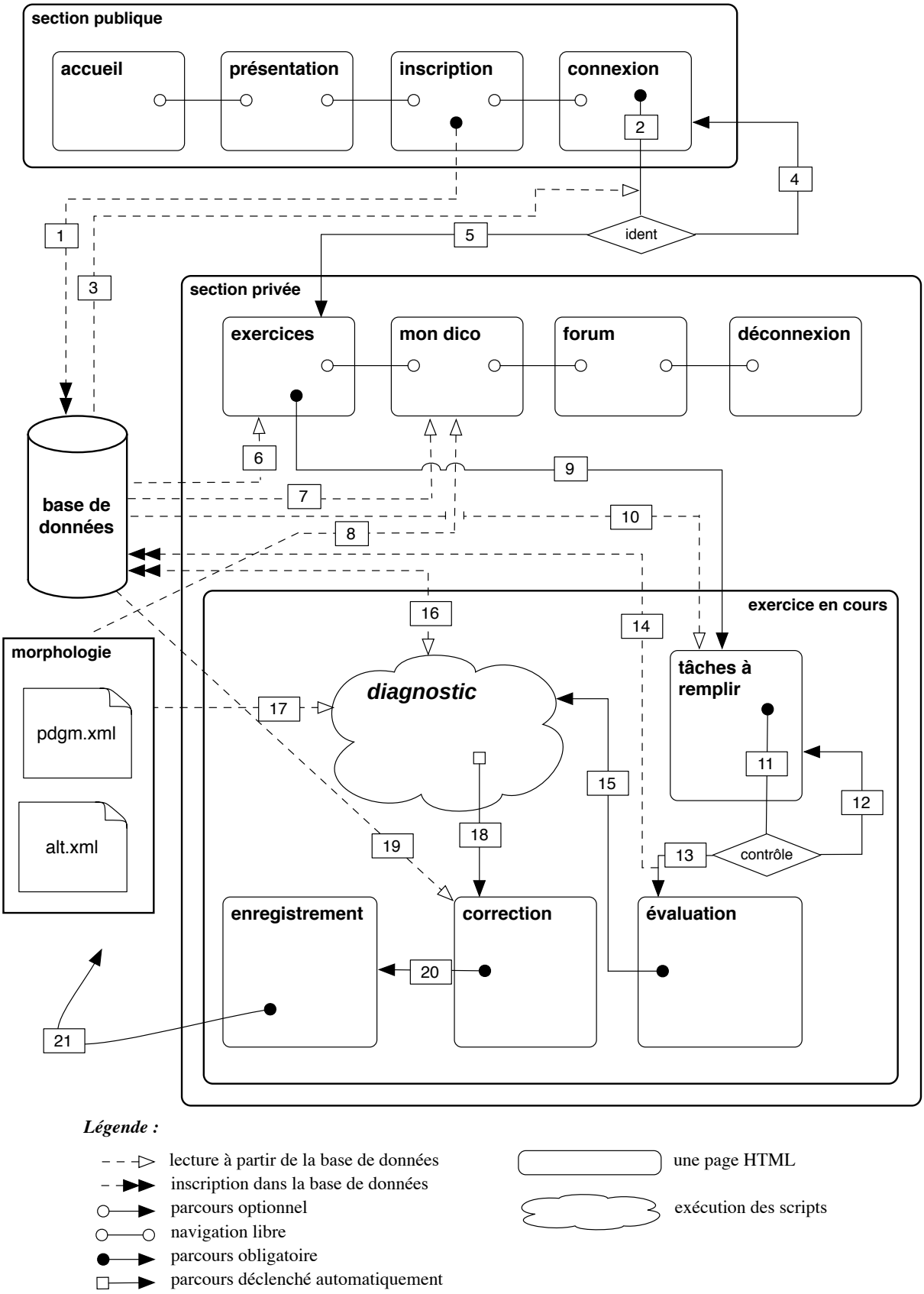



FIG. 6.4 – Architecture de la plateforme apprenant

l'aide d'un formulaire. Les informations demandées ont été présentées dans 6.2.3, p. 202. La page *connexion* permet à l'utilisateur de se connecter dans la section privée 2 , à condition que le nom d'utilisateur et le mot de passe soient vérifiés dans le processus d'identification 3 4 5 .

La section privée contient quatre pages : *exercices*, *mon dico*, *forum* et *déconnexion*. La page *exercices* (voir fig. B.1, p. 250) est la porte d'entrée de la section privée. Elle contient les exercices 6 proposés dans le cadre d'une enquête et des informations concernant la note moyenne de l'apprenant et des autres. L'en-tête de cette page peut contenir également diverses indications personnalisées en fonction des activités de l'apprenant sur le site, par exemple un mot de bienvenue après la première connexion, l'annonce du nombre de connexions dans la section privée, etc. La page *mon dico* contient une présentation de la matière linguistique 7 8 rencontrée par l'apprenant dans les exercices, elle est remplie progressivement au fur et à mesure de son avancement. La page *forum* donne une possibilité de communiquer avec l'administrateur du dispositif ou avec les autres utilisateurs. La page *déconnexion* permet de terminer la session personnelle de l'apprenant et de sortir de la section privée vers la section publique.

#### 6.3.2.2 Exercices

Les exercices sont organisés dans un parcours unique, c'est-à-dire que l'apprenant ne peut pas choisir librement un exercice mais il doit observer l'ordre dans lequel ils ont été classés au sein d'une enquête. Un exercice peut avoir trois statuts différents :

- un **exercice actif** est celui auquel l'apprenant peut accéder immédiatement.
- un **exercice terminé** est un exercice qui a déjà été effectué.
- un **exercice futur** est un exercice qui sera accessible au moment où l'apprenant aura terminé tous les exercices qui le précèdent.

A tout moment, il n'y a qu'un seul exercice actif. Les informations concernant l'exercice (voir 6.2.4, p. 206) sont présentées dans l'entête de l'exercice actif.

Après avoir démarré un exercice 9 10 , l'apprenant entre dans une série de procédures successives :

1. le remplissage des tâches
2. l'évaluation de la difficulté de l'exercice
3. la correction de l'exercice
4. l'enregistrement de l'exercice

Le déroulement de cette série est contrôlé par un mécanisme de verrouillage qui ne permet pas de revenir en arrière sans que ce retour ne soit pris en compte. En cas d'interruption illicite de cette procédure, ce mécanisme calcule le nombre d'essais nécessaires pour effectuer un exercice[10]. Par défaut, chaque apprenant n'a qu'un seul essai pour faire un exercice. Cependant, au cas où le temps réservé pour un exercice est écoulé, l'apprenant peut choisir de refaire l'exercice au prix que chaque essai en plus baisse sa note finale d'un point. Cela devrait avoir un effet dissuasif pour les interruptions illicite d'un exercice en cours ainsi que pour la répétition d'un exercice après l'écoulement du temps.

La page avec les tâches à effectuer (voir fig. B.2, p. 251) propose à l'apprenant des outils d'assistance : un clavier virtuel avec les lettres diacritées tchèques et un lexique (voir fig. B.3, p. 252). La forme à décliner (le lemme de la forme requise avec la consigne et la traduction) apparaît dans une vignette quand le pointeur de la souris est placé sur la case à remplir (voir fig. B.4, p. 253). L'exercice ne peut être envoyé à la correction 11 qu'au moment où toutes les tâches sont remplies – si le temps n'est pas écoulé, un message d'avertissement apparaît et les cases vides sont mises en relief 12 . Si le temps disponible pour l'exercice est épuisé, l'apprenant peut choisir de l'envoyer à la correction (dans ce cas, les tâches vides sont autorisées) ou de le refaire plus tard, au prix d'avoir manqué un essai ce qui se projettera dans sa note quand il aura fini l'exercice. Après avoir envoyé l'exercice à la correction 13 , les valeurs des attributs `production`, `publique` et `valid` (obtenue par la simple comparaison de la production avec la forme requise) sont enregistrées dans la table `prod_tâches` (voir la section 6.2.9, p. 217). Au cas où l'apprenant choisit l'option de refaire l'exercice, ses productions sont tout de même enregistrées dans la base de données comme des productions non-publiques. L'étape suivante 13 est l'évaluation de la difficulté de l'exercice par l'apprenant (voir fig. B.5, p. 254). Il saisit les valeurs des attributs `difficulte` et `commentaire` qui sont enregistrés dans la table `prod_exercices`.

Le diagnostic de l'erreur est effectué après la sortie de la page *évaluation* 15 et avant l'arrivée sur la page *correction* 18 . La procédure *Diagnostic* (voir la section 4.3, p. 139) prend en entrée les productions dont la valeur de l'attribut `valid` dans la base de donnée est égale à 0 16 , il détermine la nature de l'erreur en se servant éventuellement des informations morphologiques contenues dans les fichiers *pdgm.xml* et *alt.xml* 17 et il inscrit les valeurs des attributs `message_erreur` et `feedback` dans la table `prod_taches` 16 . En pratique, pour un exercice de 10 tâches, cette procédure peut durer entre 0.5 et 6 secondes

[10]En pratique, il existe des différences dans le fonctionnement de ce mécanisme dans les navigateurs différents en fonction de leur gestion de la mémoire cache (une mémoire temporaire stockant les fichiers provenant des sites web). Dans Firefox et Safari, un message avertissant l'apprenant de l'impossibilité du retour est affiché lorsque celui-ci tente de revenir, sur la page précédente. Dans MS Explorer et Opéra, ce retour est possible et l'avertissement n'est affiché qu'après le rechargement de la page. Ce problème technique n'a cependant pas d'incidence sur le calcul du nombre d'essais.

en fonction du nombre des erreurs et de leur complexité du point de vue du diagnostic et également en fonction de la vitesse de la connexion sur l'internet.

L'étape suivante est l'affichage de l'exercice corrigé (voir fig. B.6, p. 255) 18 19 . Le résultat – un résumé des informations sur l'exercice (la note, le temps de l'apprenant etc.) et la comparaison de la performance de l'apprenant avec celle de ses collègues – est affiché dans l'entête. Ces informations sont lues à partir de la table `prod_exercices` (voir la section 6.2.8, p. 215). Les tâches corrigées sont affichées en respectant le code suivant :

- les production correctes sont affichées en vert ;
- les productions erronées sont barrées et affichées en rouge ;
- la forme requise est affichée en vert à droite de cette production.

Le diagnostic de l'erreur apparaît dans une vignette au passage du pointeur de la souris au dessus de la production (pour des exemples du diagnostic voir les fig. B.7, B.8, pp. 256–257). Les informations grammaticales, lues à partir de la table `requis` et `taches` sont mises en forme pour une lecture facilitée et elles s'affichent au passage de la souris sur la forme requise ou sur la production correcte (voir les fig. B.9 et B.10, pp. 258–259).

L'organisation de la vignette avec les informations grammaticales (voir les attributs de la table `requis` dans 6.2.6, p. 211) est présentée sur la figure 6.5.

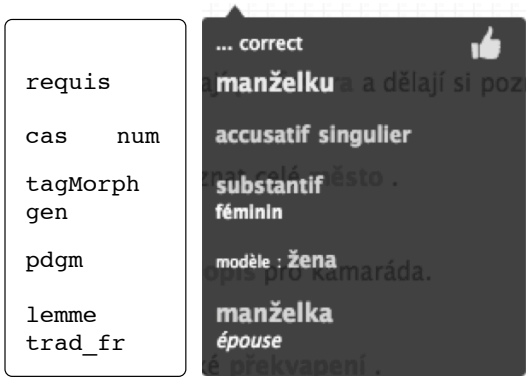



FIG. 6.5 – Vignette avec les informations grammaticales.

Le cas échéant, les valeurs des attributs `alt` et `remarque` sont affichées. La gestion de l'affichage des informations pour plusieurs formes requises au sein d'une tâche est assurée également. Au cas où les valeurs des attributs `tagMorph` et `tagLex` ne seraient pas compatibles (voir la section 3.2, p. 47), un message spécifiant cette incompatibilité est affiché. Les valeurs des attributs `tagMorph`, `tagLex`, pdgm, cas, `num`, `gen`, `alt` sont stocké dans la table `requis` de la base de données au format court (par exemple *N* | *subst* | *ph_h* | *loc* | *pl* | *m* | *ch* > *š* ). Pour être affichées sur la vignette de la plateforme auteur, elles sont retranscrites en un format lisible plus facilement (par exemple *substantif* | *pán* | *locatif* | *pluriel* |

*masculin animé* | *alternance consonantique, palatalisation A, ch > š*). Concernant le paradigme, uniquement l'appartenance au type paradigmatique est signalée. Son appartenance à un sous-type éventuel est signalée sur la page *Mon dico*.

Après avoir fini de consulter la correction de son exercice, l'apprenant passe 20 sur la page *enregistrement* (voir fig. B.11, p. 260). La fonction principale de cette page est de marquer la fin de la procédure d'un exercice en cours – au niveau technique, uniquement la valeur de l'attribut `ordre_exo_act` (le nombre de l'exercice actif de l'apprenant) est augmenté d'un point. L'action clôturant la série de procédures d'un exercice en cours est le passage sur la page *exercices* 21 . Grâce à l'incrément de la valeur `ordre_exo_act`, l'exercice que l'apprenant vient de faire change – il devient un exercice terminé. L'exercice suivant gagne le statut d'exercice actif et la boucle que nous venons d'exposer peut recommencer (voir fig. B.12, p. 261).

Pour chaque exercice terminé, l'apprenant peut consulter son résultat et les erreurs qu'il a faites (voir fig. B.13, p. 262). Comme pendant la correction de l'exercice, les erreurs et les formes requises sont accompagnées par des vignettes avec le diagnostic et les informations grammaticales (voir fig. B.14 et B.15, pp. 263–264).

#### 6.3.2.3 Mon dico

Le principe de la page *Mon dico* est la présentation des faits linguistiques qu'un apprenant rencontre au fur et à mesure dans son parcours à travers les différents exercices publiés sur la plateforme apprenant. Ceci est possible grâce à l'annotation grammaticale qui accompagne toutes les formes requises dans les tâches des exercices. Les éléments remplissent une base de données personnelle de l'apprenant qui peut être consultée à tout moment et qui peut servir comme outil d'apprentissage. Les éléments affichés sur cette page sont choisis en fonction de la valeur `ordre_exo_act` à partir des différents tables de la base `cetlef` – c'est-à-dire, uniquement les éléments liés à des exercices terminés par un apprenant donné sont affichés sur cette page. L'augmentation du volume de cette base peut être un facteur motivant pour l'apprenant. Les éléments qui nourrissent cette « base de connaissances » sont classées en quatre groupes :

- le lexique contenu dans les exercices terminés ;
- les paradigmes des formes à décliner ;
- les alternances à effectuer dans les formes à décliner ;
- le cas et le nombre des formes à décliner.

La section *lexique* (voir fig. B.16, p. 265) contient une liste alphabétique des mots tchèques et de leur traduction française qui sont inscrits dans les attributs `lemme` et `trad_fr` de la table

`taches` et dans les attributs `lex_mot_cz` et `lex_trad_fr` et de la table `lexique`. Le lien hypertexte derrière chaque entrée permet d'afficher la tâche (formatée comme une phrase en concaténant les valeurs des attributs `cntx_gauche`, `requis`, `cntx_droit` des tables `requis` et `taches`) dans laquelle cette entrée est apparue. L'apprenant a également à sa disposition un outil permetant de chercher dans les valeurs de ces quatre attributs. En plus, l'outil de recherche prend en compte le contenu lexical des attributs `cntx_gauche` et `cntx_droit`, segmentés d'une manière basique (un espace est considéré comme délimiteur des mots) et non lemmatisé.

La section *paradigmes* (voir fig. B.17, p. 266) est un outil de consultation des types paradigmatiques rencontrés dans les exercices – il s'agit uniquement des types paradigmatiques des formes requises au sein des tâches. L'apprenant peut choisir l'un des types paradigmatiques, affichés dans la colonne de gauche, et le lien hypertexte derrière le mot-modèle permet d'afficher la liste des terminaisons (voir fig. 6.6) et les mots rencontrés dans les exercices qui appartiennent à ce paradigme.

exemples : *autor, bratr, ministr, pán, pes, premiér, prezident ...*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| nom.sg. | *-#* | | | nom.pl. | *-i* | \| | *-ové* |
| gen.sg. | *-a* | | | gen.pl. | *-ů* | | |
| dat.sg. | *-u* | \| | *-ovi* | dat.pl. | *-ům* | | |
| acc.sg. | *-a* | | | acc.pl. | *-y* | | |
| voc.sg. | *-e* | | | voc.pl. | *-i* | \| | *-ové* |
| loc.sg. | *-u* | \| | *-ovi* | loc.pl. | *-ech* | | |
| inst.sg. | *-em* | | | inst.pl. | *-y* | | |

sous-types : *hoch, občan, génius, džigolo*

exceptions : *syn, bůh, host, manžel, člověk*

FIG. 6.6 – Présentation d'un type paradigmatique.

Les informations concernant le paradigme sont lues dans le fichier *pdgm.xml* (voir la section 3.4.2, p. 65). Seules les terminaisons du modèle de type paradigmatique sont présentées, les variantes du registre ne sont pas affichées. Pour chaque paradigme, l'apprenant peut consulter les exemples des mots appartenant à ce type paradigmatique, la liste des sous-types et des exceptions, distingués en fonction de la valeur de l'attribut `ouvert` correspondante (1 pour les sous-types, 0 pour les exceptions, voir la section 3.4.1, p. 56). Les différences dans la

déclinaison des mots rencontrés sont mises en évidence dans des vignettes qui apparaissent lorsque la souris est placée au dessus des mots rangés dans la colonne à droite (voir fig. 6.7 pour l'affichage d'un sous-type du paradigme *pán*).

FIG. 6.7 – Affichage des différences dans la déclinaison d'un sous-type.

La présentation de l'écart entre le paradigme du type-modèle et les mots appartenant à des sous-types ou des exceptions, est basé sur les valeurs de l'attribut `ET` (voir la section 3.4.1, p. 56). Chaque mot dans la colonne de droite est doté d'un lien hypertexte qui permet l'affichage de la tâche avec ce mot dans la section *lexique*.

La section *alternances* (voir fig. B.18, p. 267) permet de visualiser les différentes occurrences des alternances rencontrées dans les tâches des exercices. Pour afficher les exemples de la réalisation d'une alternance dans les exercices, l'apprenant est invité d'abord à choisir un type d'alternance (par exemple vocalique quantitative, palatalisation A, mouillure etc., voir la section 3.5, p. 83), puis une alternance particulière (par exemple $k > c$). Au passage de la souris sur le couple *lemme – forme alternée*, il est possible de visualiser une vignette portant des informations sur le mot alterné et l'alternance elle même (voir fig. 6.8). Les deux mots sont dotés d'un lien hypertexte avec la même fonction que dans les sections précédentes – afficher la tâche correspondante dans la section *lexique*.

FIG. 6.8 – Vignette accompagnant une alternance.

La section *cas* (voir fig. B.19, p. 268) permet une consultation des formes requises en fonction de la valeur de la catégorie grammaticale du nombre et du cas. L'apprenant choisi

d'abord le cas et puis le nombre. Comme dans les sections précédentes, il est possible d'afficher les mots dans la section *lexique* à l'aide d'un lien hypertexte.

# Chapitre 7

# Conclusion

Le travail présenté dans cette thèse a relié trois éléments : un modèle de la déclinaison du tchèque, un module de diagnostic automatique des erreurs dans des exercices de déclinaison et une plateforme d'enseignement assisté par ordinateur.

## Plateforme CETLEF

La conception et la réalisation de la plateforme CETLEF, a été le coeur de notre travail. Il a fallu tout d'abord établir un modéle de la déclinaison du tchèque qui rend possible des procédures automatiques, il permet d'annoter morphologiquement les données afin de faire des recherches ciblées et il peut être employée dans une perspective didactique.

L'affichage des informations morphologiques est conçue dans la perspective d'une description explicite de la grammaire du tchèque. Les apprenants peuvent consulter l'annotation de toutes les productions qu'ils ont rencontrées dans les exercices dans leur base de donnée personnelle qui représente, en quelque sorte, une mini-grammaire électronique. La présentation des types de déclinaison et des alternances offrent à l'apprenant une vision du système morphologique accessible et complète en même temps. Malgré le nombre important de différents sous-types paradigmatiques définis par l'annotation, leur définition par des écarts par rapport au sous-type modèle, permet une présentation simple des sous-types ouverts et fermés (des exceptions) en mettant en relief les désinences qui diffèrent dans les cas spécifiés.

Concernant l'amélioration du modèle de la déclinaison en lui-même, des techniques de réduction des sous-types paradigmatiques sont possibles et pourraient être intégrées dans des versions futures. Il s'agirait notamment de l'élaboration de règles qui assureraient la suppression des sous-types définis uniquement par les différences entre les désinences *–e* et *–ě* (par exemple les deux types *žena* et *holka* à cause du locatif singulier : *žen–ě*×*holc–e*)

ou *–i* et *–y* (les types *žena* × *skica*) dont la distribution est déterminée par la nature de la consonne finale du radical. De la même façon, les changements des désinences, liées aux alternances, pourraient être gérées par des règles. Par exemple, le locatif pluriel du type *hrad* (*hrad-ech*) prend la désinence *-ích* si la consonne finale du radical est une vélaire.

## Diagnostic des erreurs

Notre hypothèse que les erreurs de déclinaison sont calculables a été éprouvée dans le diagnostic automatique. L'application du diagnostic sur un échantillon de productions authentiques, collectées sur CETLEF, a permis de vérifier que cette hypothèse est vraie pour une grande partie de productions. La majorité des erreurs peut être interprétée automatiquement comme une combinaison du radical de la forme requise et d'une désinence.

Nous avons proposé une représentation formelle des erreurs à l'aide des structures de traits morphologiques, et nous avons établi leur typologie en fonction des attributs qui diffèrent dans la production erronée et dans la forme requise.

Pour l'évaluation globale du diagnostic automatique, il est nécessaire de considérer la situation spécifique dans laquelle les erreurs analysées ont été produites. En déclinant une forme au sein d'un exercice grammatical, l'apprenant et la machine procèdent effectivement d'une manière assez semblable au niveau de l'analyse et de la génération. De ce point de vue, il serait intéressant d'étudier les occurrences des erreurs décrites dans ce travail dans les productions libres où l'apprenant n'est pas limité à un cadre défini aussi strictement que dans la tâche d'un exercice.

L'utilité du diagnostic pour la recherche des différents types d'erreurs dans la base de données est indéniable. Grâce à l'annotation morphologique des formes requises et grâce au message d'erreur caractérisant les productions erronées, des recherches basées sur cette annotation peuvent être effectuées facilement et servir des analyses variées, comme nous l'avons illustré avec un échantillon de données recueillis sur CETLEF. Avec le nombre croissant de productions dans la base de données au cours du temps, il sera possible d'entreprendre des études d'une envergure plus grande.

L'adéquation du diagnostic du point de vue didactique est une question qui reste ouverte pour le moment. Une grande partie des messages est bien compréhensible, notamment pour les productions erronées avec une seule interprétation (par exemple les erreurs de cas, de genre, d'alternance). Quand le diagnostic propose deux interprétations différentes, le message est en général plus lourd et représente pour l'apprenant une sorte de « devinette » à résoudre, ce qui peut avoir un effet plutôt négatif. Malgré tout, nous estimons que la spécification de l'erreur donne à l'apprenant une impulsion à étudier sa propre production et à

réfléchir sur le système de déclinaison, ce qui est un point positif. Cette problématique devrait être abordée dans une étude spécifique portant sur la réception de cette information par les apprenants.

## Perspectives

Des modifications du diagnostic se révéleront nécessaires au fur et à mesure du service de CETLEF, avec le volume croissant de différentes productions erronées. Comme une des perspectives, il serait souhaitable de l'améliorer par l'intégration d'informations syntaxiques qui permettraient d'enrichir les critères pour le choix de l'interprétation la plus probable, pour identifier automatiquement des erreurs d'accord, des erreurs dans l'attribution d'une rection à un mot, etc.

L'outil CETLEF permet des analyses de volumes de données plus importantes que celles qui ont été exploitées dans notre travail. Ceci devrait permettre d'effectuer une analyse des erreurs dans l'acquisition de la déclinaison, qui serait menée non pas uniquement sur la base de critères morphologiques, comme c'est le cas dans notre travail, mais qui pourrait prendre en compte des critères plus complexes, comme les interférences entre les deux langues.

# Annexe A

# CETLEF Plateforme auteur

Cette annexe contient les vues de l'application CETLEF qui illustrent l'exposé du fonctionnement de la plateforme auteur dans le chapitre 6 *Implémentation*, p. 226. Les images ont été prises sous le système d'exploitation Mac OS X 10.4 avec le navigateur Safari 3.1.2.

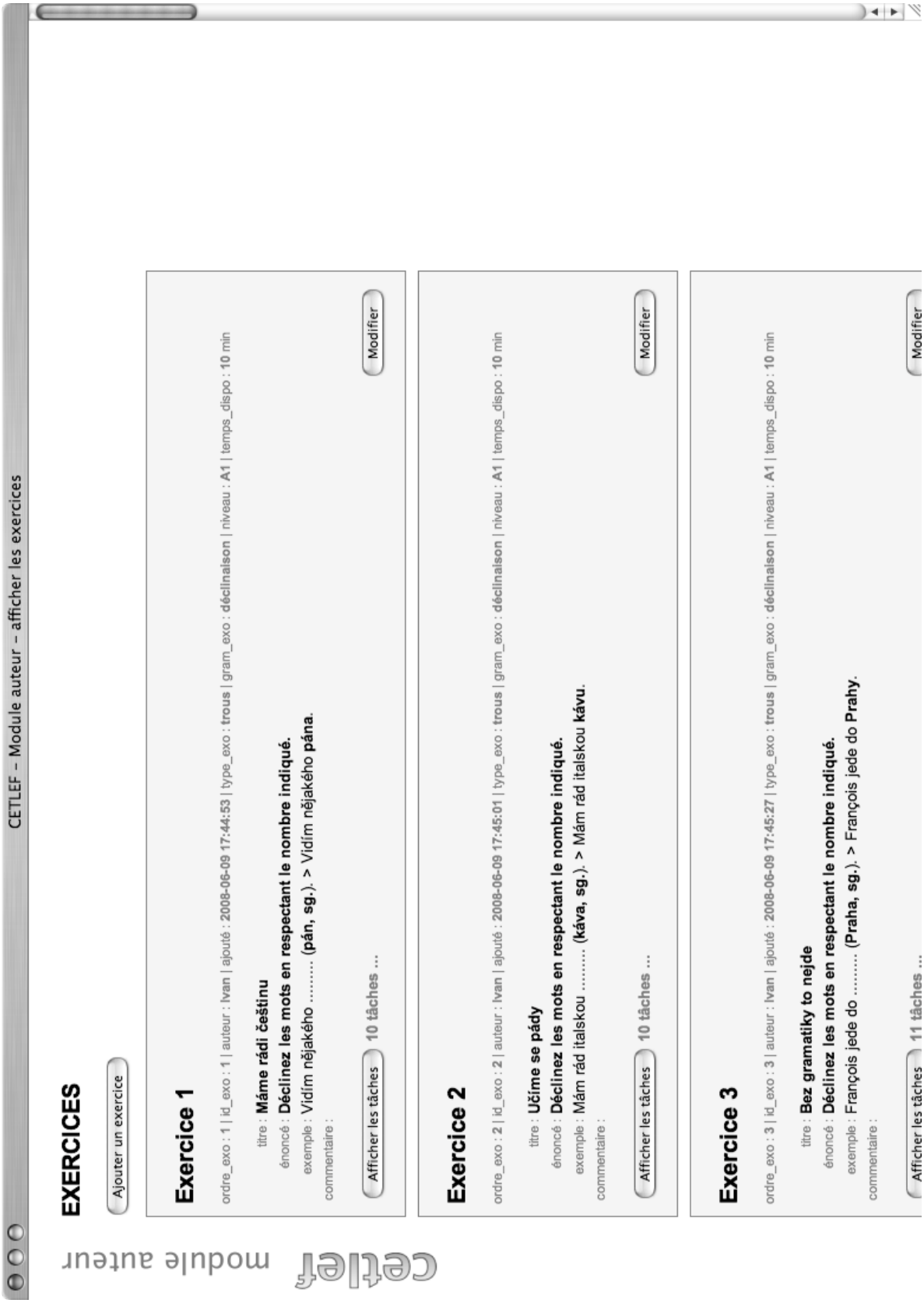



FIG. A.1 – Liste des exercices.

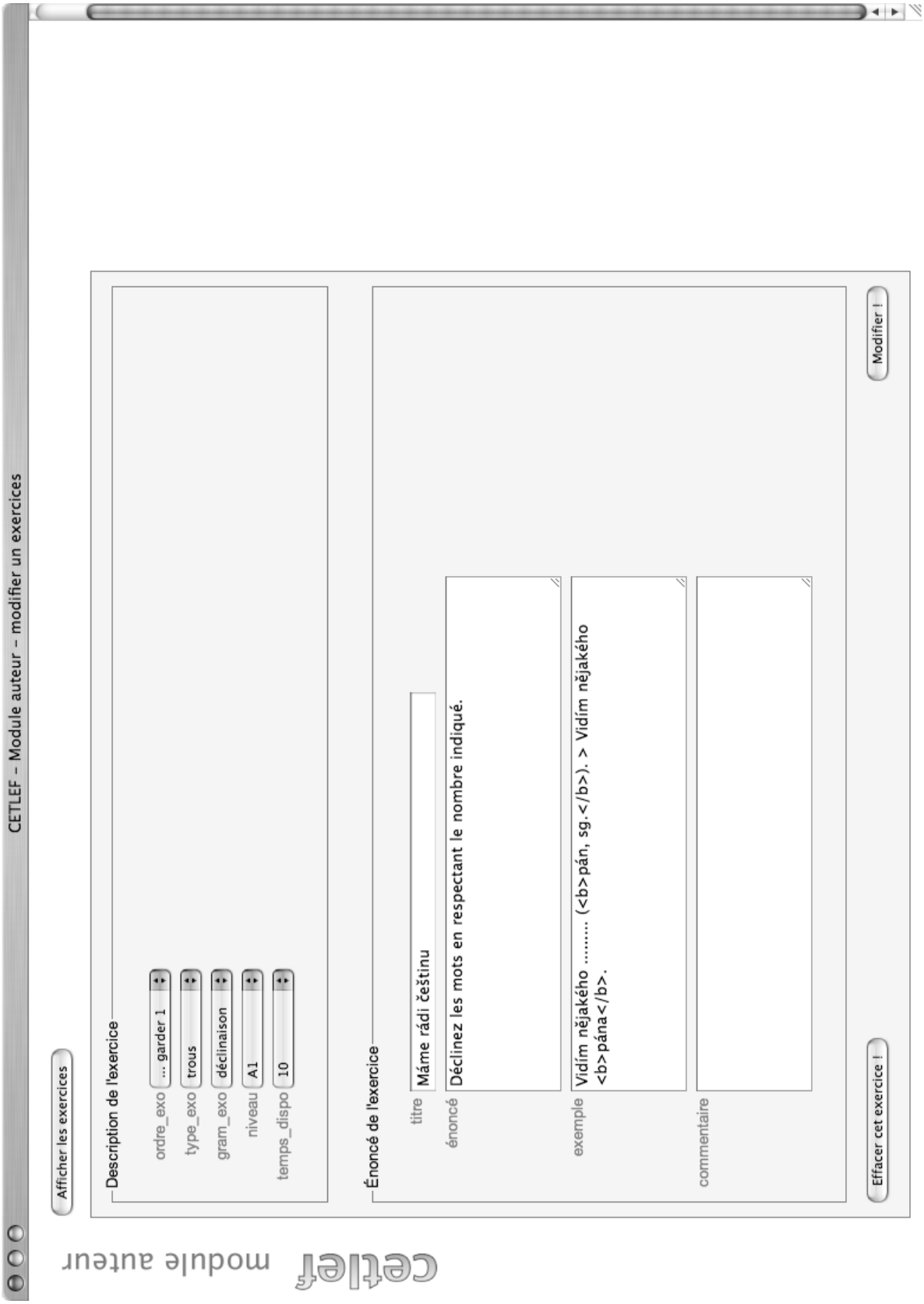



FIG. A.2 – Ajouter / modifier un exercice.

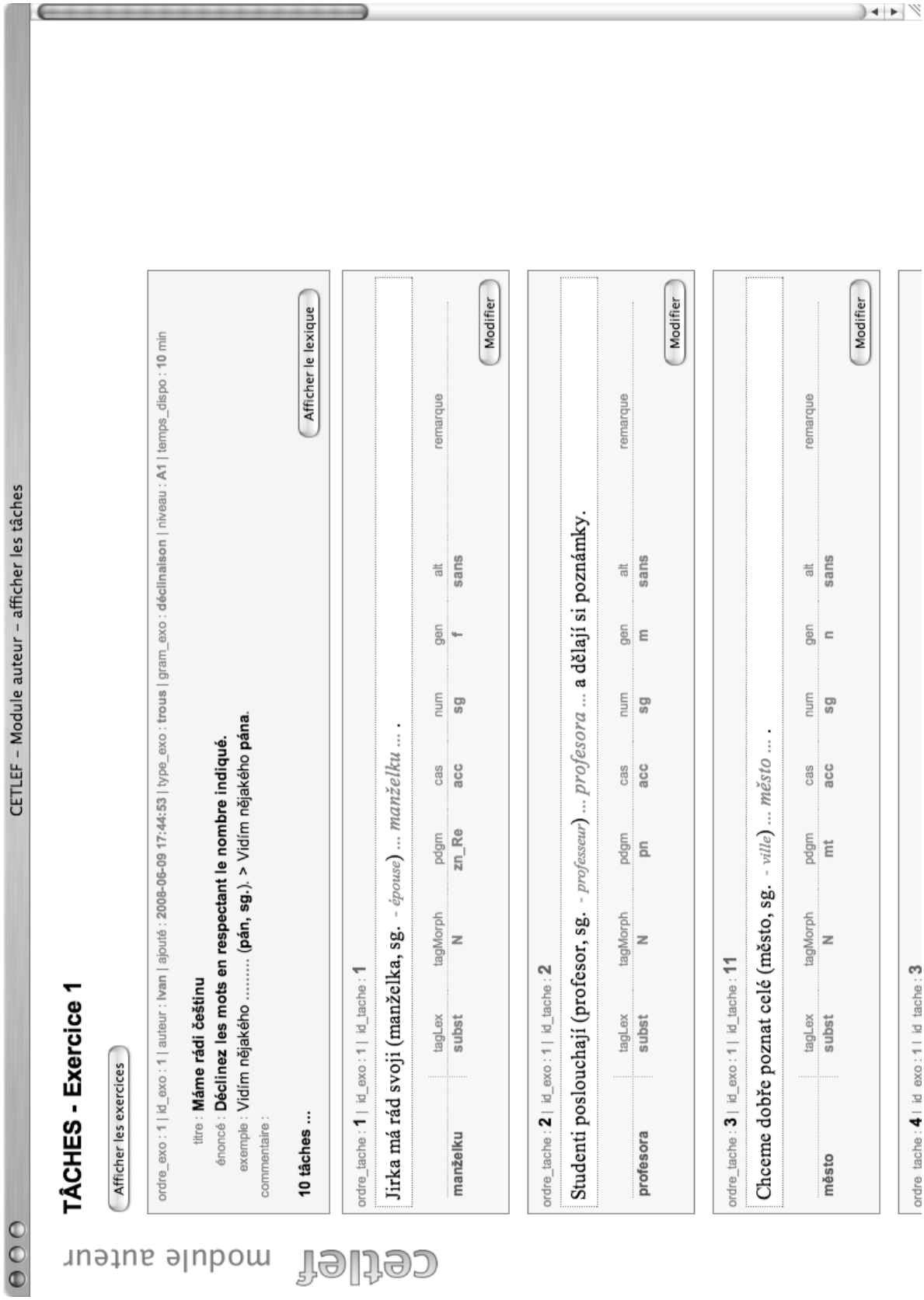



FIG. A.3 – Liste des tâches dans un exercice.

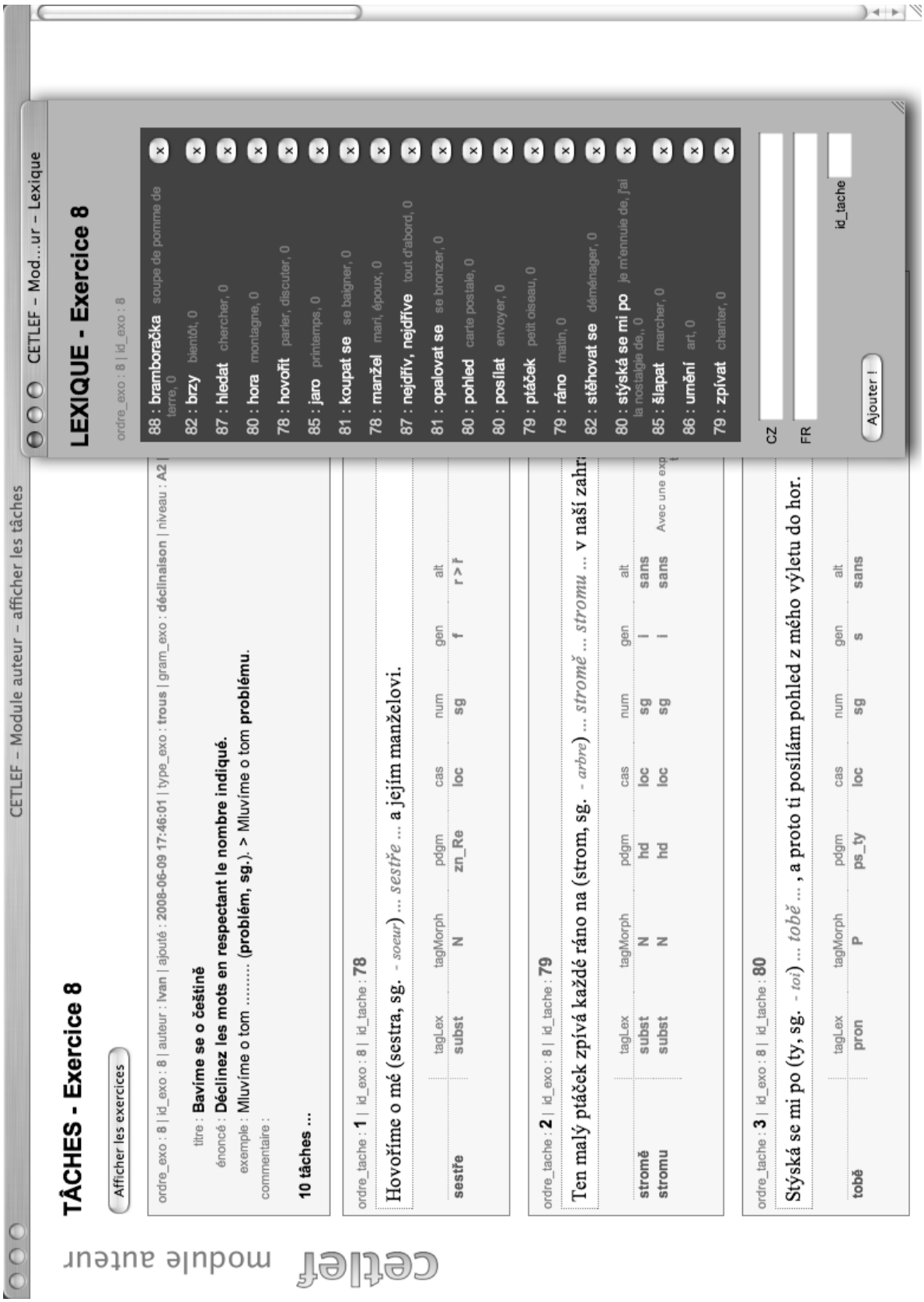



FIG. A.4 – Liste des tâches avec le lexique.

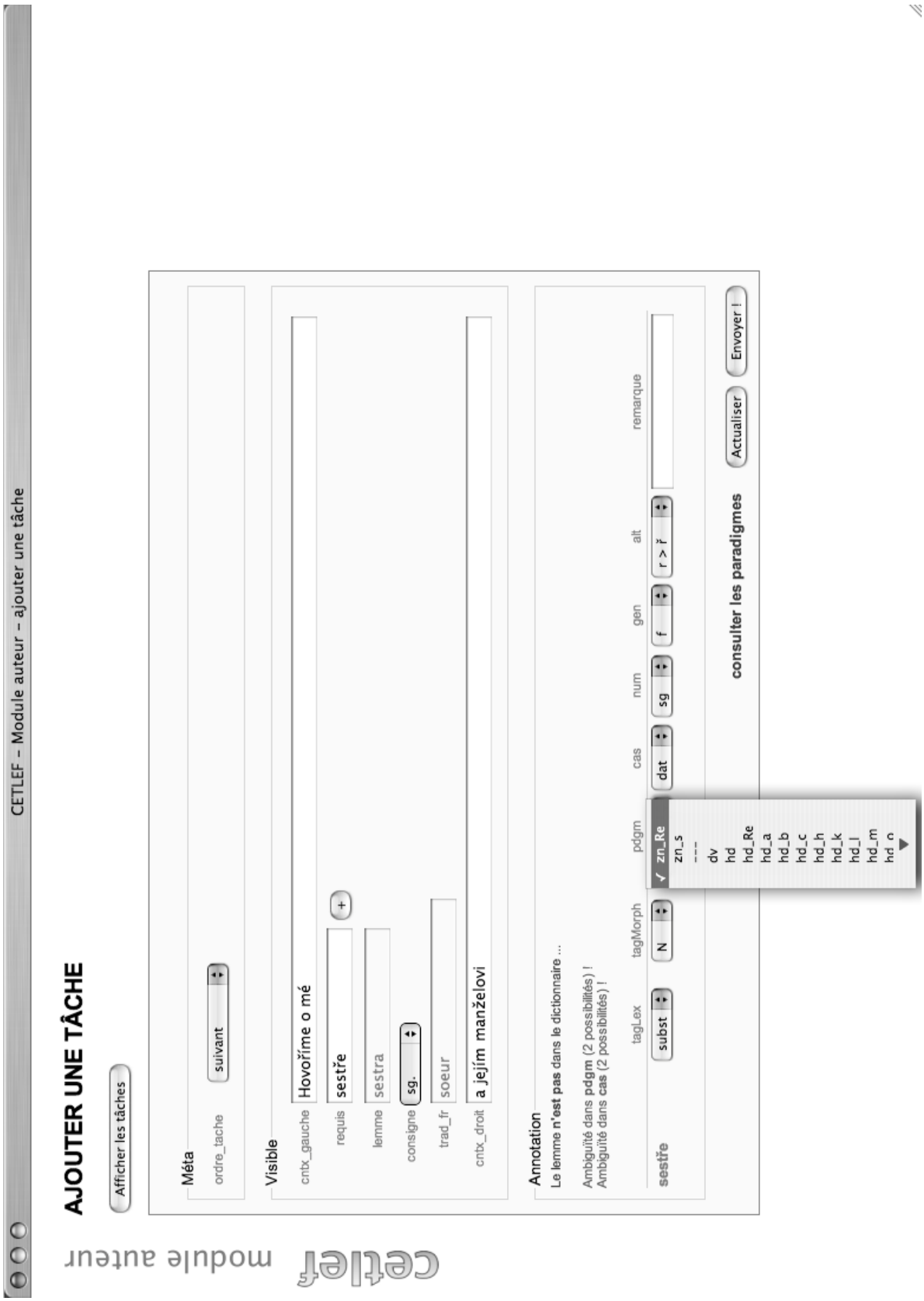



FIG. A.5 – Ajout d'une tâche avec la sortie de l'annotation.

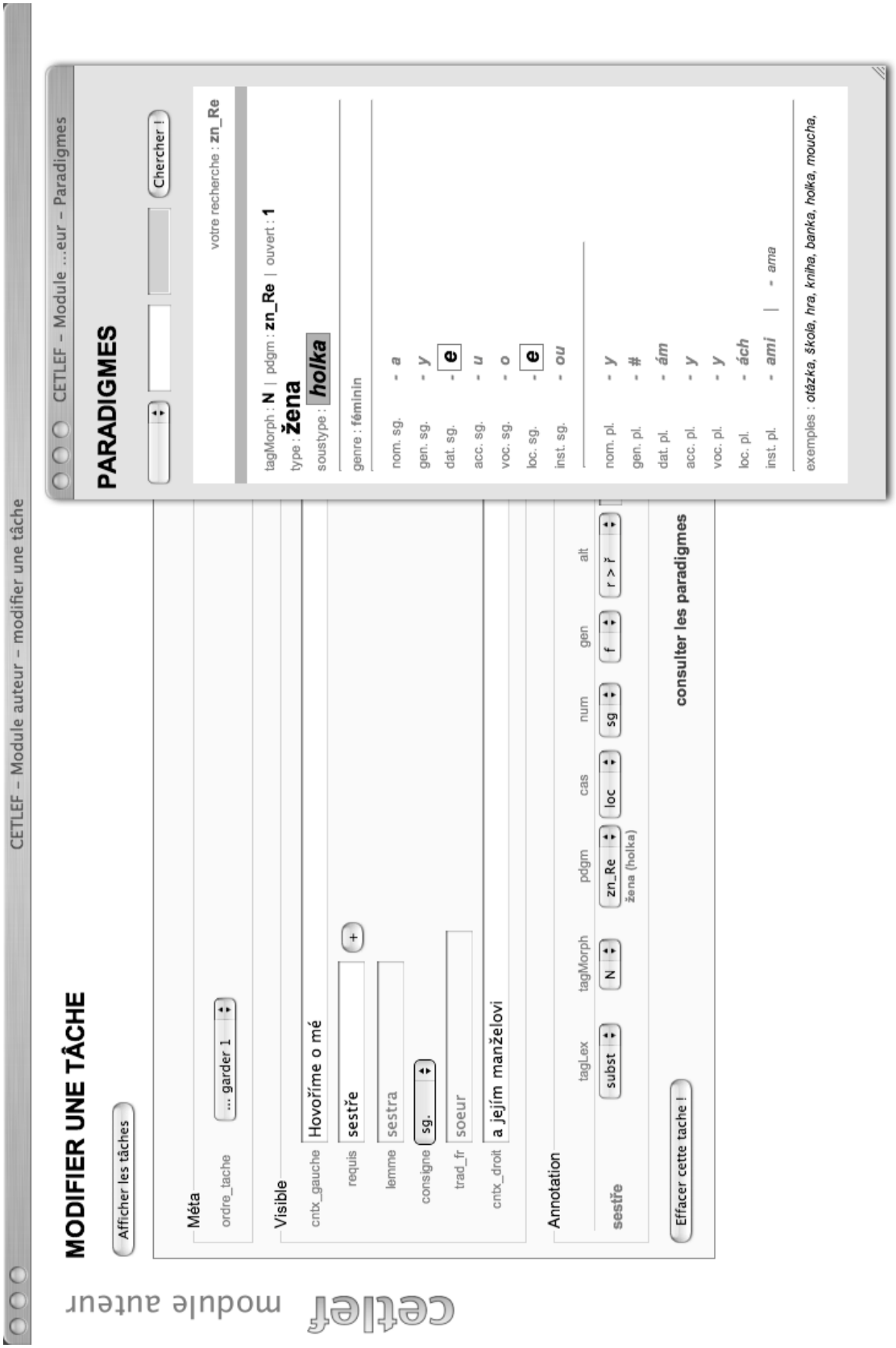



FIG. A.6 – Ajout d'une tâche avec la consultation des paradigmes.

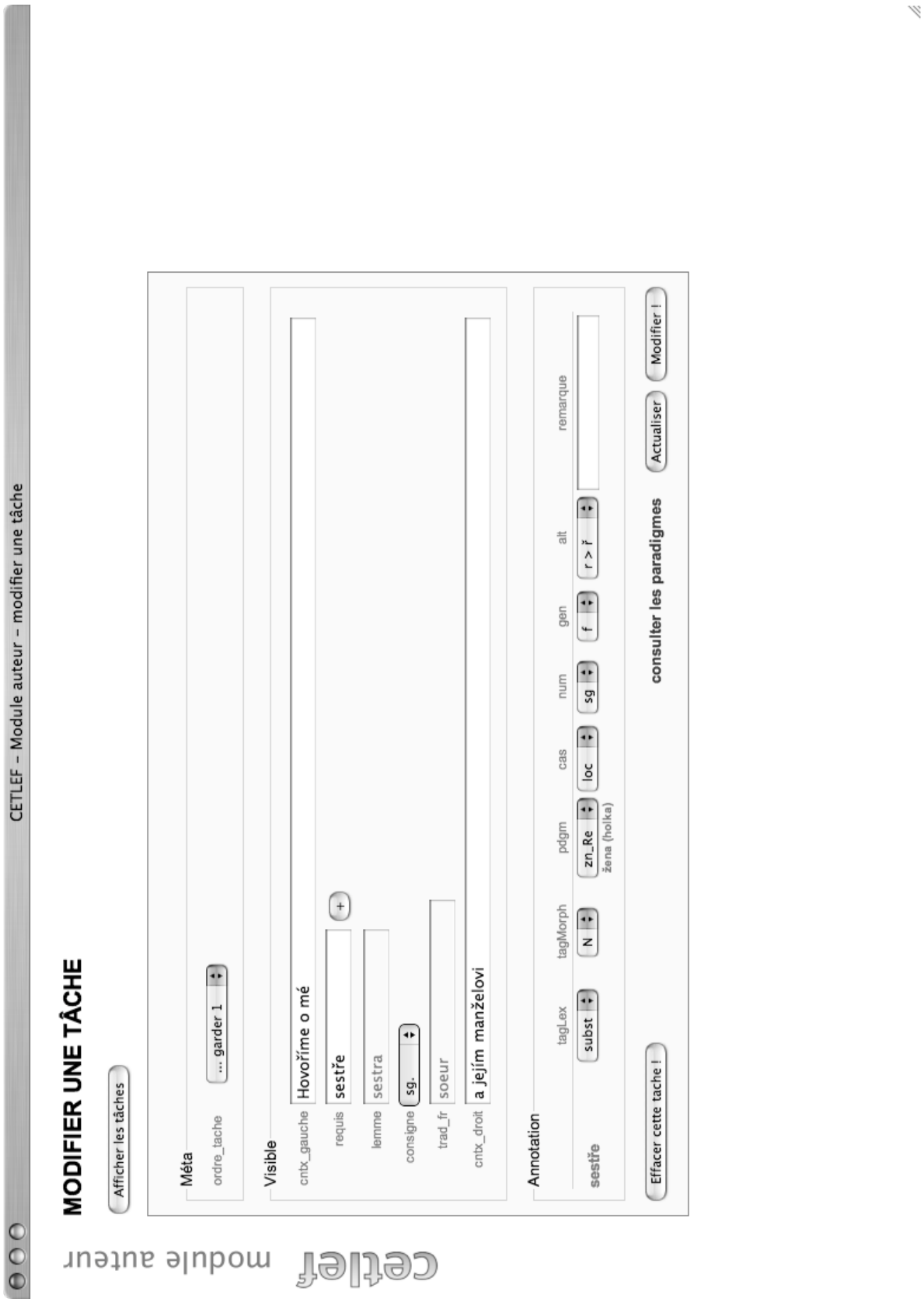



FIG. A.7 – Ajout d'une tâche complété.

# Annexe B

# CETLEF Plateforme apprenant

Cette annexe contient les vues de l'application CETLEF qui illustrent l'exposé du fonctionnement de la plateforme auteur dans le chapitre 6 *Implémentation*, p. 228. Les images ont été prises sous le système d'exploitation Mac OS X 10.4 avec le navigateur Safari 3.1.2.

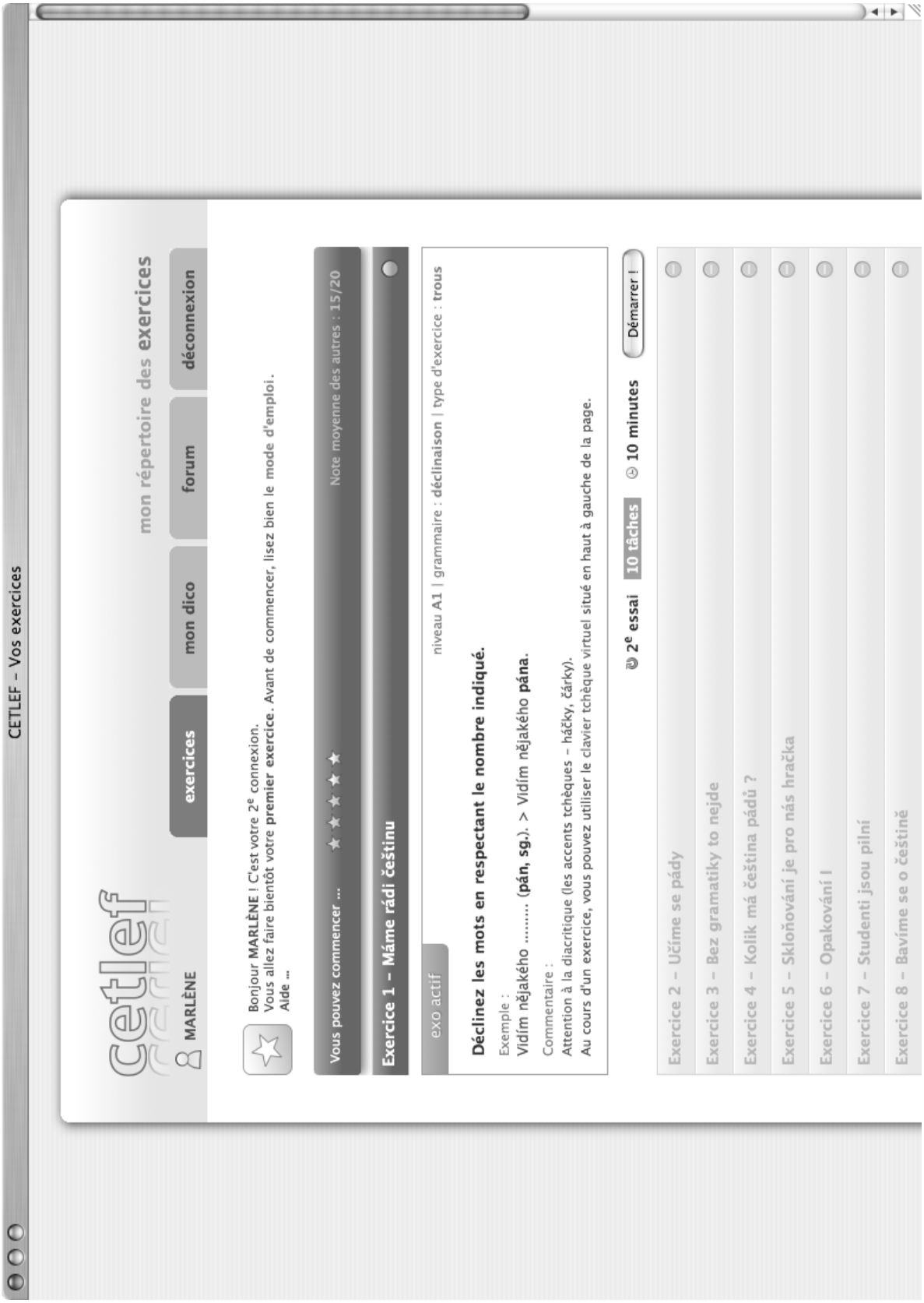



FIG. B.1 – Première connexion sur CETLEF.

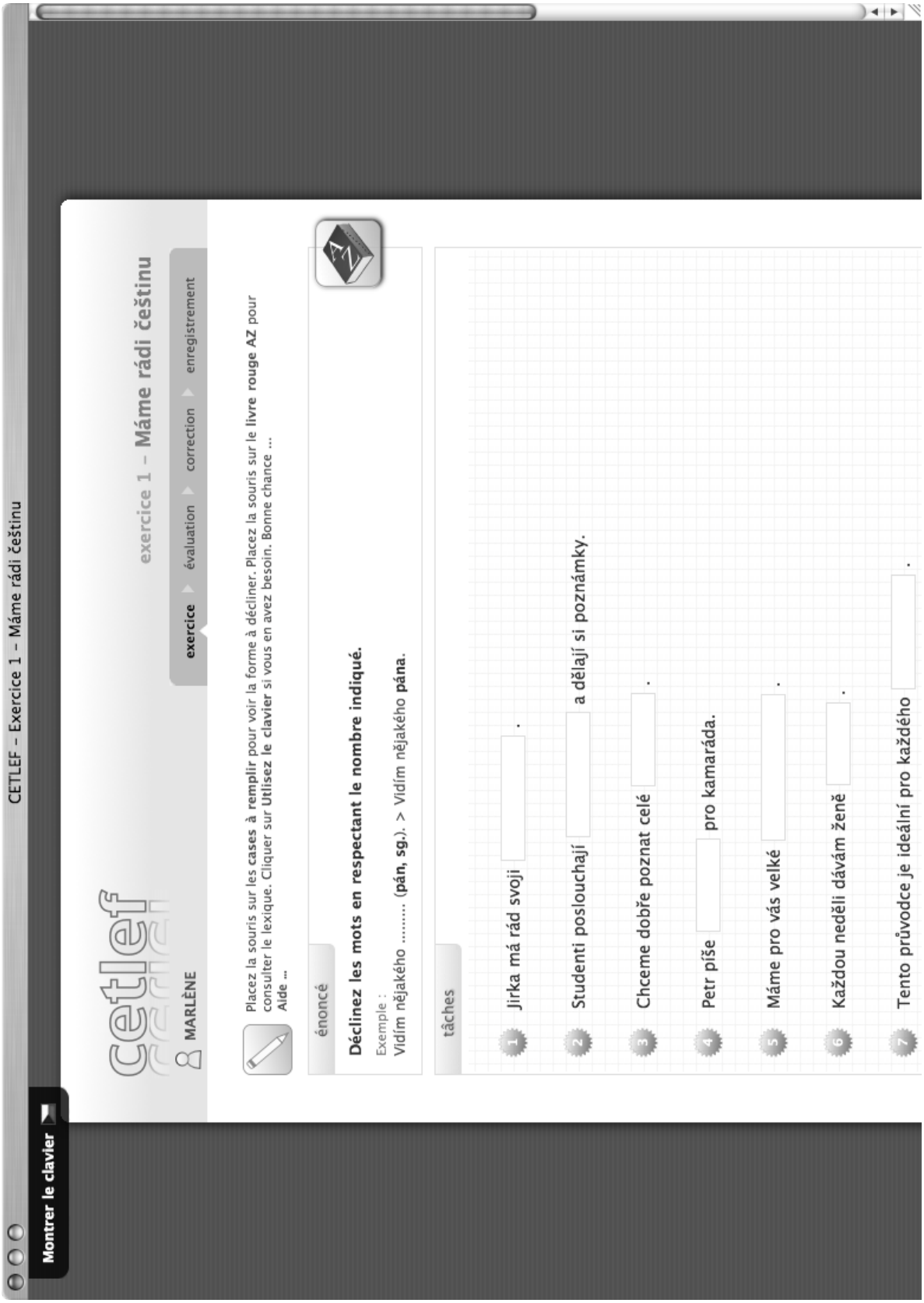



FIG. B.2 – Un exercice en cours.

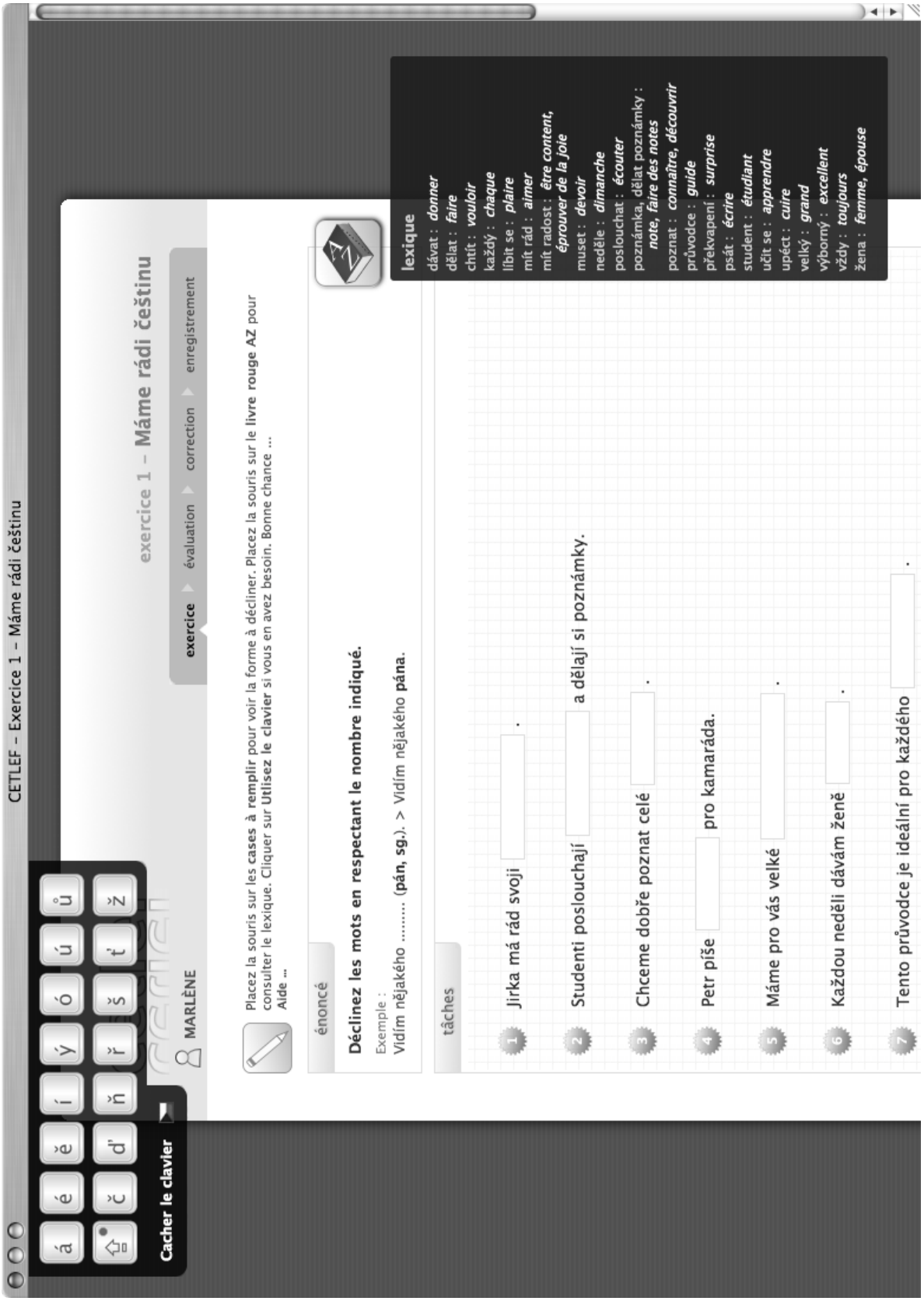



FIG. B.3 – Un exercice en cours avec le clavier et le lexique.

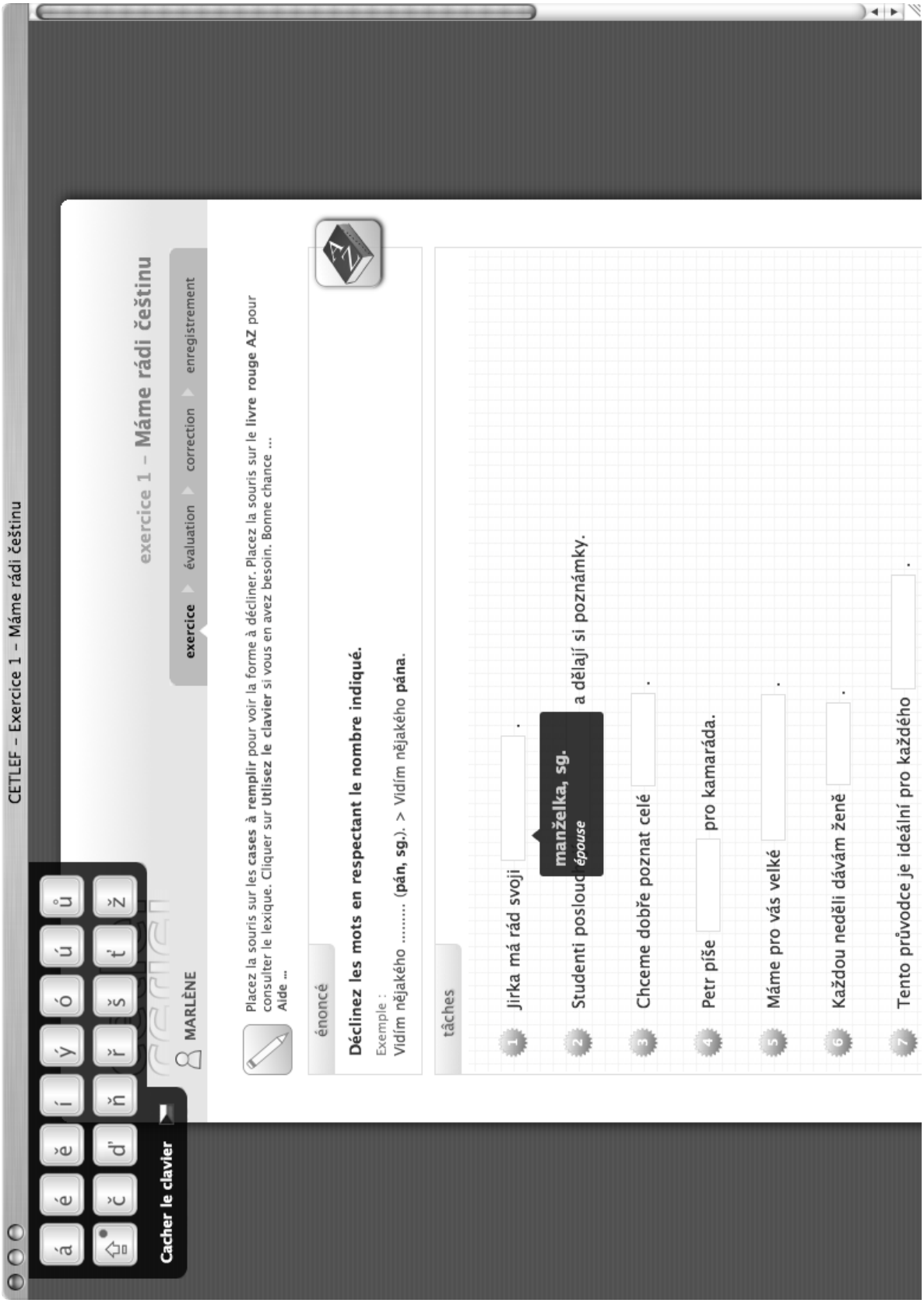



FIG. B.4 – Une tâche avec le mot à décliner.

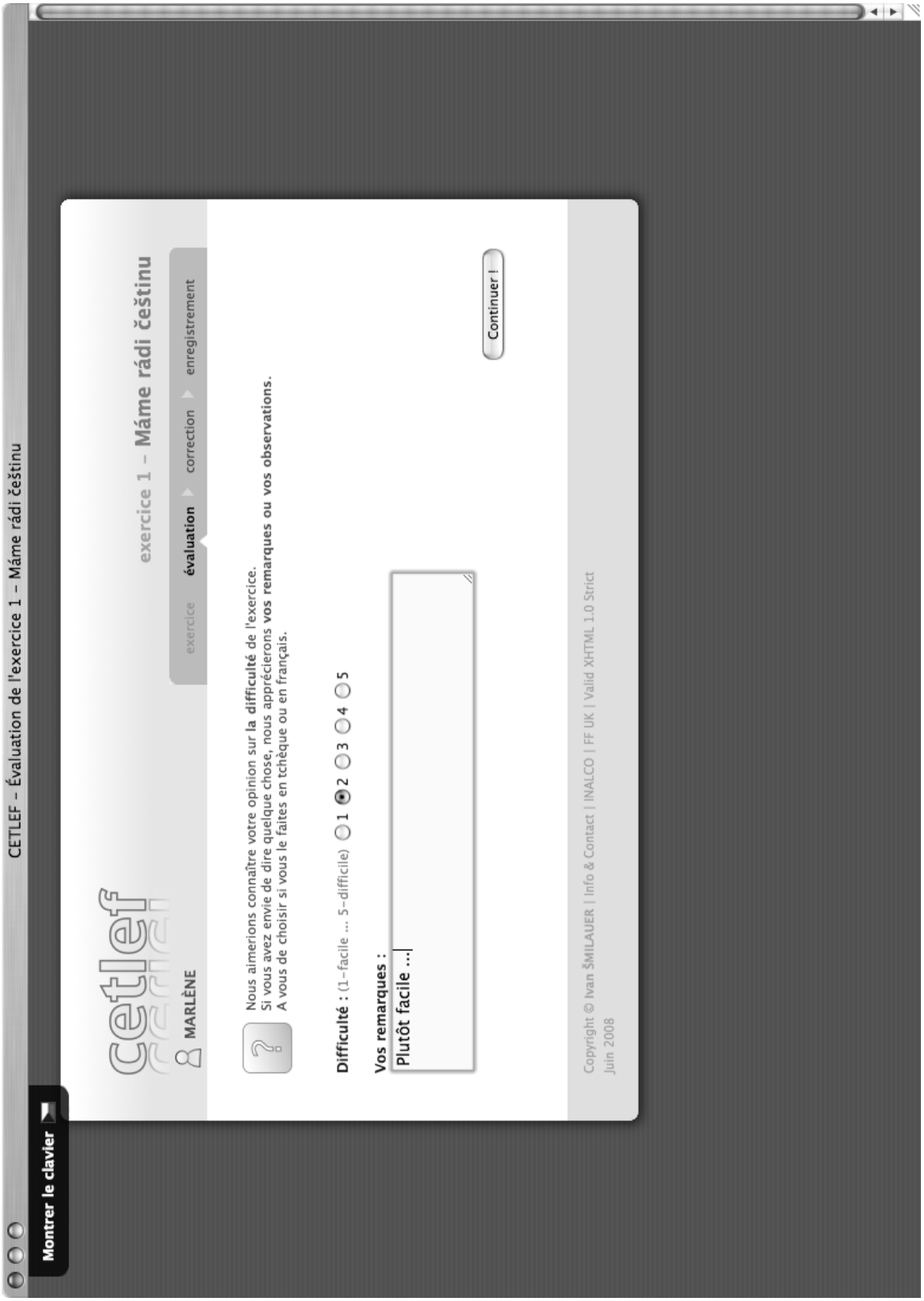



FIG. B.5 – L'apprenant évalue la difficulté de l'exercice.

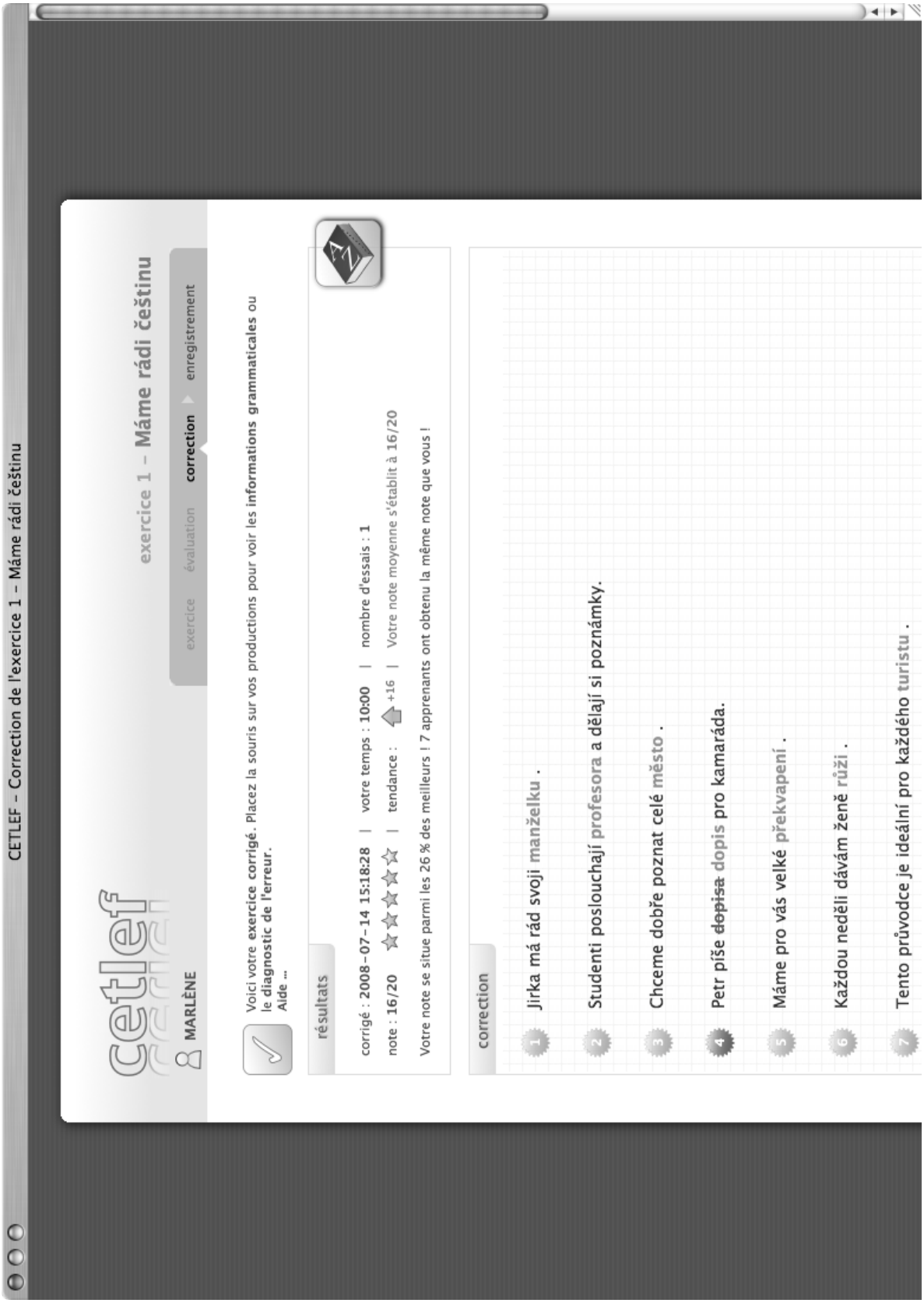



FIG. B.6 – L'exercice corrigé.

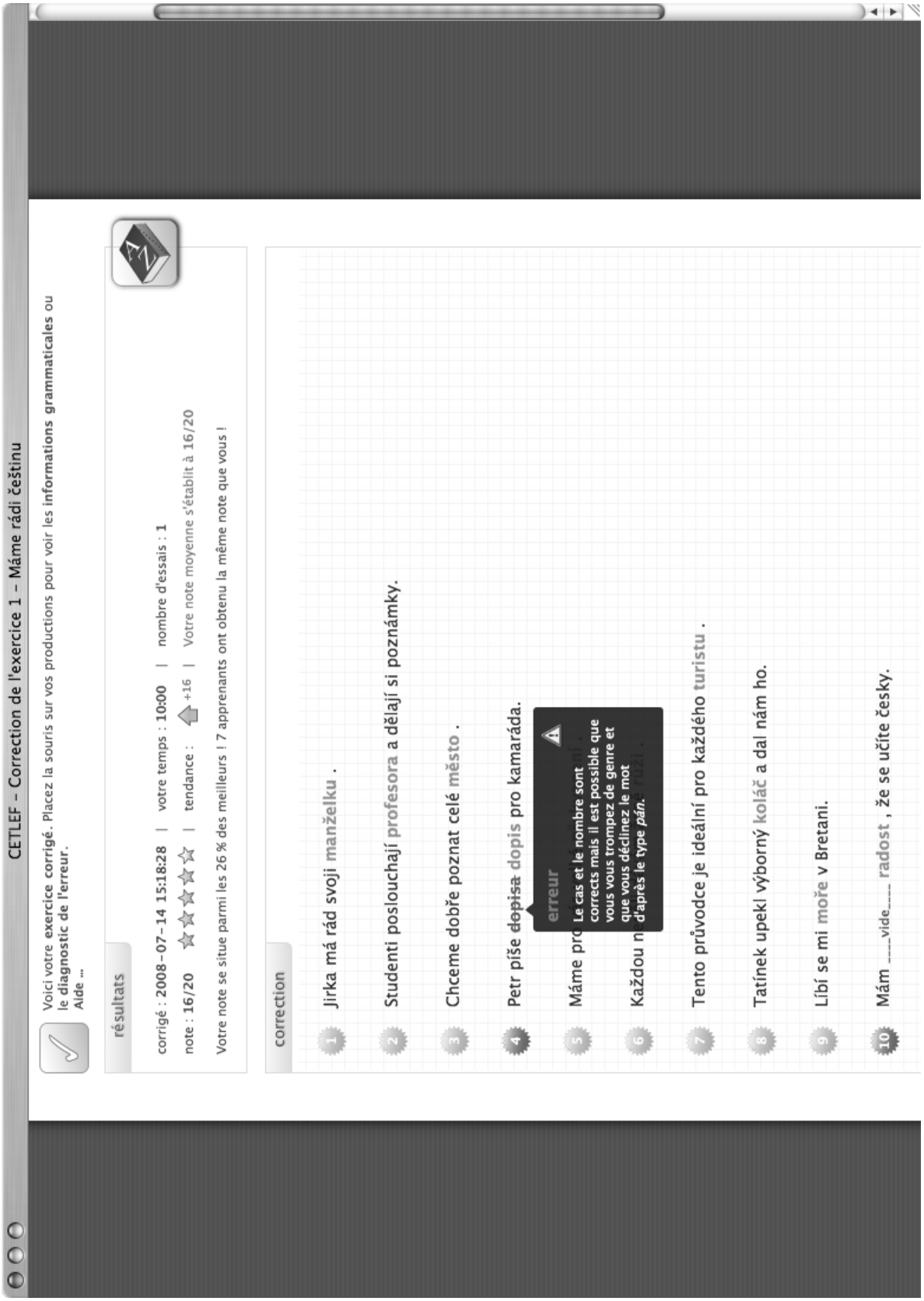



FIG. B.7 – Le diagnostic d'une erreur (exemple 1).

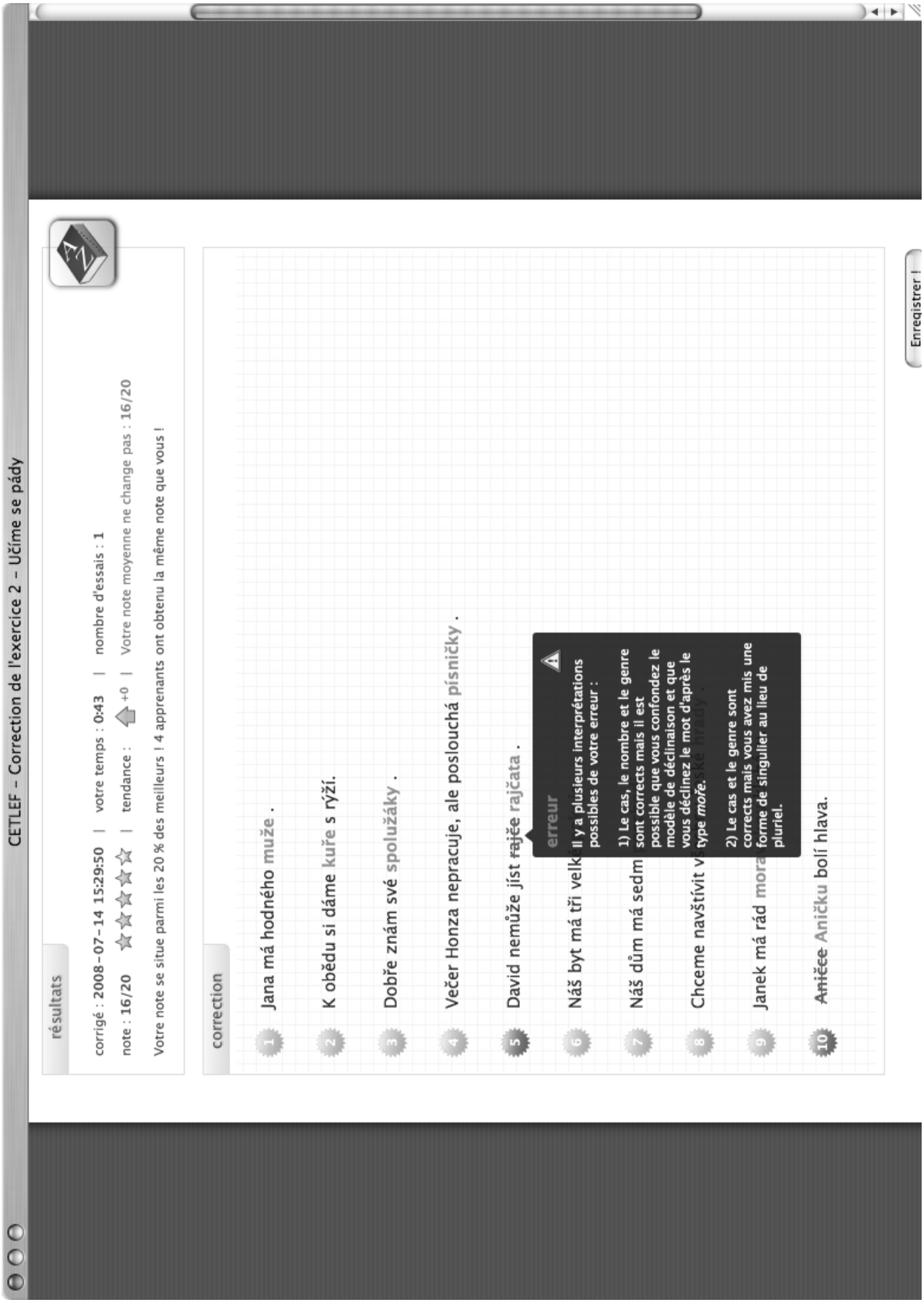



FIG. B.8 – Le diagnostic d'une erreur (exemple 2).

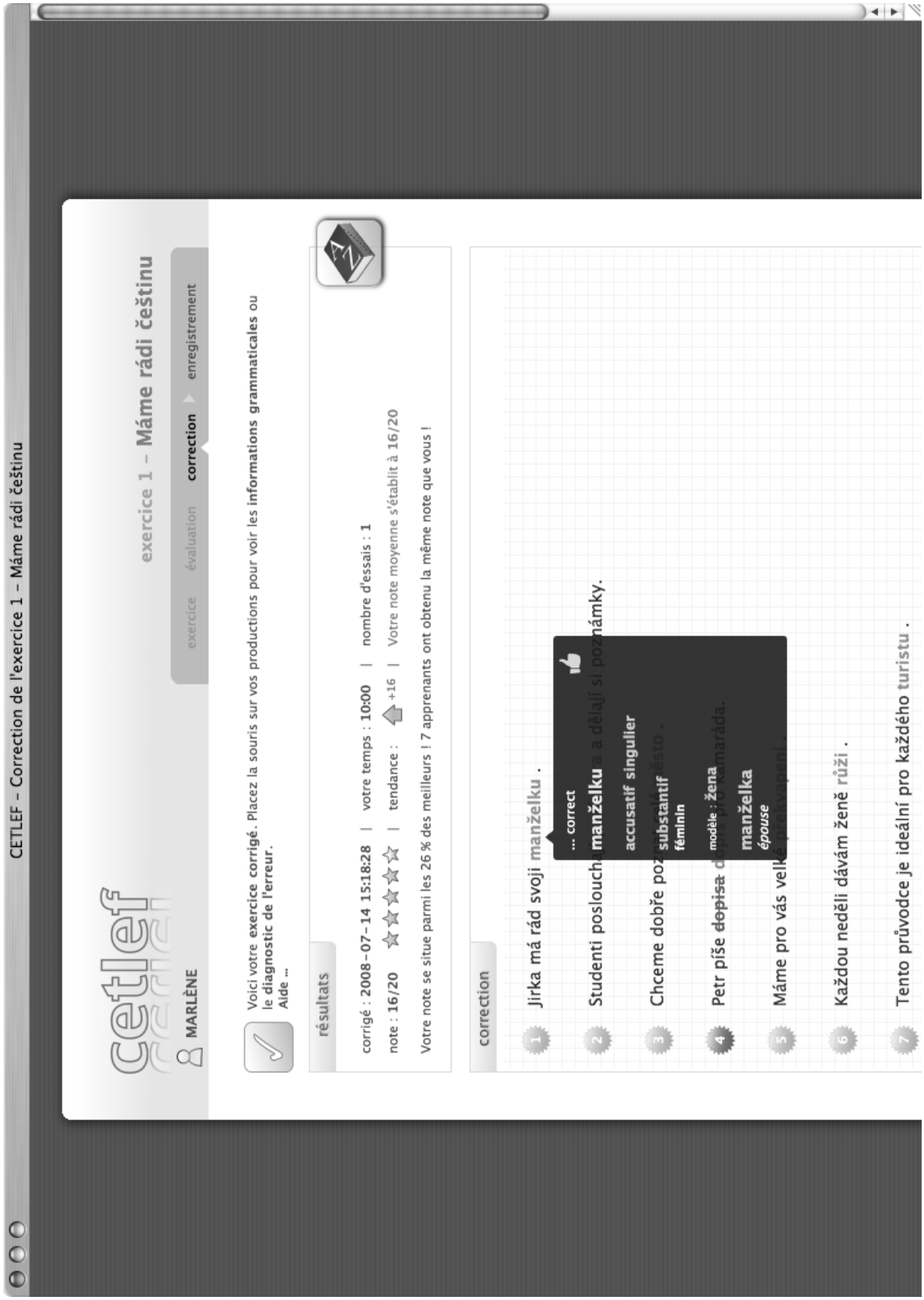



FIG. B.9 – Les informations grammaticales sur la forme requise (exemple 1).

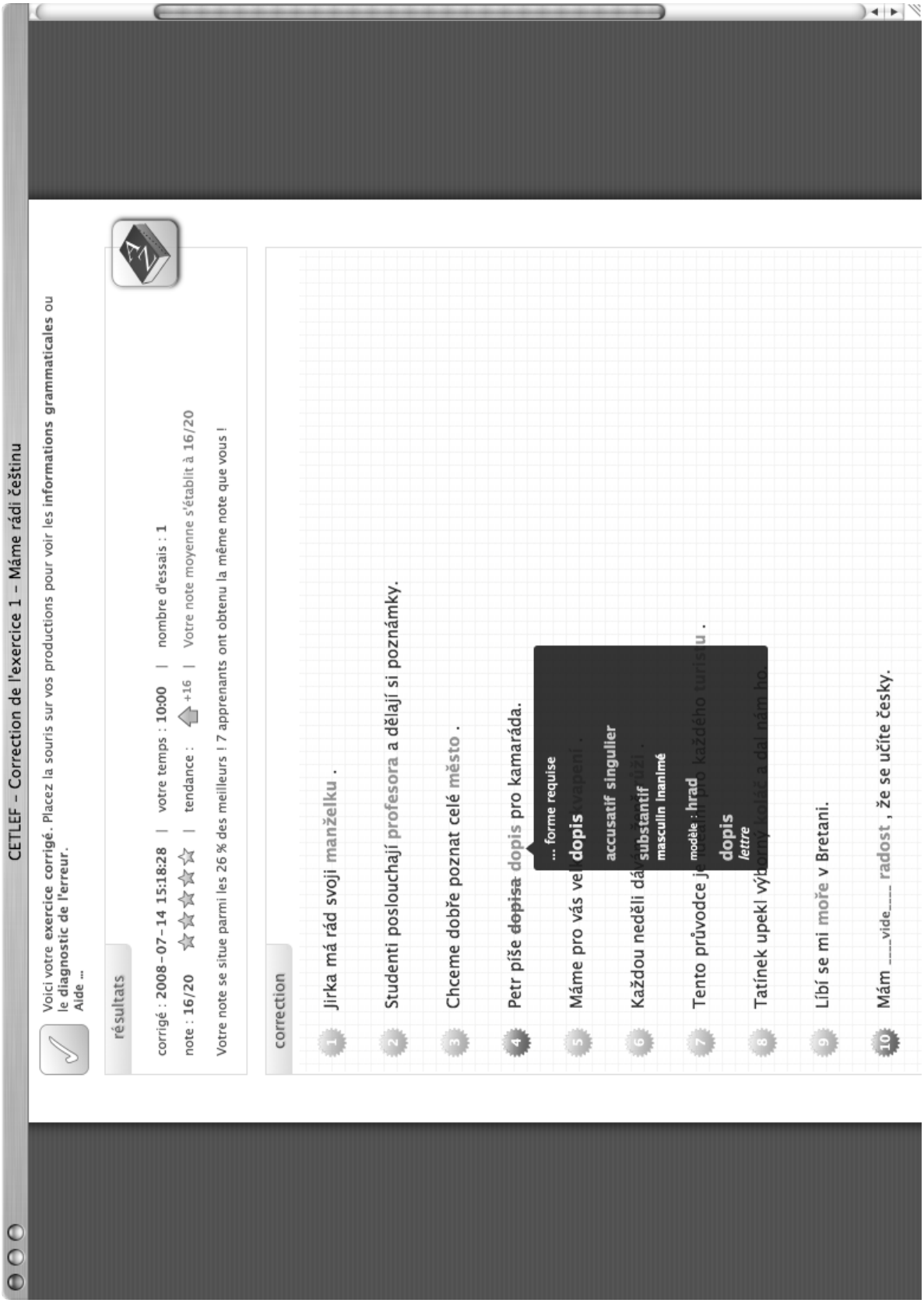



FIG. B.10 – Les informations grammaticales sur la forme requise (exemple 2).

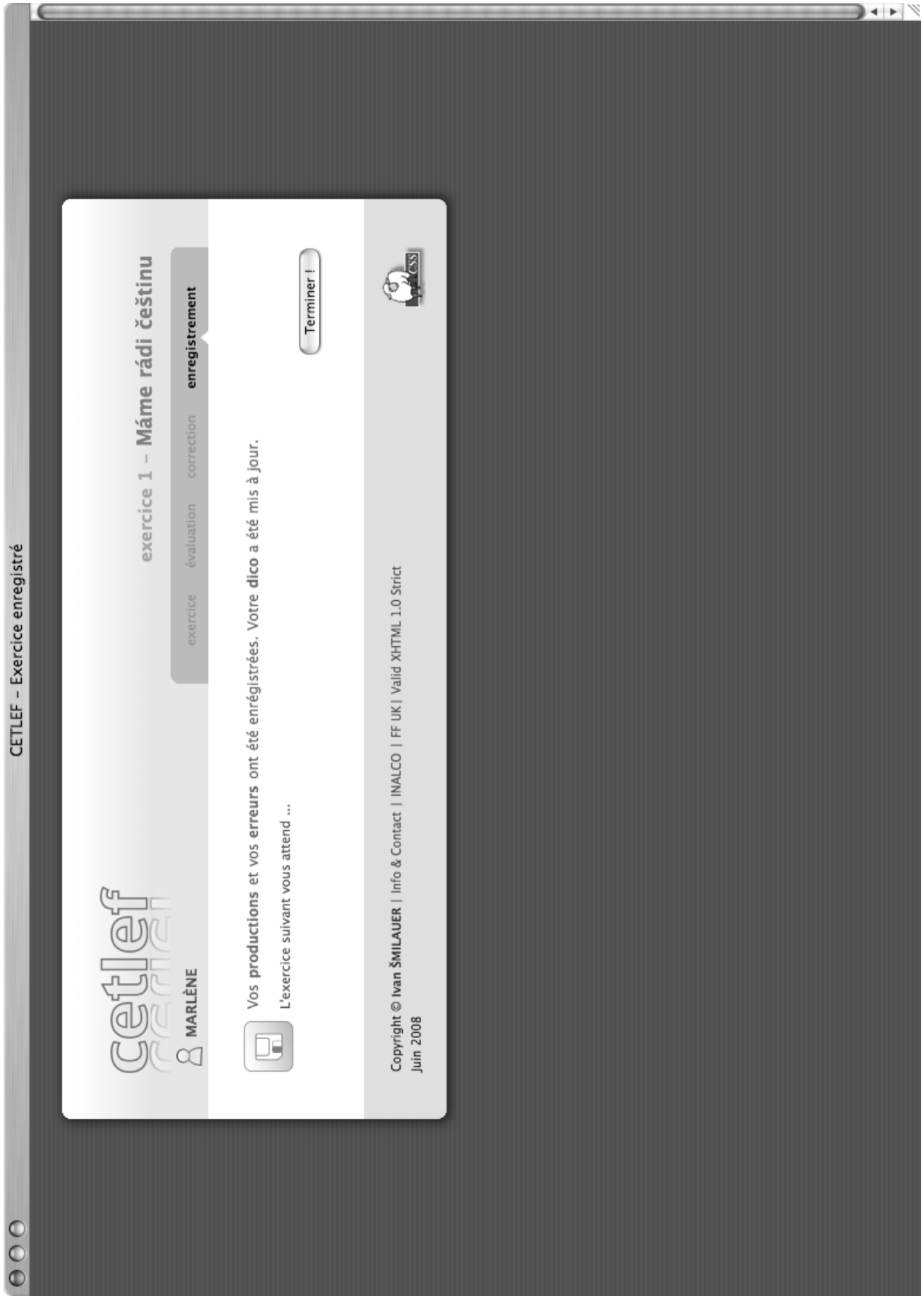



FIG. B.11 – L'enregistrement des productions dans la base de données.

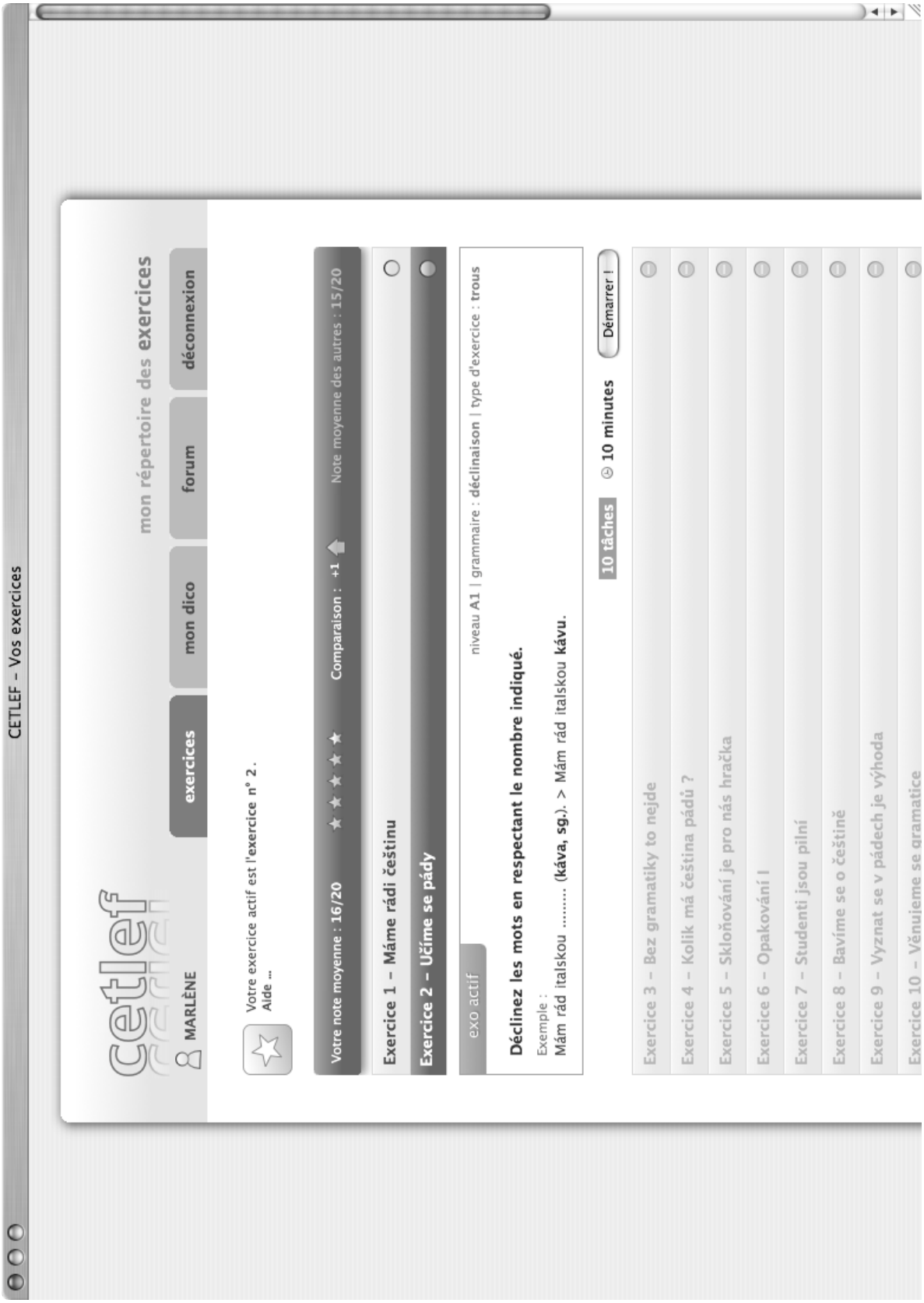



FIG. B.12 – L'exercices suivant.

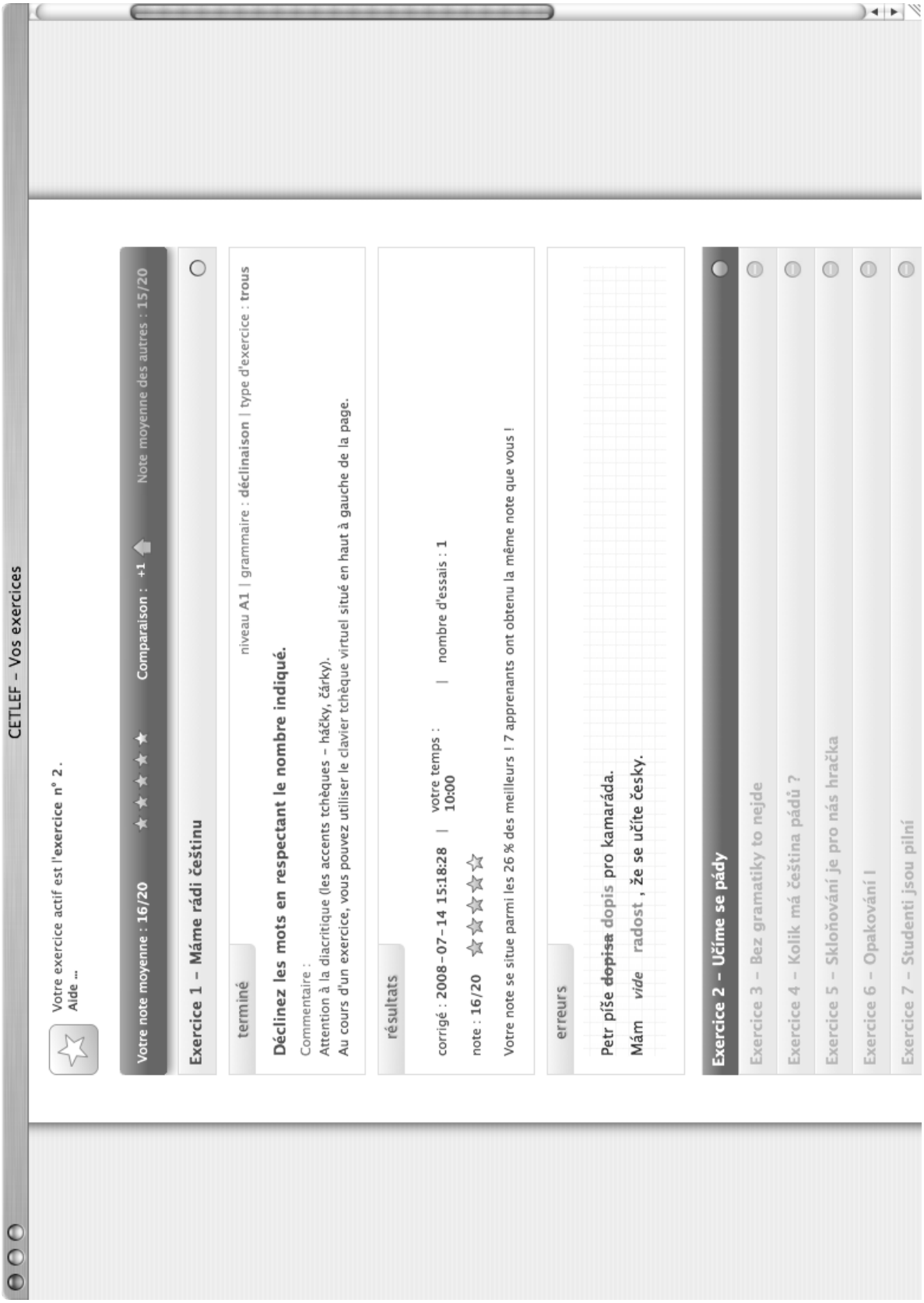



FIG. B.13 – Les informations sur un exercice terminé (exemple 1).

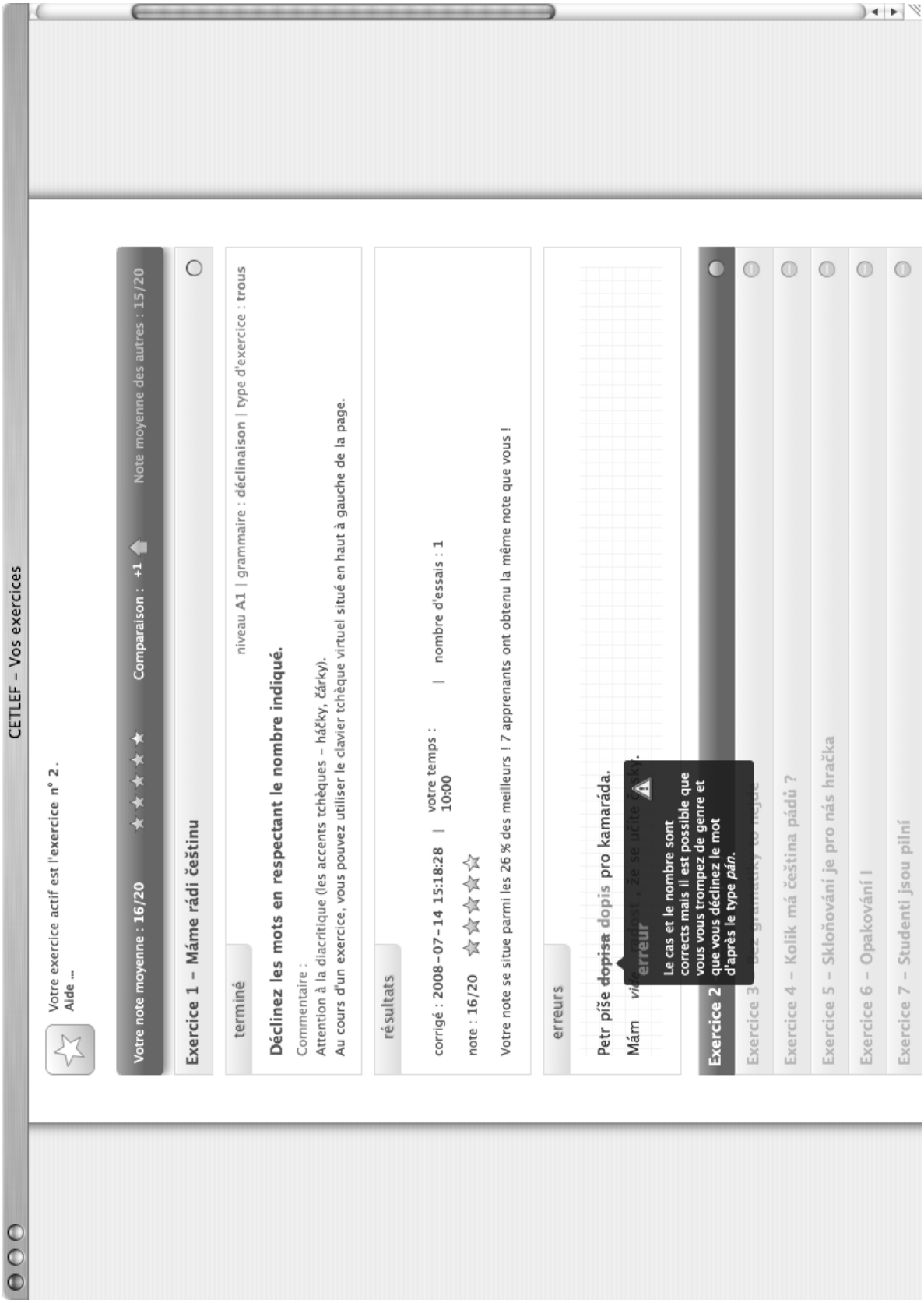



FIG. B.14 – Les informations sur un exercice terminé (exemple 2).

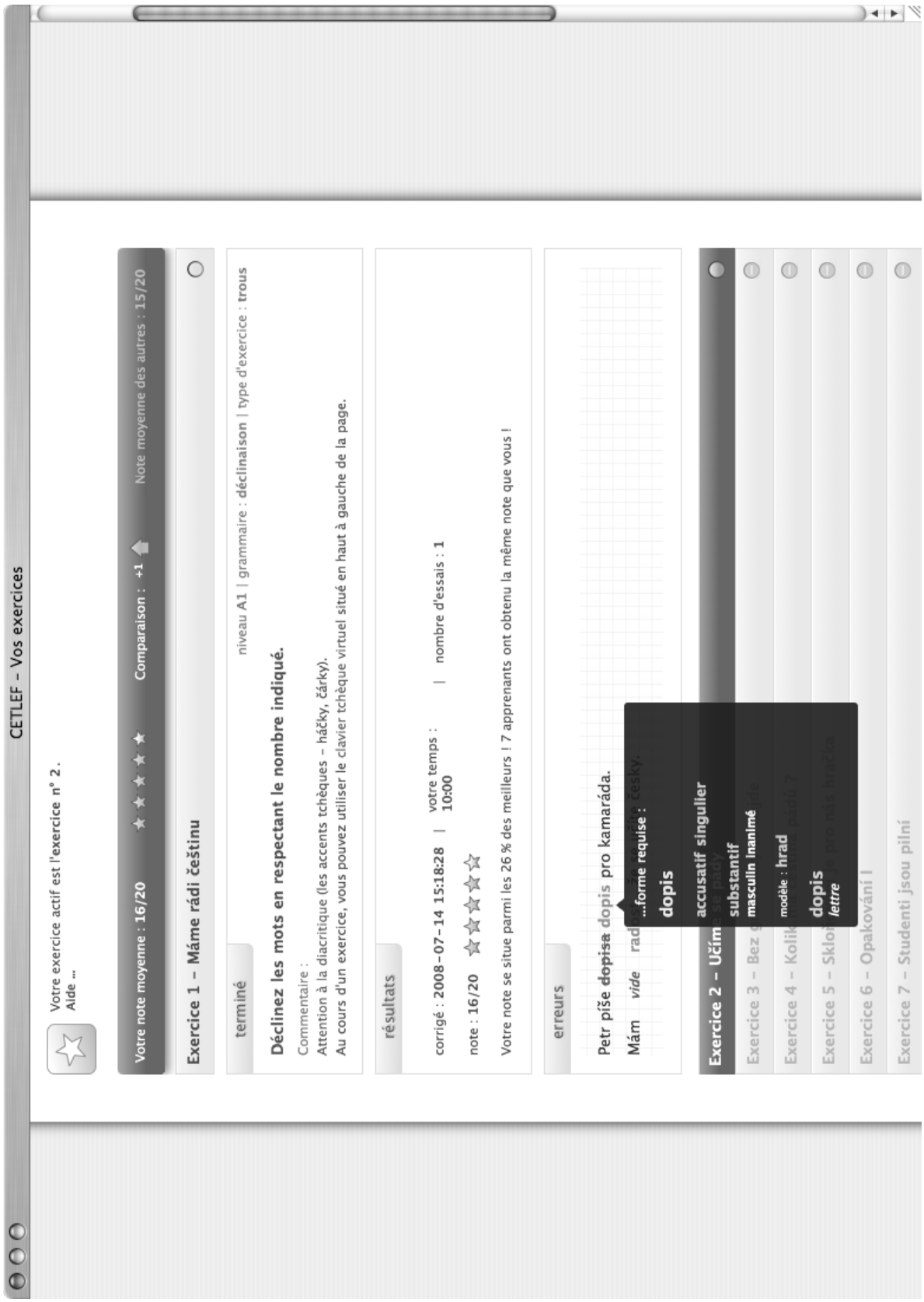



FIG. B.15 – Les informations sur un exercice terminé (exemple 3).

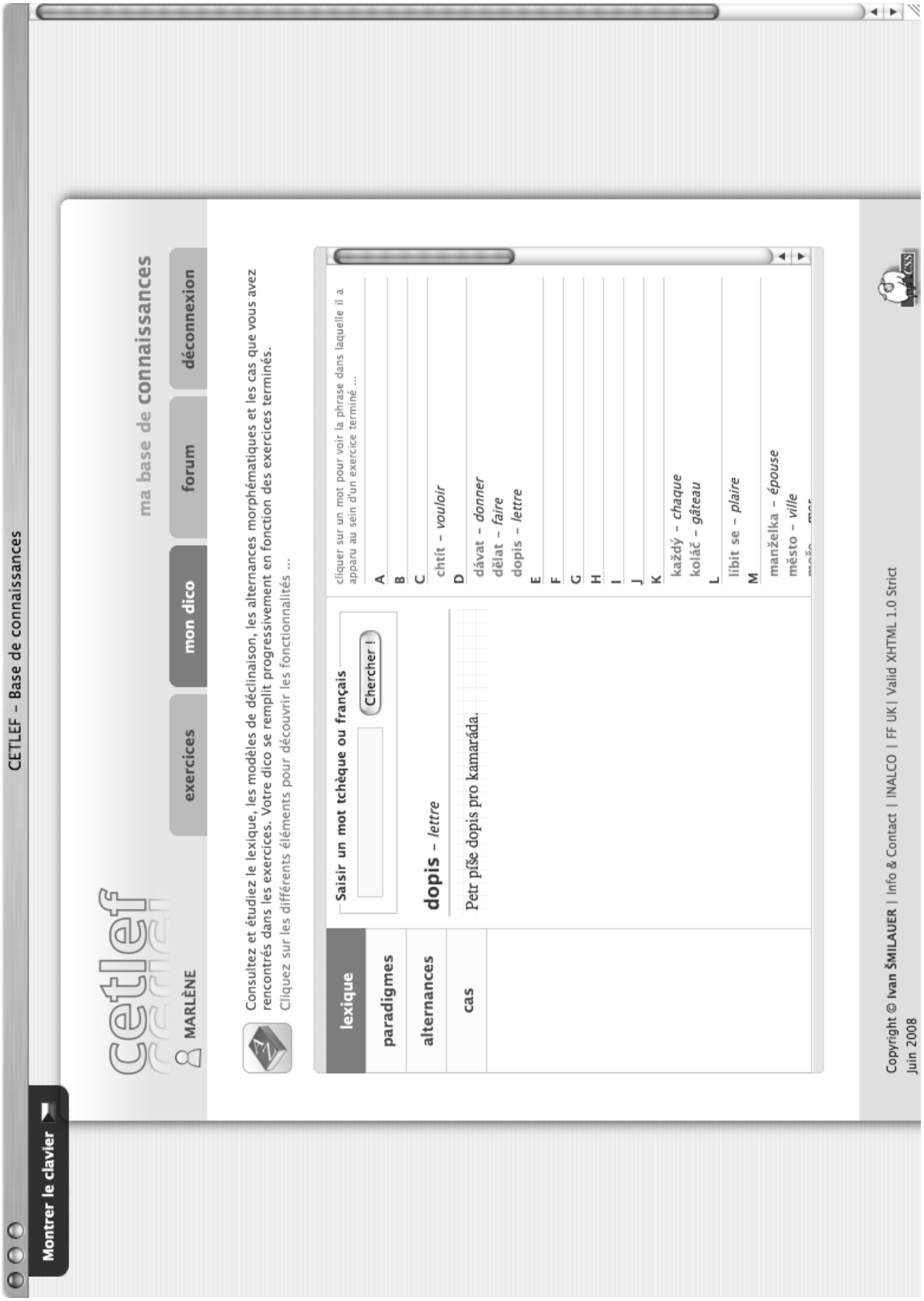



FIG. B.16 – La base de connaissances – lexique.

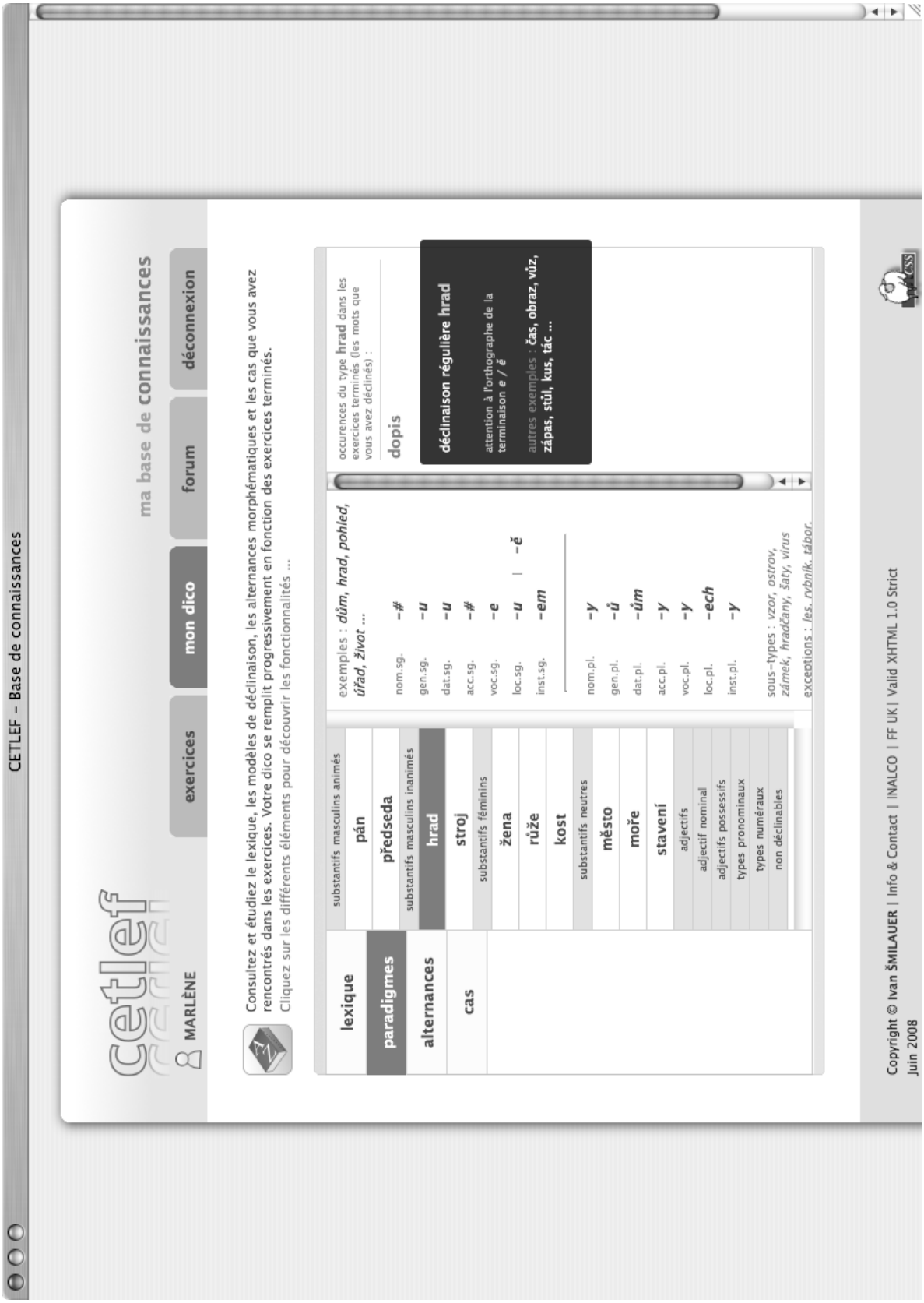



FIG. B.17 – La base de connaissances – paradigmes.

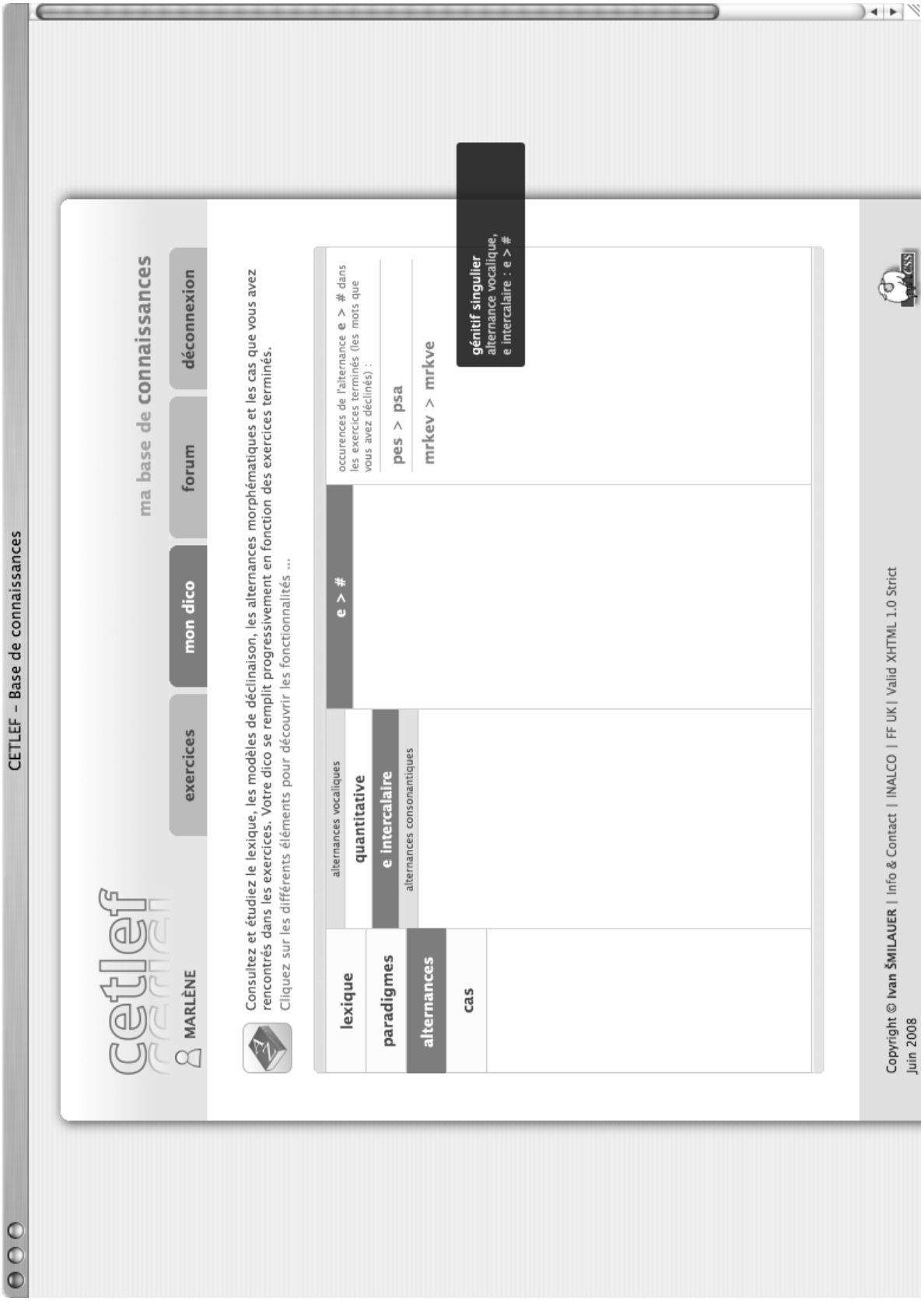



FIG. B.18 – La base de connaissances – alternances.

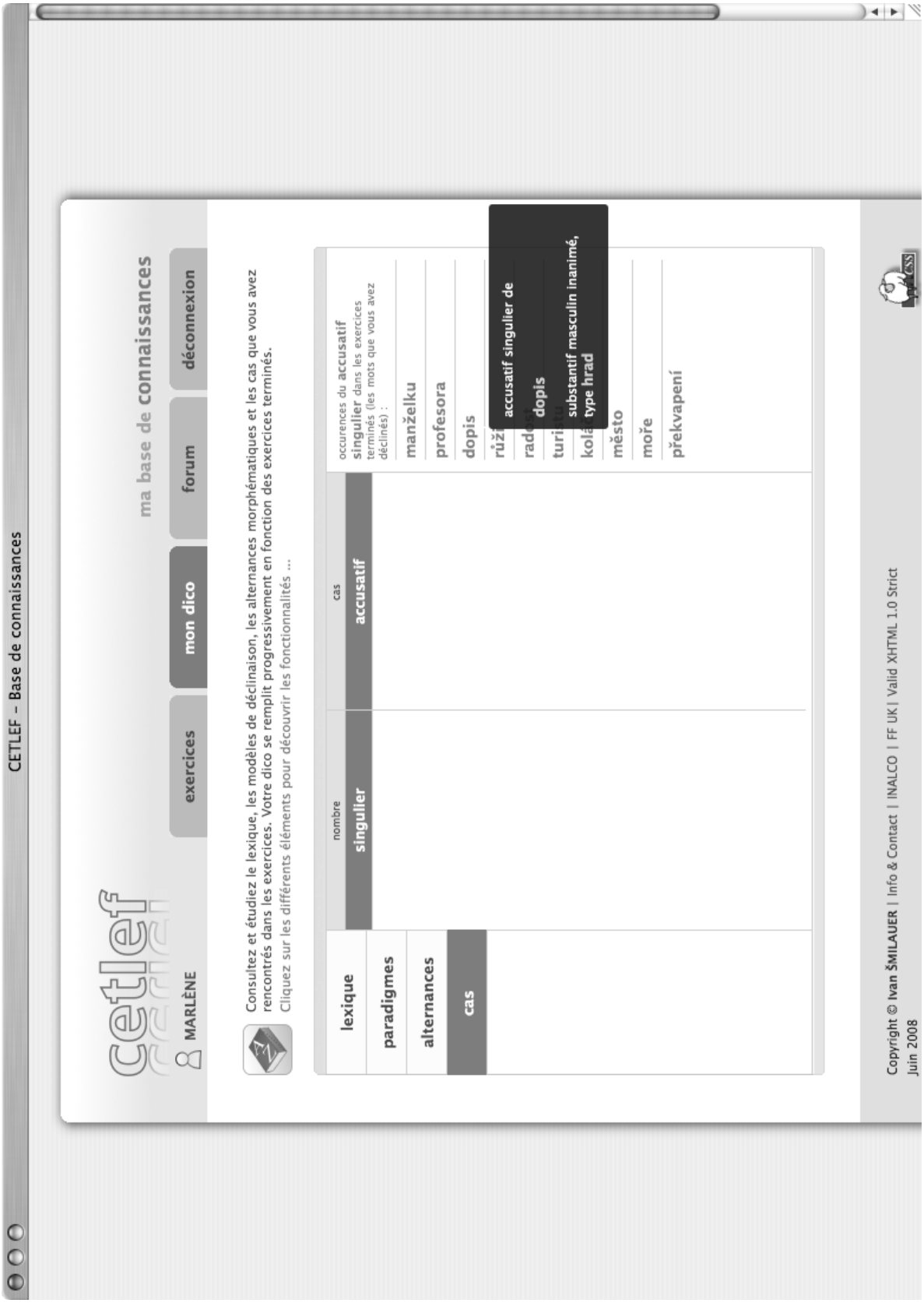



FIG. B.19 – La base de connaissances – cas.

# Annexe C

# Paradigmes

Cette annexe contient les paradigmes de déclinaison présentés dans le chapitre *Annotation*, section 3.4.3, p. 65–69. Elle a été générée automatiquement par un script qui retranscrit le fichier *pdgm.xml* (voir la section 3.4.2, p. 65) au format LaTeX. Pour la représentation des sous-types, uniquement, les désinences qui diffèrent par rapport au sous-type modèle sont affichés.

## C.1 Flexion nominale

### Type *pán*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| pán | pn | 1 | 1 | pán, ministr, prezident, autor, premiér, bratr, pes |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | a |
| dat | sg | m | fnct | u |
| dat | sg | m | fnct | ovi |
| acc | sg | m | | a |
| voc | sg | m | | e |
| loc | sg | m | fnct | u |
| loc | sg | m | fnct | ovi |
| inst | sg | m | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | i |
| nom | pl | m | fnct | ové |
| gen | pl | m | | ů |
| dat | pl | m | | ům |
| dat | pl | m | reg | um |
| acc | pl | m | | y |
| voc | pl | m | fnct | i |
| voc | pl | m | fnct | ové |
| loc | pl | m | | ech |
| inst | pl | m | | y |
| inst | pl | m | reg | ama |

TAB. C.1 – Modèle *pán*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| hoch | pn_h | 1 | 1 | hoch, kluk, číšník, zpěvák, chirurg, alkoholik, kuřák, úředník |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| voc | sg | m | | R | u |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | m | | R | ích |
| loc | pl | m | reg | A | ách |

TAB. C.2 – Sous-type *hoch*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| bůh | pn_b | 0 | 1 | bůh |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | m | | R | ích |

TAB. C.3 – Sous-type *bůh*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| člověk | pn_c | 0 | 1 | člověk |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | fnct | X | i |
| nom | pl | m | fnct | X | ové |
| gen | pl | m | | X | ů |
| dat | pl | m | | X | ům |
| dat | pl | m | reg | X | um |
| acc | pl | m | | X | y |
| voc | pl | m | fnct | X | ové |
| voc | pl | m | fnct | X | i |
| loc | pl | m | | X | ích |
| loc | pl | m | reg | X | ách |
| inst | pl | m | | X | y |
| inst | pl | m | reg | X | ama |

TAB. C.4 – Sous-type *člověk*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| syn | pn_s | 0 | 1 | syn, duch |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| voc | sg | m | | R | u |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.5 – Sous-type *syn*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| občan | pn_o | 1 | 1 | občan, host, křesťan, venkovan, dvořan |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | reg | V | i |
| nom | pl | m | | R | é |
| nom | pl | m | reg | V | ové |
| voc | pl | m | reg | V | i |
| voc | pl | m | | R | é |
| voc | pl | m | reg | V | ové |

TAB. C.6 – Sous-type *občan*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| host | pn_t | 0 | 1 | host |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | fnct | A | é |
| nom | pl | m | reg | V | ové |
| gen | pl | m | fnct | A | í |
| voc | pl | m | fnct | A | é |
| voc | pl | m | reg | V | ové |

TAB. C.7 – Sous-type *host*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| manžel | pn_m | 0 | 1 | manžel, anděl |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| voc | sg | m | | R | i |
| loc | sg | m | fnct | A | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | fnct | R | é |
| acc | pl | m | fnct | A | e |
| voc | pl | m | fnct | R | é |
| loc | pl | m | fnct | A | ích |
| inst | pl | m | fnct | A | i |
| inst | pl | m | reg | A | ema |

TAB. C.8 – Sous-type *manžel*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| génius | pn_g | 1 | 1 | génius, Celsius |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | m | | R | us |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | X | i |
| acc | pl | m | | R | e |
| voc | pl | m | | X | i |
| loc | pl | m | fnct | A | ích |
| inst | pl | m | | R | i |
| inst | pl | m | reg | A | ema |

TAB. C.9 – Sous-type *génius*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| džigolo | pn_d | 1 | 0 | džigolo, Picasso, Hugo, Silvio, Sančo |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | m | | R | o |
| voc | sg | m | | R | o |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | X | i |
| voc | pl | m | | X | i |

TAB. C.10 – Sous-type *džigolo*

## Type *hrad*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| hrad | hd | 1 | 1 | život, dům, úřad, pohled, hrad |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | u |
| dat | sg | i | | u |
| acc | sg | i | | # |
| voc | sg | i | | e |
| loc | sg | i | fnct | u |
| loc | sg | i | fnct | ě |
| inst | sg | i | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | y |
| gen | pl | i | | ů |
| dat | pl | i | | ům |
| dat | pl | i | reg | um |
| acc | pl | i | | y |
| voc | pl | i | | y |
| loc | pl | i | | ech |
| inst | pl | i | | y |
| inst | pl | i | reg | ama |

TAB. C.11 – Modèle *hrad*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| čas | hd_Re | 1 | 1 | čas, obraz, vůz, dopis, zápas, stůl, kus, tác |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | i | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.12 – Sous-type *čas*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| vzor | hd_v | 1 | 1 | problém, milión, týden, zájem, počet, program, vzor |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | i | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.13 – Sous-type *vzor*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| zámek | hd_z | 1 | 1 | výsledek, trh, začátek, prostředek, druh, úspěch, zámek |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| voc | sg | i | | R | u |
| loc | sg | i | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | i | | R | ích |
| loc | pl | i | reg | A | ách |

TAB. C.14 – Sous-type *zámek*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ostrov | hd_o | 1 | 1 | zákon, svět, kout, ostrov, domov, Pacov, hřbitov, chléb, vepřín, včelín |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | a |
| gen | sg | i | reg | V | u |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.15 – Sous-type *ostrov*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| rybník | hd_r | 1 | 1 | rok, rybník |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | a |
| gen | sg | i | reg | V | u |
| voc | sg | i | | R | u |
| loc | sg | i | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | i | reg | A | ách |
| loc | pl | i | | R | ích |

TAB. C.16 – Sous-type *rybník*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| červen | hd_m | 0 | 1 | leden, únor, březen, duben, květen, červen, srpen, říjen |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | a |
| loc | sg | i | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.17 – Sous-type *červen*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| tábor | hd_t | 0 | 1 | tábor, dvůr |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | fnct | A | a |
| loc | sg | i | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.18 – Sous-type *tábor*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| sklep | hd_s | 0 | 1 | sklep |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | a |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | i | fnct | A | ích |

TAB. C.19 – Sous-type *sklep*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| les | hd_l | 0 | 1 | les |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | a |
| loc | sg | i | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | i | fnct | A | ích |

TAB. C.20 – Sous-type *les*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kámen | hd_k | 1 | 0 | kámen, pramen, ječmen, kořen, křemen, řemen, kmen, plamen, hřeben |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | fnct | A | e |
| dat | sg | i | fnct | A | i |
| voc | sg | i | | R | i |
| loc | sg | i | fnct | A | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.21 – Sous-type *kámen*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| den | hd_d | 0 | 1 | den |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | R | e |
| dat | sg | i | fnct | A | i |
| voc | sg | i | fnct | A | i |
| loc | sg | i | fnct | A | i |
| loc | sg | i | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | i | fnct | A | i |
| gen | pl | i | fnct | A | í |
| dat | pl | i | fnct | A | ím |
| acc | pl | i | fnct | A | i |
| voc | pl | i | fnct | A | i |
| inst | pl | i | reg | A | ěma |

TAB. C.22 – Sous-type *den*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| virus | hd_u | 1 | 1 | virus, rasismus, socialismus, komunismus, bolševismus |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | i | | R | us |
| acc | sg | i | | R | us |
| loc | sg | i | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.23 – Sous-type *virus*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| šaty | hd_a | 1 | 1 | šaty, džíny, alimenty |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | i | | X | # |
| gen | sg | i | | X | u |
| dat | sg | i | | X | u |
| acc | sg | i | | X | # |
| voc | sg | i | | X | e |
| loc | sg | i | fnct | X | ě |
| loc | sg | i | fnct | X | u |
| inst | sg | i | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.24 – Sous-type *šaty*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| hradčany | hd_h | 1 | 0 | Hradčany, Modřany, Klatovy |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | i | | X | # |
| gen | sg | i | | X | u |
| dat | sg | i | | X | u |
| acc | sg | i | | X | # |
| voc | sg | i | | X | e |
| loc | sg | i | fnct | X | ě |
| loc | sg | i | fnct | X | u |
| inst | sg | i | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | i | | R | # |

TAB. C.25 – Sous-type *hradčany*

# Type *muž*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| muž | mz | 1 | 1 | muž, lékař, rodič, novinář, strýc, kuchař |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | e |
| dat | sg | m | fnct | i |
| dat | sg | m | fnct | ovi |
| acc | sg | m | | e |
| voc | sg | m | | i |
| loc | sg | m | fnct | i |
| loc | sg | m | fnct | ovi |
| inst | sg | m | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | i |
| nom | pl | m | fnct | ové |
| gen | pl | m | | ů |
| dat | pl | m | | ům |
| dat | pl | m | reg | um |
| acc | pl | m | | e |
| voc | pl | m | fnct | i |
| voc | pl | m | fnct | ové |
| loc | pl | m | | ích |
| inst | pl | m | | i |
| inst | pl | m | reg | ema |

TAB. C.26 – Modèle *muž*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| učeň | mz_Re | 1 | 1 | choť, svišť, učeň, zeť, tuleň, sršeň, sršáň, hlemýžď, sleď |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | m | | Re | ě |
| acc | sg | m | | Re | ě |
| inst | sg | m | | Re | ěm |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| acc | pl | m | | Re | ě |
| inst | pl | m | reg | Re | ěma |

TAB. C.27 – Sous-type *učeň*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| učitel | mz_u | 1 | 1 | ředitel, majitel, podnikatel, učitel, spisovatel, zaměstnavatel |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | R | é |
| nom | pl | m | reg | V | ové |
| voc | pl | m | | R | é |
| inst | pl | m | reg | A | ama |

TAB. C.28 – Sous-type *učitel*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| přítel | mz_p | 0 | 1 | přítel, obyvatel |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | R | é |
| gen | pl | m | | R | # |
| gen | pl | m | reg | V | ů |
| voc | pl | m | | R | é |
| inst | pl | m | reg | A | ama |

TAB. C.29 – Sous-type *přítel*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| otec | mz_o | 1 | 1 | otec, chodec, poslanec, umělec, chlapec, cizinec |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| voc | sg | m | | R | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.30 – Sous-type *otec*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| rodič | mz_r | 0 | 1 | rodič |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | fnct | R | e |
| voc | pl | m | fnct | R | e |

TAB. C.31 – Sous-type *rodič*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kůň | mz_k | 0 | 0 | kůň |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | m | | Re | ě |
| acc | sg | m | | Re | ě |
| inst | sg | m | | Re | ěm |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | R | ě |
| nom | pl | m | | X | ové |
| gen | pl | m | fnct | A | í |
| dat | pl | m | fnct | A | ím |
| acc | pl | m | | Re | ě |
| voc | pl | m | | Re | ě |
| inst | pl | m | fnct | A | mi |
| inst | pl | m | reg | A | ma |
| inst | pl | m | reg | Re | ěma |

TAB. C.32 – Sous-type *kůň*

## Type *stroj*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| stroj | sj | 1 | 1 | konec, měsíc, prosinec, pokoj, počítač, klíč, stroj |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | e |
| dat | sg | i | | i |
| acc | sg | i | | # |
| voc | sg | i | | i |
| loc | sg | i | | i |
| inst | sg | i | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | e |
| gen | pl | i | | ů |
| dat | pl | i | | ům |
| dat | pl | i | reg | um |
| acc | pl | i | | e |
| voc | pl | i | | e |
| loc | pl | i | | ích |
| inst | pl | i | | i |
| inst | pl | i | reg | ema |

TAB. C.33 – Modèle *stroj*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| stupeň | sj_Re | 1 | 1 | oheň, stupeň, větroň, déšť, rožeň |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | i | | Re | ě |
| inst | sg | i | | Re | ěm |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | i | | Re | ě |
| acc | pl | i | | Re | ě |
| inst | pl | i | reg | Re | ěma |

TAB. C.34 – Sous-type *stupeň*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| punčocháče | sj_p | 1 | 1 | punčocháče, papuče |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | i | | X | # |
| gen | sg | i | | X | e |
| dat | sg | i | | X | i |
| acc | sg | i | | X | # |
| voc | sg | i | | X | i |
| loc | sg | i | | X | i |
| inst | sg | i | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.35 – Sous-type *punčocháče*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| peníze | sj_n | 0 | 1 | peníze |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | i | | X | # |
| gen | sg | i | | X | e |
| dat | sg | i | | X | i |
| acc | sg | i | | X | # |
| voc | sg | i | | X | i |
| loc | sg | i | | X | i |
| inst | sg | i | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | i | | R | # |
| inst | pl | i | | R | mi |
| inst | pl | i | reg | R | ma |

TAB. C.36 – Sous-type *peníze*

## Type *předseda*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| předseda | pd | 1 | 1 | táta, předseda, vévoda, nenasyta |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | a |
| gen | sg | m | | y |
| dat | sg | m | | ovi |
| acc | sg | m | | u |
| voc | sg | m | | o |
| loc | sg | m | | ovi |
| inst | sg | m | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | ové |
| gen | pl | m | | ů |
| dat | pl | m | | ům |
| dat | pl | m | reg | um |
| acc | pl | m | | y |
| voc | pl | m | | ové |
| loc | pl | m | | ech |
| inst | pl | m | | y |
| inst | pl | m | reg | ama |

TAB. C.37 – Modèle *předseda*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| sluha | pd_s | 1 | 1 | kolega, sluha, taťka |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | m | | R | ích |

TAB. C.38 – Sous-type *sluha*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| husita | pd_h | 1 | 1 | policista, turista, terorista, cyklista, optimista, socialista, invalida, bandita, husita |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | R | é |
| nom | pl | m | reg | A | i |
| voc | pl | m | | R | é |
| voc | pl | m | reg | A | i |

TAB. C.39 – Sous-type *husita*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| paňáca | pd_p | 1 | 0 | paňáca, rikša |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | m | | Ri | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| acc | pl | m | | Ri | i |
| loc | pl | m | | R | ích |
| inst | pl | m | | Ri | i |

TAB. C.40 – Sous-type *paňáca*

## Type *soudce*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| soudce | sc | 1 | 1 | soudce, správce, důchodce, strůjce, svůdce, žalobce, výrobce |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | e |
| gen | sg | m | | e |
| dat | sg | m | fnct | i |
| dat | sg | m | fnct | ovi |
| acc | sg | m | | e |
| voc | sg | m | | e |
| loc | sg | m | fnct | i |
| loc | sg | m | fnct | ovi |
| inst | sg | m | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | i |
| nom | pl | m | fnct | ové |
| gen | pl | m | | ů |
| dat | pl | m | | ům |
| dat | pl | m | reg | um |
| acc | pl | m | | e |
| voc | pl | m | fnct | i |
| voc | pl | m | fnct | ové |
| loc | pl | m | | ích |
| inst | pl | m | | i |
| inst | pl | m | reg | ema |

TAB. C.41 – Modèle *soudce*

# Type *žena*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| žena | zn | 1 | 1 | doba, strana, cesta, žena, firma, vrána |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | a |
| gen | sg | f | | y |
| dat | sg | f | | ě |
| acc | sg | f | | u |
| voc | sg | f | | o |
| loc | sg | f | | ě |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | y |
| gen | pl | f | | # |
| dat | pl | f | | ám |
| acc | pl | f | | y |
| voc | pl | f | | y |
| loc | pl | f | | ách |
| inst | pl | f | | ami |
| inst | pl | f | reg | ama |

TAB. C.42 – Modèle *žena*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| holka | zn_Re | 1 | 1 | otázka, škola, hra, kniha, banka, holka, moucha |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | sg | f | | Re | e |
| loc | sg | f | | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.43 – Sous-type *holka*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| skica | zn_s | 1 | 0 | skica, Pája, Anča, Hanča, Monča, pizza, gejša |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | | Ri | i |
| dat | sg | f | | Re | e |
| loc | sg | f | | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Ri | i |
| acc | pl | f | | Ri | i |
| voc | pl | f | | Ri | i |

TAB. C.44 – Sous-type *skica*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| Stáňa | zn_t | 1 | 0 | Stáňa, Máňa, Káťa, Páťa, pastorkyňa |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | | Ri | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Ri | i |
| acc | pl | f | | Ri | i |
| voc | pl | f | | Ri | i |

TAB. C.45 – Sous-type *Stáňa*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| dcera | zn_d | 0 | 1 | dcera |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | sg | f | | R | i |
| dat | sg | f | reg | V | e |
| loc | sg | f | | R | i |
| loc | sg | f | reg | V | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.46 – Sous-type *dcera*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| noha | zn_n | 0 | 1 | noha |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | sg | f | | Re | e |
| loc | sg | f | | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | f | fnct | A | ou |
| dat | pl | f | reg | A | oum |
| loc | pl | f | fnct | A | ou |
| inst | pl | f | | R | ama |

TAB. C.47 – Sous-type *noha*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ruka | zn_r | 0 | 1 | ruka |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | sg | f | | Re | e |
| loc | sg | f | | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | R | e |
| gen | pl | f | | R | ou |
| dat | pl | f | reg | A | oum |
| acc | pl | f | | R | e |
| voc | pl | f | | R | e |
| loc | pl | f | fnct | A | ou |
| inst | pl | f | | R | ama |

TAB. C.48 – Sous-type *ruka*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| idea | zn_i | 0 | 0 | idea |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | fnct | A | je |
| dat | sg | f | fnct | R | ji |
| loc | sg | f | | R | ji |
| inst | sg | f | fnct | A | jí |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | fnct | A | je |
| gen | pl | f | | R | jí |
| dat | pl | f | fnct | A | jím |
| acc | pl | f | fnct | A | je |
| voc | pl | f | fnct | A | je |
| loc | pl | f | fnct | A | jích |
| inst | pl | f | fnct | A | jemi |

TAB. C.49 – Sous-type *idea*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kalhoty | zn_k | 1 | 1 | kalhoty, prázdniny, narozeniny, hodinky, nůžky, plavky |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | a |
| gen | sg | f | | X | y |
| dat | sg | f | | X | ě |
| acc | sg | f | | X | u |
| voc | sg | f | | X | o |
| loc | sg | f | | X | ě |
| inst | sg | f | | X | ou |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.50 – Sous-type *kalhoty*

## Type *růže*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| růže | rz | 1 | 1 | práce, informace, akce, organizace, policie, televize, půle, růže |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | e |
| gen | sg | f | | e |
| dat | sg | f | | i |
| acc | sg | f | | i |
| voc | sg | f | | e |
| loc | sg | f | | i |
| inst | sg | f | | í |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | e |
| gen | pl | f | | í |
| dat | pl | f | | ím |
| acc | pl | f | | e |
| voc | pl | f | | e |
| loc | pl | f | | ích |
| inst | pl | f | | emi |
| inst | pl | f | reg | ema |

TAB. C.51 – Modèle *růže*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| sukně | rz_Re | 1 | 1 | země, snídaně, kuchyně, sukně, minisukně, koupě, kápě, dýně, žíně |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | Re | ě |
| gen | sg | f | | Re | ě |
| voc | sg | f | | Re | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Re | ě |
| acc | pl | f | | Re | ě |
| voc | pl | f | | Re | ě |
| inst | pl | f | | Re | ěmi |
| inst | pl | f | reg | Re | ěma |

TAB. C.52 – Sous-type *sukně*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ulice | rz_u | 1 | 1 | chvíle, ulice, hranice, nemocnice, stanice, silnice |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | f | | R | # |

TAB. C.53 – Sous-type *ulice*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| definice | rz_p | 1 | 1 | prohibice, definice, koalice, exhibice |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | f | | R | # |
| gen | pl | f | reg | V | í |

TAB. C.54 – Sous-type *definice*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| žákyně | rz_z | 1 | 1 | přítelkyně, předsedkyně, kolegyně, ministryně, důchodkyně, plavkyně, žákyně |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | Re | ě |
| gen | sg | f | | Re | ě |
| voc | sg | f | | Re | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Re | ě |
| gen | pl | f | fnct | A | # |
| acc | pl | f | | Re | ě |
| voc | pl | f | | Re | ě |
| inst | pl | f | | Re | ěmi |
| inst | pl | f | reg | Re | ěma |

TAB. C.55 – Sous-type *žákyně*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| brýle | rz_b | 1 | 1 | brýle, housle, vidle, štafle, fazole |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | e |
| gen | sg | f | | X | e |
| dat | sg | f | | X | i |
| acc | sg | f | | X | i |
| voc | sg | f | | X | e |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.56 – Sous-type *brýle*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| dveře | rz_d | 0 | 1 | dveře |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | e |
| gen | sg | f | | X | e |
| dat | sg | f | | X | i |
| acc | sg | f | | X | i |
| voc | sg | f | | X | e |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| inst | pl | f | | R | mi |
| inst | pl | f | reg | R | ma |

TAB. C.57 – Sous-type *dveře*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| lidice | rz_l | 1 | 0 | Lidice, Černovice, Olbramovice |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | e |
| gen | sg | f | | X | e |
| dat | sg | f | | X | i |
| acc | sg | f | | X | i |
| voc | sg | f | | X | e |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | f | | R | # |

TAB. C.58 – Sous-type *lidice*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| velikonoce | rz_v | 0 | 1 | Velikonoce, Vánoce |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | e |
| gen | sg | f | | X | e |
| dat | sg | f | | X | i |
| acc | sg | f | | X | i |
| voc | sg | f | | X | e |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | f | | R | # |
| dat | pl | f | | R | ům |

TAB. C.59 – Sous-type *velikonoce*

## Type *píseň*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| píseň | ps | 1 | 1 | úroveň, žízeň, síť, daň, píseň, kuchyň, báseň, dlaň |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | # |
| gen | sg | f | | ě |
| dat | sg | f | | i |
| acc | sg | f | | # |
| voc | sg | f | | i |
| loc | sg | f | | i |
| inst | sg | f | | í |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | ě |
| gen | pl | f | | í |
| dat | pl | f | | ím |
| acc | pl | f | | ě |
| voc | pl | f | | ě |
| loc | pl | f | | ích |
| inst | pl | f | | ěmi |
| inst | pl | f | reg | ěma |

TAB. C.60 – Modèle *píseň*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| tvář | ps_Re | 1 | 1 | tvář, obec, kancelář, lahev, postel, laboratoř, nádrž, kolej, stráž, hůl, půl, klec |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Re | e |
| acc | pl | f | | Re | e |
| voc | pl | f | | Re | e |
| inst | pl | f | | Re | emi |
| inst | pl | f | reg | Re | ema |

TAB. C.61 – Sous-type *tvář*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ocel | ps_o | 1 | 1 | ocel, modř, žluč |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | fnct | Re | e |
| gen | sg | f | fnct | A | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | fnct | Re | e |
| nom | pl | f | fnct | A | i |
| acc | pl | f | fnct | A | i |
| acc | pl | f | fnct | Re | e |
| voc | pl | f | fnct | A | i |
| voc | pl | f | fnct | Re | e |
| inst | pl | f | | Re | emi |
| inst | pl | f | reg | Re | ema |

TAB. C.62 – Sous-type *ocel*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| čtvrť | ps_c | 1 | 1 | niť, pouť, trať, stať, peruť, závrať, čeleď, kapraď, záď, žerď, chuť |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | fnct | A | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | fnct | A | i |
| acc | pl | f | fnct | A | i |
| voc | pl | f | fnct | A | i |

TAB. C.63 – Sous-type *čtvrť*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| loď | ps_l | 1 | 1 | loď, suť |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | sg | f | fnct | A | i |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | fnct | A | i |
| acc | pl | f | fnct | A | i |
| voc | pl | f | fnct | A | i |
| inst | pl | f | fnct | A | mi |
| inst | pl | f | reg | A | ma |

TAB. C.64 – Sous-type *loď*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kleště | ps_k | 1 | 1 | kleště, hutě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | # |
| gen | sg | f | | X | ě |
| dat | sg | f | | X | i |
| acc | sg | f | | X | # |
| voc | sg | f | | X | i |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.65 – Sous-type *kleště*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| housle | ps_h | 1 | 1 | housle, Nusle |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | # |
| gen | sg | f | | X | ě |
| dat | sg | f | | X | i |
| acc | sg | f | | X | # |
| voc | sg | f | | X | i |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | Re | ě |
| acc | pl | f | | Re | ě |
| voc | pl | f | | Re | ě |
| inst | pl | f | | Re | ěmi |
| inst | pl | f | reg | Re | ěma |

TAB. C.66 – Sous-type *housle*

## Type *kost*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kost | kt | 1 | 1 | společnost, činnost, událost, povinnost, radost, žádost, rychlost, kost, část, věc, řeč |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | # |
| gen | sg | f | | i |
| dat | sg | f | | i |
| acc | sg | f | | # |
| voc | sg | f | | i |
| loc | sg | f | | i |
| inst | sg | f | | í |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | i |
| gen | pl | f | | í |
| dat | pl | f | | em |
| acc | pl | f | | i |
| voc | pl | f | | i |
| loc | pl | f | | ech |
| inst | pl | f | | mi |
| inst | pl | f | reg | ma |

TAB. C.67 – Modèle *kost*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| pěst | kt_p | 1 | 1 | pěst, náplast, soustrast, strast, neřest, svěžest, bolest, past, lest, pelest, probust, oběť, zeď, paměť |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | pl | f | fnct | A | ím |
| loc | pl | f | fnct | A | ích |

TAB. C.68 – Sous-type *pěst*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| odpověď | kt_o | 1 | 1 | odpověď, předpověď, myš, výpověď, step, otep, směs, lež, ves, rez, mosaz |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | pl | f | | R | ím |
| loc | pl | f | | R | ích |

TAB. C.69 – Sous-type *odpověď*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| moc | kt_m | 1 | 1 | moc, pomoc, výpomoc, svépomoc, nemoc |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | pl | f | fnct | A | ím |
| loc | pl | f | fnct | A | ích |
| inst | pl | f | | R | emi |
| inst | pl | f | reg | R | ema |

TAB. C.70 – Sous-type *moc*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| noc | kt_n | 1 | 1 | noc, půlnoc, mysl, sůl |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| dat | pl | f | | R | ím |
| loc | pl | f | | R | ích |
| inst | pl | f | | R | emi |
| inst | pl | f | reg | R | ema |

TAB. C.71 – Sous-type *noc*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| lidé | kt_l | 0 | 1 | lidé |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | f | | X | # |
| gen | sg | f | | X | i |
| dat | sg | f | | X | i |
| acc | sg | f | | X | # |
| voc | sg | f | | X | i |
| loc | sg | f | | X | i |
| inst | sg | f | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | m | | A | é |
| voc | pl | m | | A | é |

TAB. C.72 – Sous-type *lidé*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| čtyři | kt_4 | 0 | 1 | čtyři |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | s | | X | # |
| gen | sg | s | | X | i |
| dat | sg | s | | X | i |
| acc | sg | s | | X | # |
| voc | sg | s | | X | i |
| loc | sg | s | | X | i |
| inst | sg | s | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | s | reg | A | ech |

TAB. C.73 – Sous-type *čtyři*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| tři | kt_3 | 0 | 1 | tři |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | s | | X | # |
| gen | sg | s | | X | i |
| dat | sg | s | | X | i |
| acc | sg | s | | X | # |
| voc | sg | s | | X | i |
| loc | sg | s | | X | i |
| inst | sg | s | | X | í |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | s | reg | A | ech |
| inst | pl | s | | R | emi |
| inst | pl | s | reg | R | ema |

TAB. C.74 – Sous-type *tři*

## Type *město*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| město | mt | 1 | 1 | město, slovo, jméno, auto, okno, pivo, léto, kino |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | o |
| gen | sg | n | | a |
| dat | sg | n | | u |
| acc | sg | n | | o |
| voc | sg | n | | o |
| loc | sg | n | fnct | ě |
| loc | sg | n | fnct | u |
| inst | sg | n | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | a |
| gen | pl | n | | # |
| dat | pl | n | | ům |
| dat | pl | n | reg | um |
| acc | pl | n | | a |
| voc | pl | n | | a |
| loc | pl | n | | ech |
| inst | pl | n | | y |
| inst | pl | n | reg | ama |

TAB. C.75 – Modèle *město*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| tělo | mt_Re | 1 | 1 | teplo, číslo, tělo, divadlo, světlo, maso, mléko, jaro |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.76 – Sous-type *tělo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| lečo | mt_l | 1 | 0 | lečo, bendžo, pončo, mimčo, čůčo, pivčo |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| inst | pl | n | | Ri | i |

TAB. C.77 – Sous-type *lečo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| břicho | mt_r | 1 | 1 | břicho, roucho |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | fnct | Re | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | | R | ách |

TAB. C.78 – Sous-type *břicho*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| božstvo | mt_b | 1 | 1 | ráno, metro, pero, volno, pečivo, euro, bělmo, dynamo, eskymo, pižmo, božstvo |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.79 – Sous-type *božstvo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| echo | mt_e | 1 | 1 | vlhko, ticho, horko, sako, tričko, echo |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | | R | ách |

TAB. C.80 – Sous-type *echo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| středisko | mt_s | 1 | 1 | jablko, středisko, hledisko, východisko, vojsko |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | | X | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | fnct | R | ích |
| loc | pl | n | fnct | A | ách |

TAB. C.81 – Sous-type *středisko*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ústa | mt_u | 1 | 1 | ústa, záda, kamna, játra, vrata, akta, blata |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | X | o |
| gen | sg | n | | X | a |
| dat | sg | n | | X | u |
| acc | sg | n | | X | o |
| voc | sg | n | | X | o |
| loc | sg | n | fnct | X | ě |
| loc | sg | n | fnct | X | u |
| inst | sg | n | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.82 – Sous-type *ústa*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| dvířka | mt_d | 1 | 0 | dvířka, luka |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | X | o |
| gen | sg | n | | X | a |
| dat | sg | n | | X | u |
| acc | sg | n | | X | o |
| voc | sg | n | | X | o |
| loc | sg | n | fnct | X | ě |
| loc | sg | n | fnct | X | u |
| inst | sg | n | | X | em |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | | R | ách |

TAB. C.83 – Sous-type *dvířka*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| video | mt_v | 1 | 1 | video, rodeo, kakao, stereo, audio, rádio, trio, portfolio, studio, patio |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | fnct | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | n | | R | í |
| dat | pl | n | fnct | A | ím |
| dat | pl | n | reg | A | im |
| loc | pl | n | fnct | A | ích |
| inst | pl | n | | Ri | i |

TAB. C.84 – Sous-type *video*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| koleno | mt_k | 0 | 1 | koleno, rameno |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | fnct | A | ou |
| loc | pl | n | reg | A | ách |

TAB. C.85 – Sous-type *koleno*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| prso | mt_p | 0 | 1 | prso |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | | Re | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | | A | ou |
| loc | pl | n | reg | A | ách |

TAB. C.86 – Sous-type *prso*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| oko | mt_o | 0 | 1 | oko, ucho |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | sg | n | | X | e |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | n | | R | i |
| gen | pl | n | | R | í |
| dat | pl | n | | R | ím |
| dat | pl | n | reg | A | im |
| acc | pl | n | | R | i |
| voc | pl | n | | R | i |
| loc | pl | n | | R | ích |
| inst | pl | n | | R | ima |
| inst | pl | n | reg | A | imi |

TAB. C.87 – Sous-type *oko*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| sto | mt_0 | 0 | 1 | sto |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | n | fnct | A | ě |
| acc | pl | n | fnct | A | ě |
| voc | pl | n | fnct | A | ě |

TAB. C.88 – Sous-type *sto*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| centrum | mt_c | 1 | 1 | centrum, datum, vízum, antibiotikum, substantivum |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | R | um |
| acc | sg | n | | R | um |
| voc | sg | n | | R | um |
| loc | sg | n | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.89 – Sous-type *centrum*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| muzeum | mt_m | 1 | 1 | muzeum, gymnázium, stipendium, linoleum, akvarium |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | R | um |
| acc | sg | n | | R | um |
| voc | sg | n | | R | um |
| loc | sg | n | | X | ě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | n | | R | í |
| dat | pl | n | | R | ím |
| dat | pl | n | reg | V | ům |
| loc | pl | n | | R | ích |
| inst | pl | n | | Ri | i |

TAB. C.90 – Sous-type *muzeum*

## Type *moře*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| moře | mr | 1 | 1 | srdce, pole, slunce, moře, ovoce |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | e |
| gen | sg | n | | e |
| dat | sg | n | | i |
| acc | sg | n | | e |
| voc | sg | n | | e |
| loc | sg | n | | i |
| inst | sg | n | | em |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | e |
| gen | pl | n | | í |
| dat | pl | n | | ím |
| acc | pl | n | | e |
| voc | pl | n | | e |
| loc | pl | n | | ích |
| inst | pl | n | | i |
| inst | pl | n | reg | ema |

TAB. C.91 – Modèle *moře*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| letiště | mr_l | 1 | 1 | letiště, hřiště, jeviště, bydliště, parkoviště, hlediště |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | Re | ě |
| gen | sg | n | | Re | ě |
| acc | sg | n | | Re | ě |
| voc | sg | n | | Re | ě |
| inst | sg | n | | Re | ěm |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | n | | Re | ě |
| gen | pl | n | | R | # |
| acc | pl | n | | Re | ě |
| voc | pl | n | | Re | ě |
| inst | pl | n | reg | Re | ěma |

TAB. C.92 – Sous-type *letiště*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| vejce | mr_v | 0 | 1 | vejce |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| gen | pl | n | | R | # |

TAB. C.93 – Sous-type *vejce*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| poledne | mr_p | 0 | 1 | poledne |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| loc | pl | n | | R | ách |
| inst | pl | n | | Ri | y |
| inst | pl | n | reg | R | ama |

TAB. C.94 – Sous-type *poledne*

## Type *kuře*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kuře | kr | 1 | 1 | zvíře, prase, kuře, rajče, morče |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | e |
| gen | sg | n | | ete |
| dat | sg | n | | eti |
| acc | sg | n | | e |
| voc | sg | n | | e |
| loc | sg | n | | eti |
| inst | sg | n | | etem |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | ata |
| gen | pl | n | | at |
| dat | pl | n | | atům |
| dat | pl | n | reg | atum |
| acc | pl | n | | ata |
| voc | pl | n | | ata |
| loc | pl | n | | atech |
| inst | pl | n | | aty |
| inst | pl | n | reg | atama |

TAB. C.95 – Modèle *kuře*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| mládě | kr_Re | 1 | 1 | mládě, medvídě, koště, slůně, káně, nemluvně |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | Re | ě |
| gen | sg | n | | Re | ěte |
| dat | sg | n | | Re | ěti |
| acc | sg | n | | Re | ě |
| voc | sg | n | | Re | ě |
| loc | sg | n | | Re | ěti |
| inst | sg | n | | Re | ětem |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|

TAB. C.96 – Sous-type *mládě*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| dítě | kr_d | 0 | 1 | dítě |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | Re | ě |
| gen | sg | n | | Re | ěte |
| dat | sg | n | | Re | ěti |
| acc | sg | n | | Re | ě |
| voc | sg | n | | Re | ě |
| loc | sg | n | | Re | ěti |
| inst | sg | n | | Re | ětem |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | f | | R | i |
| gen | pl | f | | R | í |
| dat | pl | f | | R | em |
| acc | pl | f | | R | i |
| voc | pl | f | | R | i |
| loc | pl | f | | R | ech |
| inst | pl | f | | R | mi |
| inst | pl | f | reg | R | ma |

TAB. C.97 – Sous-type *dítě*

## Type *stavení*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| stavení | st | 1 | 1 | období, prostředí, století, zahraničí, září, umění, stavení |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | í |
| gen | sg | n | | í |
| dat | sg | n | | í |
| acc | sg | n | | í |
| voc | sg | n | | í |
| loc | sg | n | | í |
| inst | sg | n | | ím |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | í |
| gen | pl | n | | í |
| dat | pl | n | | ím |
| acc | pl | n | | í |
| voc | pl | n | | í |
| loc | pl | n | | ích |
| inst | pl | n | | ími |
| inst | pl | n | reg | íma |

TAB. C.98 – Modèle *stavení*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| úterý | st_Ri | 0 | 1 | úterý |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | sg | n | | Ri | ý |
| gen | sg | n | | Ri | ý |
| dat | sg | n | | Ri | ý |
| acc | sg | n | | Ri | ý |
| voc | sg | n | | Ri | ý |
| loc | sg | n | | Ri | ý |
| inst | sg | n | | Ri | ým |

| cas | num | gen | var | ET | term |
|---|---|---|---|---|---|
| nom | pl | n | | Ri | ý |
| gen | pl | n | | Ri | ý |
| dat | pl | n | | Ri | ým |
| acc | pl | n | | Ri | ý |
| voc | pl | n | | Ri | ý |
| loc | pl | n | | Ri | ých |
| inst | pl | n | | Ri | ými |
| inst | pl | n | reg | Ri | ýma |

TAB. C.99 – Sous-type *úterý*

## Type *klima*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| klima | kl | 1 | 1 | klima, drama, dogma, zeugma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | a |
| gen | sg | n | | atu |
| dat | sg | n | | atu |
| acc | sg | n | | a |
| voc | sg | n | | a |
| loc | sg | n | | atu |
| inst | sg | n | | atem |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | ata |
| gen | pl | n | | at |
| dat | pl | n | | atům |
| dat | pl | n | reg | atum |
| acc | pl | n | | ata |
| voc | pl | n | | ata |
| loc | pl | n | | atech |
| inst | pl | n | | aty |

TAB. C.100 – Modèle *klima*

# Type *rád*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| rád | rd | 1 | 1 | laskav, rád, nerad, jist, šťasten, dobyt |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | |
| dat | sg | m | | |
| acc | sg | m | | a |
| voc | sg | m | | |
| loc | sg | m | | |
| inst | sg | m | | |
| nom | sg | i | | # |
| gen | sg | i | | |
| dat | sg | i | | |
| acc | sg | i | | # |
| voc | sg | i | | |
| loc | sg | i | | |
| inst | sg | i | | |
| nom | sg | f | | a |
| gen | sg | f | | |
| dat | sg | f | | |
| acc | sg | f | | u |
| voc | sg | f | | |
| loc | sg | f | | |
| inst | sg | f | | |
| nom | sg | n | | o |
| gen | sg | n | | |
| dat | sg | n | | |
| acc | sg | n | | o |
| voc | sg | n | | |
| loc | sg | n | | |
| inst | sg | n | | |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | i |
| gen | pl | m | | |
| dat | pl | m | | |
| acc | pl | m | | y |
| voc | pl | m | | |
| loc | pl | m | | |
| inst | pl | m | | |
| nom | pl | i | | y |
| gen | pl | i | | |
| dat | pl | i | | |
| acc | pl | i | | y |
| voc | pl | i | | |
| loc | pl | i | | |
| inst | pl | i | | |
| nom | pl | f | | y |
| gen | pl | f | | |
| dat | pl | f | | |
| acc | pl | f | | y |
| voc | pl | f | | |
| loc | pl | f | | |
| inst | pl | f | | |
| nom | pl | n | | a |
| gen | pl | n | | |
| dat | pl | n | | |
| acc | pl | n | | a |
| voc | pl | n | | |
| loc | pl | n | | |
| inst | pl | n | | |

TAB. C.101 – Modèle *rád*

# C.2 Flexion adjectivale

## Type *mladý*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| mladý | md | 1 | 1 | nový, velký, celý, každý, mladý |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | ý |
| nom | sg | m | reg | ej |
| gen | sg | m | | ého |
| gen | sg | m | reg | ýho |
| dat | sg | m | | ému |
| dat | sg | m | reg | ýmu |
| acc | sg | m | | ého |
| acc | sg | m | reg | ýho |
| voc | sg | m | | ý |
| voc | sg | m | reg | ej |
| loc | sg | m | | ém |
| loc | sg | m | reg | ym |
| loc | sg | m | reg | ým |
| inst | sg | m | | ým |
| inst | sg | m | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | í |
| nom | pl | m | reg | ý |
| gen | pl | m | | ých |
| gen | pl | m | reg | ejch |
| dat | pl | m | | ým |
| dat | pl | m | reg | ym |
| dat | pl | m | reg | ejm |
| acc | pl | m | | é |
| acc | pl | m | reg | ý |
| voc | pl | m | | í |
| voc | pl | m | reg | ý |
| loc | pl | m | | ých |
| loc | pl | m | reg | ejch |
| inst | pl | m | | ými |
| inst | pl | m | reg | ýma |
| inst | pl | m | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | ý |
| nom | sg | i | reg | ej |
| gen | sg | i | | ého |
| gen | sg | i | reg | ýho |
| dat | sg | i | | ému |
| dat | sg | i | reg | ýmu |
| acc | sg | i | | ý |
| acc | sg | i | reg | ej |
| voc | sg | i | | ý |
| voc | sg | i | reg | ej |
| loc | sg | i | | ém |
| loc | sg | i | reg | ym |
| loc | sg | i | reg | ým |
| inst | sg | i | | ým |
| inst | sg | i | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | é |
| nom | pl | i | reg | ý |
| gen | pl | i | | ých |
| gen | pl | i | reg | ejch |
| dat | pl | i | | ým |
| dat | pl | i | reg | ym |
| dat | pl | i | reg | ejm |
| acc | pl | i | | é |
| acc | pl | i | reg | ý |
| voc | pl | i | | é |
| voc | pl | i | reg | ý |
| loc | pl | i | | ých |
| loc | pl | i | reg | ejch |
| inst | pl | i | | ými |
| inst | pl | i | reg | ýma |
| inst | pl | i | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | á |
| gen | sg | f | | é |
| gen | sg | f | reg | ý |
| dat | sg | f | | é |
| dat | sg | f | reg | ý |
| acc | sg | f | | ou |
| voc | sg | f | | á |
| loc | sg | f | | é |
| loc | sg | f | reg | ý |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | é |
| nom | pl | f | reg | ý |
| gen | pl | f | | ých |
| gen | pl | f | reg | ejch |
| dat | pl | f | | ým |
| dat | pl | f | reg | ym |
| dat | pl | f | reg | ejm |
| acc | pl | f | | é |
| acc | pl | f | reg | ý |
| voc | pl | f | | é |
| voc | pl | f | reg | ý |
| loc | pl | f | | ých |
| loc | pl | f | reg | ejch |
| inst | pl | f | | ými |
| inst | pl | f | reg | ýma |
| inst | pl | f | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | é |
| nom | sg | n | reg | ý |
| gen | sg | n | | ého |
| gen | sg | n | reg | ýho |
| dat | sg | n | | ému |
| dat | sg | n | reg | ýmu |
| acc | sg | n | | é |
| acc | sg | n | reg | ý |
| voc | sg | n | | é |
| voc | sg | n | reg | ý |
| loc | sg | n | | ém |
| loc | sg | n | reg | ym |
| loc | sg | n | reg | ým |
| inst | sg | n | | ým |
| inst | sg | n | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | á |
| nom | pl | n | reg | ý |
| gen | pl | n | | ých |
| gen | pl | n | reg | ejch |
| dat | pl | n | | ým |
| dat | pl | n | reg | ym |
| dat | pl | n | reg | ejm |
| acc | pl | n | | á |
| acc | pl | n | reg | ý |
| voc | pl | n | | á |
| voc | pl | n | reg | ý |
| loc | pl | n | | ých |
| loc | pl | n | reg | ejch |
| inst | pl | n | | ými |
| inst | pl | n | reg | ýma |
| inst | pl | n | reg | ejma |

TAB. C.102 – Modèle *mladý*

# Type *můj*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| můj | md_m | 0 | 1 | můj, tvůj, svůj |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | ůj |
| gen | sg | m | | ého |
| gen | sg | m | reg | ýho |
| dat | sg | m | | ému |
| dat | sg | m | reg | ýmu |
| acc | sg | m | | ého |
| acc | sg | m | reg | ýho |
| voc | sg | m | | ůj |
| loc | sg | m | | ém |
| loc | sg | m | reg | ym |
| loc | sg | m | reg | ým |
| inst | sg | m | | ým |
| inst | sg | m | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | í |
| nom | pl | m | fnct | oji |
| gen | pl | m | | ých |
| gen | pl | m | reg | ejch |
| dat | pl | m | | ým |
| dat | pl | m | reg | ejm |
| acc | pl | m | fnct | é |
| acc | pl | m | fnct | oje |
| acc | pl | m | reg | ý |
| voc | pl | m | fnct | í |
| voc | pl | m | fnct | oji |
| loc | pl | m | | ých |
| loc | pl | m | reg | ejch |
| inst | pl | m | | ými |
| inst | pl | m | reg | ýma |
| inst | pl | m | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | ůj |
| gen | sg | i | | ého |
| gen | sg | i | reg | ýho |
| dat | sg | i | | ému |
| dat | sg | i | reg | ýmu |
| acc | sg | i | | ůj |
| voc | sg | i | | ůj |
| loc | sg | i | | ém |
| loc | sg | i | reg | ym |
| loc | sg | i | reg | ým |
| inst | sg | i | | ým |
| inst | sg | i | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | fnct | é |
| nom | pl | i | fnct | oje |
| gen | pl | i | | ých |
| gen | pl | i | reg | ejch |
| dat | pl | i | | ým |
| dat | pl | i | reg | ejm |
| acc | pl | i | fnct | é |
| acc | pl | i | fnct | oje |
| voc | pl | i | fnct | é |
| voc | pl | i | fnct | oje |
| loc | pl | i | | ých |
| loc | pl | i | reg | ejch |
| inst | pl | i | | ými |
| inst | pl | i | reg | ýma |
| inst | pl | i | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | fnct | á |
| nom | sg | f | fnct | oje |
| gen | sg | f | fnct | é |
| gen | sg | f | fnct | ojí |
| gen | sg | f | reg | ý |
| dat | sg | f | fnct | é |
| dat | sg | f | fnct | ojí |
| dat | sg | f | reg | ý |
| acc | sg | f | fnct | ou |
| acc | sg | f | fnct | oji |
| voc | sg | f | fnct | á |
| voc | sg | f | fnct | oje |
| loc | sg | f | fnct | é |
| loc | sg | f | fnct | ojí |
| loc | sg | f | reg | ý |
| inst | sg | f | fnct | ou |
| inst | sg | f | fnct | ojí |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | fnct | é |
| nom | pl | f | fnct | ejo |
| nom | pl | f | reg | ý |
| gen | pl | f | | ých |
| gen | pl | f | reg | ejch |
| dat | pl | f | | ým |
| dat | pl | f | reg | ejm |
| acc | pl | f | fnct | é |
| acc | pl | f | fnct | oje |
| acc | pl | f | reg | ý |
| voc | pl | f | | é |
| voc | pl | f | reg | ý |
| loc | pl | f | | ých |
| loc | pl | f | reg | ejch |
| inst | pl | f | | ými |
| inst | pl | f | reg | ýma |
| inst | pl | f | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | fnct | é |
| nom | sg | n | fnct | oje |
| nom | sg | n | reg | ý |
| gen | sg | n | | ého |
| gen | sg | n | reg | ýho |
| dat | sg | n | | ému |
| dat | sg | n | reg | ýmu |
| acc | sg | n | fnct | é |
| acc | sg | n | fnct | oje |
| acc | sg | n | reg | ý |
| voc | sg | n | | é |
| voc | sg | n | reg | ý |
| loc | sg | n | | ém |
| loc | sg | n | reg | ym |
| loc | sg | n | reg | ým |
| inst | sg | n | | ým |
| inst | sg | n | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | fnct | á |
| nom | pl | n | fnct | oje |
| nom | pl | n | reg | ý |
| gen | pl | n | | ých |
| gen | pl | n | reg | ejch |
| dat | pl | n | | ým |
| dat | pl | n | reg | ejm |
| acc | pl | n | fnct | á |
| acc | pl | n | fnct | oje |
| acc | pl | n | reg | ý |
| voc | pl | n | | á |
| voc | pl | n | reg | ý |
| loc | pl | n | | ých |
| loc | pl | n | reg | ejch |
| inst | pl | n | | ými |
| inst | pl | n | reg | ýma |
| inst | pl | n | reg | ejma |

TAB. C.103 – Modèle *můj*

# Type *sám*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| sám | md_s | 0 | 1 | sám |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | ého |
| gen | sg | m | reg | ýho |
| dat | sg | m | | ému |
| dat | sg | m | reg | ýmu |
| acc | sg | m | | ého |
| acc | sg | m | reg | ýho |
| voc | sg | m | | # |
| loc | sg | m | | ém |
| loc | sg | m | reg | ym |
| loc | sg | m | reg | ým |
| inst | sg | m | | ým |
| inst | sg | m | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | í |
| nom | pl | m | reg | ý |
| gen | pl | m | | ých |
| gen | pl | m | reg | ejch |
| dat | pl | m | | ým |
| dat | pl | m | reg | ejm |
| acc | pl | m | | é |
| acc | pl | m | reg | ý |
| voc | pl | m | | í |
| voc | pl | m | reg | ý |
| loc | pl | m | | ých |
| loc | pl | m | reg | ejch |
| inst | pl | m | | ými |
| inst | pl | m | reg | ýma |
| inst | pl | m | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | ého |
| gen | sg | i | reg | ýho |
| dat | sg | i | | ému |
| dat | sg | i | reg | ýmu |
| acc | sg | i | | # |
| voc | sg | i | | # |
| loc | sg | i | | ém |
| loc | sg | i | reg | ym |
| loc | sg | i | reg | ým |
| inst | sg | i | | ým |
| inst | sg | i | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | é |
| nom | pl | i | reg | ý |
| gen | pl | i | | ých |
| gen | pl | i | reg | ejch |
| dat | pl | i | | ým |
| dat | pl | i | reg | ejm |
| acc | pl | i | | é |
| acc | pl | i | reg | ý |
| voc | pl | i | | é |
| voc | pl | i | reg | ý |
| loc | pl | i | | ých |
| loc | pl | i | reg | ejch |
| inst | pl | i | | ými |
| inst | pl | i | reg | ýma |
| inst | pl | i | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | a |
| gen | sg | f | | é |
| gen | sg | f | reg | ý |
| dat | sg | f | | é |
| dat | sg | f | reg | ý |
| acc | sg | f | | ou |
| voc | sg | f | | a |
| loc | sg | f | | é |
| loc | sg | f | reg | ý |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | é |
| nom | pl | f | reg | ý |
| gen | pl | f | | ých |
| gen | pl | f | reg | ejch |
| dat | pl | f | | ým |
| dat | pl | f | reg | ejm |
| acc | pl | f | | é |
| acc | pl | f | reg | ý |
| voc | pl | f | | é |
| voc | pl | f | reg | ý |
| loc | pl | f | | ých |
| loc | pl | f | reg | ejch |
| inst | pl | f | | ými |
| inst | pl | f | reg | ýma |
| inst | pl | f | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | o |
| nom | sg | n | reg | ý |
| gen | sg | n | | ého |
| gen | sg | n | reg | ýho |
| dat | sg | n | | ému |
| dat | sg | n | reg | ýmu |
| acc | sg | n | | o |
| voc | sg | n | reg | o |
| loc | sg | n | | ém |
| loc | sg | n | reg | ym |
| loc | sg | n | reg | ým |
| inst | sg | n | | ým |
| inst | sg | n | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | á |
| nom | pl | n | reg | ý |
| gen | pl | n | | ých |
| gen | pl | n | reg | ejch |
| dat | pl | n | | ým |
| dat | pl | n | reg | ejm |
| acc | pl | n | | á |
| acc | pl | n | reg | ý |
| voc | pl | n | | á |
| voc | pl | n | reg | ý |
| loc | pl | n | | ých |
| loc | pl | n | reg | ejch |
| inst | pl | n | | ými |
| inst | pl | n | reg | ýma |
| inst | pl | n | reg | ejma |

TAB. C.104 – Modèle *sám*

## Type *týž*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| týž | md_t | 0 | 0 | týž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | fnct | týž |
| nom | sg | m | fnct | tentýž |
| gen | sg | m | | téhož |
| gen | sg | m | reg | týhož |
| dat | sg | m | | témuž |
| dat | sg | m | reg | týmuž |
| acc | sg | m | | téhož |
| acc | sg | m | reg | týhož |
| voc | sg | m | fnct | týž |
| voc | sg | m | fnct | tentýž |
| loc | sg | m | fnct | témž |
| loc | sg | m | fnct | témže |
| loc | sg | m | fnct | tomtémž |
| inst | sg | m | fnct | týmž |
| inst | sg | m | fnct | tímtéž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | fnct | tíž |
| nom | pl | m | fnct | titíž |
| gen | pl | m | | týchž |
| dat | pl | m | | týmž |
| acc | pl | m | | tytéž |
| acc | pl | m | reg | tytýž |
| voc | pl | m | fnct | tíž |
| voc | pl | m | fnct | titíž |
| loc | pl | m | | týchž |
| inst | pl | m | | týmiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | fnct | týž |
| nom | sg | i | fnct | tentýž |
| gen | sg | i | | téhož |
| gen | sg | i | reg | týhož |
| dat | sg | i | | témuž |
| dat | sg | i | reg | týmuž |
| acc | sg | i | fnct | týž |
| acc | sg | i | fnct | tentýž |
| voc | sg | i | fnct | týž |
| voc | sg | i | fnct | tentýž |
| loc | sg | i | fnct | témž |
| loc | sg | i | fnct | témže |
| loc | sg | i | fnct | tomtémž |
| inst | sg | i | fnct | týmž |
| inst | sg | i | fnct | tímtéž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | fnct | tytéž |
| gen | pl | i | | týchž |
| dat | pl | i | | týmž |
| acc | pl | i | | tytéž |
| acc | pl | i | reg | tytéž |
| voc | pl | i | fnct | tytéž |
| voc | pl | i | fnct | titíž |
| loc | pl | i | | týchž |
| inst | pl | i | | týmiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | fnct | táž |
| nom | sg | f | fnct | tatáž |
| gen | sg | f | | téže |
| gen | sg | f | reg | týže |
| dat | sg | f | | téže |
| dat | sg | f | reg | týže |
| acc | sg | f | fnct | touž |
| acc | sg | f | fnct | tutéž |
| voc | sg | f | fnct | táž |
| voc | sg | f | fnct | tatáž |
| loc | sg | f | fnct | téže |
| inst | sg | f | fnct | touž |
| inst | sg | f | fnct | toutéž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | tytéž |
| gen | pl | f | | týchž |
| dat | pl | f | | týmž |
| acc | pl | f | | tytéž |
| acc | pl | f | reg | tytýž |
| voc | pl | f | | tytéž |
| nom | pl | f | | tytéž |
| loc | pl | f | | týchž |
| inst | pl | f | | týmiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | totéž |
| gen | sg | n | | téhož |
| gen | sg | n | reg | týhož |
| dat | sg | n | | témuž |
| dat | sg | n | reg | týmuž |
| acc | sg | n | | totéž |
| voc | sg | n | | totéž |
| voc | sg | n | fnct | tentýž |
| loc | sg | n | fnct | témž |
| loc | sg | n | fnct | témže |
| loc | sg | n | fnct | tomtémž |
| inst | sg | n | fnct | týmž |
| inst | sg | n | fnct | tímtéž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | fnct | táž |
| nom | pl | n | fnct | tatáž |
| gen | pl | n | | týchž |
| dat | pl | n | | týmž |
| acc | pl | n | fnct | táž |
| acc | pl | n | fnct | tatáž |
| voc | pl | n | fnct | táž |
| voc | pl | n | fnct | tatáž |
| loc | pl | n | | týchž |
| inst | pl | n | | týmiž |

TAB. C.105 – Modèle *týž*

# Type *jarní*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| jarní | jn | 1 | 1 | další, větší, první, státní, ostatní, jarní |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | í |
| gen | sg | m | | ího |
| dat | sg | m | | ímu |
| acc | sg | m | | ího |
| voc | sg | m | | í |
| loc | sg | m | | ím |
| loc | sg | m | reg | im |
| inst | sg | m | | ím |
| inst | sg | m | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | í |
| gen | pl | m | | ích |
| dat | pl | m | | ím |
| dat | pl | m | reg | im |
| acc | pl | m | | í |
| voc | pl | m | | í |
| loc | pl | m | | ích |
| inst | pl | m | | ími |
| inst | pl | m | reg | íma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | í |
| gen | sg | i | | ího |
| dat | sg | i | | ímu |
| acc | sg | i | | í |
| voc | sg | i | | í |
| loc | sg | i | | ím |
| loc | sg | i | reg | im |
| inst | sg | i | | ím |
| inst | sg | i | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | í |
| gen | pl | i | | ích |
| dat | pl | i | | ím |
| acc | pl | i | | í |
| voc | pl | i | | í |
| loc | pl | i | | ích |
| inst | pl | i | | ími |
| inst | pl | i | reg | íma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | í |
| gen | sg | f | | í |
| dat | sg | f | | í |
| acc | sg | f | | í |
| voc | sg | f | | í |
| loc | sg | f | | í |
| inst | sg | f | | í |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | í |
| gen | pl | f | | ích |
| dat | pl | f | | ím |
| dat | pl | f | reg | im |
| acc | pl | f | | í |
| voc | pl | f | | í |
| loc | pl | f | | ích |
| inst | pl | f | | ími |
| inst | pl | f | reg | íma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | í |
| gen | sg | n | | ího |
| dat | sg | n | | ímu |
| acc | sg | n | | ího |
| voc | sg | n | | í |
| loc | sg | n | | ím |
| loc | sg | n | reg | im |
| inst | sg | n | | ím |
| inst | sg | n | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | í |
| gen | pl | n | | ích |
| dat | pl | n | | ím |
| dat | pl | n | reg | im |
| acc | pl | n | | í |
| voc | pl | n | | í |
| loc | pl | n | | ích |
| inst | pl | n | | ími |
| inst | pl | n | reg | íma |

TAB. C.106 – Modèle *jarní*

# C.3 Flexion mixte

## Type *otcův*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| otcův | ot | 1 | 1 | otcův, pánův |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | a |
| dat | sg | m | | u |
| acc | sg | m | | a |
| voc | sg | m | | # |
| loc | sg | m | | ě |
| inst | sg | m | | ým |
| inst | sg | m | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | i |
| gen | pl | m | | ých |
| gen | pl | m | reg | ejch |
| dat | pl | m | | ým |
| dat | pl | m | reg | ym |
| dat | pl | m | reg | ejm |
| acc | pl | m | | y |
| voc | pl | m | | i |
| loc | pl | m | | ých |
| loc | pl | m | reg | ejch |
| inst | pl | m | | ými |
| inst | pl | m | reg | ýma |
| inst | pl | m | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | a |
| dat | sg | i | | u |
| acc | sg | i | | # |
| voc | sg | i | | # |
| loc | sg | i | | ě |
| inst | sg | i | | ým |
| inst | sg | i | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | y |
| gen | pl | i | | ých |
| gen | pl | i | reg | ejch |
| dat | pl | i | | ým |
| dat | pl | i | reg | ym |
| dat | pl | i | reg | ejm |
| acc | pl | i | | y |
| voc | pl | i | | y |
| loc | pl | i | | ých |
| loc | pl | i | reg | ejch |
| inst | pl | i | | ými |
| inst | pl | i | reg | ýma |
| inst | pl | i | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | a |
| gen | sg | f | | y |
| dat | sg | f | | ě |
| acc | sg | f | | u |
| voc | sg | f | | a |
| loc | sg | f | | ě |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | y |
| gen | pl | f | | ých |
| gen | pl | f | reg | ejch |
| dat | pl | f | | ým |
| dat | pl | f | reg | ym |
| dat | pl | f | reg | ejm |
| acc | pl | f | | y |
| voc | pl | f | | y |
| loc | pl | f | | ých |
| loc | pl | f | reg | ejch |
| inst | pl | f | | ými |
| inst | pl | f | reg | ýma |
| inst | pl | f | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | o |
| gen | sg | n | | a |
| dat | sg | n | | u |
| acc | sg | n | | o |
| voc | sg | n | | o |
| loc | sg | n | | ě |
| inst | sg | n | | ým |
| inst | sg | n | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | a |
| nom | pl | n | reg | y |
| gen | pl | n | | ých |
| gen | pl | n | reg | ejch |
| dat | pl | n | | ým |
| dat | pl | n | reg | ym |
| dat | pl | n | reg | ejm |
| acc | pl | n | | a |
| acc | pl | n | reg | y |
| voc | pl | n | | a |
| voc | pl | n | reg | y |
| loc | pl | n | | ých |
| loc | pl | n | reg | ejch |
| inst | pl | n | | ými |
| inst | pl | n | reg | ýma |
| inst | pl | n | reg | ejma |

TAB. C.107 – Modèle *otcův*

## Type *matčin*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| matčin | mn | 1 | 1 | matčin, ženin |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | a |
| dat | sg | m | | u |
| acc | sg | m | | a |
| voc | sg | m | | # |
| loc | sg | m | | ě |
| inst | sg | m | | ým |
| inst | sg | m | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | i |
| gen | pl | m | | ých |
| gen | pl | m | reg | ejch |
| dat | pl | m | | ým |
| dat | pl | m | reg | ym |
| dat | pl | m | reg | ejm |
| acc | pl | m | | y |
| voc | pl | m | | i |
| loc | pl | m | | ých |
| loc | pl | m | reg | ejch |
| inst | pl | m | | ými |
| inst | pl | m | reg | ýma |
| inst | pl | m | reg | ejma |
| inst | pl | m | reg | ami |
| inst | pl | m | reg | ama |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | a |
| dat | sg | i | | u |
| acc | sg | i | | # |
| voc | sg | i | | # |
| loc | sg | i | | ě |
| inst | sg | i | | ým |
| inst | sg | i | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | y |
| gen | pl | i | | ých |
| gen | pl | i | reg | ejch |
| dat | pl | i | | ým |
| dat | pl | i | reg | ym |
| dat | pl | i | reg | ejm |
| acc | pl | i | | y |
| voc | pl | i | | y |
| loc | pl | i | | ých |
| loc | pl | i | reg | ejch |
| inst | pl | i | | ými |
| inst | pl | i | reg | ýma |
| inst | pl | i | reg | ejma |
| inst | pl | i | reg | ami |
| inst | pl | i | reg | ama |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | a |
| gen | sg | f | | y |
| dat | sg | f | | ě |
| acc | sg | f | | u |
| voc | sg | f | | a |
| loc | sg | f | | ě |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | y |
| gen | pl | f | | ých |
| gen | pl | f | reg | ejch |
| dat | pl | f | | ým |
| dat | pl | f | reg | ym |
| dat | pl | f | reg | ejm |
| acc | pl | f | | y |
| voc | pl | f | | y |
| loc | pl | f | | ých |
| loc | pl | f | reg | ejch |
| inst | pl | f | | ými |
| inst | pl | f | reg | ýma |
| inst | pl | f | reg | ami |
| inst | pl | f | reg | ama |
| inst | pl | f | reg | ejma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | o |
| gen | sg | n | | a |
| dat | sg | n | | ě |
| acc | sg | n | | o |
| voc | sg | n | | o |
| loc | sg | n | | ě |
| inst | sg | n | | ým |
| inst | sg | n | reg | ym |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | a |
| gen | pl | n | | ých |
| gen | pl | n | reg | ejch |
| dat | pl | n | | ým |
| dat | pl | n | reg | ym |
| dat | pl | n | reg | ejm |
| acc | pl | n | | a |
| voc | pl | n | | a |
| loc | pl | n | | ých |
| loc | pl | n | reg | ejch |
| inst | pl | n | | ými |
| inst | pl | n | reg | ýma |
| inst | pl | n | reg | ejma |
| inst | pl | n | reg | ami |
| inst | pl | n | reg | ama |

TAB. C.108 – Modèle *matčin*

# C.4 Flexion pronominale

## Type *já*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| já | ps_ja | 0 | 1 | já |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | s | | já |
| gen | sg | s | fnct | mě |
| gen | sg | s | fnct | mne |
| dat | sg | s | fnct | mi |
| dat | sg | s | fnct | mně |
| acc | sg | s | fnct | mě |
| acc | sg | s | fnct | mne |
| voc | sg | s | | |
| loc | sg | s | | mně |
| inst | sg | s | | mnou |

TAB. C.109 – Modèle *já*

## Type *my*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| my | ps_my | 0 | 1 | my |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | s | | my |
| gen | pl | s | | nás |
| dat | pl | s | | nám |
| acc | pl | s | | nás |
| voc | pl | s | | |
| loc | pl | s | | nás |
| inst | pl | s | | námi |
| inst | pl | s | reg | náma |

TAB. C.110 – Modèle *my*

## Type *ty*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ty | ps_ty | 0 | 1 | ty |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | s | | ty |
| gen | sg | s | fnct | tě |
| gen | sg | s | fnct | tebe |
| dat | sg | s | fnct | ti |
| dat | sg | s | fnct | tobě |
| acc | sg | s | fnct | tě |
| acc | sg | s | fnct | tebe |
| voc | sg | s | | te |
| loc | sg | s | | tobě |
| inst | sg | s | | tebou |

TAB. C.111 – Modèle *ty*

## Type *vy*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| vy | ps_vy | 0 | 1 | vy |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | s | | vy |
| gen | pl | s | | vás |
| dat | pl | s | | vám |
| acc | pl | s | | vás |
| voc | pl | s | | vy |
| loc | pl | s | | vás |
| inst | pl | s | | vámi |
| inst | pl | s | reg | váma |

TAB. C.112 – Modèle *vy*

## Type *se*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| se | ps_r | 0 | 1 | se |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | s | | |
| gen | sg | s | | sebe |
| dat | sg | s | fnct | si |
| dat | sg | s | fnct | sobě |
| acc | sg | s | fnct | se |
| acc | sg | s | fnct | sebe |
| voc | sg | s | | |
| loc | sg | s | | sobě |
| inst | sg | s | | sebou |

TAB. C.113 – Modèle *se*

## Type *ten*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| ten | tn | 0 | 1 | ten, tamten, támhleten, onen, jeden |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | en |
| gen | sg | m | | oho |
| dat | sg | m | | omu |
| acc | sg | m | | oho |
| voc | sg | m | | |
| loc | sg | m | | om |
| inst | sg | m | | ím |
| inst | sg | m | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | i |
| gen | pl | m | | ěch |
| dat | pl | m | | ěm |
| acc | pl | m | | y |
| voc | pl | m | | |
| loc | pl | m | | ěch |
| inst | pl | m | | ěmi |
| inst | pl | m | reg | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | en |
| gen | sg | i | | oho |
| dat | sg | i | | omu |
| acc | sg | i | | en |
| voc | sg | i | | |
| loc | sg | i | | om |
| inst | sg | i | | ím |
| inst | sg | i | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | y |
| gen | pl | i | | ěch |
| dat | pl | i | | ěm |
| acc | pl | i | | y |
| voc | pl | i | | |
| loc | pl | i | | ěch |
| inst | pl | i | | ěmi |
| inst | pl | i | reg | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | a |
| gen | sg | f | | é |
| gen | sg | f | reg | ý |
| dat | sg | f | | é |
| dat | sg | f | reg | ý |
| acc | sg | f | | u |
| voc | sg | f | | |
| loc | sg | f | | é |
| loc | sg | f | reg | ý |
| inst | sg | f | | ou |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | y |
| gen | pl | f | | ěch |
| dat | pl | f | | ěm |
| acc | pl | f | | y |
| voc | pl | f | | |
| loc | pl | f | | ěch |
| inst | pl | f | | ěmi |
| inst | pl | f | reg | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | o |
| gen | sg | n | | oho |
| dat | sg | n | | omu |
| acc | sg | n | | o |
| voc | sg | n | | |
| loc | sg | n | | om |
| inst | sg | n | | ím |
| inst | sg | n | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | a |
| gen | pl | n | | ěch |
| dat | pl | n | | ěm |
| acc | pl | n | | a |
| voc | pl | n | | |
| loc | pl | n | | ěch |
| inst | pl | n | | ěmi |
| inst | pl | n | reg | ěma |

TAB. C.114 – Modèle *ten*

## Type *kdo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kdo | tn_k | 0 | 1 | kdo, někdo, kdosi, kdopak |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | s | | kdo |
| gen | sg | s | | koho |
| dat | sg | s | | komu |
| acc | sg | s | | koho |
| voc | sg | s | | |
| loc | sg | s | | kom |
| inst | sg | s | | kým |

TAB. C.115 – Modèle *kdo*

## Type *náš*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| náš | ns | 0 | 1 | náš, váš |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | # |
| gen | sg | m | | eho |
| dat | sg | m | | emu |
| acc | sg | m | | eho |
| voc | sg | m | | # |
| loc | sg | m | | em |
| inst | sg | m | | ím |
| inst | sg | m | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | i |
| gen | pl | m | | ich |
| dat | pl | m | | im |
| acc | pl | m | | e |
| voc | pl | m | | i |
| loc | pl | m | | ich |
| inst | pl | m | | imi |
| inst | pl | m | reg | ima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | # |
| gen | sg | i | | eho |
| dat | sg | i | | emu |
| acc | sg | i | | # |
| voc | sg | i | | # |
| loc | sg | i | | em |
| inst | sg | i | | ím |
| inst | sg | i | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | e |
| gen | pl | i | | ich |
| dat | pl | i | | im |
| acc | pl | i | | e |
| voc | pl | i | | e |
| loc | pl | i | | ich |
| inst | pl | i | | imi |
| inst | pl | i | reg | ima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | e |
| gen | sg | f | | í |
| dat | sg | f | | í |
| acc | sg | f | | i |
| voc | sg | f | | # |
| loc | sg | f | | í |
| inst | sg | f | | í |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | e |
| gen | pl | f | | ich |
| dat | pl | f | | im |
| acc | pl | f | | e |
| voc | pl | f | | e |
| loc | pl | f | | ich |
| inst | pl | f | | imi |
| inst | pl | f | reg | ima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | e |
| gen | sg | n | | eho |
| dat | sg | n | | emu |
| acc | sg | n | | e |
| voc | sg | n | | e |
| loc | sg | n | | em |
| inst | sg | n | | ím |
| inst | sg | n | reg | im |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | e |
| gen | pl | n | | ich |
| dat | pl | n | | im |
| acc | pl | n | | e |
| voc | pl | n | | e |
| loc | pl | n | | ich |
| inst | pl | n | | imi |
| inst | pl | n | reg | ima |

TAB. C.116 – Modèle *náš*

## Type *co*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| co | ns_c | 0 | 1 | co, něco, cosi, copak |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | co |
| gen | sg | n | | čeho |
| dat | sg | n | | čemu |
| acc | sg | n | | co |
| voc | sg | n | | |
| loc | sg | n | | čem |
| inst | sg | n | | čím |

TAB. C.117 – Modèle *co*

## Type *nic*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| nic | ns_n | 0 | 1 | nic, pranic, prachnic |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | nic |
| gen | sg | n | | ničeho |
| dat | sg | n | | ničemu |
| acc | sg | n | | nic |
| voc | sg | n | | |
| loc | sg | n | | ničem |
| inst | sg | n | | ničím |

TAB. C.118 – Modèle *nic*

## Type *on*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| on | ns_o | 0 | 1 | on |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | on |
| gen | sg | m | fnct | jeho |
| gen | sg | m | fnct | jej |
| gen | sg | m | fnct | něho |
| gen | sg | m | fnct | něj |
| dat | sg | m | fnct | jemu |
| dat | sg | m | fnct | mu |
| acc | sg | m | fnct | jeho |
| acc | sg | m | fnct | jej |
| acc | sg | m | fnct | něho |
| acc | sg | m | fnct | něj |
| acc | sg | m | fnct | ho |
| acc | sg | m | fnct | ň |
| voc | sg | m | | |
| loc | sg | m | | něm |
| inst | sg | m | fnct | jím |
| inst | sg | m | fnct | ním |
| inst | sg | m | reg | jim |
| inst | sg | m | reg | nim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | oni |
| gen | pl | m | fnct | jich |
| gen | pl | m | fnct | nich |
| dat | pl | m | fnct | jim |
| dat | pl | m | fnct | nim |
| acc | pl | m | fnct | je |
| acc | pl | m | fnct | ně |
| voc | pl | m | | |
| loc | pl | m | | nich |
| inst | pl | m | fnct | jimi |
| inst | pl | m | fnct | nimi |
| inst | pl | m | reg | jima |
| inst | pl | m | reg | nima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | on |
| gen | sg | i | fnct | jeho |
| gen | sg | i | fnct | jej |
| gen | sg | i | fnct | něho |
| gen | sg | i | fnct | něj |
| dat | sg | i | fnct | jemu |
| dat | sg | i | fnct | mu |
| acc | sg | i | fnct | jej |
| acc | sg | i | fnct | něj |
| acc | sg | i | fnct | ho |
| acc | sg | i | fnct | ň |
| voc | sg | i | | |
| loc | sg | i | | něm |
| inst | sg | i | fnct | jím |
| inst | sg | i | fnct | ním |
| inst | sg | i | reg | jim |
| inst | sg | i | reg | nim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | ony |
| gen | pl | i | fnct | jich |
| gen | pl | i | fnct | nich |
| dat | pl | i | fnct | jim |
| dat | pl | i | fnct | nim |
| acc | pl | i | fnct | je |
| acc | pl | i | fnct | ně |
| voc | pl | i | | |
| loc | pl | i | | nich |
| inst | pl | i | fnct | jimi |
| inst | pl | i | fnct | nimi |
| inst | pl | i | reg | jima |
| inst | pl | i | reg | nima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | ona |
| gen | sg | f | fnct | jí |
| gen | sg | f | fnct | ní |
| dat | sg | f | fnct | jí |
| dat | sg | f | fnct | ní |
| acc | sg | f | fnct | ji |
| acc | sg | f | fnct | ni |
| acc | sg | f | fnct | jí |
| voc | sg | f | | |
| loc | sg | f | | ní |
| inst | sg | f | fnct | jí |
| inst | sg | f | fnct | ní |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | ony |
| gen | pl | f | fnct | jich |
| gen | pl | f | fnct | nich |
| dat | pl | f | fnct | jim |
| dat | pl | f | fnct | nim |
| acc | pl | f | fnct | je |
| acc | pl | f | fnct | ně |
| voc | pl | f | | |
| loc | pl | f | | nich |
| inst | pl | f | fnct | jimi |
| inst | pl | f | fnct | nimi |
| inst | pl | f | reg | jima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | ono |
| gen | sg | n | fnct | jeho |
| gen | sg | n | fnct | jej |
| gen | sg | n | fnct | něho |
| gen | sg | n | fnct | něj |
| dat | sg | n | fnct | jemu |
| dat | sg | n | fnct | mu |
| acc | sg | n | fnct | jeho |
| acc | sg | n | fnct | jej |
| acc | sg | n | fnct | něho |
| acc | sg | n | fnct | něj |
| acc | sg | n | fnct | ho |
| voc | sg | n | | |
| loc | sg | n | | něm |
| inst | sg | n | fnct | jím |
| inst | sg | n | fnct | ním |
| inst | sg | n | reg | jim |
| inst | sg | n | reg | nim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | ona |
| gen | pl | n | fnct | jich |
| gen | pl | n | fnct | nich |
| dat | pl | n | fnct | jim |
| dat | pl | n | fnct | nim |
| acc | pl | n | fnct | je |
| acc | pl | n | fnct | ně |
| voc | pl | n | | |
| loc | pl | n | | nich |
| inst | pl | n | fnct | jimi |
| inst | pl | n | fnct | nimi |
| inst | pl | n | reg | jima |
| inst | pl | n | reg | nima |
| inst | pl | n | reg | nima |

TAB. C.119 – Modèle *on*

# Type *jenž*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| jenž | ns_j | 0 | 0 | jenž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | jenž |
| gen | sg | m | fnct | jehož |
| gen | sg | m | fnct | něhož |
| gen | sg | m | fnct | jejž |
| gen | sg | m | fnct | nějž |
| dat | sg | m | fnct | jemuž |
| acc | sg | m | fnct | jehož |
| acc | sg | m | fnct | něhož |
| acc | sg | m | fnct | jejž |
| acc | sg | m | fnct | nějž |
| voc | sg | m | | |
| loc | sg | m | | němž |
| inst | sg | m | fnct | jímž |
| inst | sg | m | fnct | nímž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | již |
| gen | pl | m | fnct | jichž |
| gen | pl | m | fnct | nichž |
| dat | pl | m | fnct | jimž |
| dat | pl | m | fnct | nimž |
| acc | pl | m | fnct | jež |
| acc | pl | m | fnct | něž |
| voc | pl | m | | |
| loc | pl | m | | nichž |
| inst | pl | m | fnct | jimiž |
| inst | pl | m | fnct | nimiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | jenž |
| gen | sg | i | fnct | jehož |
| gen | sg | i | fnct | něhož |
| gen | sg | i | fnct | jejž |
| gen | sg | i | fnct | nějž |
| dat | sg | i | fnct | jemuž |
| acc | sg | i | fnct | jehož |
| acc | sg | i | fnct | něhož |
| acc | sg | i | fnct | jejž |
| acc | sg | i | fnct | nějž |
| voc | sg | i | | |
| loc | sg | i | | němž |
| inst | sg | i | fnct | jímž |
| inst | sg | i | fnct | nímž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | jež |
| gen | pl | i | fnct | jichž |
| gen | pl | i | fnct | nichž |
| dat | pl | i | fnct | jimž |
| dat | pl | i | fnct | nimž |
| acc | pl | i | fnct | jež |
| acc | pl | i | fnct | něž |
| voc | pl | i | | |
| loc | pl | i | | nichž |
| inst | pl | i | fnct | jimiž |
| inst | pl | i | fnct | nimiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | | jež |
| gen | sg | f | fnct | jíž |
| gen | sg | f | fnct | níž |
| dat | sg | f | fnct | jíž |
| dat | sg | f | fnct | níž |
| acc | sg | f | | již |
| voc | sg | f | | |
| loc | sg | f | | níž |
| inst | sg | f | fnct | jíž |
| inst | sg | f | fnct | níž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | jež |
| gen | pl | f | fnct | jichž |
| gen | pl | f | fnct | nichž |
| dat | pl | f | fnct | jimž |
| dat | pl | f | fnct | nimž |
| acc | pl | f | fnct | jež |
| acc | pl | f | fnct | něž |
| voc | pl | f | | |
| loc | pl | f | | nichž |
| inst | pl | f | fnct | jimiž |
| inst | pl | f | fnct | nimiž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | | jež |
| gen | sg | n | fnct | jehož |
| gen | sg | n | fnct | něhož |
| gen | sg | n | fnct | jejž |
| gen | sg | n | fnct | nějž |
| dat | sg | n | fnct | jemuž |
| acc | sg | n | fnct | jež |
| acc | sg | n | fnct | něž |
| voc | sg | n | | |
| loc | sg | n | | němž |
| inst | sg | n | fnct | jímž |
| inst | sg | n | fnct | nímž |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | jež |
| gen | pl | n | fnct | jichž |
| gen | pl | n | fnct | nichž |
| dat | pl | n | fnct | jimž |
| dat | pl | n | fnct | nimž |
| acc | pl | n | fnct | jež |
| acc | pl | n | fnct | něž |
| voc | pl | n | | |
| loc | pl | n | | nichž |
| inst | pl | n | fnct | jimiž |
| inst | pl | n | fnct | nimiž |

TAB. C.120 – Modèle *jenž*

# Type *všechen*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| všechen | ns_v | 0 | 1 | všechen |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | m | | všechen |
| gen | sg | m | | všeho |
| dat | sg | m | | všemu |
| acc | sg | m | | všeho |
| voc | sg | m | | všechen |
| loc | sg | m | | všem |
| inst | sg | m | | vším |
| inst | sg | m | reg | všim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | všchni |
| gen | pl | m | | všech |
| dat | pl | m | | všem |
| acc | pl | m | | všechny |
| voc | pl | m | | všchni |
| loc | pl | m | | všech |
| inst | pl | m | | všemi |
| inst | pl | m | reg | všema |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | i | | všecek |
| gen | sg | i | | všeho |
| dat | sg | i | | všemu |
| acc | sg | i | | všecek |
| voc | sg | i | | všecek |
| loc | sg | i | | všem |
| inst | sg | i | | vším |
| inst | sg | i | reg | všim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | fnct | všechny |
| nom | pl | i | fnct | všecky |
| gen | pl | i | | všech |
| dat | pl | i | | všem |
| acc | pl | i | fnct | všechny |
| acc | pl | i | fnct | všecky |
| voc | pl | i | fnct | všecky |
| voc | pl | i | fnct | všechny |
| loc | pl | i | | všech |
| inst | pl | i | | všemi |
| inst | pl | i | reg | všema |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | f | fnct | všechna |
| nom | sg | f | fnct | všecka |
| gen | sg | f | | vší |
| dat | sg | f | | vší |
| acc | sg | f | fnct | vší |
| acc | sg | f | fnct | všechnu |
| voc | sg | f | fnct | všechna |
| voc | sg | f | fnct | všecka |
| loc | sg | f | | vší |
| inst | sg | f | | vší |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | fnct | všechny |
| nom | pl | f | fnct | všecky |
| gen | pl | f | | všech |
| dat | pl | f | | všem |
| acc | pl | f | fnct | všechny |
| acc | pl | f | fnct | všecky |
| voc | pl | f | fnct | všechny |
| voc | pl | f | fnct | všecky |
| loc | pl | f | | všech |
| inst | pl | f | | všemi |
| inst | pl | f | reg | všima |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | sg | n | fnct | vše |
| nom | sg | n | fnct | všechno |
| nom | sg | n | fnct | všecko |
| gen | sg | n | | všeho |
| dat | sg | n | | všemu |
| acc | sg | n | fnct | vše |
| acc | sg | n | fnct | všechno |
| acc | sg | n | fnct | všecko |
| voc | sg | n | fnct | všechno |
| voc | sg | n | fnct | všecko |
| loc | sg | n | | všem |
| inst | sg | n | | vším |
| inst | sg | n | reg | všim |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | fnct | všechna |
| nom | pl | n | fnct | všecky |
| gen | pl | n | | všech |
| dat | pl | n | | všem |
| acc | pl | n | | všechna |
| acc | pl | n | | všecky |
| voc | pl | n | fnct | všechna |
| voc | pl | n | fnct | všecky |
| loc | pl | n | | všech |
| inst | pl | n | | všemi |
| inst | pl | n | reg | všema |

TAB. C.121 – Modèle *všechen*

# C.5 Flexion numérale

## Type *dva*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| dva | dv | 0 | 1 | dva, oba |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | a |
| gen | pl | m | | ou |
| dat | pl | m | | ěma |
| acc | pl | m | | a |
| voc | pl | m | | a |
| loc | pl | m | | ou |
| inst | pl | m | | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | a |
| gen | pl | i | | ou |
| dat | pl | i | | ěma |
| acc | pl | i | | a |
| voc | pl | i | | a |
| loc | pl | i | | ou |
| inst | pl | i | | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | ě |
| gen | pl | f | | ou |
| dat | pl | f | | ěma |
| acc | pl | f | | ě |
| voc | pl | f | | ě |
| loc | pl | f | | ou |
| inst | pl | f | | ěma |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | ě |
| gen | pl | n | | ou |
| dat | pl | n | | ěma |
| acc | pl | n | | ě |
| voc | pl | n | | ě |
| loc | pl | n | | ou |
| inst | pl | n | | ěma |

TAB. C.122 – Modèle *dva*

## Type *pět*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| pět | nu_i | 1 | 1 | pět, deset, devadesát, jedenáct, třicet |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | # |
| gen | pl | m | | i |
| dat | pl | m | | i |
| acc | pl | m | | # |
| voc | pl | m | | # |
| loc | pl | m | | i |
| inst | pl | m | | i |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | # |
| gen | pl | i | | i |
| dat | pl | i | | i |
| acc | pl | i | | # |
| voc | pl | i | | # |
| loc | pl | i | | i |
| inst | pl | i | | i |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | # |
| gen | pl | f | | i |
| dat | pl | f | | i |
| acc | pl | f | | # |
| voc | pl | f | | # |
| loc | pl | f | | i |
| inst | pl | f | | i |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | # |
| gen | pl | n | | i |
| dat | pl | n | | i |
| acc | pl | n | | # |
| voc | pl | n | | # |
| loc | pl | n | | i |
| inst | pl | n | | i |

TAB. C.123 – Modèle *pět*

## Type *kolik*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| kolik | nu_a | 0 | 1 | kolik, tolik |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | # |
| gen | pl | m | | a |
| dat | pl | m | | a |
| acc | pl | m | | # |
| voc | pl | m | | # |
| loc | pl | m | | a |
| inst | pl | m | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | # |
| gen | pl | i | | a |
| dat | pl | i | | a |
| acc | pl | i | | # |
| voc | pl | i | | # |
| loc | pl | i | | a |
| inst | pl | i | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | # |
| gen | pl | f | | a |
| dat | pl | f | | a |
| acc | pl | f | | # |
| voc | pl | f | | # |
| loc | pl | f | | a |
| inst | pl | f | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | # |
| gen | pl | n | | a |
| dat | pl | n | | a |
| acc | pl | n | | # |
| voc | pl | n | | # |
| loc | pl | n | | a |
| inst | pl | n | | a |

TAB. C.124 – Modèle *kolik*

## Type *málo*

| mnemo | pdgm | ouvert | A1A2 | exemples |
|---|---|---|---|---|
| málo | nu_o | 0 | 1 | málo, mnoho |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | m | | o |
| gen | pl | m | | a |
| dat | pl | m | | a |
| acc | pl | m | | o |
| voc | pl | m | | o |
| loc | pl | m | | o |
| inst | pl | m | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | i | | o |
| gen | pl | i | | a |
| dat | pl | i | | a |
| acc | pl | i | | o |
| voc | pl | i | | o |
| loc | pl | i | | o |
| inst | pl | i | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | f | | o |
| gen | pl | f | | a |
| dat | pl | f | | a |
| acc | pl | f | | o |
| voc | pl | f | | o |
| loc | pl | f | | o |
| inst | pl | f | | a |

| cas | num | gen | var | term |
|---|---|---|---|---|
| nom | pl | n | | o |
| gen | pl | n | | a |
| dat | pl | n | | a |
| acc | pl | n | | o |
| voc | pl | n | | o |
| loc | pl | n | | o |
| inst | pl | n | | a |

TAB. C.125 – Modèle *málo*

# Annexe D

# Alternances

Cette annexe contient les alternances morphématiques présentées dans le chapitre *Annotation*, section 3.4, p. 69–98. Elle a été générée automatiquement par un script PHP qui retranscrit le fichier *alt.xml* (voir la section 3.5.5, p. 98) au format LaTeX.

## D.1 Alternances consonantiques

### Mouillure

| @mnemo | @dep | @arr |
|---|---|---|
| d > t | d | t |
| t > t | t | t |
| n > n | n | n |
| r > ř | r | ř |

### Alternance graphique

| @mnemo | @dep | @arr |
|---|---|---|
| d > ď | d | ď |
| ď > d | ď | d |
| n > ň | n | ň |
| ň > n | ň | n |
| t > ť | t | ť |
| ť > t | ť | t |

### Dépalatalisation

| @mnemo | @dep | @arr |
|---|---|---|
| ď > d | ď | d |
| ť > t | ť | t |

### Palatalisation A

| @mnemo | @dep | @arr |
|---|---|---|
| k > c | k | c |
| g > z | g | z |
| h > z | h | z |
| ch > š | ch | š |

### Palatalisation B

| @mnemo | @dep | @arr |
|---|---|---|
| k > č | k | č |
| c > č | g | č |
| h > ž | h | ž |

### Palatalisation de groupe

| @mnemo | @dep | @arr |
|---|---|---|
| sk > št | sk | št |
| ck > čt | ck | čt |

## D.2 Alternances vocaliques

### Alternance qualitatives

| @mnemo | @dep | @arr | <except> | |
|---|---|---|---|---|
| a > e | a | e | @neg | <u> |
| | | | oui | Jan |
| e > a | e | a | @neg | <u> |
| | | | oui | vejce |
| í > á | í | á | @neg | <u> |
| | | | oui | přítel |

## Alternance quantitatives

| @mnemo | @dep | @arr | <except> | | | |
|---|---|---|---|---|---|---|
| á > a | a | e | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | blána, brána, bába, čára, dáma, dráha, chvála, jáma, kráva, pára, rána, skála, tráva, vrána, váha, záda, žába, práce |
| | | | oui | NbSyllabes | 1 | hrách, mráz, práh, pán, náš, váš |
| é > e | é | e | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | jméno, léto, péro |
| | | | oui | NbSyllabes | 1 | chléb |
| í > i | í | i | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | lípa, mísa, slíva, síla, síra, žíla |
| | | | oui | NbSyllabes | 1 | líh |
| í > ě | í | ě | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | bída, díra, míra, víra, dílo, svíce, peníze, dítě, vítr |
| | | | oui | NbSyllabes | 1 | sníh |
| ou > u | ou | u | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | bouda, houba, hrouda, louka, smlouva, strouha, trouba |
| ů > o | ů | o | @neg | @att_spec | @val_spec | <u> |
| | | | oui | NbSyllabes | 2 | hůra, fůra, smůla |
| | | | non | NbSyllabes | 1 | kůl, kůr, půl, růž, trůn, tůň |

# Alternance de *-e-* intercalaire

| @mnemo | @dep | @arr | <except> | | | |
|---|---|---|---|---|---|---|
| | | | @neg | @att_spec | @val_spec | <u> |
| e > # | e | # | non | PeriphLemme | k | bek, flek, jek, klek, rek, skřek, šnek, špek, šek, vlek, vztek, dotek, oblek, poklek, průsek, úlek, úsek, výsek, zásek |
| | | | non | PeriphLemme | c | hec, kec, klec, pec, plec |
| | | | non | PeriphLemme | ň | dřeň, žeň, holeň, zeleň, tuleň, červeň |
| | | | oui | PeriphLemme | m | dojem, nezájem, nájem, pojem, pronájem, podnájem, příjem, zájem, průjem |
| | | | oui | PeriphLemme | v | broskev, brukev, církev, konev, koroptev, korouhev, lahev, láhev, mrkev, pánev, plástev, ploutev, podešev, rakev, ředkev, tykev, větev, krev, lev, šev |
| | | | oui | PeriphLemme | n | blázen, březen, buben, červen, den, duben, hrozen, jícen, květen, leden, osten, půlden, říjen, sen, srpen, svícen, týden, jeden |
| | | | oui | PeriphLemme | š | veš, faleš |
| | | | oui | PeriphLemme | l | ďábel, chumel, kachel, kašel, kel, kotel, kozel, kyčel, obratel, orel, osel, posel, pytel, uhel, uzel, úhel, živel |
| | | | oui | PeriphLemme | t | bezpočet, kmitočet, letopočet, počet, přepočet, odpočet, součet, výpočet, účet, zápočet, výčet, dehet, loket, nehet, ocet, ret, tucet, půltucet |
| | | | oui | PeriphLemme | s | pes, kočkopes, ves, náves |
| | | | oui | PeriphLemme | r | uher |
| | | | oui | PeriphLemme | b | pohřeb |
| | | | oui | PeriphLemme | ď | zeď |
| | | | oui | PeriphLemme | ž | lež |

| @mnemo | @dep | @arr | <except> | | | |
|---|---|---|---|---|---|---|
| | | | @neg | @att_spec | @val_spec | <u> |
| # > e | # | e | non | PeriphLemme | la | bulla, mulla |
| | | | non | PeriphLemme | ma | gamma |
| | | | non | PeriphLemme | ra | |
| | | | non | PeriphLemme | ka | banka, odmlka, pomlka, zámlka |
| | | | non | PeriphLemme | na | bavlna, vlna, lajna, moderna, postmoderna, obrna, skvrna, srna |
| | | | non | PeriphLemme | ba | bomba, sexbomba, gamba, plomba, rumba, samba |
| | | | non | PeriphLemme | va | brva, konzerva, mrva, rezerva, salva, želva |
| | | | oui | PeriphLemme | ta | buchta, jachta, kruchta, plachta, plichta, rychta, šachta, šichta, kuchta, karta |
| | | | oui | PeriphLemme | pa | výspa, zácpa, škarpa |
| | | | oui | PeriphLemme | sa | kapsa |
| | | | non | PeriphLemme | da | mzda |
| | | | non | PeriphLemme | lo | cello, cemballo |
| | | | non | PeriphLemme | ko | brko, vojsko, disko, manko |
| | | | oui | PeriphLemme | mo | bělmo, jařmo, pižmo, pásmo, písmo |
| | | | non | PeriphLemme | no | porno, storno, terno, zrno, lejno, hejno |
| | | | non | PeriphLemme | ro | maestro |
| | | | non | PeriphLemme | vo | |
| | | | oui | PeriphLemme | ce | vejce |

# Bibliographie


ABEILLÉ, A. (1993). *Les nouvelles syntaxes. Grammaires d'unification et analyse du français*. Armand Colin, Paris.

ADAM, H. (1996). Problematika spisovnosti a nespisovnosti z hlediska češtiny jako cizího jazyka. *In Spisovnost a nespisovnost dnes*, pages 254–255. Masarykova univerzita, Brno.

ADAMOVIČOVÁ, A., BISCHOFOVÁ, J., CVEJNOVÁ, J., GLADKOVA, H., HASIL, J., HRDLIČKA, M., MAREŠ, P., NEKVAPIL, J., PALKOVÁ, Z., ŠÁRA, M. et HOLUB, J. (2005). *Čeština jako cizí jazyk. Úroveň B2*. Ministerstvo školství, mládeže a tělovýchovy, Univerzita Karlova, Ústav bohemistických studií, Praha.

ANTONIADIS, G., ECHINARD, S., KRAIF, O., LEBARBÉ, T., LOISEAU, M. et PONTON, C. (2004). NLP-based scripting for CALL activities. *In eLearning for Computational Linguistics and Computational Linguistics for eLearning International Workshop in Association with COLING 2004*, pages 18–25, Genève.

BAUTIER-CASTAING, E. (1977). Acquisition comparée de la syntaxe du français par des enfants francophones et non francophones. Étude expérimentale de quelques stratégies d'apprentissage. *Étude de linguistique appliquée*, 27:19–41.

BÉMOVÁ, A. et KRÁLÍKOVÁ, K. (1988). K otázkám automatického zpracování českého tvarosloví. *Slovo a slovesnost*, 49(4):285–295.

BERMEL, N. (2000). Spisovnost a nespisovnost v učebnicích češtiny pro cizince. *Čeština doma a ve světě*, VIII(1):52–63.

BERTIN, J.-C. (2001). *Des outils pour des langues. Multimédia et Apprentissage*. Ellipses Éditions, Paris.

BESSE, H. et PORQUIER, R. (1991). *Grammaires et didactiques des langues*. Hatier / Didier, Paris.

BIDLAS, V., CONFORTIOVÁ, H., TURZÍKOVÁ, M. et ČADSKÁ, M. (2005). *Čeština jako cizí jazyk. Úroveň A2*. Ministerstvo školství, mládeže a tělovýchovy, Univerzita Karlova, Ústav jazykové a odborné přípravy, Praha.

BRUGÈRE, T. et MOLLARD, A. (2003). *Mathématiques à l'usage des informaticiens*. Ellipses Éditions, Paris.

BRUILLARD, E. (1997). *Les machines à enseigner*. Hermès, Paris.

BUBENÍKOVÁ, L., KOLLMANNOVÁ, L. et SKÁLOVÁ, E. (1975). *Teorie a praxe jazykového testování*. Academia, Praha.

CAMERON, K., éditeur (1999). *CALL : Media, Design and Applications*. Swets & Zeitlinger, Lisse.

ČECHOVÁ, E. et KOL. (1993). *Pour les francophones qui veulent parler tchéque*. Česká Lípa.

ČECHOVÁ, M., HLAVSA, Z., DOKULIL, M., HRUŠKOVÁ, Z. et HRBÁČEK, J. (2000). *Čeština - řeč a jazyk*. ISV, Praha.

CECRL (2001). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Conseil de l'Europe, Division des Langues Vivantes, Didier, Paris.

CEJPEK, J., HAJIČOVÁ, E., KIRSCHNER, Z., KRÁLÍKOVÁ, K., PASTOREK, K. et SGALL, P. (1982). *Automatické vyhl'adavanie z úplného textu*. Bratislava.

ČEMUSOVÁ, J. et HOLÁ, L., éditeurs (2007). *Sborník Asociace učitelů češtiny jako cizího jazyka 2006-2007*. Akropolis, Praha.

ČERMÁK, F. (1985). Čeština jako cizí jazyk. Metodologické aspekty oboru a disciplíny. *In* TAX, J., éditeur : *Čeština jako cizí jazyk. Materiály z první metodologické konference USS FF UK*, pages 57–77. Univerzita Karlova, Praha.

ČERMÁK, F. (1997). Czech national corpus : A case in many contexts. *International Journal of Corpus Linguistics*, 2(2):181–197.

ČERMÁK, F. (2000). Index a indexálnost v jazyce : některé poznámky. *In* HLADKÁ, Z. et KARLÍK, P., éditeurs : *Čeština, univerzália a specifika 2*, pages 115–121, Brno. Masarykova univerzita.

ČERMÁK, F. et HOLUB, J. (2005). *Syntagmatika a paradigmatika českého slova. Valence a kolokabilita*. Karolinum, Praha.

ČERMÁK, F., HOLUB, J., HRONEK, J., ŠÁRA, M. et SHORT, D. (1993). *Czech. A multi-level course for advanced learners*. Brno.

ČERMÁK, F. et KŘEN, M. (2004). *Frekvenční slovník češtiny*. Nakladatelství Lidové Noviny, Praha.

CHAMPION, J. (1974). *Les langues africaines et la francophonie. Essai d'une pédagogie du français en Afrique noire par une analyse typologique des fautes.* Mouton, Paris.

CLARK, E. V. (2001). Morphology in language acquisition. *In* SPENCER, A. et M.ZWICKY, A., éditeurs : *The Handbook of Morphology*, pages 374–389. Blackwell publishers, Oxford.

CODD, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.

CONFORTIOVÁ, H. (1993). Relations between written and oral forms of language in teaching czech foreigners. *Slovo a slovesnost*, 54(3):165–168.

CONFORTIOVÁ, H. et MICHÁLKOVÁ, L. (1972). *Tchéque pour les étudiants parlant français*. Státní pedagogické nakladatelství, Praha.

CONNOLY, T. et BEGG, C. (2005). *Systèmes de bases de données*. Eyrolles, Québec.

CONTENSIN, M. (2004). *Bases de données et internet avec PHP et MySQL.* Dunod, Paris.

CORDER, S. P. (1981). *Error Analysis and Interlanguage*. Oxford University Press.

CORNU, E. (1994). Evaluating second language grammar checkers. *TRANEL (Travaux neuchâtelois de linguistique)*, 21:195–204.

CVRČEK, V. (2007). Rozdíl mezi psanou a mluvenou češtinou – kde je místo obecné češtiny ? *In* ČEMUSOVÁ, J. et ŠTINDLOVÁ, B., éditeurs : *Sborník Asociace učitelů češtiny jako cizího jazyka 2006-2007*, pages 171–179. Akropolis, Praha.

DAGNEAUX, E., DENNESS, S. et GRANGER, S. (1998). Computer-aided error analysis. *System*, 26(2):163–174.

DELVIGNA, C. (2005). Introduction aux bases de données et aux fondements de l'exemple relationnel. *Mathématiques et sciences humaines*, 169(1):5–41.

DEMAIZIÈRE, F. (2007). Didactique des langues et TIC : les aides à l'apprentissage. *ALSIC : Apprentissage des Langues et Systèmes d'Information et de Communication*.

DESMET, P. (2006). L'enseignement/apprentissage des langues à l'ère du numérique : tendances récentes et défis. *Revue française de linguistique appliquée*, XI(1):119–138.

DOCA, G. (1981). *Analyse psycholinguistique des erreurs faites lors de l'apprentissage d'une langue étrangère*. Publications de la Sorbonne, Paris.

DODIGOVIC, M. (2005). *Artificial Intelligence in Second Language Learning*. Multilingual Matters, Clevedon.

DOKTER, D., NERBONNE, J., SCHURCKS-GROZEVA, L. et SMIT, P. (1998). Glosser-rug : A user study. *In Language Teaching and Language Technology*, pages 167–176. Swets & Zeitlinger, Lisse.

DRESSLER, W. U. (1985). On the predicativeness of natural morphology. *Journal of Linguistics*, 21:321–337.

DRESSLER, W. U., MAYERTHALER, W., PANAGL, O. et WURZEL, W. U. (1987). *Leitmotifs in Natural Morphology*. Benjamins, Amsterdam / Philadelphia.

DUCHÁČEK, O. (1949). *Tchèque pour Français*. Orbis, Praha.

DUCHÁČEK, O. (1971). *Langue tchèque : histoire et norme actuelle*. Vander, Louvain.

DULAY, H., BURT, M. et KRASHEN, S. (1982). *Language Two*. Oxford University Press.

DUŠKOVÁ, L. (1969). On sources of errors in foreign language learning. *Internation Rewiev of Applied Linguistics*, 7:11–36.

ELLIS, R. (1994). *The Study of Second Language Acquisition*. Oxford University Press.

ELLIS, R. (1997). *Second language acquisition*. Oxford University Press.

FALTIN, A. V. (2003). *Syntactic Error Diagnosis in the context of Computer Assisted Language Learning*. Thèse de doctorat, Faculté des lettres, Université de Genève.

FAYOL, M., éditeur (2002). *Production du langage*. Hermès Science, Paris.

FERRAND, L. (2001). La production du langage : une vue d'ensemble. *Psychologie française*, 46(1):3–15.

FILIPEC, J. et DANEŠ, F., éditeurs (1978). *Slovník spisovná češtiny pro školu a veřejnost*. Academia, Praha.

FORTMANN, C. et FORST, M. (2004). An LFG grammar checker for CALL. *In Proceedings of InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems*, Venice.

FROULÍKOVÁ, L. (2002). *Zahrada českého jazyka*. Academia, Praha.

GAONAC'H, D. (1991). *Théories d'apprentissage et acquisition d'une langue étrangère*. Hatier / Didier, Paris.

GIACOBBE, J. (1990). Le recours à la langue première. *In Acquisition et utilisation d'une langue étrangère*, pages 115–123. Hachette, Paris.

GIACOBBE, J. (1992). *Acquisition d'une langue étrangère*. CNRS, Paris.

GRANGER, S., éditeur (1998). *Learner English on Computer*. Addison Wesley Longman, Austin, TX, USA.

GRANGER, S. (2003). Error-tagged learner corpora and CALL : A promising synergy. *CALICO*, 20(3):465–480.

GRANGER, S. (2004). Computer learner corpus research : Current status and future prospects. *In* CONNOR, U. et UPTON, T. A., éditeurs : *Applied Corpus linguistics : A multidimensional Perspective*, pages 123–145. Amsterdam and Atlanta.

GRANGER, S., HUNG, J. et PETCH-TYSON, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Benjamins, Amsterdam.

GRANGER, S., VANDEVENTER, A. et HAMEL, M.-J. (2001). Analyse de corpus d'apprenants pour l'ELAO basé sur TAL. *Traitement automatique des langues*, 42(2):609–622.

HÁDKOVÁ, M., LÍNEK, J. et VLASÁKOVÁ, K. (2005). *Čeština jako cizí jazyk. Úroveň A1*. Ministerstvo školství, mládeže a tělovýchovy, Univerzita Palackého v Olomouci, Katedra bohemistiky Filozofické fakulty.

HAJIČ, J. (2004). *Disambiguation of Rich Inflection. Computational Morphology of Czech*. Karolinum, Praha.

HAJIČ, J., HAJIČOVÁ, E., PAJAS, P., PANEVOVÁ, J., SGALL, P. et HLADKÁ, B. (2001). *The Prague Dependency Treebank. CDROM LDC2001T10*. Linguistic Data Consortium, Philadelphia : University of Pennsylvania.

HAJIČ, J. et HLADKÁ, B. (1997). Morfologické značkování korpusu českých textů stochastickou metodou. *Slovo a slovesnost*, 58(4):288–304.

HAJIČOVÁ, E. (2006). Využití korpusu pro ověřování lingvistických hypotéz. *In* ČERMÁK, F. et BLATNÁ, R., éditeurs : *Korpusová lingvistika. Stav a modelové přístupy*, pages 118–130. Nakladatelství Lidové Noviny, Ústav českého národního korpusu, Praha.

HAJIČOVÁ, E., PANEVOVÁ, J. et SGALL, P. (2002a). K nové úrovni bohemistické práce : Využití anotovaného korpusu (1. část). *Slovo a slovesnost*, 63(3):161–177.

HAJIČOVÁ, E., PANEVOVÁ, J. et SGALL, P. (2002b). K nové úrovni bohemistické práce : Využití anotovaného korpusu (2.část). *Slovo a slovesnost*, 63(4):241–262.

HAJIČOVÁ, E., PANEVOVÁ, J. et SGALL, P. (2004). Deep syntactic annotation : Tectogrammatical representation and beyond. *In* MEYERS, A., éditeur : *HLT-NAACL 2004 Workshop : Frontiers in Corpus Annotation*, pages 32–38. Association for Computational Linguistics, Boston.

HANA, J., ZEMAN, D., HAJIČ, J., HANOVÁ, H., HLADKÁ, B. et JEŘÁBEK, E. (2005). Manual for morphological annotation. Rapport technique 2005-27, ÚFAL.

HANSON-SMITH, E. (2003). A brief history of CALL theory. *CATESOL Journal*, 15(1):21–30.

HAVRÁNEK, B. et JEDLIČKA, A. (2002). *Stručná mluvnice česká*. Nakladatelství Fortuna, Praha.

HAWKINS, R. (2001). *Second language syntax : a generative introduction*. Blackwell publishers, New York.

HEIFT, T. (2003). Type or drag, but don't click : A study on the effectiveness of different call exercise types. *Canadian Journal of Applied Linguistics*, 6(3):69–87.

HEIFT, T. et SCHULZE, M. (2003). Error diagnosis and error correction in CAL : Introduction. *CALICO*, 20(3):433–436.

HEIFT, T. et SCHULZE, M. (2007). *Errors and Intelligence in Computer-Assisted Language Learning : Parsers and Pedagogues*. Routledge, UK.

HLAVÁČKOVÁ, D. (2001). Korpus mluvené češtiny z brněnského prostředí a jeho morfologické značkování. *Slovo a slovesnost*, 62(1):53–70.

HLAVÁČOVÁ, J. et SAVICKÝ, P. (2002). Measures of word commonness. *Journal of Quantitative Linguistics*, 9(3):215–213.

HOBZOVÁ, D. (1993). *Le tchèque tout de suite*. Pocket, Paris.

HOLLAND, V. M. et KAPLAN, J. D. (1995). NLP techniques in CALL : Status and instructional issues. *Instructional Science*, 23:352–380.

HRDLIČKA, M. (2000). *Předložky ve výuce češtiny jako cizího jazyka*. Karolinum, Praha.

HRDLIČKA, M. (2002). *Cizí jazyk čeština*. ISV, Praha.

HRDLIČKA, M. (2005). Jak vykládat cizincům aspekt. *In* ČEMUSOVÁ, J. et HOLÁ, L., éditeurs : *Sborník Asociace učitelů češtiny jako cizího jazyka 2006-2007*, pages 21–26. Akropolis, Praha.

HRDLIČKA, M. et HRDLIČKOVÁ, H. (1997). *Apprenez le tchèque avec nous*. ISV, Praha.

HRONEK, J. (1985). Čeština jako cizí jazyk. stav, problematika, potřeby. *In* TAX, J., éditeur : *Čeština jako cizí jazyk. Materiály z první metodologické konference USS FF UK*, pages 47–56. Univerzita Karlova, Praha.

HRONOVÁ, K. (1993). *Čeština jako cizí jazyk*. Karolinum, Praha.

HUBÁČKOVÁ, M. (1994). *Communiquer en tchèque*. (vlastním nákladem), Praha.

IZUMI, E., UCHIMOTO, K. et ISAHARA, H. (2005). Error annotation for corpus of japanese learner english. *In Proceedings of the 6th International Workshop on Linguistically Annotated Corpora 2005 (LINC 2005), Korea*, pages 71–80.

JELÍNEK, J., BEČKA, J. V. et TĚŠITELOVÁ, M. (1961). *Frekvence slov, slovních druhů a tvarů*. Státní pedagogické nakladatelství, Praha.

JOHANNESSEN, J. B., HAGEN, K. et LANE, P. (2002). The performance of a grammar checker with a deviant language input. *In Proceedings of the 19th international conference on Computational linguistics*, volume 2, pages 1–8.

KARLÍK, P., NEKULA, M. et PLESKALOVÁ, J., éditeurs (2002). *Encyklopedický slovník češtiny*. Nakladatelství Lidové Noviny, Praha.

KARLÍK, P., NEKULA, M. et RUSÍNOVÁ, Z., éditeurs (1995). *Příruční mluvnice češtiny*. Nakladatelství Lidové Noviny, Praha.

KARTTUNEN, F. (1986). A linguist looks at computer-assisted instruction. *In* FREUDENSTEIN, R. et VAUGHAN, J. C., éditeurs : *Confidence Through Competence in Modern Language Learning. CILT Reports & Papers 25*.

KASTLER, C. (1995). *La langue tchèque : nouvelle grammaire tchèque pratique et raisonnée*. Ophrys, Paris.

KILANI-SCHOCH, M. (1988). *Introduction à la morphologie naturelle*. Peter Lang, Berne.

KIRSCHNER, Z. (1983). A method of automatic extraction of significant terms from texts. *Explizite Beschreibung der Sprache und automatische Textbearbeitung*, X.

KLEIN, W. (1989). *L'acquisition de langue étrangère*. Armand Colin, Paris.

KOMÁREK, M. (1956). K otázce slovních druhů v češtině. *Slovo a slovesnost*, 17(3):160–168.

KOMÁREK, M. (1978). *Příspěvky k české morfologii*. Státní pedagogické nakladatelství, Praha.

KRAIF, O., ANTONIADIS, G., ECHINARD, S., LOISEAU, M., LEBARBÉ, T. et PONTON, C. (2004). NLP tools for CALL : the simpler, the better. *In Proceedings of InSTIL / ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems*, Venice.

KRAIF, O. et PONTON, C. (2007). Du bruit, du silence et des ambiguïtés : que faire du TAL pour l'apprentissage des langues ? *In Actes de TALN 2007*, Toulouse (France).

KUČERA, O. (2006). Pražský závislostní korpus jako elektronická cvičebnice češtiny. *In Proceedings of the 4th Student Research Competition in Informatics and Information Technologies (finalists papers)*, pages 41–47, Praha.

LADO, R. (1957). *Linguistics across culture*. University of Michigan Press, Ann Arbor.

LAST, R. W. (1989). *Artificial Intelligence Techniques in Langauge Learning*. Horwood, Chichester.

LEVENSHTEIN, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710.

LEVY, M. (1997). *Computer-Assisted Language Learning : Context and Conceptualization*. Clarendon Press, Oxford.

L'HAIRE, S. et VANDEVENTER-FALTIN, A. (2003). Diagnostic d'erreurs dans le projet FreeText. *ALSIC : Apprentissage des Langues et Systèmes d'Information et de Communication*, 6(2):21–37.

LIPKOVÁ, J. et KISSLING, P. (1991). *Malý most, une initiation au tchèque*. Lausanne.

LUKÁŠOVÁ, J. (2007). Obecná čeština z pohledu studentů – českých i nerodilých mluvčí. *In* ČEMUSOVÁ, J. et ŠTINDLOVÁ, B., éditeurs : *Sborník Asociace učitelů češtiny jako cizího jazyka 2006-2007*, pages 181–187. Akropolis, Praha.

LYER, S. (1963). *Manuel de tchèque*. Státní pedagogické nakladatelství, Praha.

MALINOVSKÝ, M. (1995). Redukce gramatiky. *Slovo a slovesnost*, 56:218–222.

MAN, O. (1986). *Initiation à la langue tchèque*. Státní pedagogické nakladatelství, Praha.

MARTINCOVÁ, O. et KOL. (1993). *Pravidla českého pravopisu*. Nakladatelství Fortuna, Praha.

MARTINEZ, P. (2004). *La didactique des langues étrangères*. Que sais-je ? Presse Universitaire de France, Paris.

MEL'ČUK, I. (1993). *Cours de morphologie générale*, volume I - Introduction et Première partie : Mot. Les Presses de l'Université de Montréal, Montréal.

MEY, J. (1966). Word-classes in automatic syntactic analysis. *Slovo a slovesnost*, 27:220–235.

MILDE, J. et KOL. (1948). *Le tchèque sans savoir*. Praha.

MILLET, Y. (1968). Difficulté pour un français entre le verbe déterminé et verbe indéterminé en tchèque. *Revue des études slaves*, 47:109–121.

MITKOV, R., éditeur (2003). *The Oxford Handbook of Computational Linguistics*. Oxford University Press, New York.

NATION, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

NEKULA, M. (2007). Systém a úzus. K výuce české deklinace se zřetelem k substantivům. *In* ČEMUSOVÁ, J. et HOLÁ, L., éditeurs : *Sborník Asociace učitelů češtiny jako cizího jazyka 2006-2007*, pages 23–47. Akropolis, Praha.

NERBONNE, J. (2003). Natural language processing in computer-assisted language learning. *In* MITKOV, R., éditeur : *The Oxford Handbook of Computational Linguistics*, pages 670–698. Oxford University Press.

NERBONNE, J. et DOKTER, D. (1999). An intelligent word-based language learning assistant. *Traitement automatique des langues*, 40(1):125–142.

NERBONNE, J., JAGER, S. et van ESSEN, A., éditeurs (1998). *Language Teaching and Language Technology*. Swets & Zeitlinger, Lisse.

NERBONNE, J. et SMIT, P. (1996). Glosser-rug : In support of reading. *In COLING '96*, pages 830–835.

OSOLSOBĚ, K. (1996). *Algoritmický popis formální morfologie a strojový slovník češtiny*. Thèse de doctorat, Filozofická fakulta Masarykovy univerzity, Brno.

OSOLSOBĚ, K., PALA, K. et RYCHLÝ, P. (1998). Frekvence vzorů českých substantiv. *Sborník prací Filozofické fakulty brněnské univerzity*, Řada jazykovědná(A 46):77–94.

PALÍKOVÁ, E. (2001). *Tchèque Express*. Éditions du Dauphin, Paris.

PEIRCE, C. S. (1987). *Textes fondamentaux de sémiotique*. Méridiens Klincksieck, Paris.

PETR, J., DOKULIL, M., HORÁLEK, K., HŮRKOVÁ, J. et KNAPOVÁ, M., éditeurs (1986a). *Mluvnice češtiny (1) Fonetika, Fonologie, Morfonologie a morfemika, Tvoření slov*. Academia, Praha.

PETR, J., KOMÁREK, M., KOŘENSKÝ, J. et VESELKOVÁ, J., éditeurs (1986b). *Mluvnice češtiny (2) Tvarosloví*. Academia, Praha.

POGNAN, P. (1972). *Analyse morphosyntaxique de textes médicaux tchèques. Extraction automatique du verbe*. Thèse de doctorat, Université de Paris 3.

POGNAN, P. (1988). Analyse automatique du tchèque : Termes de qualité en -ost : détermination de leur origine adjectivale. *Revue des études slaves*, LX(Fascicule 3):673–680.

POGNAN, P. (1992). Automatické zpracování češtiny pro vědecko-technické informace. *In 16th World Congress of SVU (Společnost pro Vědu a Umění)*, Prague.

POLDAUF, I. et ŠPRUŇK, K. (1968). *Čeština jazyk cizí*. Státní pedagogické nakladatelství, Praha.

PORQUIER, R. (1977). L'analyse des erreurs. Problèmes et perspectives. *Étude de linguistique appliquée*, 25:23–43.

PRAVEC, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26:81–114.

PROŠEK, M. (2007). Konstanty a proměnné morfologických dotazů v jazykové poradně. *Naše řeč*, 90(4):174–194.

RAGUSICH, N.-C. (1977). *Contribution à l'étude de problème de la difficulté en langue étrangère*. Québec.

RYNDOVÁ, J., ČEMUSOVÁ, J. et HOLÁ, L., éditeurs (2005). *Sborník Asociace učitelů češtiny jako cizího jazyka 2003-2005*. Akropolis, Praha.

RYNDOVÁ, J., ČEMUSOVÁ, J. et HOLÁ, L., éditeurs (2006). *Sborník Asociace učitelů češtiny jako cizího jazyka 2005-2006*. Akropolis, Praha.

SCHNEIDER, D. et MCCOY, K. F. (1998). Recognizing syntactic errors in the writing of second language learners. *In COLING-ACL*, pages 1198–1204.

SCHULZE, M. (2008). Modeling SLA processes using NLP. *In* CHAPELLE, C. A., CHUNG, Y. et XU, J., éditeurs : *Towards adaptive CALL : Natural language processing for diagnostic language assessment*, pages 149–166, Iowa State University.

SEDLÁČEK, R. et SMRŽ, P. (2001). A new czech morphological analyser ajka. *In Proceedings of the 4th International Conference on Text, Speech and Dialogue*, pages 100–107, London UK. Springer-Verlag.

SELINKER, L. (1972). Interlanguage. *IRAL*, 10(3):209–231.

SERRJ (2001). *Společný evropský referenční rámec pro jazyky. Jak se učíme jazykům, jak je vyučujeme a jak v jazycích hodnotíme*. Council of Europe, Univerzita Karlova, Ústav bohemistických studií, Praha.

SGALL, P. (1960). Soustava pádových koncovek v češtině. *Acta Universitatis Carolinae - Philologica 2, Slavica Pragensia II*, pages 65–84.

SGALL, P. (1967). *Generativní popis jazyka a česká deklinace*. Academia, Praha.

SGALL, P. (2006). *Language in its multifarious aspects*. Karolinum, Praha.

SGALL, P., HAJIČOVÁ, E. et PANEVOVÁ, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. D. Reidel Publishing Company, Praha – Amsterdam.

SGALL, P., HAJIČOVÁ, E. et PANEVOVÁ, J. (1988). Thème et rhème dans les langues slaves. *Revue des études slaves*, LX(Fascicule 3):657–663.

SGALL, P., HAJIČOVÁ, E. et PANEVOVÁ, J. (2003). *Úvod do teoretické a počítačové lingvistiky. I.svazek - Teoretická lingvistika.* Karolinum, Praha.

SGALL, P. et HRONEK, J. (1992). *Čeština bez příkras*. H & H, Praha.

SGALL, P. et PANEVOVÁ, J. (2004). *Jak psát a jak nepsat česky*. Karolinum, Praha.

ŠIMANDL, J. (2000). Morfologická problematika v jazykové poradně. *Naše řeč*, 83:57–76, 113–131, 169–192, 225–242.

SKALIČKA, V. (1935). *Zur ungarischen Grammatik*. Praha.

SKALIČKA, V. (1951). *Typ češtiny*. Slovanské nakladatelství, Praha.

SKALIČKA, V. (2004). *Souborné dílo*, volume 1. Karolinum, Praha.

SKINNER, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts, New York.

SLAMA-CAZACU, T. (1981). *Psycholinguistique appliquée : problèmes de l'enseignement des langues*. Nathan, Paris.

SLAVÍČKOVÁ, E. (1975). *Retrográdní morfematický slovník češtiny s připojenými inventárními slovníky českých morfémů kořenových, prefixálních a sufixálních*. Academia, Praha.

ŠMILAUER, V. (1972). *Nauka o českém jazyku*. Státní pedagogické nakladatelství, Praha.

SPILAR, O. (1994). *Le tchèque sans peine*. Assimil, Paris.

STUMP, G. T. (2001). Inflection. *In* SPENCER, A. et M.ZWICKY, A., éditeurs : *The Handbook of Morphology*, pages 13–43. Blackwell publishers, Oxford.

TAX, J., éditeur (1985). *Čeština jako cizí jazyk. Materiály z první metodologické konference USS FF UK*. Univerzita Karlova, Praha.

TĚŠITELOVÁ, M., PETR, J. et KRÁLÍK, J. (1986). *Retrográdní slovník současné češtiny*. Academia, Praha.

TONO, Y. (2003). Learner corpora : design, development and applications. *In* ARCHER, D., RAYSON, P., WILSON, A. et MCENERY, T., éditeurs : *Proceedings of the Corpus Linguistics 2003 conference*, pages 800–809.

TSCHICHOLD, C. (2006). Intelligent CALL : The magnitude of the task. *In* MERTENS, P., FAIRON, C., DISTER, A. et WATRIN, P., éditeurs : *Verbum ex machina. Actes de la 13e conférence sur le Traitement automatique des langues naturelles*, pages 806–814, Louvain-la-Neuve. Presses universitaires de Louvain.

VLASÁK, V., ŠÁRA, M. et ŠÁROVÁ, J. (1967). *Le tchèque sur la base du français*. Státní pedagogické nakladatelství, Praha.

WHITE, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge University Press, Cambridge.

ZOCK, M. (1994). Language in action, or learning a language by watching it work. *In 7th Twente Workshop on Language Technology : Computer-Assisted Language Learning*, Twente.

ZOCK, M. (1996). Computational linguistics and its use in real world : The case of computer assisted-language learning. *In Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, pages 1002–1004.