Plate-forme CETLEF-slovaque : application Web dynamique proposant des exercices de grammaire slovaque sur la valence verbale

I. Šmilauer

1 Présentation de la plate-forme et sa motivation

La volonté de diversifier les outils pédagogiques proposés aux utilisateurs d'une future grammaire du slovaque pour les francophones nous a amené à envisager une composante complémentaire utilisant un médium électronique. En effet, l'importance et la popularité croissantes des TICE (technologies de l'information et de la communication pour l'enseignement) sont aujourd'hui un facteur crucial dans l'articulation des besoins au cours de la conception des outils pédagogiques. Dans cette perspective, la plate-forme CETLEF-Slovaque – une application Web dynamique – est proposée ici en tant que complément de la grammaire diffusée sur support papier. Par un ensemble de références qui seraient insérées dans le texte de l'ouvrage imprimé, cette plate-forme permettra aux apprenants d'acquérir les notions théoriques moyennant des exercices grammaticaux et offrira également un accès alternatif à certaines informations grammaticales exposées dans la grammaire.

La plate-forme présentée dans ce chapitre est une modification de la plate-forme CETLEF [1] qui est développée à l'INALCO depuis 2008 et qui propose des exercices de déclinaison tchèque accompagnés d'un module de visualisation de paradigmes de déclinaison et du lexique pédagogique de base. La particularité de cet outil, le distinguant de la majorité des outils TICE présents sur le marché, est l'intégration d'un diagnostic automatique des erreurs permettant une rétroaction d'ordre didactique destiné à l'apprenant. Bien que ce diagnostic ne se limite pour le moment qu'aux formes nominales isolées issues des exercices à trous, son utilité et son efficacité se confirment par la présence de centaines d'utilisateurs qui se servent de la plate-forme pour accompagner leur apprentissage du tchèque.

1. Voir Šmilauer (2008) et Šmilauer (2011). CETLEF est un acronyme pour *Comprendre et corriger les erreurs en tchèque langue étrangère*. La plate-forme est accessible à l'adresse http ://www.cetlef.fr.

Du point de vue de l'ingénierie linguistique, le diagnostic des erreurs, possible grâce à l'intégration des techniques du traitement automatique des langues, est basé sur un modèle formel de la flexion nominale tchèque composé d'une partie déclarative contenant les listes de désinences casuelles et d'une composante procédurale contenant la définition des règles d'alternances morpho-phonologiques. Pour la mutation slovaque, la question de la conception d'un modèle similaire s'est posée et nous discuterons plus bas les solutions déjà adoptées ou envisageables dans les travaux futurs.

Le diagnostic est effectué à partir de la comparaison des formes attendues avec les formes saisies par l'apprenant au sein d'une tâche de déclinaison et il permet de déterminer, dans plus de 80 % de cas, une interprétation linguistique de l'erreur. Cette interprétation est basé exclusivement sur les informations d'ordre morphologique, voire orthographique, et permet de produire par la suite un retour didactique formulé en langage naturel. Du point de vue pédagogique, l'intégration de ce procédé automatique dans les activités complémentaires à l'enseignement en présentiel permet de déléguer les tâches répétitives de correction des exercices à l'ordinateur et de se concentrer sur les activités nécessitant une interaction entre l'enseignant et l'apprenant.

Le tchèque et le slovaque étant deux langues dont les systèmes de flexion nominale sont très semblables, le fait de travailler sur une mutation slovaque de la plate-forme CETLEF a été une excellente occasion de vérifier les solutions adoptées préalablement et d'envisager des améliorations, notamment en vue du développement d'une version générique, indépendante d'une langue spécifique, dont le fonctionnement serait fondé sur un paramétrage du diagnostic avec des données externes indépendantes.

Dans ce qui suit, nous allons d'abord découvrir la plate-forme telle qu'elle se présente aux utilisateurs – à l'apprenant et à la personne qui est chargée de la création des exercices – puis nous allons exposer les solutions sous-jacentes à sa réalisation de point de vue linguistique et technologique.

2 Structure et contenu de la plate-forme

La valence verbale étant choisie comme pivot de la conception de notre grammaire, l'objectif principal de la plate-forme dans son état actuel est de proposer des exercices orientés vers la forme casuelle des compléments verbaux exprimés par des substantifs ou des adjectifs dans un contexte syntaxique déterminé par les éléments d'une « phrase à trous » – le trou représente la position du complément et il est à remplir par l'apprenant à partir du lemme de la forme requise.

L'objectif secondaire de la plate-forme est de fournir à l'apprenant une présentation des données métalinguistiques (annotation morphologique, paradigmes de déclinaison, dictionnaire des valences verbales) afin de favoriser son acquisition de la langue étrangère par une compréhension explicite des structures linguistiques afférentes.

La plate-forme peut être utilisée par toute personne connectée sur Internet et dont le terminal (ordinateur, Smartphone, tablette) est doté d'un navigateur Web (MS Explorer, Firefox, Safari, etc.). Deux environnements différents ont été mis en place : la plate-forme apprenant, divisée en section publique et section privée, accessible

uniquement après l'inscription, et contenant des exercices et différentes présentations d'informations grammaticales ; la plate-forme auteur destinée à l'auteur des exercices qui peut accéder à l'ensemble des données nécessaires pour le fonctionnement de la plate-forme ainsi qu'aux données fournies par les apprenants.

Au sein de la plate-forme apprenant, la section publique donne une idée générale des fonctionnalités et propose un formulaire permettant d'obtenir un accès à la zone privée avec les exercices. Les informations saisies par l'apprenant lors de l'inscription permettent d'avoir une image approximative de ses caractéristiques sociolinguistiques (l'âge, la catégorie socio-professionnelle, l'autoévaluation de son niveau de slovaque, les autres langues maîtrisées, la motivation pour apprendre le slovaque, etc.). Ces informations peuvent être exploitées lors de l'analyse des données langagières recueillies dans les exercices. La section privée de la plateforme, accessible avec un nom d'utilisateur et un mot de passe attribués à chaque apprenant pendant la création de son compte individuel, est composée de trois parties (exercices, dictionnaire et Vallex) que nous allons détailler ci-dessous.

2.1 Exercices

Dans cette section, l'utilisateur peut d'abord choisir quel est son niveau de maîtrise de la langue slovaque d'après l'échelle établie par le Cadre commun européen de référence pour les langues (A1 – A2 pour débutants, B1 – B2 pour intermédiaires ou C1 – C2 pour avancés). Les exercices sont rassemblés en batteries en fonction de ces niveaux qui restreignent les compétences indispensables pour la bonne exécution des tâches de déclinaison demandées. La grammaire imprimée devra inclure des références vers les exercices en spécifiant le niveau ciblé de l'apprenant. Le prototype contient quatre exercices pour le niveau A1 et A2 (compléments des verbes *byť, ďakovať, chodiť* et *robiť*), le niveau B1 – B2 en contient six (compléments des verbes *byť, ďakovať, doraziť, chodiť, obdivovať, povedať*). La différenciation des exercices d'après les niveaux est faite notamment par le choix des unités lexicales utilisées et la complexité des cadres verbaux dont la maîtrise est testée.

Pour chaque exercice, un ensemble d'informations comme le titre, l'énoncé, les exemples et la durée disponible sont présentés. Chaque exercice est constitué d'une dizaine de tâches qui sont des phrases à trous avec les lemmes des formes requises. Afin de se concentrer principalement sur la grammaire, l'apprenant a à sa disposition un lexique contenant les mots utilisés dans les phrases. L'ensemble de ces mots est défini par l'auteur des exercices en fonction de leur difficulté présumée. Pour éliminer l'influence du dispositif technique qui pourrait affecter les productions des apprenants, un clavier virtuel permettant de saisir les caractères slovaques diacrités est intégré à la page affichée dans le navigateur Web.



Figure 1 – Vue d'une batterie d'exercices



Figure 2 – Un exercice en cours

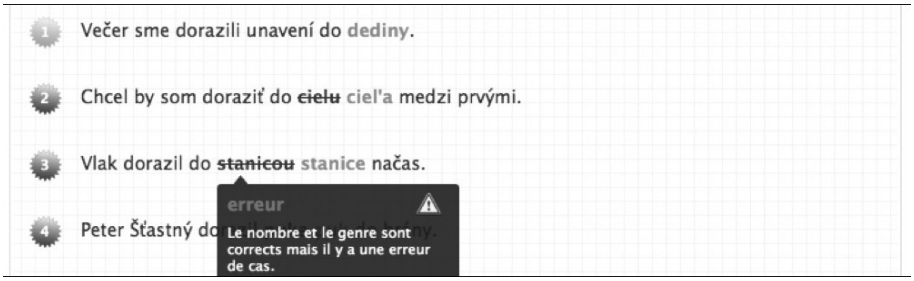


Figure 3 – Exemple d'un exercice corrigé

La correction de l'exercice se fait à la demande de l'apprenant ou après que le délai fixé par l'auteur des exercices soit écoulé. Un ensemble des informations incluant le pourcentage des réponses correctes et la comparaison de la performance de l'apprenant avec celles de ses collègues déjà enregistrées dans la base est affichée dans l'entête de la section contenant la correction. Pour la correction des différentes tâches, les formes requises sont accompagnées par leur annotation linguistique contenant la catégorie lexicale, les catégories morphologiques, le type de déclinaison et la traduction française. Si l'apprenant a commis une erreur, la forme erronée est accompagnée de son diagnostic.

La correction est enregistrée dans la base de données et elle peut être à tout moment affichée, pour chaque exercice terminé, sur la page d'accueil de la section avec les exercices. Pour le moment, un exercice déjà fait ne peut pas être refait. Cette mesure permet à l'utilisateur de garder conscience des problèmes rencontrés et cibler leur élimination dans les exercices à venir.

2.2 Dictionnaire

La section Dictionnaire pourrait être caractérisée comme une sorte de « mémoire » grammaticale de chaque apprenant car elle se remplit au fur et à mesure que l'apprenant avance dans les exercices. Les phénomènes rencontrés (le lexique, les paradigmes de déclinaison, les alternances et les cas du singulier et du pluriel) sont ajoutés progressivement dans les sections correspondantes. Ce conditionnement est censé avoir un effet motivant pour l'apprenant, l'incitant à faire le maximum d'exercices.

La section lexique contient les différentes unités lexicales accompagnées de leur traduction française, correspondante à la signification de l'unité dans la tâche donnée. Dans le futur, il serait utile d'envisager l'intégration d'un dictionnaire mieux fourni, qui présenterait toutes les significations possibles comme dans un dictionnaire bilingue standard. Chaque mot renvoie à la phrase dans laquelle il a été utilisé. Un moteur de recherche basique est proposé pour accélérer le travail de l'apprenant. Les paradigmes, caractérisés par un mot modèle issu de la tradition grammaticale slovaque, sont présentés sous la forme classique des tables de désinences. Chaque paradigme est accompagné de quelques exemples et des substantifs représentant des sous-types. Ainsi, par exemple pour le type *dub*, les sous-types mentionnés sont *med, roh, tunel,*

strom[2]. La section alternances permet de revoir les alternances morpho-phonologiques rencontrées dans les exercices. Finalement, la section cas permet d'accéder aux différentes formes demandées selon la valeur de la catégorie grammaticale de cas et de nombre. Chaque forme listée dans ces présentations renvoie à la phrase dans laquelle elle a été utilisée.

La présentation explicite de ces données grammaticales est censée renforcer la conscience de l'apprenant du système linguistique slovaque dont la compréhension, comme pour toutes langues à flexion complexe, devrait être jugée comme une étape nécessaire au cours de son apprentissage.

2.3 Vallex

La section Vallex propose un accès au prototype du dictionnaire des valences verbales du slovaque élaboré d'après le modèle du Vallex tchèque[3]. Ce dictionnaire, dont la structure a été exposée dans les chapitres précédentes, contient les cadres pour 19 verbes choisis (*chodiť, chutiť, cítiť, ďakovať, doraziť, existovať, hovoriť, myslieť, obdivovať, ochutnať, povedať, povedať si, robiť, tvoriť, tvoriť sa, vedieť, volať, voňať, žiť*).

En utilisant les éléments de la mise en page adoptée par le Vallex pragois, cette visualisation permet d'avoir une idée claire des valences des différents verbes, qui sont numérotés et accompagnés par les informations similaires concernant leurs équivalents sémantiques tchèques et français. En ce qui concerne l'exploitation pédagogique de ces éléments, la présentation des cadres verbaux nécessite une bonne maîtrise du métalangage utilisé, notamment pour les noms raccourcis des différents foncteurs (les noms utilisés pour désigner les classes sémantiques des différents types de compléments verbaux). Néanmoins, les exemples permettent à l'apprenant d'avoir une bonne vision du cadre en question et de mieux comprendre la notation (qui garde son aspect technique quelque peu opaque pour un non-spécialiste). L'adaptation pédagogique des données du Vallex à un public d'apprenants reste un défi à relever car il s'agît d'informations très utiles, voire indispensables pour l'apprentissage d'une langue étrangère.

2.4 Plate-forme auteur

Cette plate-forme permet de créer de nouveaux exercices et de modifier les exercices existants. Son accès est limité à l'équipe pédagogique et actuellement, ne se fait qu'en local sur une machine privée. Toutes les informations (titres, énoncés, exemples) et les contenus (phrases, formes requises, lexique) doivent être saisis à l'aide d'une interface de formulaires qui permettent de remplir la base de données sous-jacente à la plate-forme.

Afin de rendre possible le diagnostic des erreurs et l'affichage des informations métalinguistiques sur la plate-forme apprenant, les formes requises dans les exercices doivent être annotées morphologiquement. Les catégories renseignées sont la catégorie lexicale, le type morphologique (basé uniquement sur les critères formels, contrairement à la catégorie lexicale qui est traditionnellement conçue comme un amalgame de critères

2. Cette section est fournie par le modèle de la déclinaison slovaque dont on parlera plus loin.
3. Voir Lopatková *et al.* (2008).

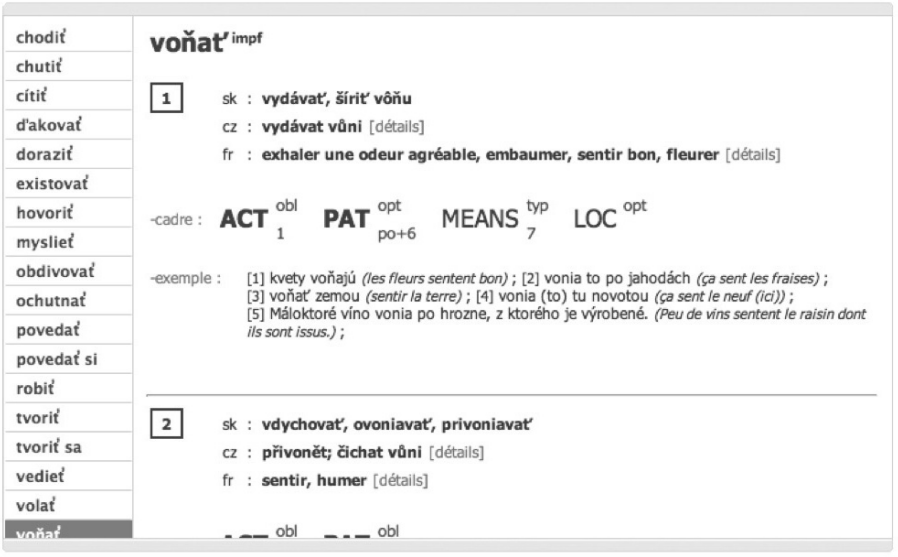


Figure 4 – L'affichage du Vallex slovaque sur la plate-forme apprenant

formels, syntaxiques et sémantiques), le paradigme de déclinaison spécifiant le sous-type éventuel, le cas, le nombre, le genre et éventuellement la présence d'une alternance morpho-phonologique. Afin de raccourcir le temps nécessaire pour la conception d'un exercice et pour éviter des erreurs d'inattention (fautes de frappes, oublis, etc.), l'annotation morphologique de la forme requise est faite semi-automatiquement à l'aide d'un analyseur morphologique rudimentaire qui travaille avec le modèle intégré de la déclinaison slovaque et qui fourni un ensemble restreint d'annotations possibles d'une forme requise. La désambiguïsation (le choix de l'annotation correspondante) doit être faite manuellement par l'auteur de l'exercice.

3 Modèle de la déclinaison slovaque

Dans le système que nous proposons, le diagnostic des erreurs au sein des exercices est effectué par la génération des formes hypothétiques à partir du radical de la forme requise et l'ensemble des désinences existantes dans le système de déclinaison donné. Si une correspondance est trouvée entre une des formes générées et la forme produite par l'apprenant, l'écart entre les valeurs des catégories grammaticales de cette forme et de la forme requise est utilisé comme l'interprétation de l'erreur.

Afin de pouvoir tester le diagnostic automatique des erreurs sur le slovaque et pour l'enrichissement de la plate-forme par une présentation de différents paradigmes de déclinaison, il a été nécessaire de mettre en place un modèle de la déclinaison slovaque. Ce modèle doit être composé des ensembles de désinences casuels définissant les paradigmes de déclinaison et les règles pour les alternances morpho-phonologiques, se produisant dans des configurations particulières de certains radicaux avec certaines désinences casuelles.

En ce qui concerne les paradigmes de déclinaison, nous avons adopté le format utilisé pour la version tchèque et nous avons structuré la présentation des paradigmes de déclinaison établie par Diana Jamborová-Lemay dans Boisserie, Jamborová et Křečková (2009). Pour un système nécessitant une couverture plus exhaustive, il aurait été plus judicieux de se servir de présentations plus complètes, comme celle de Sokolová (2007). Néanmoins, la visée pédagogique de la plate-forme, conjuguée avec le caractère restreint du lexique utilisé, ne nécessite pas, pour l'instant, de tels efforts. Il est souhaitable que le modèle utilisé sur la plate-forme corresponde exactement à celui exposé dans la grammaire.

Chacun des types et des sous-types de déclinaison slovaque est modélisé par un ensemble de désinences accompagnées des valeurs de leurs catégories grammaticales (cas, nombre, genre) et représenté par un mot modèle. Nous reprenons les types et les sous-types présentés par D. Jamborová Lemay :

a. substantifs
   - o masculins : chlap + brat, syn, agronóm, genius, Aristoteles ; hrdina + socialista, poeta ; dub + med, roh, tunel, strom ; stroj + čas, násobiteľ, delenec, kôň ;
   - o féminins : žena + idea, láska ; ulica + studnia, revolúcia ; dlaň + koľaj ; kosť ;
   - o neutres : mesto + mlieko, krídlo, rádio, centrum ; srdce + more, pľúca ; umenie ; dievča ;

b. adjectifs : pekný, cudzí, otcov.

La déclinaison des pronoms n'est pas modélisée dans le prototype élaboré dans le cadre de notre projet.

Les règles contextuelles pour les alternances morpho-phonologiques (par exemple *žena* > *žien*) n'ont pas été élaborées dans le cadre de ce projet mais le système proposé permet leur intégration relativement rapide à l'aide des fonctions prenant comme argument la configuration morpho-phonologique requise pour une alternance et la liste des exceptions (positives ou négatives). Il est cependant possible d'annoter les formes requises avec une étiquette correspondante à l'alternance éventuelle ce qui permet l'affichage des formes alternées, classées d'après la nature de l'alternance, dans le dictionnaire personnel de l'apprenant. Les alternances listées dans notre modèle sont les suivantes :

a. alternances vocaliques : á > a (ex. mráz > mrazu), a > á (ex. hlava > hláv), u > ú (ex. ruka > rúk), i > í (ex. ulica > ulíc), a > ia (ex. žaba > žiab), ia > a (ex. peniaze > peňazí), ä > ia (ex. mäso > mias), e > ie (ex. žena > žien), ie > e (ex. chlieb > chlebom), o > ô (ex. voda > vôd), ô > o (ex. kôň > koňa) ;

b. e-mobile : e > # (ex. majster > majstra), # > e (ex. sídlo > sídel), ie > # (ex. veniec > venca), # > ie (ex. mydlo > mydiel), o > # (ex. svedok > svedka), # > o (ex. otázka > otázok).

Dans notre représentation, le sens de l'alternance est établi toujours en partant du lemme vers la forme déclinée.

Les données caractérisant les paradigmes sont structurées à l'aide du langage XML. Pour donner un exemple de cette structuration, voici les informations concernant le type *chlap*[4] :

```
<soustype mnemo="chlap" pdgm="ch">
<term cas="nom" num="sg" gen="m">#</term>
<term cas="gen" num="sg" gen="m">a</term>
<term cas="dat" num="sg" gen="m">ovi</term>
<term cas="acc" num="sg" gen="m">a</term>
<term cas="loc" num="sg" gen="m">ovi</term>
<term cas="inst" num="sg" gen="m">om</term>
<term cas="nom" num="pl" gen="m">i</term>
<term cas="gen" num="pl" gen="m">ov</term>
<term cas="dat" num="pl" gen="m">om</term>
<term cas="acc" num="pl" gen="m">ov</term>
<term cas="loc" num="pl" gen="m">och</term>
<term cas="inst" num="pl" gen="m">mi</term>
<ex ouvert="1" A1A2="1">
<u>chlap</u><u>majster</u>
</ex>
</soustype>
```

En fonction de la différence des valeurs des catégories grammaticales de la forme requise et la forme produite par l'apprenant, ce modèle de déclinaison slovaque permet d'identifier les erreurs de cas (confusion de désinence au nombre identique, par exemple *chlapa* au lieu de *chlapovi*), de nombre (désinence correspondante à la forme au nombre opposé, par exemple *chlapom* au lieu de *chlapovi*), leur combinaison, puis les erreurs relevant de la confusion des types de déclinaison (par exemple*, brati* au lieu de *bratia* par analogie avec le type *chlap*). Un message caractérisant ces erreurs est produit au cours de la correction et il est affiché pendant la correction. Étant donnée l'absence des règles pour les alternances dans le système proposé dans le cadre de ce projet, les erreurs relevant de ce domaine ne sont pas identifiées. Une évaluation du diagnostic basé sur ce modèle sera possible une fois que le volume de données obtenu de la part des apprenants sera statistiquement significatif.

En ce qui concerne la présentation du système de déclinaison sur la plate-forme, des listes de désinences pour chaque type de déclinaison rencontré dans les exercices sont affichées dans le dictionnaire personnel de l'apprenant.

4 Quelques aspects techniques et perspectives

La plate-forme CETLEF-Slovaque est une application Web dynamique hébergée à l'adresse http ://cetlef-sk.fr. Elle a été programmée en langage PHP qui permet de générer le contenu dynamique des pages HTML à partir de deux types de sources de données : (1) une base de données relationnelles utilisant le moteur MySQL contenant les informations relatives aux exercices (énoncés, tâches, annotation des formes

4. La définition détaillée de la structure du fichier XML contenant les paradigmes peut être retrouvée dans Šmilauer (2008).



Figure 5 – Affichage des paradigmes sur la plate-forme apprenant

requises) et aux apprenants (leur profil et production) ; (2) des fichiers XML contenant le modèle de la déclinaison et le contenu du dictionnaire de valences verbales Vallex (qui a été importé et formaté en HTML à partir d'un fichier élaboré à l'aide du logiciel MS Access). Le choix d'un format statique pour les fichiers contenant les modèles est justifié par leur caractère pérenne, tandis que le contenu de la base de données peut changer au cours du temps.

Le diagnostic des erreurs est effectué à l'aide des fonctions, développées de manière *ad hoc*, qui permettent de produire les alternances nécessaires lors de la génération des formes pour le diagnostic. L'ensemble des opérations mises en place est décrit dans Smilauer (2008) et Šmilauer (2011).

Comme nous l'avons dit auparavant, pour que le diagnostic employé sur la plate-forme actuelle soit complet, il serait nécessaire de formuler des règles d'alternance morphophonologiques. Or faire ceci à l'aide des fonctions *ad hoc*, comme c'est le cas pour la plate-forme élaborée pour le tchèque, est réalisable mais nécessite une intervention directement au niveau du code source de l'application.

Dans cette perspective, afin de rendre le diagnostic des erreurs mieux adaptable aux autres langues, à commencer par le slovaque, nous avons étudié les possibilités de baser le diagnostic sur l'implémentation de la morphologie à l'aide d'un transducteur à états finis, composé d'une cascade de transducteurs élémentaires – filtre morphotaxique, dictionnaire des morphèmes et règles contextuelles pour la réalisation des alternances – combinés à l'aide de l'opération régulière de composition. Pour obtenir une analyse des formes erronées, il serait possible d'utiliser une série de transducteurs « intermédiaires » avec relâchement des contraintes dans le filtre morphématique et dans les règles d'alternance. L'emploi de cette méthode impliquerait donc une refonte du diagnostic des erreurs morphologiques afin de le rendre plus général, applicable aux autres langues flexionnelles, et notamment aux autres langues slaves proches du tchèque comme le slovaque ou le polonais.