

Mémoire

Générateur automatique des exercices à partir d'un corpus étiqueté. Applications et didactique pour le Géorgien

Natia DAVITASHVILI

Sous la direction de **Monsieur Damien Nouvel**

oo

oooooooooooooooooooooooooooooooooooo

Co-directrice : **Madame Hélène Gérardin**

Institut National des Langues et Civilisations Orientales

Master 2 Traitement Automatique des Langues

Spécialité : Technologie de la Traduction et Traitement des Données Multilingues

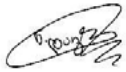
2019-2020



Déclaration sur l'honneur

Je, soussigné, **Natia DAVITASHVILI**, déclare avoir rédigé ce travail sans aide extérieure ni sources autres que celles qui sont citées. Toutes les utilisations de textes préexistants, publiés ou non, y compris en version électronique, sont signalées comme telles. Ce travail n'a été soumis à aucun autre jury d'examen sous une forme identique ou similaire, que ce soit en France ou à l'étranger, à l'université ou dans une autre institution, par moi-même ou par autrui.

Paris, le 22 novembre 2020



Remerciements

La réalisation de ce mémoire, dans le cadre du Master 2 Traitement Automatique des Langues spécialité Technologie de la Traduction et Traitement des Données Multilingues, a été rendue possible grâce à la collaboration et au soutien de certaines personnes auxquelles je souhaite faire part de mes sincères remerciements.

En premier lieu, mes remerciements s'adressent à mon directeur du mémoire, Monsieur Damien Nouvel, pour son soutien perpétuel, ses précieuses directives et ses idées scientifiques. Qu'il soit assuré de mon très grand respect et du plaisir que j'ai eu à travailler avec lui.

Je souhaite également exprimer toute ma reconnaissance à Helene Gérardin qui m'a inspirée dans une grande partie de ce travail, par ses encouragements et ses contributions pertinentes.

Je tiens aussi à exprimer ma profonde gratitude à Monsieur François Stuck pour le partager de son expérience et pour ne pas avoir hésité à me soutenir grâce à ses conseils importants dès le début de mon travail.

Je voulais surtout remercier Monsieur Florent Mell, Madame Valérie Jardin et à Monsieur Geoffrey Royer, pour leur relecture, ainsi que pour m'avoir motivée tout au long de cette année et pour leur soutien en ce moment si particulier de ma vie.

Enfin, j'adresse mes plus sincères remerciements à toute équipe l'ERTIM qui m'a beaucoup inspirée par son travail sérieux, accompagnée et encouragée pendant ces dernières mois.

SOMMAIRE

DECLARATION SUR L'HONNEUR - p. 2

REMERCIEMENTS - p. 3

INTRODUCTION - p. 8

PREMIERE PARTIE: L'état de l'art – p.10

DEUXIEME PARTIE : Problématique et architecture -p.18

TROISIEME PARTIE : Résultats et mise en pratique – p.27

CONCLUSION - p. 37

ANNEXE 1 . - p. 39

ANNEXE 2 – p. 40

ANNEXE 3 – p. 41

ANNEXE 4 – p. 42

BIBLIOGRAPHIE - p. 45

TABLE DES MATIERES - p. 51

Résumé

Ce travail se focalise sur le développement d'outils de traitement automatique des langues (TAL) pour l'apprentissage du géorgien dédiés aux francophones, dans lequel nous décrivons nos recherches.

L'objet de ce mémoire est de proposer un générateur d'exercices sur la plateforme Moodle (avec un plugin H5P) pour les enseignants, dans des exercices de conjugaison de verbes géorgiens. Notre travail présente la réalisation de ce générateur d'exercices pour mettre en ligne des exercices à trous et propose la conception de différents niveaux de complexité.

Ce document contient la présentation de la sociolinguistique du géorgien, ainsi que de brèves explications sur la structure grammaticale du géorgien et les problèmes didactiques de cette langue. Nous ferons un bilan des quelques avancées en traitement automatique de la langue géorgienne. Ce travail propose une double problématique : théorique et technique, qui sont liées avec la générateur d'exercices.

Les résultats obtenus montrent la faisabilité pour générer des exercices en géorgien à partir d'un corpus et de critères. Nous dresserons quelques perspectives sur ce travail en particulier et plus généralement sur la didactique numérique à partir de corpus.

Mots clés : didactique des langues, apprentissage du géorgien, générateur d'exercices, Moodle, Traitement automatique des langues

მოკლე შინაარსი

წინამდებარე ნაშრომი ეხება ბუნებრივი ენის კომპიუტერული დამუშავების თემას, კონკრეტულად კი ამ კუთხით ქართული ენის ფრანკოფონებისათვის შესწავლასთან დაკავშირებულ საკითხებსა და ჩვენს კვლევებს.

ნამუშევრის მიზანია წარმოაჩინოს სავარჯიშოების გენერატორი, რომელიც იმუშავებს Moodle-ის პლატფორმაზე H5P პლაგინის ჩართვით. ნაშრომში წაიკითხავთ სავარჯიშოების გენერატორის აღწერას, მის ქსელში ჩართვასა და სავარჯიშოების შედგენის სხვადასხვა სირთულეს.

ნამუშევარში წარმოდგენილია ქართული ენის სოციოლინგვისტური საკითხები, გრამატიკული სტრუქტურის მოკლე აღწერა და სწავლებასთან დაკავშირებული პრობლემები. ასევე აღწერილი იქნება ქართულის, როგორც ბუნებრივი ენის ტექსტების კომპიუტერული დამუშავების თემა. აგრეთვე განხილული იქნება ქართული ენის სავარჯიშოების საკითხის თეორიული და პრაქტიკული მხარე. და ბოლოს, გავეცნობით შედეგებს და ვისაუბრებთ პერსპექტივებზე.

Liste de sigles et abréviations :

ELAO - l'Enseignement des Langues Assistées par Ordinateur

IA - Intelligence artificielle

Moodle - Modular Object-Oriented Dynamic Learning Environment (en anglais)

TAL - Traitement automatique des langues

Introduction

Tandis que la démographie est beaucoup plus dynamique au sud qu'au nord de la planète et que les conflits et la pauvreté y sont plus grands, les migrations sont devenues une part de l'actualité majeure depuis ces dernières années. En 2019 l'évaluation quantitative des flux migratoires géorgiens en France, quand la Géorgie était le premier pays d'origine pour les demandes d'asile en France, loin devant les Syriens, a obligé le ministre de l'Intérieur Français de l'époque, Christophe Castaner, à se rendre en Géorgie pour cette raison qui est une « anomalie » selon lui. (Figaro, 2019) D'une part il y a une longue liste de relations politiques qui ont émergé entre les deux pays depuis le début de XXe siècle, et d'autre part, depuis la situation sanitaire actuelle liée à la COVID19, nous voyons que la demande pour l'apprentissage du géorgien par les francophones a fortement augmenté : les cours à distance pour cette langue, proposés par l'Institut Français de Géorgie, ont eu un tel succès qu'il ne cesse d'ajouter de nouveaux groupes encore aujourd'hui. Cette popularité augmente pour différentes raisons : il y a des motifs personnels et professionnels (les linguistes, les artistes, les politiciens, les travailleurs humanitaires, par exemple Médecins sans frontières etc...). Néanmoins, les méthodes et les outils pour l'apprentissage du géorgien comme deuxième langue pour les francophones restent encore très loin des attentes des enseignants ainsi que des apprenants, car ils ne répondent pas parfaitement à leurs besoins pour le moment. Enfin, nous précisons qu'il existe peu de matériaux sur la didactique du géorgien, et très peu en français et enfin, parmi eux, il n'en existe aucun à notre connaissance qui fonctionne grâce au numérique. Donc notre contribution représente le premier travail sur le sujet.

Problématique

Alors que la langue géorgienne suscite un intérêt par de plus en plus de francophones, comment mettre en place une méthode numérique pour l'apprentissage de cette langue, avec une structure grammaticale si différente de la langue française, et peut-on s'appuyer sur un corpus pour générer des exercices ayant le niveau le plus adapté possible?

Hypothèses de recherche

Pour développer cette problématique nous avons un questionnement sur plusieurs niveaux : dans un premier temps, nous nous intéresserons à l'état de l'art, en présentant à quoi ressemble linguistiquement le géorgien. Puis nous décrirons la problématique qui est focalisée sur 2 aspects: technique et théorique. Et enfin, nous verrons les résultats de notre recherche, nous les discuterons et nous ferons un point sur le champ des possibilités qu'ils nous ouvrent.

Nous nous questionnerons donc sur le contexte sociolinguistique et grammatical du géorgien. Nous nous intéresserons ainsi aux raisons de l'image que l'on se fait souvent du géorgien comme étant une langue très difficile à maîtriser parfaitement par les étrangers.

- Est-ce que le contexte social externe a un impact sur le géorgien moderne ?
- Est-ce que la structure de la grammaire géorgienne diffère beaucoup de la structure française ?
- Est-ce que la raison pour laquelle le géorgien est considéré comme une langue très complexe à apprendre vient du défaut de méthodes d'apprentissage ?
- Est-ce que les géorgiens portent un intérêt au TAL ?

Ensuite, le questionnement se concentrera sur les méthodes actuelles de l'apprentissage de la langue, sur les aspects théoriques et techniques du générateur d'exercices.

- La présence de l'ordinateur peut-elle faciliter l'apprentissage ?
- Est-ce qu'aujourd'hui les professeurs sont obligés d'avoir des connaissances techniques en informatique pour mener leurs cours ?

- Existe-il des solutions pour un enseignant pour déposer son exercice en ligne sans savoir coder ?
- Est-ce que la fréquence d'un mot peut être un indicateur de sa facilité ?

Enfin, nous questionnerons nos résultats :

- Est-ce que notre choix de corpus et le type d'exercices que nous avons choisis sont les meilleurs et les plus pertinents?

Sens pratique du travail

Ce travail a du sens d'un point de vue pratique, car le générateur d'exercices proposé a été testé sur la langue géorgienne et ses fonctionnalités ont été vérifiées, ce qui veut dire qu'il est utilisable.

Procédures méthodologiques

Corpus

Pour notre travail nous avons tiré un sous corpus issu du Corpus National Géorgien¹ (ქართული ეროვნული კორპუსი). Sa création date de l'année 2012. C'est un corpus déjà entièrement étiqueté grammaticalement pour partie manuellement et pour partie automatiquement contenant toutes les étapes de son développement historique. Notre corpus s'appelle « tavisufleba.tsv »

Verbe

Pour le choix de la didactique, on se focalisera sur les verbes. L'exercice propose de trouver et de conjuguer le verbe caché dans un texte à trous.

Plateforme

¹ Georgien National Corpus (GNC) lien : <http://gnc.gov.ge/gnc/page?page-id=gnc-main-page>

Le corpus, déjà étiqueté, est ensuite traité dans un programme python que nous avons créé, et sera envoyé sur la plateforme MOODLE en utilisant le plugin H5P.

Annonce du plan

La première partie esquisse un état de l'art contenant la sociolinguistique du géorgien, la structure grammaticale, le problème de la didactique du géorgien et une observation du TAL géorgien. Dans la partie suivante, nous allons détailler notre problématique et l'architecture de l'exercice. Enfin, la troisième partie sera dédiée aux résultats et à leur discussions.

Partie 1. L'état de l'art

Avant de commencer à discuter sur la problématique, il nous semble important en premier lieu d'évoquer les aspects sociolinguistique et grammatical du géorgien. Qui parle le géorgien et où est-il parlé? À quelle famille appartient cette langue ?

1.1. Présentation de la sociolinguistique géorgienne

Le géorgien² représente la langue officielle de la République de Géorgie, un pays qui est situé dans le centre du Caucase. Sur le territoire de ce pays elle est parlée par à peu près 3,7 millions de personnes selon les données du 1er janvier 2020³, en plus des migrants Géorgiens en Turquie, Iran, Russie, États-Unis, dans les pays Européens et enfin, par ceux qui ne sont pas ethniquement géorgiens et qui l'ont apprise. Les chiffres exacts sont compliqués à déterminer car il n'existe pas de données fiables dessus, mais les chercheurs estiment à cinq millions le nombre de locuteurs.

² En géorgien : ქართული ენა

³ Office national des statistiques de Géorgie(GeoStat)
http://www.geostat.ge/?action=page&p_id=472&lang=geo

En réalité, le géorgien est une langue de la famille caucasique du Sud, appelée aussi « kartvélienne ». Elle n'est pas la seule à appartenir à cette famille, il existe aussi trois autres langues : le svane, le mingrélien et le laze. Il faudrait noter que dans chaque région de la Géorgie, sur certains territoires de l'Iran (Féréidani), de la Turquie et de l'Azerbaïdjan (Balakan, Zaqatala et Qax) qui sont peuplés depuis des siècles par les géorgiens, les dialectes sont différents. Pour l'enseignement, l'usage officiel et l'écrit soutenu sont construits sur le dialecte de la région de la capitale du pays – Tbilissi (Gérardin, 2016) .

D'ailleurs la langue géorgienne a une longue tradition littéraire attestée depuis au moins le Vème siècle après J.-C. Cela nous permet d'observer le développement de la langue à travers les siècles. Même si dans les textes anciens il peut se trouver des mots que nous n'utilisons plus aujourd'hui, ou bien certaines formes de déclinaisons ou de conjugaisons qui ont changé depuis, dans l'ensemble ils restent compréhensibles pour les lecteurs même de nos jours. Ainsi ce n'est pas seulement la langue littéraire des géorgiens, mais aussi des autres peuples du Caucase.

Comme la langue est un organisme vivant et qu'elle se transforme avec l'histoire des peuples, le lexique de la langue géorgienne n'est pas exclu de ces types de changements depuis des siècles. De nos jours, il est rempli de barbarismes et néologismes de plusieurs pays : perses, turcs, russes (les mots provenant de cette langue restent surtout des barbarismes et rarement des néologismes), français et plus récemment anglais. Elle s'est enrichie notamment avec le grec depuis les périodes anciennes. (Akaki, 1953) Enfin, depuis deux millénaires elle est en péril mais la langue a survécu. (MELOUA, 2009)

1.2. Présentation de la structure grammaticale du géorgien

En ce qui concerne la structure grammaticale du géorgien, qui est analysée par les chercheurs et les linguistes depuis des siècles (Gérardin, Contribution à l'étude des structures réfléchies en géorgien, 2011), nous n'avons pas pour ambition de décrire tous les détails de la grammaire de cette langue, qui est apparue dans les 5 langues les plus compliquées à apprendre (Seelinger,

2017). Il s'agit juste d'une présentation linguistique rapide des éléments essentiels pour la compréhension de ce travail.

Pour commencer, les géorgiens écrivent avec leur propre alphabet nommé მხედრული - mkhedruli. C'est une écriture unicamérale. Avant le mkhedruli, les géorgiens utilisaient l'écriture : asomtavruli (ასომთავრული) et nouskhuri (ნუსხური). Selon les chroniques géorgiennes, le premier alphabet géorgien date du III siècle avant J.-C.⁴

Les alphabets géorgiens sont représentés par les caractères Unicode U+10A0 à U+10FC et U+2D00 à U+2D25.

Nous avons 10 parties du discours pour le géorgien : noms, adjectifs, déterminants numériques, pronoms, verbes, adverbes, prépositions, conjonctions, particules et intermédiaires.

Pour aller plus loin sur la grammaire, comme Hélène Gérardin l'évoque dans sa thèse de doctorat publiée en 2016 : « le géorgien est une langue agglutinante à ordre des mots SOV⁵ très souple, avec les caractéristiques suivantes : absence de genre, l'ordre déterminant-déterminé, postpositions. Les relations grammaticales y sont marquées à la fois sur les constituants nominaux sous forme de cas et de postpositions, et sur le verbe sous forme d'indices actanciels aussi bien préfixés que suffixés ».

La morphologie nominale a l'aspect d'une déclinaison, avec des cas. Les principaux cas du géorgien moderne sont:

1. სახელობითი / sakhelobiti «nominatif»
2. მოთხრობითი / motkhrobiti «ergatif» (littéralement: «narratif»)
3. მიცემითი / mitsemiti «datif»
4. ნათესაობითი / natesaobiti «génitif»
5. მოქმედებითი / mokmedebiti «instrumental»
6. ვითარებითი / vitarebiti «adverbial»

⁴ Ils désignent conventionnellement le principal recueil de textes historiques médiévaux de Géorgie Kartlis Tskhovreba (en géorgien ქართლის ცხოვრება), littéralement « Vie du Kartli »

⁵ En typologie syntaxique, une langue dont les phrases suivent, généralement, un ordre sujet-objet-verbe

7.წოდებობა / tsodebiti «vocatif»

Quand nous commençons à parler de «cas», la difficulté de décrire la langue géorgienne commence à s'éclaircir, car il s'agit bien d'un système suffixal et non d'un système de désinences. (COULIE, 2013)

D'un autre côté, les catégories de temps, d'aspect et de mode s'expriment toujours par la morphologie verbale. Quand nous disons que le géorgien est une langue agglutinante, on peut préciser qu'un verbe peut comporter de nombreux préfixes indiquant le sujet, le temps, l'objet direct/indirect et ainsi de suite. Nous avons onze tiroirs (ensemble de formes verbales)⁶ verbaux qui comprennent le temps, l'aspect et le mode. Chaque tiroir a trois séries.

En géorgien, les infinitifs et les participes sont considérés comme des formes nominales du verbe et pas comme des formes verbales; étant des formes nominales, elles produisent une «flexion» comme les noms (Shanidze, 1962). La conjugaison du géorgien a recours à une forme appelée -საწყობო /Satskisi, souvent utilisée dans un emploi verbal correspondant à l'infinitif des langues indo-européennes (COULIE, 2013).

La langue géorgienne se démarque par son verbe «superpuissant». Les philologues et les chercheurs évoquent souvent dans leurs travaux qu'il est «la colonne vertébrale» de la phrase géorgienne, donc trouver la forme correcte du verbe est un vrai défi pour un apprenant. La morphologie fait intervenir des préfixes, des infixes et des suffixes. (Fähnrich, 1994) Les phrases géorgiennes peuvent être construites par un seul mot – verbe, par exemple, «წვიმს⁷», «მაბედნიერებ⁸». Le verbe géorgien transformé en phrase d'un mot peut raconter une petite histoire, par exemple, «შემომეხარჯა», qui veut dire que «par accident j'ai tout dépensé mais je ne l'ai pas voulu et j'attends votre compréhension à ce sujet»⁹; Il est capable de montrer le caractère des actants, comme «მომატყუა»¹⁰. Et cela est une liste très courte des aspects

⁶ En géorgien 'msk'riv-i'. En français, le terme est traduit comme «tiroir» à Damourette et Pichon (1936). Il a été appliqué au géorgien par Denis Creissels et René Lacroix. Ibid. 3

⁷ [tswims] - «Il pleut»

⁸ [mabedniereb] –«Tu me rends heureux»

⁹ [shemomekhardja] –la traduction de ce mot était prise et traduit du poste de Twitter de Justin McKenzie Smith - L'Ambassadeur de Grande Bretagne en Géorgie

¹⁰ [momatkua] –«Il m'a menti»

que le verbe géorgien est capable de montrer à lui seul sans l'ajout d'une autre partie du discours. Donc, dans tous les exemples donnés dans ce paragraphe nous voyons que le verbe arrive à montrer morphologiquement la corrélation entre les actants.

D'ailleurs, la catégorisation de la classe grammaticale et son expression morphologique sont caractéristiques de la structure grammaticale des langues ibéro-caucasiennes (Chikobava, 1942). Et le géorgien étant l'une d'entre elles, au fur et à mesure les changements des noms et des verbes par les classes grammaticales ont été remplacés par les déclinaisons des noms et des conjugaisons des verbes en les accordant. (Mélikishvili, 2014)

Contrairement aux langues indo-européennes, et notamment le français, la principale caractéristique du verbe géorgien est qu'il est polypersonnel, (Mélikishvili, 2014) c'est-à-dire il ne comprend pas que le sujet, mais aussi le complément d'objet, et cela est la raison de sa structure et de sa mécanique syntaxique. Il s'agit du mécanisme de l'accordance du nom et du verbe par un actant, par déclinaison, par conjugaison et par l'accordance en nombre.

Pour conclure, nous avons pu voir que la plus complexe partie du discours de cette langue est le verbe. C'est justement pour cette raison qu'il attire notre attention principale dans les recherches et dans la suite de ce travail.

1.3. La didactique des langues par l'ELAO (avec aide du TAL).

Dès professeurs au tableau jusqu'au professeurs sur ordinateurs

Tout d'abord il faudrait dire que parmi les nombreuses disciplines affiliées à l'ELAO (l'Enseignement des Langues Assistées par Ordinateur) se trouvent : la linguistique appliquée, la didactique des langues, l'informatique, le traitement automatique des langues, l'intelligence artificielle en dernier, la moins important pour nous et etc... Avec l'augmentation de l'accessibilité des produits informatiques, beaucoup d'outils sont utilisés dans l'optique d'enseigner des langues dans le cadre académique ou industriel. Cet enjeu est devenu la

problématique de nombreux chercheurs participant à des activités d'associations spécialisées comme CALICO¹¹, EUROCALL¹² et beaucoup d'autres (Šmilauer, 2008).

En ce qui concerne l'intégration des techniques de traitement automatique des langues (TAL) dans l'ELAO, d'après les chercheurs comme Karttunen (1986), Zock (1996), Nerbonne (2003), l'ELAO est un cercle idéal pour la vérification des fonctionnalités des techniques de TAL, car la tâche d'accueillir un apprenant dans son apprentissage implique virtuellement tous les objectifs visés par cette discipline, comme l'évoque Šmilauer dans sa thèse de doctorat soutenue en 2008 (Šmilauer, 2008).

D'ailleurs, il n'y a pas si longtemps, la profession d'enseignant était associée à un professeur dans sa classe, au tableau, devant ses élèves. Mais aujourd'hui avec les évolutions technologiques et didactiques, le domaine de la profession d'enseignant est partagé avec les machines. De nos jours, il est possible d'apprendre en classe, mais aussi sur ordinateur – les deux en même temps est également une possibilité- les limites ont fortement diminué. La capacité des machines devient si puissante qu'il y a 30 ans, cela aurait été associé au genre de la science-fiction. Actuellement, les humains se posent la question, de savoir si un jour les professeurs seront remplacés par les robots. Mais avant que cela n'arrive (ou pas), une partie des enseignants a commencé à utiliser l'aide des ordinateurs pour être plus efficace dans leur travail. Les outils technologiques viennent soutenir l'enseignant à l'apprentissage (Marcel, 2007). Des tableaux interactifs et des plateformes en ligne leur permettent de gagner du temps et de l'énergie et facilitent la création de ressources pédagogiques complexes. Nous voudrions surtout discuter de la façon dont les ordinateurs aident les professeurs. Il s'agit d'outils qui prennent en compte leurs besoins spécifiques pour l'apprentissage du géorgien pour la génération de matériels pédagogiques pour créer un exercice. L'exercice est une activité d'application, d'apprentissages antérieurs dans un but de maîtrise, de perfectionnement, de consolidation, de rappel ou de contrôle de connaissances (Legendre, 2005).

Enfin, il faudrait préciser que quand nous discutons des exercices générés à l'aide d'une machine pour accompagner les enseignants, eux, les professeurs, n'ont pas besoin de connaissances techniques en programmation, mais ils doivent être quand même capables de

¹¹ <https://calico.org/>

¹² <http://siglp.eurocall-languages.org/>

guider les apprenants dans ces sources d'informations, soit par leur présence, soit par la production de documents descriptifs en accompagnement. L'enseignement de langues étrangères en général, et le géorgien, parmi elles, pose des problématiques spécifiques, liées à la particularité d'une langue et à ses caractéristiques (Martinez, 2017).

1.4. Traitement automatique de langues pour le géorgien

En ce qui concerne le TAL(Traitement Automatique des Langues) géorgien, comme la langue est morphologiquement riche, avec l'ordre des mots SOV très souple et avec une grande complexité de la syntaxe, elle a besoin d'approches très spécifiques et de programmes puissants pour être analysée. Les articles sur le TAL sont assez récents, le plus ancien date de 2003.¹³ Cela montre que les géorgiens ont commencé à appliquer le TAL géorgien assez tardivement (les raisons sont sans doute liées avec la situation socio-économique des années 90s). Mais dès le début (Mishelashvili, 2006) ils se sont montrés assez actifs car ils l'appliquent dans beaucoup de domaines : reconnaissance de parole pour les dictionnaires, correcteur automatique de fautes d'orthographe, traduction électronique, système de recherche, analyseur morphologique, syntaxique et sémantique. Ensuite, ils l'utilisent pour la didactique de la langue en créant des exercices, ce qui devient de plus en plus populaire.

Comme analyseur pour le corpus géorgien ils utilisent souvent Sparklis¹⁴. Pour l'analyse syntaxique ils utilisent également Synpathy. Et comme analyseur morphologique XEROX FST tools. Nous précisons que ces informations sont tirées d'articles et non de notre propre expérience (Oleg Kapanadze, 2020). Par ailleurs, en dehors de quelques analyseurs connus, comme ceux déjà cités, pour le géorgien il est souvent utilisé des analyseurs faits par des informaticiens qui ne sont pas connus sur le marché, par exemple, celui créé par Davit Mishelashvili. Nous voudrions surtout cité l'analyseur syntaxique Corpuscle, (Meurer, 2012)avec lequel avait été annoté le Corpus National du Géorgien.

¹³ Listes des articles articles publiés sur le TAL géorgien :

<http://www.ice.ge/conferenciabi/Bunebriv%20enata%20damushaveba.html>

¹⁴<http://www.irisa.fr/LIS/ferre/sparklis/?title=%0A%09%09%20%20Core%20English%20DBpedia%0A%09%09&endpoint=http%3A//servolis.irisa.fr/dbpedia/sparql>

Finalement, nous pouvons confirmer que même si le TAL géorgien ne date pas de très longtemps, sa croissance est forte et l'intérêt qui lui est porté est grandissant.

Partie 2. Problématique et architecture

2.1 Problème de la didactique du géorgien

La didactique d'une langue qui a une structure complètement différente de celle qu'un apprenant francophone connaît, bien évidemment, complique la suite. En général, l'apprentissage d'une langue étrangère soulève des problématiques spécifiques, pour reprendre des mot de Cuq pour délimiter son objet, nous pouvons dire que : « Les langues étrangères s'inscrivent au sein de la Didactique des langues dont elles constituent des sous-ensembles particuliers » (Jean-Pierre Cuq, Isabelle Gruca , 2005). En l'absence du milieu naturel, les étudiants abandonnent souvent la langue jusqu'au moment de faire leurs devoirs. Qu'en est-il d'une langue étrangère pour un apprenant ? Elle n'est pas la langue de première socialisation, ni la première dans l'ordre de ses appropriations linguistiques. Donc le rôle d'une didactique, qui est complexe, est souvent contesté (Verdelhan-Bourgade, 2002). Des sources existantes sur la didactique du géorgien posent plusieurs principes généraux qui restent pour la plupart remis en question et d'actualité.

En effet, la grammaire est souvent associée à l'examen des structures linguistiques de la langue, plutôt que comme un outil de matérialisation de la compétence communicative (Uwizeye, 2011).

Tout d'abord, le fait que l'enseignement du géorgien pour les étrangers soit fréquemment choisi avec un rapprochement de l'anglais reste toujours la cause profonde du problème. Parce que l'apprenant francophone doit associer tel ou tel aspect de la grammaire géorgienne avec sa première langue (Uwizeye, 2011).

Quand on parle de didactique du géorgien, il est impossible d'éviter de discuter des problèmes liés à la théorie de l'enseignement de cette langue. Les linguistes évoquent plusieurs types de difficultés (Mélikishvili, 2014):

D'abord, les identifications de catégories grammaticales ne sont pas basées sur les spécificités de la langue d'apprentissage, ni sur les résultats de ses analyses systémiques et statistiques.

Ensuite, le problème est le suivant : le principe de la classification ne s'appuie pas sur les critères morphologiques et syntaxiques.

Puis, pendant le processus d'apprentissage, à la place de la terminologie de la grammaire et de la linguistique, sont souvent utilisés les termes logique ou sémantique, alors que la structure grammaticale doit être toujours décrite par les notions morphosyntaxiques. Cela, pour enlever toute ambiguïté et tout malentendu.

En réalité, deux méthodes sont utilisées pour étudier la grammaire du géorgien: l'induction et la déduction. L'enseignement est inductif, lorsque le matériel de l'enseignement passe de l'analyse de faits particuliers et concrets, au général, à la règle. Et si la formation va dans le sens inverse, c'est-à-dire la règle grammaticale est d'abord enseignée, et ensuite des exemples concrets sont analysés - la manière d'enseigner est déductive. En fait, ces deux modes d'enseignement sont intimement liés et se complètent, leur séparation est quelque peu conditionnelle. Pourtant, quelques chercheurs (Mélikishvili, 2014); (Chikobava, 1942) précisent que le processus d'apprentissage doit toujours commencer par l'induction et puis aller vers la déduction, et jamais à l'inverse.

Les scientifiques évoquent les difficultés qu'ils confrontent dans la théorie, ainsi que dans la pratique. Il s'agit de faits morphosyntaxiques et terminologiques en général, pour :

1. Tous les niveaux quand il s'agit de tiroirs ('msk'riv-i')
2. La transitivité de verbes polypersonnels
3. Le problème d'inversion et les difficultés liées aux marquages pour définir les fonctions logiques/sémantiques.

En définitive, Mélikishvili croit que la cause de ce dernier défaut peut être liée à l'utilisation de principes de classification dédiés aux langues européennes, appliqués au géorgien.

En outre, pour la didactique du géorgien nous utilisons toujours la terminologie de Akaki Shanidze (linguiste et philologue géorgien (1887 – 1987)). Elle diffère de la terminologie universelle, ce qui complique l'enseignement.

Comme nous avons déjà pu l'évoquer, le fait que, contrairement aux langues indo-européennes, la principale caractéristique du verbe géorgien est qu'il est polypersonnel; pour la didactique du géorgien, c'est l'une des plus importantes difficultés qui peut émerger.

Hélène Gérardin dans son article « Problèmes de l'enseignement de la morphosyntaxe verbale en géorgien : deux difficultés et quelques suggestions » (Gérardin, ზმნური მორფოსინტაქსის სწავლების საკითხები: ორი სირთულე და რამდენიმე შეთავაზება, 2015) publié à Tbilissi en 2015, évoque deux problèmes que rencontrent les apprenants du géorgien en tant que langue étrangère. La recherche avait été effectuée sur la base de comparaisons d'expériences de quelques étudiants. Gérardin indique que la raison la plus importante du problème d'apprentissage est la différence morphosyntaxique verbale géorgienne de celle des langues que maîtrisent souvent des apprenants.

De plus, elle précise que les regroupements des verbes ne sont pas bien représentés dans la littérature et on confronte alors deux types de problèmes : théorique et pratique.

Enfin, comme la langue est un système et n'est pas un matériel isolé de tout, pour les chercheurs, le problème important reste l'apprentissage de la langue sans oublier que tout est systémique.

2.2 .Aspects théoriques – différentes méthodes en ligne et choix de l'analyse

Introduire les technologies et l'enseignement en ligne dans les pratiques des enseignants est un défi que nous nous sommes engagés à relever. Notre objectif est de concevoir sur une

plateforme d'apprentissage d'aide semi-automatisée, la création de matériels pédagogiques de langue géorgienne. Pourquoi semi-automatisé? Parce que même si le générateur est automatique, la plateforme sur laquelle nous allons mettre nos exercices, permet des post-éditions. Ceci doit tout d'abord, davantage répondre aux besoins d'un enseignant, et pas un apprenant, car il n'en est pas l'utilisateur direct. Deuxièmement, prendre en compte les particularités de la langue géorgienne de niveau académique adapté pour un niveau souhaitable des exercices.

Aujourd'hui il existe quelques ressources en ligne pour l'apprentissage du géorgien par les francophones :

www.17-minute-languages.com (consulté 19 /11 /2020)

utalk.com/fr/store/georgian(consulté 19 /11 /2020)

www.50languages.com/phrasebook/fr/ka/(consulté 19 /11 /2020)

www.languageliveschool.com/fr/land/georgien/(consulté 19 /11 /2020)

www.goethe-verlag.com/book2/FR/FRKA/FRKA002.HTM (consulté 19 /11 /2020)

www.evaneos.fr/georgie/voyage/vocabulaire-de-survie/(consulté 19 /11 /2020)

Tous ces liens proposent le vocabulaire de base pour un touriste pour survivre. Ils sont plutôt centrés sur l'expression orale d'un apprenant de niveau basique et aucun parmi eux ne soumet d'exercices pour un niveau avancé pour apprentissage académique du géorgien. C'est pour cette raison que nous nous intéressons à trouver une solution pour les apprenants du géorgien en générant des exercices pour les professeurs de langue avec un niveau adapté.

De nos jours, il existe plusieurs outils qui donnent la possibilité de créer différents types d'exercices en ligne. Par outil, nous entendons ce qui permet de mettre en place le suivi de l'apprentissage (Bourda, 2001). Nous voudrions donner quelques exemples de logiciels qui existent aujourd'hui :

Par exemple, *Hot Potatoes*¹⁵, qui est un logiciel simple d'utilisation qui a été conçu par l'université de Victoria au Canada. Il permet aux professeurs de réaliser six différents types d'exercices au format html. Pourtant aucune connaissance spécifique du code html n'est requise. Ces exercices peuvent être publiés en ligne ou transférés vers les postes-élèves.

¹⁵ Lien vers la page : <http://hotpot.uvic.ca/>

Ensuite, l'outil *COURSERA*, qui propose différents types de formations en ligne (apprentissage des langues étrangères parmi elles), toutes fondées par les professeurs d'informatique Andrew Ng et Daphne Koller de l'université Stanford à Californie.

Enfin, parmi les plus populaires se trouve *Moodle (Modular Object-Oriented Dynamic Learning Environment)*. C'est une plateforme d'apprentissage en ligne qui est en source libre. Elle est développée à partir de principes pédagogiques. En plus de la création de cours à l'aide d'outils intégrés à l'usage des enseignants, Moodle propose des possibilités d'organisation des cours sous forme de filières, qui lui donnent également des caractéristiques propres à la mise en place de dispositifs complets d'enseignement. Il ajoute aussi de nombreux outils d'interactions pédagogiques et il a un accès aux plugins développés par sa communauté pour permettre l'extension de ses fonctionnalités et de répondre ainsi à d'autres besoins. C'est grâce à tous ces aspects que son utilisation devient de plus en plus appréciée par les professionnels et il a d'ailleurs été intégré au Socle interministériel de logiciels libres¹⁶ de l'État français en mai 2020. À la faveur de ses larges listes de possibilités et les actualités optimistes sur cette plateforme, nous avons décidé de commencer à y exercer notre expérimentation.

Tenant compte des spécificités de langue visée, les exercices peuvent être structurés sur différents niveaux: phonétique, morphologique, morphosyntaxique... Ainsi nous rencontrons d'autres types d'exercices, qui sont destinés à analyser (ils identifient et différencient les unités constitutives de la phrase); à représenter de manière abstraite (ils schématisent); à manipuler (ils réduisent ou déplacent); à reconstruire des phrases (ils remplissent, remplacent ou corrigent) etc... Selon les compétences linguistiques visées, par exemple la compréhension de la grammaire et de l'écrit, les exercices se distinguent en tenant compte de leurs aboutissements: de répétition, d'expansion, de reformulation, structuraux, de reconstitution, à trous (ARMĀSAR, 2013). Nous voyons que Moodle parmi ses choix varié propose des exercices à trous, liste à choix, rédaction du text. Pour notre travail nous avons choisi de générer un exercice à trous.

¹⁶ Le socle interministériel de logiciels libres (SILL) est un ensemble de logiciels libres préconisés par l'État français depuis 2016, dans le cadre de la modernisation globale de ses systèmes d'informations Lien vers la page : <https://sill.etalab.gouv.fr/fr/software>

Nous avons tenté de mettre en place, pour la création d'un premier type d'exercices intégré à la plateforme, **des exercices à trous sur les verbes**. Le choix des verbes a été effectué d'après des échanges avec des linguistes et enseignants du géorgien, ainsi qu'après avoir consulté la littérature scientifique liée au sujet de la grammaire géorgienne. Ce n'est pas surprenant que le verbe ait attiré le plus l'attention des chercheurs, car comme nous l'évoquions, nous touchons à la fois au sujet le plus compliqué de la langue et à la spécificité du géorgien en général, comme le souligne Helene Gérardin (Gérardin 2015).

L'idée est de préparer semi-automatiquement des **corpus géorgiens étiquetés**, corpus que nous avons choisis, pour en faire des exercices à trous. Nous allons revenir sur les détails du corpus dans la partie suivante de ce travail.

En effet, en dehors des aspects purement techniques que nous allons expliquer en détail dans notre propos un peu plus tard, ce qui est surtout important pour ce type d'outil, ce sont les critères appliqués au choix des tâches à inclure dans les exercices. Il est préférable d'avoir une intention didactique basée sur des paramètres linguistiques du géorgien, c'est cela qui permettra à un enseignant de choisir un niveau adapté pour un apprenant.

Quels peuvent être les meilleurs critères permettant d'aider l'utilisateur (un professeur dans ce cas-là) à distinguer les niveaux de difficulté des mots, plus concrètement les verbes, pour créer le meilleur exercice avec le niveau le plus adapté. Pour l'utilisation des verbes nous proposons différentes options :

- verbes fréquents dans le corpus (faire la liste des verbes selon leur fréquence)
- ceux qui sont au présent dans le corpus

Pourquoi soulevons-nous la question de la fréquence d'un mot en même temps que le choix de niveau d'un exercice ? En fait, il est apparu que la fréquence d'un mot dans un corpus indique qu'il est plus facile que les autres (Averil Coxhead Paul Nation, 2000,2001).

Si nous partons de l'idée que les mots faciles sont ceux qui sont les plus fréquents (Coxhead 2000, Nation 2001), nous pouvons faire la listes des verbes :

- très faciles ;
- faciles ;
- compliqués ;
- très compliqués .

L'utilisateur pourra choisir telle ou telle liste pour un niveau adapté et adéquat pour l'apprenant.

Mais est-il réellement correct d'un point de vue linguistique que les mots fréquents soient les mots les plus faciles ? Ce n'est pas sûr. C'est donc pour cette raison que nous proposons de faire une liste des verbes au présent et de travailler en considérant que les verbes au présent sont plus faciles pour les apprenants que les verbes des autres temps. Surtout que pendant l'apprentissage le temps du présent est toujours mis en avant (Robert., 1985). Ou choisir quel temps est grammaticalement plus pertinent à un moment donné.

Enfin, ces critères peuvent être choisis individuellement, ou tous ensembles.

En ce qui concerne le niveau de difficulté de l'exercice, il est avancé, car il s'agit de la conjugaison de verbes, qui ne prennent pas seulement en compte la compréhension de la signification de mots, mais surtout leur utilisation en prenant en compte différents critères. Donc il est destiné à un étudiant qui a déjà un niveau assez élevé dans la langue, car rien que le vocabulaire du corpus, qui est fondé sur les conversations politiques, le rend complexe. C'est pour faciliter la tâche que nous avons proposé quelques critères, pour que l'apprenant ne passe pas trop de temps sur chaque phrase et ne se décourage pas pendant son travail. Les exercices sont une certaine mécanique intellectuelle, ils ont pour rôle de générer de la motivation chez l'apprenant, en les organisant par paliers de difficulté de plus en plus important (Robert., 1985).

En guise de conclusion, parmi les nombreuses plateformes existant en ligne, nous avons choisi Moodle pour ses nombreuses fonctionnalités, qui sont assez claires et compréhensibles par les usagers, et pour sa forte croissance en nombre d'utilisateurs.

D'autre part, nous avons décidé de travailler sur un corpus géorgien déjà étiqueté, ce qui est une contrainte forte, heureusement nous avons un corpus adéquat. Nous avons choisi à travailler sur les verbes selon un(des) critère(s) choisi(s) pour créer un exercice à trous sur Moodle.

2. 3. Aspects techniques de l'apprentissage - l'architecture de l'exercice

Le travail d'enseignement de n'importe quelle langue étrangère ouvre de multiples possibilités d'intégrer des mécanismes de langue dans différentes méthodes d'apprentissage. Afin d'équiper les enseignants de langue géorgienne, en prenant en compte leurs besoins spécifiques nous proposons de concevoir et de développer sur la plateforme Moodle la génération de matériels pédagogiques. Pour répondre à cet objectif, nous avons effectué une première expérimentation concernant la génération automatique d'exercices d'apprentissage du géorgien.

Après avoir fait le choix de la plateforme, nous avons choisi notre corpus. Donc dans un premier temps, la ressource nécessaire pour générer cette activité est un corpus géorgien déjà étiqueté.

Corpus

Pour notre travail nous avons choisi un Corpus National Géorgien ¹⁷(ქართული ეროვნული კორპუსი). Sa création date de l'année 2012. C'est un corpus complet de langue géorgienne contenant toutes les étapes de son développement historique. En réalité, il est toujours en cours de mise au point. Il contient des sous-corpus de géorgien ancien, moyen et moderne :

- GNC Old Georgian
- GNC Middle Georgian
- GNC Modern Georgian

De plus, ce dernier contient deux sous-corpus de transcriptions de l'enregistrement de la langue courante :

- le corpus de dialectes géorgiens (GDC et SSGG)
- le corpus du projet sur la situation sociolinguistique de la Géorgie actuelle

Dans ce cadre, des corpus de textes mingréliens et svans sont également en cours de réalisation.

¹⁷ Géorgien National Corpus (GNC) lien : <http://gnc.gov.ge/gnc/page?page-id=gnc-main-page>

Une grande partie de ce corpus est entièrement annoté grammaticalement (formes, lemmes ainsi que des caractéristiques morphosyntaxiques), ainsi tous les textes de ce sous-corpus disposent de métadonnées complètes.

Les textes utilisés pour notre expérimentation sont des textes tirés du sous corpus de GNC moderne : « tavisupleba » en format tsv. La taille du fichier est 1,53 GB.

Les phrases sont segmentées avec les ponctuations. Leur nombre total est de 4 157 587. En ce qui concerne les séparateurs : le nombre de points est de 1 059 556 et le nombre de points d'exclamations est de 11 069.

Le corpus « tavisupleba.tsv » contient 1101877 phrases au total et donc si nous allons faire la sommes du nombre de ponctuations points et du nombre de ponctuations exclamations, le résultat total de la phrase est plutôt crédible.

Dans le fichier tsv il y a 4 colonnes : form, lemma, POS et morph (annotation morphosyntaxique).

Comme nous l'avons déjà évoqué dans la partie précédente, Moodle permet la création de cours à l'aide d'outils intégrés (ressources et activités). Pour créer un exercice, nous avons choisi le plugin H5P¹⁸. C'est un logiciel libre de création de contenus interactifs, basé sur HTML5 et JavaScript. D'ailleurs, hormis la plateforme E-Learning Moodle, H5P peut être interfacé dans Drupal et WordPress et il est même possible de s'inscrire sur le site de son éditeur <https://h5p.org> pour créer des exercices et en récupérer les codes d'intégration "Embed" pour les insérer dans une page web ou ailleurs. Il comprend un éditeur de contenu intuitif qui permet de générer assez facilement des contenus interactifs.

H5P dans le cadre d'une activité peut prendre différentes formes: cela peut être des types de contenus multimédias et des questionnaires, des modules sur le principe du glisser/déposer, on peut soit utiliser uniquement du texte ou jumeler texte et image, ou tout simplement uniquement des images. Il y a ensuite un enregistreur audio et plusieurs autres choix de ce type. Pour tester,

¹⁸ <https://h5p.org/>

nous avons choisi les exercices à trous (Fill in the Blanks), comme cela a déjà été évoqué dans notre propos.

En précision, les mots dans ces trous ne se remplissent pas automatiquement, c'est à l'apprenant de remplir les mots manuellement.

Du point de vue technique, notre exercice Moodle qui sera localisé dans le serveur web et contiendra un ZIP au format H5P. Ce ZIP comprendra des fichiers au format json, ainsi que des scripts en python et les résultats reçus après l'exécution de ces scripts.

Enfin, après les démarches théoriques liées à la problématique du choix de quel type d'exercices de grammaire est plus intéressant à travailler pour nous, et après leur réalisation technique, nous sommes arrivés au point de téléversement de notre corpus au format H5P sur Moodle.

Partie 3 Mise en pratique et résultats

3.1. Implémentations et mises en pratique

L'intégration du corpus pour une utilisation souhaitable a nécessité divers types de traitements. Donc nous avons effectué plusieurs traitements sur le corpus nommé « tasifuleba.tsv ». Tous les prétraitements et post-traitements ont été faits automatiquement avec des programmes écrits en Python. Pour commencer nous avons effectué le prétraitement avec le script `selectcorpus.py`¹⁹. L'algorithme de ce script est le suivant :

- effacer toutes les balises html qui gênaient l'exécution du texte
- extraire des informations pour chaque phrase (taille, morpho-syntaxe, etc.)

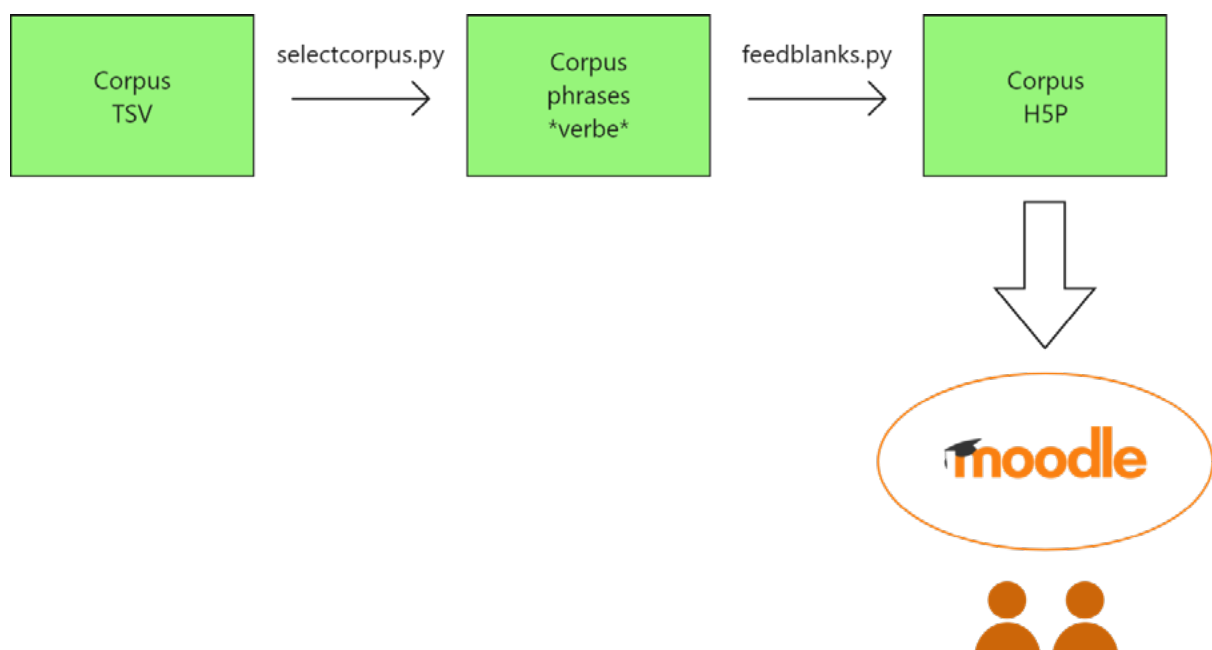
¹⁹ Tous les scripts nommés dans ce travail seront mise à l'annexe

- si la phrase correspond aux filtres (taille, présence de verbe)
 - mettre chaque phrase sur une ligne
 - trouver tous les verbes
 - les mettre entre deux astérisques * verbe*
- afficher la phrase en sortie

Ensuite, avec le script nommé `feedblanks.py` nous avons pu ajouter plusieurs questions depuis le corpus, et de cette manière, alimenter un nouveau *corpus H5P*. Pour cela nous avons d'abord écrit une première phrase manuellement dans Moodle et l'avons téléchargée sur notre ordinateur. Le fichier téléchargé était au format h5p et correspondait à un ZIP.

- unzipper le fichier H5P
- ouvrir le fichier `content.json`
 - ajouter une ou plusieurs questions depuis le corpus
 - rezipper le fichier
- afficher le nouveau fichier ZIP

Alors, d'un côté nous avons le corpus, avec tous ces traitements indiqués ci-dessus, qui est prêt, et de l'autre nous devons mettre ce corpus dans Moodle. Schématiquement, cela peut ressembler à ceci :



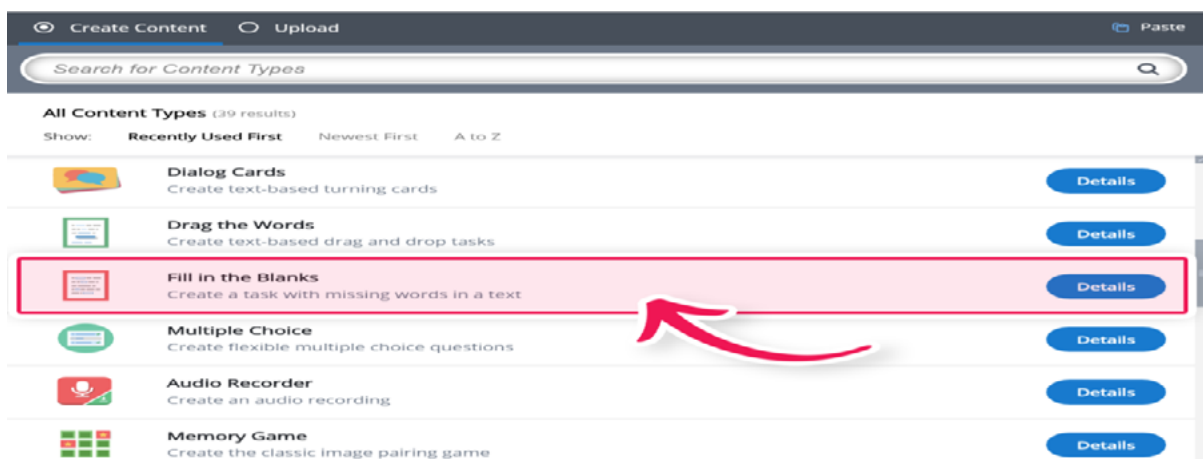
Enfin, avec le script nommé feedblanks.py nous avons pu ajouter plusieurs questions depuis le corpus et finalement créer un ZIP au format H5P pour les exercices. Ce ZIP contient des fichiers au format json. Pour avoir la base de format H5P nous avons téléchargé celle-ci depuis Moodle.

Finalement, nous avons effectué le téléversement des données depuis notre ordinateur vers le site web de Moodle.

Sur le site, du point de vue de l'interface entre le système et l'utilisateur, nous avons deux plateformes distinctes sur Moodle:

- la plateforme d'un auteur, qui est destinée à la création des exercices, où il est possible de téléverser le générateur au format H5P. Elle est utilisée par un enseignant.
- la plateforme qui est visible et utilisable par un apprenant

Donc pour implémenter ce contenu, le processeur est censé télécharger ou occasionnellement activer le plugin H5P. Concrètement pour «Fill in the blanks»(l'espace de H5P est dans sa langue d'origine c'est-à-dire - en anglais) dans son espace Moodle.



Il ne restera qu'à «téléverser » le corpus préparé auparavant. D'ailleurs, tous les champs (pour mettre les scores, comment il/elle voudrait que les résultats soient affichés et ainsi de suite) c'est au professeur de les remplir manuellement, ils sont très faciles à remplir, et ils n'ont pas besoin de connaissances techniques spécifiques.

Désormais, nous pouvons faire la mise en pratique ainsi que voir les résultats de ce travail.

3.2. Résultats du travail et mises en pratique

Nous avons choisi de créer le générateur d'exercices à trous basés sur les phrases tirées d'un corpus de textes authentiques. Les résultats de ces premières expérimentations montrent la faisabilité de la génération d'exercices à trous à partir de listes de vocabulaire et d'un corpus intégrés à la plateforme, ainsi que de textes géorgiens étiquetés.

Les résultats visuels après les implémentations sont les suivants :

Pour tester, nous avons pris et rempli les 4 phrases.

დღეისთვის ვაკანტურად მხოლოდ სახელმწიფო ქონებისა და გარემოს დაცვის მინისტრების თანამდებობები რჩება

გასულ კვირას და 4 დეკემბერს განხილულ მინისტრობის კანდიდატებს პარლამენტმა ერთად უყარეს კენჭი .
გამყრელიძის ვარაუდით , წელს სახელმწიფო ბიუჯეტის გარდვევა 150 მილიონი მაინც იქნება .
რამდენ ხანს გასტანს ამ განტვირთვით გამოწვეული შვება , მალე გაირკვევა .

✓ Vérifier

↻ Réutiliser <> Embarqué H:P

Tout simplement, comme le verbe dans la deuxième phrase n'était pas correct, le programme a affiché l'erreur et le score :

დღეისთვის ვაკანტურად მხოლოდ სახელმწიფო ქონებისა და გარემოს დაცვის მინისტრების თანამდებობები რჩება ✓

გასულ კვირას და 4 დეკემბერს განხილულ მინისტრობის კანდიდატებს პარლამენტმა ერთად უყარეს ✗ კენჭი .
გამყრელიძის ვარაუდით , წელს სახელმწიფო ბიუჯეტის გარდვევა 150 მილიონი მაინც იქნება ✓ .
რამდენ ხანს გასტანს ამ განტვირთვით გამოწვეული შვება , მალე გაირკვევა .

4/5

👁 Voir la correction

🔄 Recommencer

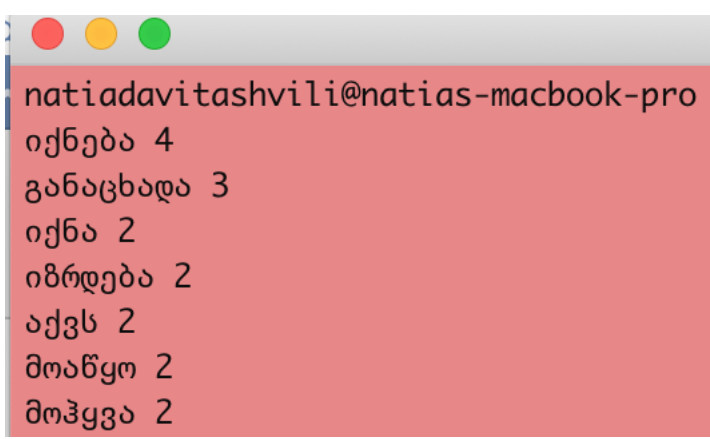
↻ Réutiliser <> Embarqué H:P

D'un côté nous avons constaté la faisabilité de ce travail, mais que cela peut-il donner comme résultat? Pour cela, discutons en prenant un exemple.

Tout d'abord, nous avons décidé de choisir les 100 premières phrases, avec les critères suivants : la taille de la phrase doit être supérieure ou égale à 10 mots et inférieure ou égale à 20 mots. Voici une partie des résultats :

4 დეკემბერს , გვიან ღამით , ბოლოს და ბოლოს , მეტ-ნაკლებად *დასრულდა* ახალი მთავრობის დაკომპლექტების პროცესი .
4 დღისთვის ვაკანტურად მხოლოდ სახელმწიფო ქონებისა და გარემოს დაცვის მინისტრების თანამდებობები *ჩრქება* .
გასულ კვირას და 4 დეკემბერს განხილულ მინისტრობის კანდიდატებს პარლამენტმა ერთად *უყარა* კენჭი .
გამყრელიძის ვარაუდით , წელი სახელმწიფო ბიუჯეტის გარღვევა 150 მილიონი მაინც *იქნება* .
ზოგადად , ზურაბ ნოლიძის კანდიდატურის განხილვა საპარლამენტო ოპოზიციის დადგენის საშუალებად *იქნა* აღქმული .
მუსტად ამდენმა კანონმდებელმა არ *დაუჭირა* მხარი ზურაბ ნოლიძის , რასაც ხელი არ შეუშლია ფინანსთა მინისტრად მისი დამტკიცებისთვის .
მოკლედ , პარლამენტმა და საზოგადოების ნაწილი მთავრობის „ გადახადისებით “ თავისებური ფსიქოლოგიური განტვირთვა *მოიწყო* .
რამდენ ხანს *გადასტავს* ამ განტვირთვით გამომწვეული შვება , მალე გაიჩვენება .
საქართველოს პარლამენტის ვიცე-სპიკერმა ვახტანგ რჩეულიშვილმა სამართალდამცავ ორგანოებს აღდამოვის საკითხის გარკვევა *მოსთხოვა* .
აღდამოვი რჩეულიშვილს *ურჩევს* , თუ რუსეთში სერიოზობა მოენატრა , ოჯახთან ერთად „ ბარგი-ბარხანა “ ჩააღვოს და მოსკოვში წავიდეს .
რჩეულიშვილმა აღდამოვის საქმიანობაში ეჭვი *შეიტანა* და მას ნარკომაფიისა და აღამიანის გამტაცებლებთან შუამავლობა დააბრადა .
ამ ბრალდების სისწორისა რა *მოგახსენოთ* , მაგრამ ჩვენთვის ნარკომაფიისა და აღამიანის გამტაცებლებთან შუამავლობის მიზნით .
ამიტომაც ესეთი აღამიანები მხოლოდ და მხოლოდ *ამწვავებენ* უთითრობას და ამიტომაც ვამბობ , ჩვენ არ დავუშვებთ ამას .
“ რჩეულიშვილის პოზიციებს *ამყარებს* საქართველოში რუსეთის საგანგებო და სრულუფლებიანი ელჩი ვლადიმერ გულევი .
[*გულევის* ხმა] „ ამ აღამიანის საქართველოს ტერიტორიაზე საქმიანობა უკანონოა .
ეს კი სურვილის შემთხვევაში , საქართველოს ხელისუფლებას პოლიტიკურად მიუღებელი ფიგურის ჩამოცილების არაპოლიტიკურ საფუძველს *აძევებს* .
გიორგაძის ამ სურვილის ხორციელებისთვის კი ისევ ადგილობრივი პროკომუნისტურ-პრორუსული ძალები *იჩვენებენ* .
გიორგაძე თავისუფალია , ის მის მიმართ წაყენებული ბრალდებებისგან მისმა უდიდებულესობამ ღრმ *გაათავისუფლა* .
უშიშროების ყოფილ მინისტრს ბრალდებებისთვის არ *უპასუხია* და , სხვათა შორის , არც შევარდნაძე განიცდის უხერხულობას ამ ბრალდებების გამო .
ოღონდ დაბრუნების შემთხვევაში ის ვერ *გადაურჩება* დაკითხვას იმ ბრალდებებზე , რომლებიც მის მიმართ არსებობს .
გიორგაძის დამცველი ქეთევან ჯაბუკიანი კი *იჩვენება* , რომ ეს უძომისი ბრალდებები მხოლოდ სამ იძულებით ჩვენებას ეფუძნება და უსაფუძვლოა .
პასუხი ცალსახაა : დაბრუნების შემთხვევაში გიორგაძეს გამოძიების , შემდეგ კი სასამართლოს წინაშე წარდგომა *მოუწევს* .
მტაცებელმა საილუსტრაციოდ *მოიხმეს* ამბავი , რომელიც რამდენიმე თვის წინ ჩვენს ეთერშიც გავიდა , ოღონდ სხვაგვარი ინტერპრეტაციით .
ვალდებულ , თამარ ჩხაიძის ქალიშვილის , ვიოლა ხურცილავას თქმით , კატეგორიული უარია *განაცხადა* სასამართლოს გადაწყვეტილების აღსრულებაზე .
ამის გამო ვაკე-საბურთალოს გამგეობის იურისტზე ახლა სისხლის სამართლის საქმეა აღძრული .
ამ მოვლენას საქართველოს ნავთობის საერთაშორისო კორპორაციამ პროექტის საზოგადოებრივი მხარდაჭერის უპრეცედენტო კამპანია *მიუძღვნა* .

Ensuite, nous avons vérifié la fréquence de chaque verbe dans les 100 phrases.



```
natiadavitashvili@nati-as-macbook-pro
იქნება 4
განაცხადა 3
იქნა 2
იზრდება 2
აქვს 2
მოაწყო 2
მოჰყვა 2
```

Ci-dessus : les verbes les plus utilisés dans le corpus.

Le verbe le plus fréquent იქნება²⁰, est le verbe basique « être », conjugué au futur à la 3ème personne du singulier. Effectivement, il ne s’agit pas d’un verbe au temps présent. Donc quand nous avons posé l’hypothèse que si un mot facile est celui qui est le plus fréquemment utilisé ou celui au présent, nous pouvons désormais répondre grâce à notre résultat d’après cet exemple.

En ce qui concerne les autres verbes comme „განაცხადა“²¹, „იქნა“²², „იზრდება“²³, „აქვს“²⁴, „მოჰყვა“²⁵, certains qui sont plus rarement utilisés sont plus faciles que les autres. Par exemple, „აქვს“(2 fois), qui correspond au verbe « avoir » est plus simple, que „განაცხადა“(3 fois) – qui signifie “a déclaré”, tout simplement parce que les verbes comme « être » et « avoir » sont des verbes basiques qu’on apprend en premier dans l’apprentissage de n’importe quelle langue étrangère, et le géorgien est parmi elles (Gabunia, 2016).

Pour aller plus loin, expérimentons de sortir les phrases qui contiennent les verbes « ყოფნა » (« être »), « ქონა » (« avoir quelque chose »), « ყოლა » (« avoir quelqu’un »)²⁶. Pour essayer prenons les 25 phrases qui contiennent ces verbes basiques dans l’apprentissage d’une langue, avec la taille des phrases entre 3 et 7.

```
natiadavitashvili@natias-MacBook-Pro generateur % python3 selectcorpus.py -n 25 -mins 3 -maxs 7  
-lem ყოფნა,ქონა,ყოლა tavisupleba.tsv resultats.txt  
Number of sentences: 25
```

Les résultats reçus correspondent parfaitement aux nos attentes :

²⁰ :ikneba – Il sera

²¹ :ganatskhada – Il a déclaré

²² :ikna – Il était

²³ :izrdeba’ – Il grandis

²⁴ :akvs -Il a

²⁵ :mohkva – il a commencé à..

²⁶ Dans la langue géorgienne il n’existe pas un verbe commun pour ‘avoir’ pour vivants et non-vivants

რამდენიმე მინისტრის პოსტზეც *არიან* ქალები გათვადისწინებული .
 გუბერნატორები , გამგებლები *მყავს* მხედველობაში .
 იარაღის ორი ვრთველი დატენილი *იყოს* .
 ნორმან გრანცი არ *ყოფილა* ჯაზის მუსიკოსი .
 ეს არ *არის* ქრისტიანობა .
 მათ *ჰქონდათ* პოზიტიური , კონსტრუქციული მოღაპარაკებები .
 საქსოვი მანქანა თითქმის ყველა ოჯახს *აქვს* .
 შეტევისთვის მზად *არიან* ავღანეთა რაზმებიც .
 მერყეობს ხორცი , 107-დან 129-მდე *გვყვია* .
 მე კანონის უზენაესობის მომხრე *ვარ* .
 და ეს *არის* დიდი სხვაობა .
 და დღეს ისინი დამარცხებულები *არიან* “ .
 დევნილები უკმაყოფილონი *არიან* ქართულ-აფხაზური დიალოგის მიმდინარეობითაც .
 ყრილობაზე უაღრესად ნერვული ატმოსფერო *იყოს* .
 ისაკაძე *იყოს* მისი მფარველი !
 2005 წლისთვის მიღსადენი მზად უნდა *იყოს* .
 ეუთოს მეთვადყურენი არ *არიან* შეიარაღებული .
 საჭირო *იყოს* მინიმუმ 226 .
 მაგრამ ისინი განვრთნილი *იყვენენ* .
 ამოტრიალებული სიტუაცია *არის* “ .
 ინტელიგენტი აქაც კასტრირებული *იქნება* .
 ქვეყნიდან წასვლა დეპუტატს გადაწყვეტილი არასოდეს *ჰქონია* .
 ეს *არის* ბორტმოქმედების , განუკითხაობის წაქეზება .
 რუმინეთში არაფერი ამის მაგვარი არ *არის* .
 ვიზიტი მომავალი წლისთვის *არის* დაგეგმილი .

Cela montre bien que notre générateur peut être assez facilement adaptable aux besoins d’un professeur.

D’ailleurs, en n’indiquant aucun paramètre, nous avons reçu 254 805 phrases qui contiennent les mots « ყოფნა , «ქონა», « ყოლა » dans le corpus « tavisupleba.tsv.

```
natiadavitashvili@nati-MacBook-Pro generateur % python3 selectcorpus.py -n 0 -mins 0 -maxs 0 -
lem ყოფნა,ქონა,ყოლა tavisupleba.tsv resultats.txt
Number of sentences: 254805
```

Ensuite, nous allons compter la quantité de phrases avec les mêmes verbes, avec une taille entre 4 et 8. Le résultat révèle 16 789 phrases dans le corpus.

```
natiadavitashvili@nati-MacBook-Pro generateur % python3 selectcorpus.py -n 0 -mins 4 -maxs 8 -
lem ყოფნა,ქონა,ყოლა tavisupleba.tsv resultats.txt
Number of sentences: 16789
```

Puis, regardons combien de phrases ont une taille comprise entre 8 et 11.

```
natiadavitashvili@natis-MacBook-Pro generateur % python3 selectcorpus.py -n 0 -mins 8 -maxs 11  
-pos V tavisupleba.tsv resultats.txt  
Number of sentences: 134465
```

Comme nous le voyons, le programme a pu compter 134 465 phrases avec ces paramètres.

Après nous pouvons changer encore les paramètres et choisir les phrases entre 5 et 15 mots.

Dans ce tableau se trouvent les résultats de notre test :

Toutes phrases entre 5 et 15 formes	394 313
au moins un POS V	355 035
au moins un POS N	382 771
au moins un POS Adv	186 448
au moins un lemme ყოფნა	48 615
au moins un lemme ქონა	14 240
au moins un lemme ყოლა	2 005

Enfin, comme nous l'avons déjà indiqué, ces exercices sont plutôt destinés aux élèves avec un niveau de géorgien élevé, qui peuvent se débrouiller avec la terminologie de la presse, donc nous pensons que pour eux cela peut être un bon exercice. Surtout, que les utilisateurs peuvent utiliser différents paramètres et filtres linguistiques pour personnaliser leurs phrases et puis, les exercices, de ce point de vue notre modèle ressemble au modèle de Pilan, Volodina et Johansson pour l'apprentissage du Suédois pour le niveau L2 (Pilan, Volodina, Johansson, 2014).

En fait, notre cas n'est pas un exercice de lexique et nous nous rendons compte qu'il ne faut pas mélanger les objectifs. C'est pour cela qu'il faut bien définir les paramètres afin qu'ils soient adaptés au niveau de l'apprenant. Enfin, *à priori*, il suffit juste de bien ajuster les critères par un apprenant pour savoir ce qu'il voudrait comme option : verbes au temps concret comme

le présent, verbes les plus utilisés ou les moins utilisés dans le corpus, les verbes longs ou courts.

3.3. Discussions et perspectives

Le générateur d'exercices embarqué sur Moodle fonctionne, cependant, de notre expérimentation découlent quelques interrogations car, à notre avis, il a mis en avant quelques manquements.

Tout d'abord, pourquoi ce corpus ? C'est un langage de presse avec des sujets politiques, ce qui veut dire qu'on exclut certaines catégories d'apprenants, par exemple les enfants. Ils peuvent avoir un niveau élevé et ils peuvent être prêt à se lancer dans les entraînements de conjugaisons de verbes, mais ils vont perdre énormément de temps pour comprendre le lexique de la phrase. Donc nous ajoutons en précision, que le public de ces exercices est surtout composé d'adultes. Mais nous croyons qu'en changeant certains critères, il est possible de rendre ces exercices faciles malgré tout. Par exemple cela peut donner des exercices très faciles:

- choisir une phrase où il y a un nom et un verbe
- pas de déterminant
- un verbe commun pour tous
- mettre un trou à la place de ce verbe

Puis, il est possible à faire niveau moyen avec ces critères :

- choisir une phrase où il y a au moins un déterminant
- il y a un mot commun facile
- mettre un trou à la place d'un verbe concret

Ensuite, pour un niveau un peu plus élevé, il est possible de faire :

- choisir une phrase avec n'importe quel déterminant
- il y a que des verbes faciles
- mettre un trou à la place de ce verbe

Deuxièmement, nous nous demandons pourquoi les exercices à trous ? S'il n'y a presque pas d'indications sur les verbes, les apprenants vont parfois tenter de les remplir par des synonymes, le sens de la phrase sera correct, mais à la fin il vont avoir un mauvais résultat dans le test, et il n'y aura pas de justification crédible leur expliquant pourquoi leur solution est considérée comme erronée. Dans ce cas-là, pour ce type d'exercice sur les verbes, il vaudrait peut-être mieux utiliser un autre module de H5P qui est basé sur le principe du glisser/déposer. Cela aurait été plus clair pour les apprenants pour savoir quel verbe doit être choisi. D'un autre côté, nous n'avons jamais dit que ces exercices pouvaient servir à un examen, donc si un apprenant est prévenu qu'il faut utiliser des mots spécifiques, il va essayer plusieurs fois et cela lui permettra de réfléchir aux synonymes.

Ensuite, peut-être que les options et les caractéristiques du choix de verbe peuvent être ajustées individuellement avec un professeur, car ce n'est pas une approche universelle, ni parfaite.

Puis, les mots dans ces trous ne se remplissent pas automatiquement, ce qui rend encore plus compliqué l'exercice. Peut-être que le remplissage semi-automatique après le début du mot pourrait être plus encourageant pour un apprenant. Et cela lui permettrait de ne pas « marcher dans le vide ».

D'autre part, afin de générer des exercices à trous, il faudrait peut-être mieux avoir des suggestions de réponses qui pourraient être données à un apprenant cherchant à deviner le verbe et sa forme correcte dans une phrase à trous.

Donc nous voyons bien que les discussions et les perspectives sur ce sujet peuvent être très longues, mais le fait est que la générateur fonctionne, et il ne fonctionne pas uniquement pour la langue géorgienne, mais aussi pour n'importe quelle autre langue.

Enfin, nous nous rendons compte que dans la structuration ordinaire de l'apprentissage de langue, l'exercice est central, c'est une démarche didactique qui porte sur un savoir ou un savoir-faire identifié est c'est très délicat. *A priori*, s'il est bien formulé, il concerne la découverte, l'approfondissement et l'évaluation des contenus. Donc il n'est pas surprenant, et l'approche historique de la question (Savatovsky, 1998) ; (Jey, 1997) ; (Jey M. , 1998) montre que la critique peut être très vaste sur ce sujet.

C'est sans doute ce qui justifie l'importance de beaucoup de travail et de la formation sur cette question, surtout quand il s'agit de l'apprentissage d'une langue. Nous mettons en avant la complexité de la question de l'apprentissage des langues, car dans la littérature il est souvent considéré que le genre de l'exercice est moins problématique dans certaines disciplines (comme les mathématiques, physique), alors qu'en ce qui concerne l'enseignement/apprentissage de la grammaire, c'est moins le cas (Masseron, 2001) ; (Denizot, 2015a) ; (Denizot, L'exercice dans l'enseignement de la littérature, 2015b)

Conclusion

Notre but était de proposer une première version de générateur d'exercices en ligne pour l'apprentissage de la langue géorgienne. Nous visions à aider les utilisateurs (des enseignants du géorgien, avec des connaissances informatiques limitées ou quasi-inexistantes) sur la plateforme Moodle, à manipuler des concepts et des notions qui ne font, *à priori*, pas partie de leurs problématiques, n'étant pas en lien avec l'apprentissage de la langue. Les enseignants du géorgien au lieu d'exprimer des réponses pédagogiques à l'aide d'outils pour leur discipline, sont obligés de chercher des solutions informatiques qui les font souvent s'éloigner, voire abandonner des solutions pédagogiques en informatique. Cela parce qu'ils ne savent pas les exprimer en termes informatiques. L'informatique, avec ses propres contraintes, a la capacité de fournir différents outils pour résoudre cette problématique. L'appel à d'autres connaissances en dehors de la linguistique et de la didactique est indispensable. L'utilisation des procédures du TAL peut apporter des solutions pour les professeurs de géorgien. Le travail en collaboration

entre didacticiens de la langue et spécialistes du TAL devrait apporter toutes les réponses pour ce champ encore vide.

Nous ne prétendons pas que ce travail est une solution irréprochable, et évidemment il peut toujours être modifié ou ajusté avec le temps et cela était notre but : apporter une modeste contribution à la didactique de l'apprentissage du géorgien pour les francophones, qui ouvrira la porte à d'autres projets de TAL entre les deux pays.

Annexe 1

Map de Géorgie ²⁷



²⁷ Les photos sont téléchargées de <https://svegeorgie.wordpress.com/georgie-ce-drole-de-pays/la-georgie-cest-ou/>

Annexe 2

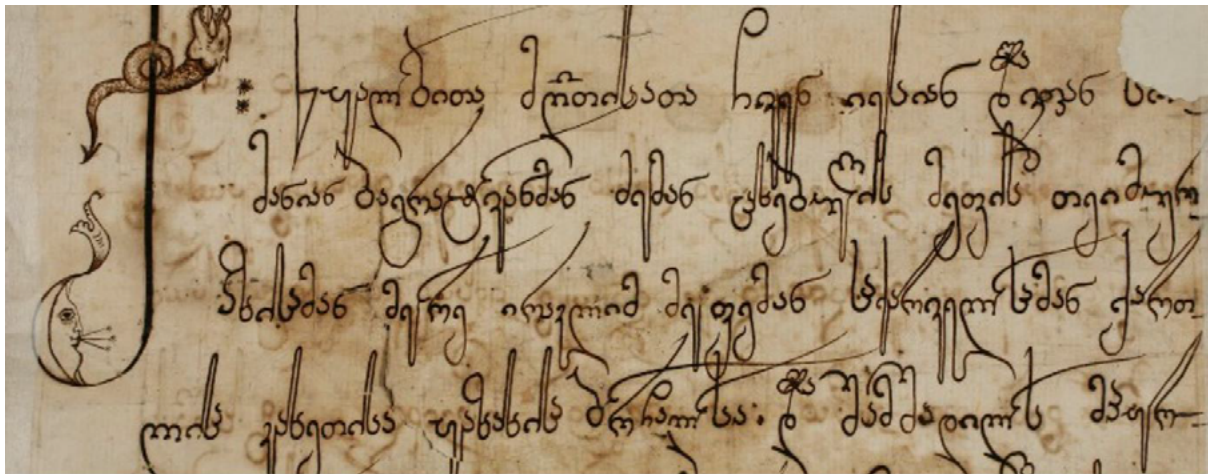
Alphabet géorgien²⁸

ა	ბ	გ	დ	ე	ვ	ზ	თ	ი	კ	ლ
a	b	g	d	e	v	z	t	i	k'	l
[a]	[b]	[g]	[d]	[e]	[v]	[z]	[t ^h]	[i]	[k']	[l]
მ	ნ	ო	პ	ჟ	რ	ს	ტ	უ	ფ	კ
m	n	o	p'	zh	r	s	t'	u	p	k
[m]	[n]	[o]	[p']	[ʒ]	[r]	[s]	[t']	[u]	[p ^h]	[k ^h]
ყ	ყ	შ	ჩ	ც	ძ	წ	ჭ	ხ	ჯ	ჰ
gh	q'	sh	ch	ts	dz	ts'	ch'	kh	j	h
[ɣ]	[q']	[ʃ]	[tʰ]	[ts ^h]	[dz]	[ts']	[tʃ']	[x]	[dʒ]	[h]

²⁸ <https://www.itinari.com/fr/georgian-script-one-of-the-oldest-in-the-world-rnuv>

Annexe 3

a) Exemple d'un manuscrit en géorgien²⁹



b) Extrait d'un livre géorgien³⁰



²⁹ <https://manuscript.ge/>

³⁰ <http://chqarosnebi.blogspot.com/2017/09/blog-post.html>

Annexe 4

a)selectcorpus.py

```
selectcorpus.py
1  #!/usr/bin/env python3
2  # coding=utf-8
3
4  import csv
5  import re
6  import argparse
7  import sys
8
9  parser = argparse.ArgumentParser()
10 parser.add_argument('tsv', help='input tsv file')
11 parser.add_argument('txt', help='output txt file with stars (blanks)')
12 parser.add_argument('-n', '--number', help="number of sentences", type=int, default=20)
13 parser.add_argument('-mns', '--minsize', help="minum size of sentences", type=int, default=10)
14 parser.add_argument('-maxs', '--maxsize', help="maximum size of sentences", type=int, default=20)
15 parser.add_argument('-pos', '--posselect', help="maximum size of sentences", type=str, default=None)
16 parser.add_argument('-lem', '--lemlst', help="maximum size of sentences", type=str, default=None)
17 args = parser.parse_args()
18 # print(args.number)
19
20 file = open(args.tsv)
21 reader = csv.reader(file, delimiter = '\t')
22
23 def cleanhtml(raw_html):
24     cleanr = re.compile('<.*>')
25     cleantext = re.sub(cleanr, '', raw_html)
26     cleantext = cleantext.strip()
27     return cleantext
28
29 def terminate(sents, outfile):
30     print('Number of sentences: '+str(len(sents)))
31     with open(outfile, 'w') as txtout:
32         for sent in sents:
33             txtout.write(sent+'\n')
34     txtout.close()
35     sys.exit(0)
36
37 len_phrase = 0 #taille de la phrase
38 starred = None
39 tokens = '' #debut De la phrase
40 lemlst = None
41 if args.lemlst:
42     lemlst = args.lemlst.split(',')
43
44 numbersent=0
45 sentsout = []
46 exit = False
47 for ligne in reader:
48     if not exit:
49         #print(ligne[0])
50         #print(ligne[2])
51         tk = ligne[0]
52         lem = ligne[1][1:-1]
53         if not starred:
54             if args.posselect and ligne[2] == args.posselect:
55                 starred = True
56                 tk = '*' +tk+ '*'
57             elif lemlst and lem in lemlst:
58                 starred = True
59                 tk = '*' +tk+ '*'
60         tokens += tk+' '
61         len_phrase += 1
62         if ligne[2] == 'Punct' and lem in '.!?':
63             if not args.posselect and not lemlst:
64                 starred = True
65             if starred and (not args.minsize or args.minsize <= len_phrase) and (not args.maxsize or len_phrase <= args.maxsize) :
66                 if not args.number or len(sentsout) < args.number:
67                     sentsout.append(cleanhtml(tokens))
68             else:
69                 terminate(sentsout, args.txt)
70         len_phrase = 0 #taille de la phrase
71         starred = None
72         tokens = '' #debut De la phrase
73
74 terminate(sentsout, args.txt)
75
```

Line 8, Column 1

Tab Size: 4

Python

b)freq.py

```
freq.py
1  #!/usr/bin/env python3
2  # coding=utf-8
3
4
5
6  import sys
7  import re
8
9  file = open("tavisupleba.txt", "r")
10 #content = file.read()
11
12 listV_counted = {}
13 tFaciles = []
14 faciles = []
15 difficiles = []
16 tDifficiles = []
17
18 #listVerbs = content.split()
19 #file.close()
20
21 for phrase in file:
22     #print(phrase)
23     matchV = re.search(r"\*(.*)\*", phrase)
24     if matchV :
25         #print("Verbs : ", matchV.group(1))
26         verb=matchV.group(1)
27         if verb in listV_counted:
28             listV_counted[verb] += 1
29         else:
30             listV_counted[verb] = 1
31
32
33 listV_sorted = sorted(listV_counted.items(),key=lambda item: item[1],reverse=True)
34
35 for verb,count in listV_sorted:
36     print(verb,count)
37
38
```

c) feedblanks.py

```
feedblanks.py x
1 |!usr/bin/env python3
2 # coding=utf-8
3 import sys, zipfile, json, tempfile, os
4 import argparse
5
6 parser = argparse.ArgumentParser()
7 parser.add_argument('h5pfile', help='h5p file')
8 parser.add_argument('sentsfile', help='sentences file')
9 args = parser.parse_args()
10 h5pfilename = args.h5pfile
11 tmpfolder = tempfile.TemporaryDirectory()
12
13 print('Unzipping H5P file: ' + h5pfilename)
14 with zipfile.ZipFile(h5pfilename) as h5pzipfile:
15     h5pzipfile.extractall(tmpfolder.name)
16
17 # Changing name of exercise
18 with open(tmpfolder.name + '/h5p.json') as h5pfile:
19     h5pjson = json.load(h5pfile)
20     h5pjson['title'] = h5pjson['title'] + ' - generated'
21     h5pfile.close()
22     with open(tmpfolder.name + '/h5p.json', 'w') as h5pfile:
23         json.dump(h5pjson, h5pfile)
24
25 with open(tmpfolder.name + '/content/content.json') as contentfile:
26
27     print('Opening json content')
28     contentjson = json.load(contentfile)
29     contentfile.close()
30
31     print('Adding questions')
32     questions = contentjson['questions']
33
34     # ICI ON AJOUTE UNE OU PLUSIEURS (50) QUESTIONS DEPUIS LE CORPUS
35     with open(args.sentsfile, 'r') as sentsfile:
36         for phrase in sentsfile:
37             phrase = phrase.strip()
38             print(phrase)
39             questions.append(phrase)
40
41     with open(tmpfolder.name + '/content/content.json', 'w') as contentfile:
42         # print(contentjson)
43         json.dump(contentjson, contentfile)
44         contentfile.close()
45
46     print('Zipping file')
47     h5pfilenameout = h5pfilename[:-4] + '-generated' + '.h5p'
48     with zipfile.ZipFile(h5pfilenameout, 'w') as h5pfilenameoutzip:
49         for dirname, subdirs, files in os.walk(tmpfolder.name):
50             # h5pfilenameout.write(dirname, os.path.relpath(dirname, tmpfolder.name))
51             for filename in files:
52                 h5pfilenameoutzip.write(os.path.join(dirname, filename), os.path.relpath(os.path.join(dirname, filename), tmpfolder.name))
53
54
55 Line 1, Column 1 Tab Size: 4 Python
```

Bibliographie :

Livres

Arnauld Chikobava `სახელის ფუძის უძველესი აგებულება ქართველურ ენებში", თბ., 1942.

KARTTUNEN, FA 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. . (1986).

Heinz Fähnrich "Grammatik der altgeorgischen Sprache" 1994. VIII, 270 Seiten. 978-3-87548-065-8. Gebunden

Kakha Gabunia "მარტივი წინადადების ანალიზის პრინციპები თანამედროვე ქართულში" 2016 გამომცემელი: „სამოქალაქო ინტეგრაციისა და ეროვნებათშორისი ურთიერთობების ცენტრი“ ISBN 978-9941-0-8554-3

Legendre R., Dictionnaire actuel de l'éducation, Montréal : Guérin, 2005, 3e éd.

Lebrun Marcel., 2007a, 2e éd., Théories et méthodes pédagogiques pour enseigner et apprendre : quelle place pour les TIC dans l'éducation ? Bruxelles : De Boeck Université.

NERBONNE, J. (2003). Natural language processing in computer-assisted language learning. In MITKOV, R., éditeur : The Oxford Handbook of Computational Linguistics. Oxford University Press.

Shanidze Akaki ქართული გრამატიკის საფუძვლები, მორფოლოგია, ტომი I, თბილისი 1953

Shanidze Akaki ქართული ენის გრამატიკა: მორფოლოგია. I 1962

ZOCK, M. 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. (1996)

Articles :

Ioana-Paula ARMĂȘAR « QUELQUES TYPES D'EXERCICES DANS LE COURS DE F.L.E.» Bulletin of the Transilvania University of Brașov • Series IV • Vol. 6 (55) No. 2 - 2013

Bourda Y., 2001, « Objets pédagogiques, vous avez dit objets pédagogiques? », Cahiers GUTenberg, N°39-40 - Mai 2001, pp. 71-79.

Bernard COULIE, Bastien KINDT et Tamar PATARIDZE, Lemmatisation automatique des sources en géorgien ancien, dans Le Muséon : revue d'études orientales, 126/1-2 (2013), p. 161-201.

Jean-Pierre Cuq, Isabelle Gruca « Cours de didactique du français langue étrangère et seconde » Collection Fle Français langue étrangère et maternelle Français langue étrangère Presses universitaires de Grenoble, 2005

Averil Coxhead "A New Academic Word List" Victoria University of Wellington Wellington, New Zealand 2000 et Paul Nation "Learning Vocabulary in Another Language" Cambridge

University Press Online publication date: October 2012 Print publication year: 2001 Online ISBN: 9781139524759 DOI:

Denizot, N. (2015a). Un exercice peut-il être un contenu disciplinaire ? Le cas de la dissertation. Dans B. Daunay, C. Fluckiger et R. Hassan (dir.), *Contenus d'enseignement et d'apprentissage. Approches didactiques* (p. 45-58). Pessac : Presses universitaires de Bordeaux.

Denizot, N. (2015b). L'exercice dans l'enseignement de la littérature. Dans C. Masseron, J.-M. Privat et Y. Reuter (dir.), *Littérature, linguistique et didactique du français. Les travaux pratiques d'André Petitjean* (p. 107-115). Villeneuve-d'Ascq : Presses universitaires du Septentrion.

Hélène Gérardin « ზმნური მორფოსინტაქსის სწავლების საკითხები: ორი სირთულე და რამდენიმე შეთავაზება » Paris 2015 National Institute of Oriental Languages and Civilizations

Hélène Gérardin. Les verbes intransitifs primaires et dérivés en géorgien : description morphosyntaxique, sémantique et dérivationnelle. *Linguistique*. Université Sorbonne Paris Cité, 2016. Français. ffNNT : 2016USPCF025ff. fftel-02020819f <https://tel.archives-ouvertes.fr/tel-02020819/document>

Jey, M. (1997). La lecture comme exercice : sa place au tournant du siècle dernier. *Le français aujourd'hui*, 118, 11-18.

Jey, M. (1998). *La littérature au lycée : Invention d'une discipline (1880-1925)*. Metz : université de Metz.

Oleg Kapanadze, Gideon Kotzé, Thomas Hanneforth "Building Resources for Georgian Treebanking-based NLP" June 2020 Conference: TbiLLC2019Proc Project: Bootstrapping a Probabilistic Context-Free Grammar from a Georgian Syntactically Annotated TeeBank.

Masseron, C. (2001). Note critique sur les exercices de grammaire. *Pratiques*, 111/112, 209-236.

DOI : [10.3406/prati.2001.2433](https://doi.org/10.3406/prati.2001.2433)

Pierre Martinez "La problématique de l'enseignement" dans *La didactique des langues étrangères* (2017), pages 9 à 45

Damana Mélikishvili "ქართული ზმნის სისტემური მორფო- სინტაქსური ანალიზი" 2014 თბილისი

Paul Meurer: *Morphosyntactic Analysis of Georgian* Uni Research Computing, Bergen, Norway 2012

Pilan, Volodina, Johansson "Rule-based and machine learning approaches for second language sentence-level readability" *Sprakbanken*, University of Gothenburg ° Box 200, Gothenburg, 2014 Sweden

Vivès Robert. *Lexique-grammaire et didactique du français langue étrangère*. In: *Langue française*, n°68, 1985. *Descriptions pour le français langue étrangère*, sous la direction de Jean-Claude Beacco. pp. 48-65

Savatovsky, D. (1998). *Les exercices de grammaire à l'agrégation. Éléments d'une mémoire disciplinaire (1852-1869)*. *Langue française*, 117, 36-50.

Oscar Valenzuela « *La didactique des langues étrangères et les processus d'enseignement/apprentissage* » Université de Playa Ancha (Valparaiso-Chili) 2010

Michèle Verdelhan-Bourgade *Chapitre IV. Problèmes d'une didactique spécifique Dans Le Français de scolarisation* (2002), pages 63 à 78

Mémoires et travail des doctorants

Hélène Gérardin « Contribution à l'étude des structures réfléchies en géorgien 2011 (Mémoire M1)

Davit Mishelashvili დისერტაცია ინსტრუმენტული საშუალებები ბუნებრივი ენის ტექსტების სინტაქსური და მორფოლოგიური ანალიზატორის შესადგენად 2006

Ivan Šmilauer « Acquisition du tchèque par les francophones : analyse automatique des erreurs de déclinaison » Thèse de doctorat 2008 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

Marie-Louisa Uwizeye "Les exercices de grammaire et le développement de la compétence communicative : une typologie" Våren 2011 UNIVERSITETET I OSLO (mémoire M II)

Site d'internet

Le projet «Armazi, Caucasian Languages and Cultures»: <http://armazi.uni-frankfurt.de/framee.htm>.

Le Figaro(lien vers l'article cité) <https://www.lefigaro.fr/flash-actu/castaner-veut-lutter-contre-l-anomalie-de-la-demande-d-asile-georgienne-20190510>

Georgien National Corpus (GNC) <http://gnc.gov.ge/gnc/page?page-id=gnc-main-page>

Hot potatoes <http://hotpot.uvic.ca/>

H5P <https://h5p.org/>

Sparklis

<http://www.irisa.fr/LIS/ferre/sparklis/?title=%0A%09%09%20%20Core%20English%20DBpedia%0A%09%09&endpoint=http%3A//servolis.irisa.fr/dbpedia/sparql>

GeoStat http://www.geostat.ge/?action=page&p_id=472&lang=geo

Regard-EST <http://regard-est.com/georgie-une-langue-et-une-nation-en-peril-depuis-des-millennaires>

Etalab <https://sill.etalab.gouv.fr/fr/software>

The culture trip <https://theculturetrip.com/asia/japan/articles/the-10-most-difficult-languages-in-the-world/>

Table des matières

Résumé – 5

Résumé en géorgien - 6

Introduction - 8

Partie 1. L'état de l'art - 10

1.1. Présentation de la sociolinguistique géorgienne - 11

1.2. Présentation de la structure grammaticale - 12

1.3. La didactique des langues par ELAO avec aide du TAL. Dès professeurs au tableau jusqu'au professeurs sur ordinateurs - 15

1.4. Traitement automatique des langues pour le géorgien - 17

Partie 2. Problématique et architecture - 18

2.1 Problème de la didactique du géorgien - 18

2.2 .Aspects théoriques – différentes méthodes en ligne et choix de l'analyse - 20

2. 3. Aspects techniques de l'apprentissage - l'architecture de l'exercice
les générateurs d'exercices - 25

Partie 3 Résultats et mise en pratique - 27

3.1. Implémentations et mises en pratique - 27

3.2. Résultats du travail – 30

3.3. Discussions et perspectives - 35

Conclusion – 37