Atelier d’écriture scientifique

Y4GET732

Abréviations :

L1 Locuteurs dont la langue maternelle coïncide avec la langue dominante du pays où ils résident. L2 Locuteurs ayant appris une langue donnée comme langue étrangèreLH Langue d’héritageLLH Locuteur de langue d’héritage LLHT Locuteur de langue d’héritage turque

Paragraphe état de l’art :

Penchons-nous à présent sur les compétences des LLH. De manière générale, les domaines où les locuteurs de LH ressemblent le plus aux L1 sont la phonologie et la phonétique. Être en contact avec une langue assez tôt durant l’enfance semble être un avantage non négligeable lors de la production, de la perception et de la catégorisation des sons : lorsque les compétences phonologiques et phonétiques des LHL sont comparées à celles des L2, celles de la première catégorie de locuteurs sont significativement plus similaires aux compétences des L1 (Montrul 2010, Oh et. al. 2003). Toutefois, malgré leurs compétences phonologiques et phonétiques assez semblables à celles des L1, des traits propres à la langue dominante des LLH, ou ce que Polinsky (2018b : 116-123) qualifie d’ « accent d’héritage » font que les LLH sont assez facilement identifiés par les L1 comme n’étant pas des L1 (Polinsky & Scontras 2019). Il est également évident que la langue du locuteur sera analogue à la langue à laquelle il a été exposé. Ainsi, les parents du LLH font eux-mêmes partie de cette catégorie de locuteurs et/ou s’ils parlent un dialecte bien particulier, la langue du LLH peut se distinguer des variétés contemporaines et normatives. Les différences entres les traits phonologiques du LLH et celles des L1 se fera encore plus ressentir (Bullock 2009)

Bibliograhie :

Bullock, B. E. (2009). Prosody in contact in French: A case study from a heritage variety in the USA. International Journal of Bilingualism, 13(2), 165–194. Montrul, S. (2010) “Current Issues in Heritage Language Acquisition,” Annual Review of Applied Linguistics, Cambridge University Press, 30, 3–23. Oh, J. et. al., (2003). Holding on to childhood language memory. Cognition, 86(3), B53- B64. Polinsky M. (2018b). Phonetics and Phonology. Dans Heritage Languages and their Speakers, Cambridge: Cambridge University Press: 114–163. Polinsky, M., & Scontras, G. (2019). Understanding heritage languages. Bilingualism: Language and Cognition, 23(1), 4-20.

Notre travail s’est axé autour des objectifs suivants :

* Sélectionner les outils numériques d’aide à l’écriture en L2 ;
* - Élaborer un dispositif didactique sous forme de padlets pour préparer les élèves à l’activité de production écrite ;
* - Identifier les usages diversifiés des traducteurs automatiques par les lycéens grâce à un corpus textuel recueilli ;
* - Dresser un état des lieux de la fréquence d’utilisation des traducteurs en ligne dans chaque étape de la production écrite ;
* - Analyser les pratiques observées de ces élèves afin de leur proposer une meilleure utilisation de Google Traduction et de développer de nouvelles littératies numériques et académiques chez eux ;
* - Analyser les textes produits sans et avec le traducteur Google Traduction pour identifier les apports des outils d’aide à la traduction.

"Pymorthy3 est une librairie gratuite et Open-Source de traitement automatique de la langue russe et ukrainienne. Cette librairie fournit principalement des fonctions d’analyse morphologique, notamment la lemmatisation des mots et la détection de catégories grammaticales tel que le nombre, le genre, le cas, la partie du discours, etc.

Le développement de la première version de pymorthy a commencé en 2009 . L'analyseur pymorthy3 est implémenté en langage de programmation Python avec des extensions C++ qui rendent le fonctionnement plus rapide. Pour le russe pymorthy est basé sur le dictionnaire d’OpenCorpora . Ce dictionnaire source contient 5 000 000 formes de mots et 390 000 lemmes avec leurs étiquettes, regroupées par des lexèmes. Un lexème est constitué de toutes les formes fléchies d'un mot, avec des informations grammaticales pour chaque forme.

Dans pymorphy3 le dictionnaire est converti du format XML en une représentation compacte optimisée pour les tâches d'analyse morphologiques. Pour économiser de la mémoire vive et permettre des analyses rapides des mots qui sont présents dans le dictionnaire, ainsi que des mots inconnus, pymorphy3 extrait les paradigmes des lexèmes, convertit l'information en chiffres et encode les mots en DAFSA – Deterministic Acyclic Finite State Automaton (Daciuk, 2000). Le paradigme dans pymorphy est un modèle de déclinaison d'un lexème. Pour trouver des paradigmes pymorphy détermine les préfixes, les radicaux et les suffixes des mots, puis les encode en tableaux de nombres pour un stockage compact.

Selon (Korobov, 2015), il n'est pas pratique d'inclure tous les mots dans le dictionnaire, car il existe une longue liste de mots rarement utilisés, d’entités nommées, de termes spécifiques et de nouveaux mots. Il suffit d'avoir des paradigmes de mots et de savoir par quel modèle un mot doit se conjuguer ou se décliner.

1. Daciuk, J., Watson, B.W., Mihov, S., Watson, R.E. 2000. Incremental Construction of Minimal Acyclic Finite-State Automata. Computational Linguistics.

2. Korobov M.2015. Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts, pp 320-332.

1. https://github.com/pymorphy2/pymorphy2

2. http://opencorpora.org/"

Le TAL (Traitement automatique des langues naturelles) est une branche dans le domaine de l’intelligence artificielle, il cherche à faire l’ordinateur comprendre les langues humaines. À mesure que le nombre de personnes utilisant Internet augmente, de plus en plus d’internautes diffusent des opinions, des commentaires sur les réseaux sociaux, tels que Weibo. À l’aide du TAL, nous pouvons analyser le sentiment des utilisateurs des plateformes. Pour les entreprises, le recours à l’analyse de sentiment permet de connaître les préférées de leurs clients et d’aider les entreprises à lancer de nouveaux produits et à améliorer leurs services; dans le domaine de la psychothérapie, l’analyse de sentiment pourrait servir à reconnaître des joies et des peines des utilisateurs à partir des textes écrits par les utilisateurs, l’ordinateur serait un compagnon émotionnel pour les personnes souffrant de maladies mentales.

"En outre, l’analyse de la communication dans les jeux vidéo, et en particulier dans CSGO, a également fait l’objet de plusieurs recherches. Cink (2020) et Oksala (2022) ont notamment proposé une analyse de la communication à l’intérieur du jeu soulignant l’importance de la clarté, de la concision, de la rapidité et de la précision. Oksala (2022) a mené des interviews auprès de joueurs semi-professionnels de CSGO pour évaluer l’importance de la communication dans une équipe. Les joueurs ont indiqué que la description des actions, la connaissance des positions spécifiques de chaque carte et le fait de ne pas parler en même temps étaient essentiels pour une communication efficace dans CSGO.

Cink (2020) a appuyé ses résultats en envoyant des questionnaires à des joueurs lambda de CSGO. Avec les réponses des questionnaires, Cink (2020) a pu conclure que la communication sur CSGO consiste souvent en une simple phrase, voire un seul mot, ce qui permet à la communication d’être aussi rapide que le jeu l’est. Un participant lui indique ceci : « La communication dans CSGO est une question d'efficacité et de mots faciles à prononcer que tout le monde peut comprendre. Rendre la communication efficace et savoir quand parler est probablement la chose la plus difficile, car la communication peut aussi bloquer directement les sons du jeu. » . La communication est donc quelque chose qui doit être simple, rapide et intelligent.

La communication dans CSGO nécessite la maîtrise de termes spécifiques, la plupart étant en anglais. Flores (2021) a ainsi étudié le phénomène du code-switching dans CSGO, qui oblige les joueurs non anglophones à apprendre des mots anglais pour pouvoir communiquer efficacement. Poplack (1980) défini le code-switching comme l’alternance de deux langages au sein d’un même discours, d’une même phrase ou d’un même constituant. Dans son étude, Flores (2021) distingue deux types de code-switching : l’utilisation des termes anglais officiels et la transformation des termes anglais en leur équivalent dans la langue native du joueur. Flores qualifie ce dernier phénomène « d’aportuguesamento » (portugaisation), où les joueurs modifient les termes anglais pour les rendre plus compréhensibles. On peut parler de l’exemple du verbe « to smoke » (lancer un fumigène) en anglais, qui, en portugais, se transforme en le verbe « smokar » qui décrit l’action de lancer une grenade fumigène, ou encore de l’exemple du verbe « to flash » (lancer une grenade flash) en anglais qui en portugais se transforme en « flashar » qui décrit donc l’action de lancer une grenade flash. Nous pouvons souligner que ce phénomène est aussi présent en français, « to smoke » devient « smoker » et « to flash » devient « flasher ».

Enfin, ces termes spécifiques, parfois, ont un autre sens que celui que l’on trouve dans les dictionnaires classiques. Sund (2020) propose l’analyse des sens des termes spécifiques dans CSGO. Il propose trois catégories de sens : le « Contextual meaning » qui concerne les termes qui possèdent le même sens dans un dictionnaire classique et dans le jeu, comme « rank », « bait » ou encore « kill » qui signifient respectivement « rang », « appâter » et « élimination ». Ces termes possèdent le même sens dans les deux cas. Le « New meaning » (nouveau sens) qui concerne les mots existant dans les dictionnaires classiques, mais qui, dans CSGO possède un autre sens. Sund (2020) nous donne des termes comme « Spawn » qui en anglais classique signifie « pondre/frayer » et qui, dans CSGO définit l’endroit où apparaissent les joueurs de chaque équipe, on peut aussi parler du mot « long » qui en anglais classique signifie « long » mais qui dans CSGO définit un endroit spécifique de la carte, un joueur dira « Il est long ». Enfin, il définit la catégorie « New words » (nouveaux mots) qui concerne les termes créés spécifiquement pour le jeu et qui possèdent leur propre sens, Sund (2020) nous propose le terme « CT » qui est l’endroit où l’équipe qui défend, les « Anti-Terroristes », apparaît.

Bibliographie :

Flores, L. S. (2021). UNIVERSIDADE DE CAXIAS DO SUL ÁREA DE CONHECIMENTO DAS HUMANIDADES CURSO DE LICENCIATURA EM LETRAS - INGLÊS.

Oksala, A. (s. d.). The importance of communication in esports : An ethnographic interview with a Finnish Counter-Strike : Global Offensive team. 55.

Sund, J. (s. d.). Words and Meaning in Gaming : ‘World of Warcraft’ and ‘Counterstrike Global Offensive’.

Poplack, S. (1980). Sometimes I’ll start a sentence in Spanish Y TERMINO EN ESPAÑOL : Toward a typology of code-switching1. Linguistics, 18(7 8). https://doi.org/10.1515/ling.1980.18.7-8.581

"La Figure 7 présente la question posée à ChatGPT, et la réponse donnée par le système. Nous constatons alors que ChatGPT remet en cause la réponse à cette question, en y apportant des justifications qui ne sont pas nécessaires et qui portent à confusion.

En effet, ChatGPT explique que l’expression « reculer la jambe extérieure » n’est « pas précise ». ChatGPT poursuit et justifie cette explication en écrivant que la jambe extérieure sert à « maintenir l’impulsion du cheval dans le départ au galop ». Or, nous avons expliqué précédemment que ceci est faux, puisque c’est le rôle de la jambe intérieure. ChatGPT continue en ajoutant que la jambe extérieure « ne doit pas être reculée de manière excessive car cela peut nuire à la communication entre le cavalier et le cheval ». Cet ajout est non seulement incorrect, mais également pas nécessaire : ChatGPT ajoute des nuances là où il n’est pas question d’en ajouter. Un cavalier ne doit pas s’inquiéter de reculer sa jambe extérieure de manière excessive, mais plutôt de reculer suffisamment sa jambe pour permettre une bonne communication avec son cheval et éviter une incompréhension de la part de sa monture.

Même si la plupart des éléments de la réponse de ChatGPT sont incorrects ou imprécis, nous étudions un autre exemple. ChatGPT soutient que l’instruction « bout du nez à l’intérieur » est incorrecte, et explique que « le cheval doit être légèrement incurvé vers l’intérieur, mais pas au point de tourner complètement la tête ». Encore une fois, cette justification n’est pas nécessaire, car l’expression « bout du nez à l’intérieur » ne signifie pas que le cheval à la tête tournée ou tordue vers l’intérieur, mais plutôt que son encolure est incurvée vers l’intérieur et que sa tête est légèrement orientée vers l’intérieur. ChatGPT ajoute une nuance inexistante à la réponse du cavalier, et insinue donc que l’élément de réponse « bout du nez à l’intérieur » signifie que la tête du cheval est tournée, alors que ce n’est pas le cas.

Figure 7 : Capture d’écran de ChatGPT corrigeant une question d’examen théorique du Galop 3.

"

" Une autre de ces ressources numériques s’intitule DerivBase.RU, et elle a été développée par Vodolazsky en 2020. DerivBase.RU est un lexique dérivationnel numérique riche pour le russe. Il est composé de familles dérivationnelles, à savoirs des groupes de lexèmes dérivés. Ce lexique numérique russe a été inspiré par la version originale allemande DerivBase de Zeller et al (2013). DerivBase.RU (2020) se distingue par sa couverture exhaustive des schémas de dérivation russes et par son corpus de mots récents. La morphologie des mots russes est étudiée depuis des décennies, mais très peu d’entre elle ont établi un regroupement en familles de mots, en dehors des différentes éditions du dictionnaire de Tikhonov (1985, 2014) et Lopatin et Ulukhanov (2016).

DerivBase.Ru (2020) offre une vaste collection de lexèmes dérivés, classés en fonction de leurs racines et de leurs affixes. Par exemple, Volodazsky (2020) mentionne quelques exemples provenant de Shvedova (1980) à la page 3938 de « Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020) » :

«  – ukho (an ear) → ushko (a small ear) (velar— sibilant alternation),

– kosmos (space) → kosmicheskiy (space, adj.) (stem deletion),

– drama (a drama) → dramaticheskiy (dramatic) (stem epenthesis),  »

Dans cette extrait, Vodolazsky (2020) mentionne les procédés de formation des mots. Le nom « ухо » /oukho/ (« une oreille ») peut dériver en nom « ушко » /ouchko/ (« une petite oreille ») par l’alternance de la consonne vélaire « x » [х] et la consonne post-alvéolaire « ш » [ʃ].

Le substantif « космос » /kosmos/ (« l’espace ») peut former l’adjectif « космический » /kosmitcheskiy/ (« spatial ») par la suppression de « -ос » /-os/ et l’ajout du suffixe « -ический » /-itcheskiy/ permettant de créer un adjectif.

Le substantif « драма » /drama/ (« un drame ») peut former l’adjectif « драматический » /dramatitcheskiy/ (« dramatique ») par l’ajout du suffixe « тический » /-titcheskiy/ directement à la base « драма » /drama/.

Bibliographie :

Lopatin, V. V. et Ulukhanov, I. S., (2016), Slovar' slovoobrazovatel'nykh affiksov sovremnogo russkogo yazyka, Moscow, Publishing Center “Azbukovnikˮ, 812 pages

Shvedova, N., (1980), Russkaja grammatika, Number t.1, In : Russkaja grammatika, Izd-vo Nauka.

Tikhonov, A., (1985), Slovoobrazovatel’ny Slovar Russkogo Jazyka, Number t.1, In Slovoobrazovatel’ny Slovar Russkogo Jazyka, (Word-Formational Dictionary of Russian Language), M.: Russkij Jazyk.

Tikhonov, A., (2014), Novy Slovoobrazovatel’ny Slovar Russkogo Jazyka Dlya Vseh, Kto Hochet Byt’ Gramotnym, LitRes.

Vodolazsky, D., (2020), DerivBase.Ru: a Derivational Morphology Resource for Russian, In : Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), Marseille, 11–16 mai 2020, pp. 3937–3943.

Zeller, B., Snajder, J., et Pado, S., (2013), DErivBase: Inducing and evaluating a derivational morphology resource for German, In : Proceedings of ACL 2013, Sofia, Bulgarie, pages 1201–1211. »

"Dans le domaine de l'automatisation, il est possible de tirer parti des systèmes existants pour annoter notre propre corpus. Deux systèmes notables sont FrameNet, développé par Baker et al. (1998), et WordNet, conçu par Miller (1995).

FrameNet constitue une base de données lexicale permettant une annotation succincte de chaque mot, telle que sa catégorie grammaticale. Cette approche rend les mots plus ""lisibles"" pour les machines. FrameNet repose lui-même sur des corpus de phrases déjà annotées, ce qui renforce sa fiabilité et sa précision.

Quant à WordNet, il s'apparente à un dictionnaire offrant des annotations plus détaillées par rapport à FrameNet. Il fournit des informations approfondies sur chaque mot et ses synsets. Ainsi, WordNet répertorie pour un mot ses synonymes, ses antonymes, ses hyponymes, ses hypernymes, ses mémonymes et ses holonymes, élargissant ainsi la compréhension et la contextualisation des mots dans un langage naturel.

Miller G. A., “WordNet: A Lexical Database for English,” Communications of the ACM, Vol. 38, No. 11, 1995, pp. 39-41. http://dx.doi.org/10.1145/219717.219748

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. ""The Berkeley FrameNet Project"", In 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1, 1998, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics."