HAL
open science

DeepFLE : la plateforme pour évaluer le niveau d'un texte selon le CECRL

Simona Ruggia, Laurent Vanni

To cite this version:

Simona Ruggia, Laurent Vanni. DeepFLE : la plateforme pour évaluer le niveau d'un texte selon le CECRL. Dialogues et cultures, 2021, Dialogues et cultures, 66, pp.235-254. hal-03494844

HAL Id: hal-03494844

https://hal.science/hal-03494844v1

Submitted on 2 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DeepFLE :
la plateforme pour évaluer le niveau d'un texte selon le CECRL

Simona RUGGIA
Université Côte d'Azur, CNRS, BCL (UMR 7320)
Simona.Ruggia@univ-cotedazur.fr

Laurent VANNI
Université Côte d'Azur, CNRS, BCL (UMR 7320)
Laurent.Vanni@univ-cotedazur.fr

Résumé
Cette contribution présente la plateforme DeepFLE, un outil conçu pour tous les acteurs du français langue étrangère (FLE), qui est le résultat d'une recherche en cours, dont l'approche interdisciplinaire engage un dialogue entre la didactique du français langue étrangère (FLE), le *deep learning* et l'analyse des données textuelles (ADT). DeepFLE permet d'évaluer le niveau d'un texte en français selon les échelles du *Cadre européen commun de référence pour les langues* (CERCL). Plus précisément, cette plateforme propose une analyse à la fois prédictive et descriptive du niveau d'un texte grâce au modèle novateur de *deep learning* : le *Text Deconvolution Saliency* (TDS) (Vanni *et al.* 2018a ; 2018b ; 2021) qui opère une extraction automatique des saillances qui marquent un changement de niveau en distinguant les marqueurs qui contribuent le plus fortement à l'attribution à un niveau.
Mots-clés : didactique du FLE, *deep learning*, niveaux de langue, Cadre européen commun de référence pour les langues, évaluation du niveau d'un texte.

Introduction

Les questions d'évaluation et de description des niveaux de langue selon les échelles du *Cadre européen commun de référence pour les langues* (CERCL) (Conseil de l'Europe 2001 ; 2018) touchent tous les acteurs du français langue étrangère (FLE). Ainsi, dans le cadre de nos travaux de recherche, nous nous sommes demandé s'il était possible de créer un outil capable à la fois de prédire le niveau d'un texte et de fournir une analyse descriptive de ses spécificités.

Pour ce faire, nous menons depuis deux ans une recherche[1] qui fait dialoguer la didactique du FLE, le *deep learning* et l'analyse des données textuelles (ADT). Elle s'appuie sur un corpus numérique constitué de 6 classes[2] correspondant aux 6 niveaux du CERCL. Plus précisément, la didactique du FLE nous permet d'une part de distinguer, les niveaux de compétences selon « un axe horizontal des catégories qui décrivent les différentes activités et les aspects de la compétence[3] [à communiquer langagièrement] » (Conseil de l'Europe 2018 : 35) et « un axe vertical qui représente les progrès de la compétence dans ces

[1] Ce travail a bénéficié d'une aide du gouvernement français, gérée par l'Agence Nationale de la Recherche au titre du projet Investissements d'Avenir UCA^{JEDI} portant la référence n° ANR-15-IDEX-01.
[2] En *deep learning* on appelle « classes » les différentes parties du corpus à identifier.
[3] Ces derniers sont décrits sous forme de « je peux faire… ».

catégories » (*ibid.*) ; d'autre part, d'identifier les réalisations linguistiques correspondant à chaque niveau (de A1 à C2), listées par les *Référentiels pour le français*. En somme, en didactique du FLE les outils du Conseil de l'Europe et plus particulièrement les échelles de niveaux « sont autant utiles pour les trois activités qui sont citées dans les sous-titres du CERCL : apprendre, enseigner, évaluer » (Gouiller 2019 : 66).

Pourquoi alors faire appel au *deep learning* et aux méthodes statistiques ? Parce que les recherches en *deep learning*, qui est une technologie d'apprentissage et de classification basée sur des réseaux de neurones artificiels, ont récemment conduit à des résultats très satisfaisants en matière de prédiction et surtout de description des classes de textes. Ainsi, nous avons exploité un modèle novateur de *deep learning* : le *Text Deconvolution Saliency* (TDS) (Vanni *et al.* 2018a ; 2018b ; 2020)[4] qui « implémente l'analyse prédictive du *deep learning* à l'analyse descriptive grâce à une extraction des passages-clés » (Ruggia 2019 : 83) en fournissant « une évaluation de leur pertinence interprétative » (Vanni *et al*. 2018b : 460). Par ailleurs, l'ADT permet à la fois une analyse qualitative et quantitative de corpus numériques et a une finalité interprétative car, en statistique comme en *deep learning*, « les résultats […] ne valent que s'ils sont interprétés » (Lebart *et al*. 2019 : 18). Cette méthode nous a permis de comparer la distribution statistique des marqueurs détectés par le *deep learning* ainsi que celle des réalisations linguistiques des inventaires des *Référentiels pour le français* en fonction des niveaux. Les premiers résultats de cette recherche nous ont amenés à réaliser la plateforme DeepFLE.

1. Cadre méthodologique

L'approche méthodologique que nous adoptons s'inscrit dans les axes de recherche de notre équipe « Logométrie. Corpus, Traitements, Modèles » qui travaille des corpus textuels grâce à des traitements linguistiques qui ont recours à l'informatique et à la statistique textuelle. Plus particulièrement, cette approche est le pivot de l'axe 3 : « Corpus et didactique des langues », dont l'originalité des travaux menés tient à la mobilisation des méthodes d'exploration de corpus particulières et innovantes que développe collectivement l'ensemble des membres de l'équipe. Il s'agit d'une approche tout à fait singulière et novatrice dans le champ de la didactique des langues.

Notre projet de recherche est né de l'hypothèse suivante : « le *Text Deconvolution Saliency* est capable d'extraire les caractéristiques de textes en français et il est plus précisément capable d'extraire les saillances qui marquent un changement de niveau selon le CERCL » (Ruggia 2019 : 82). Cette hypothèse nous a paru pertinente car le TDS est un modèle de *deep learning* capable d'extraire les caractéristiques qui donnent une empreinte unique du texte. En d'autres termes, c'est une application qui détecte les saillances du texte à différents niveaux linguistiques (lexique, grammaire, morphosyntaxe, etc.) grâce au travail préalable de lemmatisation[5].

---

[4] Le TDS a été développé par Laurent Vanni, ingénieur du CNRS au sein de l'équipe « Logométrie. Corpus, Traitements, Modèles » de notre laboratoire « Bases, corpus et langage » UCA, CNRS, UMR 7320, France. Le TDS est exploité par plusieurs membres de notre équipe et deux autres plateformes ont été réalisées : la première, « Mesure du discours » (<http://mesure-du-discours.unice.fr>), est un observatoire du discours politique français outillé par la statistique. La deuxième, « Deeptext » (<http://deeptext.unice.fr>), permet aux utilisateurs de tester la fonction de prédiction de reconnaissance du lexique des sentiments d'un texte en anglais, de l'auteur d'un texte en latin et de la tendance politique (gauche ou droite) d'un texte en français.

[5] Effectuée avec TREE TAGGER.

Le TDS est établi à partir d'une base de données textuelles[6] qui constitue le corpus d'entraînement nécessaire pour l'apprentissage profond. Pour que le TDS atteigne un niveau de précision suffisant, il faut que le corpus soit homogène en taille et composé d'une masse critique de données pour chaque classe, soit 100 000 occurrences[7] minimum. Le modèle TDS est implémenté dans la plateforme d'analyse statistique des données textuelles Hyperbase web[8] et se nomme *hyperdeep*. À l'aide d'Hyperbase web, le chercheur peut charger un corpus, entrainer un modèle[9] et lancer des analyses prédictives et descriptives en soumettant un texte au système. Cette dernière fonctionnalité permet de visualiser le pourcentage de reconnaissance ainsi que les saillances détectées pour la reconnaissance, en d'autres termes les occurrences jugées comme étant les plus pertinentes pour la reconnaissance.
Par extension, DeepFLE est une implémentation de la fonction *hyperdeep* dédiée au FLE.

1.1. Le corpus d'apprentissage

L'élaboration d'un corpus d'entraînement est un travail long et méticuleux qui doit répondre à des critères bien définis. Ainsi, notre corpus est un corpus échantillonné, constitué de textes oraux (monologues et interactions[10]) issus de nombreux manuels d'ensembles didactiques de FLE[11] qui s'inscrivent dans l'approche actionnelle du CECRL, qui couvrent plusieurs niveaux, qui ont été publiés récemment[12] et aussi d'ouvrages conçus pour la préparation aux diplômes du DELF et du DALF. Nous avons ajouté ce dernier critère car la plupart des méthodes ne couvrent pas tous les niveaux.
Pour notre recherche, nous avons établi successivement trois corpus d'apprentissage. Tout d'abord, un corpus constitué de deux classes de niveaux : A1 et B2, ensuite un corpus de quatre classes de niveaux : A1, A2, B1 et B2, enfin un corpus de six classes de niveaux : A1, A2, B1, B2, C1, C2. L'étendue du troisième corpus est de 595 980 occurrences, et le tableau *infra* présente les occurrences de chaque classe :

NIVEAU CECRL	OCCURRENCES
A1	101923
A2	115030
B1	117949

[6] Le format des fichiers composant une base de données est le format « txt ».
[7] Les occurrences correspondent aux formes graphiques (mots) et aux ponctèmes (signes de ponctuation).
[8] Hyperbase web est l'évolution d'Hyperbase, un logiciel qui a vu le jour en 1989 grâce à l'œuvre d'Etienne Brunet au sein du laboratoire « Bases, Corpus et Langage », et qui permet de réaliser des bases hypertextuelles avec les textes qu'on lui fournit. Hyperbase, qui combine des fonctions documentaires et statistiques pour décrire, caractériser, classer et interpréter les textes, n'a cessé d'évoluer et depuis 2015 une version web a été développée par Laurent Vanni qui a aussi intégré, en 2017, la fonction prédictive du *deep learning* (<hyperbase.unice.fr>).
[9] L'entraînement du modèle nécessite une certaine expertise dans le domaine du *deep learning*. Dans le cas de notre étude l'entraînement a été effectué par Laurent Vanni.
[10] Nous avons d'abord établi un corpus de textes oraux et nous sommes en train d'établir un corpus de textes écrits.
[11] Voir bibliographie : Ensembles pédagogiques de FLE.
[12] Les ouvrages sélectionnés ont été publiés entre 2005 et 2018.

B2	97961
C1	99407
C2	97175

Tableau 1 : Étendue de chaque classe du corpus d'apprentissage

1.2. Les trois grandes étapes du projet de recherche

Afin de créer la plateforme DeepFLE, nous avons organisé notre recherche en trois grandes étapes correspondant à l'élaboration des trois corpus d'apprentissage décrits *supra* et à l'analyse des résultats de la prédiction et de la description du TDS en fonction de chaque corpus.

Ainsi, la première étape nous a permis de vérifier notre hypothèse de recherche à savoir que le *deep learning* est capable de classer efficacement les textes en se servant des marqueurs des niveaux du CECRL. Pour ce faire, nous avons établi un premier corpus constitué de deux classes de niveau, A1 et B2, et nous avons analysé et vérifié les résultats du TDS. D'abord, en mettant en correspondance les saillances détectées avec les inventaires du *Référentiel Niveau A1 pour le français* (Beacco et Porquier 2007) et avec ceux du niveau B2 (Beacco *et al.* 2004a). Ensuite, en cherchant la distribution statistique des marqueurs[13] le plus fortement attribués à une classe grâce à l'ADT[14]. Il est à noter que le TDS attribue aussi un taux de score de reconnaissance. Pour cette première étape, les taux étaient très élevés ; à titre d'exemple, 97,37 % pour un texte de niveau B2 et jusqu'à 100 % pour un texte de niveau A1. Dans cette première expérience, le corpus d'apprentissage était constitué par deux classes de niveaux très différents, ce qui explique le taux très élevé des scores.

La deuxième étape a permis de tester le TDS à l'aide d'un corpus de quatre classes de niveaux : A1, A2, B1 et B2, et la troisième à l'aide d'un corpus de six classes, couvrant tous les niveaux du CECRL[15]. Le protocole méthodologique adopté a été celui décrit *supra*, néanmoins il est à signaler que pour les deux premières étapes nous avons exploité la première version du TDS[16] et pour la troisième phase[17] la dernière version du TDS, à savoir le TDS pondéré qui « attribue un score à chaque mot (chaque *token*) pour chaque classe » (Vanni *et al.* 2021 : 7). Ainsi, le TDS de chaque mot ou *token* « peut être soit positif soit négatif selon la classe observée en sortie et en fonction du fait que le *token* a servi ou au contraire desservi cette classe » (*ibid.*)[18].

2. La plateforme DeepFLE

[13] Les marqueurs correspondent aux passages-clés détectés par le TDS. Un passage-clé correspond à « une unité de surcroît textométrique ; c'est-à-dire une unité dont la pertinence est calculable et l'extraction automatique » (Vanni *et al.* 2018b : 461).
[14] Ruggia 2019.
[15] Ruggia 2021.
[16] Vanni *et al.* 2018a.
[17] Vanni *et al.* 2021.
[18] Nous illustrerons cette fonction à l'aide du tableau de la Figure 4 dans les paragraphes suivants.

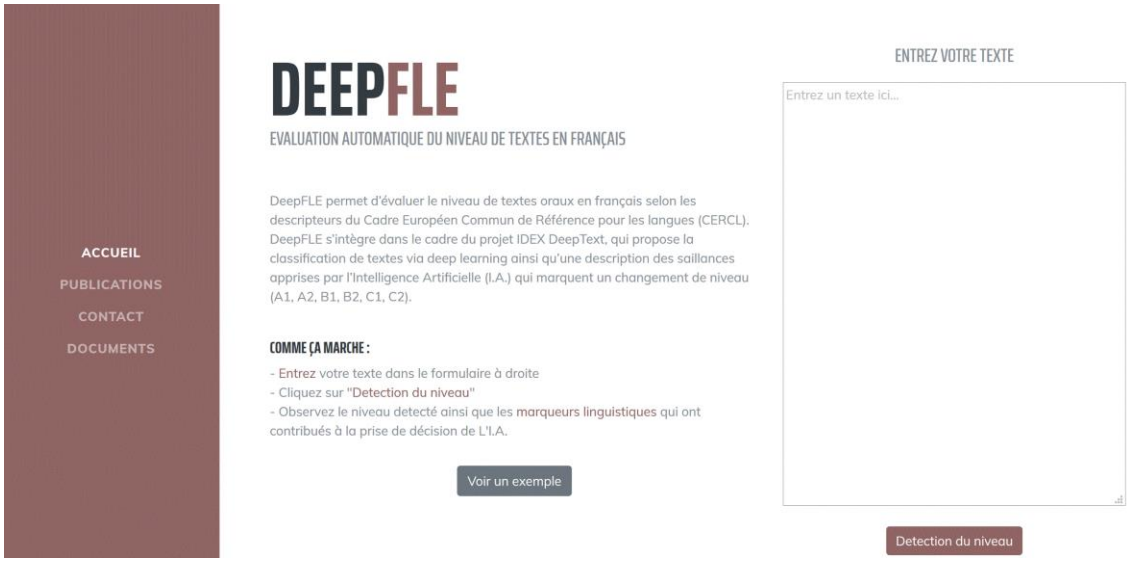

Figure 1 : Page d'accueil de la plateforme DeepFLE

DeepFLE[19] permet d'évaluer en quelques secondes le niveau d'un texte en français et de visualiser les résultats de l'analyse. Il suffit de copier/coller un texte dans la fenêtre « entrez votre texte » et de cliquer sur « détection du niveau ». Les utilisateurs peuvent aussi afficher un exemple d'analyse d'un texte prédéfini en cliquant sur « voir un exemple ». Par ailleurs, dans la page « publications » on peut télécharger des articles qui décrivent en détail le fonctionnement du TDS, ainsi que son utilisation avec divers types de corpus. Dans la page « documents » : on trouve l'échelle globale des niveaux du CECRL (Conseil de l'Europe 2001 : 25) ainsi que les grilles des objectifs concernant les fonctions, notions et catégories morphosyntaxiques de tous les niveaux (Beacco *et al*. 2004b : 33, 45-46, 59, 72-73, 84-85, 96).
Enfin, sous « contact » il y a les noms des membres du comité scientifique du projet : S. Ruggia, D. Mayaffre, L. Vanni.

2.1. Prédiction et description du niveau

[19] <http://deeptext.unice.fr/FLE>.

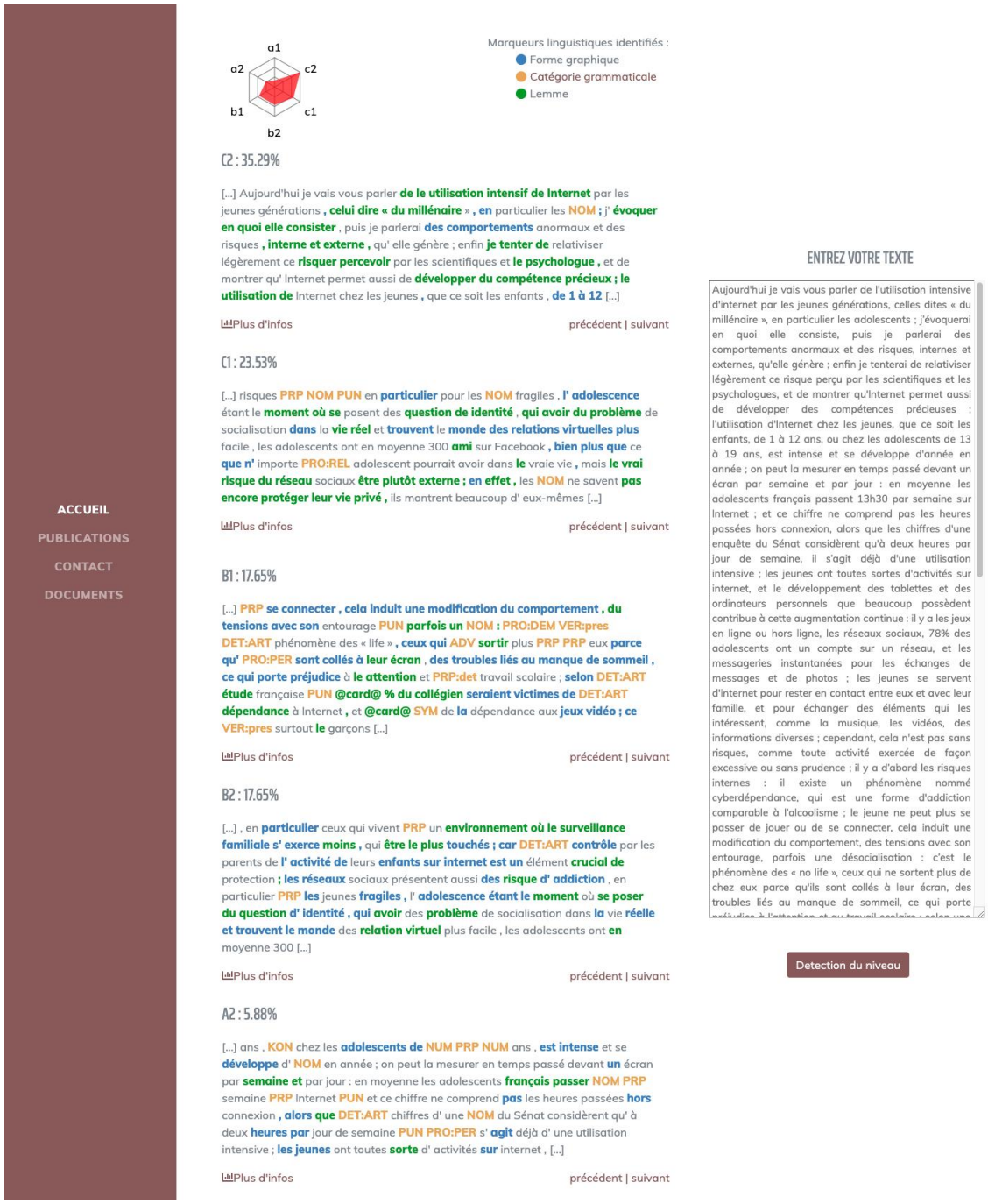

Figure 2 : Prédiction et description du niveau d'un texte

La Figure 2 illustre l'interface de la plateforme lorsqu'on demande la détection du niveau d'un texte. Le TDS opère une analyse en prenant en compte le cotexte et, pour chaque niveau détecté, affiche un score ainsi qu'un passage représentatif et, plus précisément, un passage-clé qui est « un morceau de texte jugé suffisamment parlant, notamment par sa taille qui gagne à dépasser le mot » (Vanni *et al*. 2018b : 460). Dans ce sens, un passage-clé correspond à « une unité de surcroît textométrique ; c'est-à-dire une unité dont la pertinence est calculable et l'extraction automatique » (Vanni *et al.* 2018b : 461).

Comment lire les résultats ? La prédiction du niveau est affichée aussi bien sous forme de diagramme type radar que de score de reconnaissance. Pour ce qui est du texte de la Figure 2 le système reconnaît 35,29 % de passages-clés de niveau C2 ; 23,53 % C1 ; 17,65 % B1 ; 17,65 B2 et 5,88 % A2. En d'autres mots, le système est capable de détecter les marqueurs spécifiques des différents niveaux du CECRL d'un texte, ce qui est particulièrement utile pour des textes longs de niveaux B et C. De plus, cette analyse permet de décrire l'organisation enchâssée des niveaux qui, comme l'illustre le digramme concentrique de la Figure 3, « sont inclus les uns dans les autres » (Beacco *et al.* 2008 : 15).

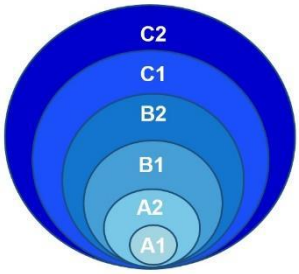

Figure 3 : Diagramme concentrique des niveaux du CECRL

En ce qui concerne l'analyse descriptive des passages-clés effectuée aussi grâce à la lemmatisation préalable du texte, les résultats sont visibles, grâce aux couleurs attribuées aux marqueurs linguistiques identifiés[20]. Ainsi, le bleu indique qu'il s'agit d'une occurrence que le système reconnaît en tant que mot, autrement dit une forme graphique, l'orange indique la catégorie grammaticale et des codes précisent le type de la catégorie (DET:ART : articles, VER : verbe…)[21], et le vert signale qu'il s'agit d'un lemme (en ce sens le système met en évidence la caractéristique la plus saillante : par exemple, un verbe à l'infinitif, un nom au singulier). L'utilisateur peut parcourir tous les passages-clés classés par niveau en cliquant sur « suivant » et revenir en arrière en cliquant sur « précédent ». Enfin, en cliquant sur « plus d'info » l'utilisateur peut visualiser un tableau du taux d'activation des marqueurs que nous allons expliquer à l'aide de l'analyse de l'exemple suivant.

Le texte que nous avons soumis (*cf.* Figure 4) a été attribué à la classe correspondant au niveau A2 avec un score de 100%. Comme nous l'avons précisé *supra* les scores des niveaux A sont souvent très élevés, mais bien évidemment les scores dépendent des marqueurs linguistiques utilisés, autrement dit on peut rédiger un texte avec des marqueurs A1 et A2 par exemple et obtenir divers scores voire même 50 % et 50 % pour chaque niveau. Ceci dit, dans l'analyse des résultats les données les plus pertinentes concernent le taux d'activation attribué à chaque marqueur, ce qui est possible grâce à la dernière version du TDS, autrement dit le TDS pondéré[22]. Ainsi, comme on peut le voir sur la Figure 4, le

[20] *Cf.* figure 2.

[21] En cliquant sur « catégories grammaticales » l'utilisateur peut afficher la liste des codes grammaticaux utilisés par le système.

[22] Vanni *et al.*, 2021.

graphique du taux d'activation permet de visualiser les marqueurs qui ont fortement contribué, ou pas, à l'attribution à une classe.

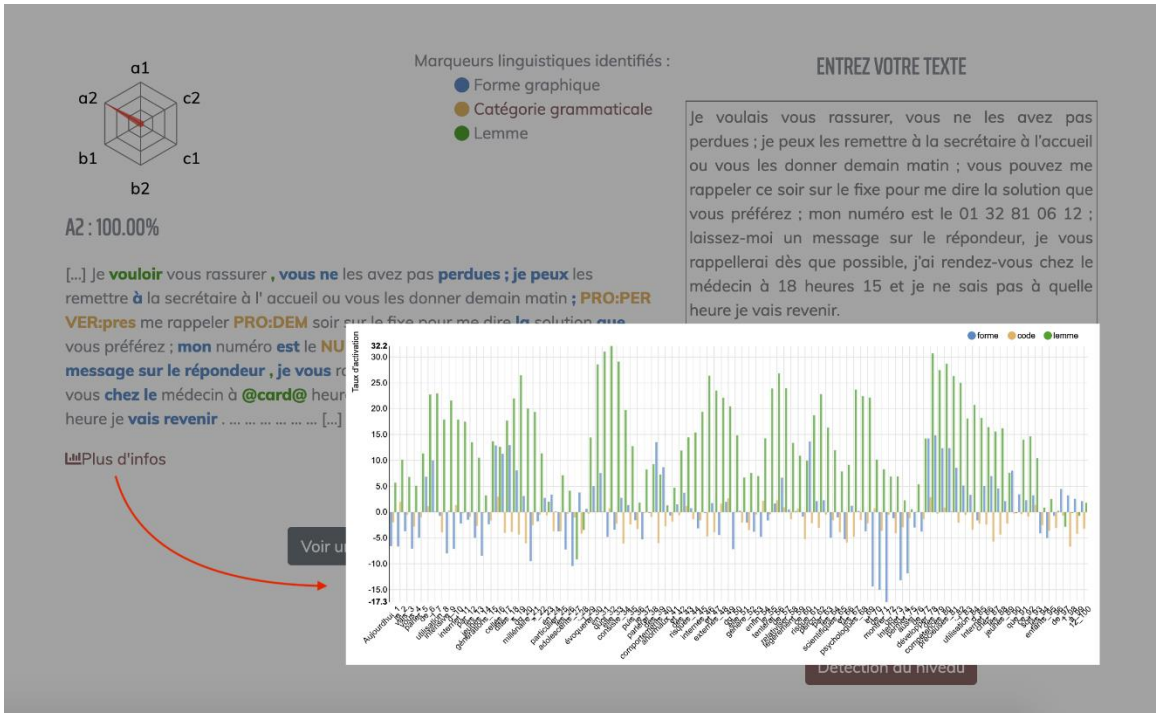

Figure 4 : Taux d'activation des marqueurs des passages-clés A2

Les résultats du TDS et donc de l'analyse de la plateforme DeepFLE peuvent être confirmés par l'analyse des données textuelles, méthode qui, comme nous l'avons précisé, a été utilisée pour notre projet de recherche[23].

À titre d'exemple, le graphique de la distribution statistique de l'enchaînement syntaxique « Pronom personnel – Verbe au présent – Pronom personnel – Verbe à l'infinitif » (Figure 5), qui est une des suites des marqueurs le plus fortement identifiés par le TDS pour la reconnaissance du niveau A2, montre qu'il est typique de ce niveau. L'analyse statistique de l'enchaînement syntaxique a été effectuée avec Hyperbase web sur le corpus d'apprentissage. Ainsi, l'extrait « vous pouvez m'appeler » du texte A2 correspondant à la structure « Pronom personnel – Verbe au présent – Pronom personnel – Verbe à l'infinitif » a été « traduit » en utilisant les codes grammaticaux employés par Hyperbase web : « PRO:PER VER:pres PRO:PER VER:infi »[24].

[23] *Cf*. 2.2. Les trois grandes étapes du projet de recherche.
[24] Les mêmes codes sont utilisés par DeepFLE.

Figure 5 : Distribution statistique des marqueurs le plus fortement identifiés par le TDS pour le niveau A2

Conclusion et perspectives

Le projet de conception et de construction de la plateforme DeepFLE qui résulte du croisement de la didactique du FLE, du *deep learning* et de l'analyse des données textuelles nous a permis de concevoir un premier outil capable d'évaluer, classer et décrire le niveau d'un texte oral en français selon les échelles du CECRL. Il s'agit d'un outil en accès libre et destiné aux acteurs du FLE : didacticiens, concepteurs de méthodes, enseignants et également apprenants. Ces premiers résultats ainsi que l'évolution de la performance du *deep learning* et tout particulièrement du TDS pondéré nous amènent à poursuivre nos recherches afin d'optimiser l'analyse prédictive et descriptive des niveaux ainsi que les fonctionnalités de la plateforme DeepFLE.

Références bibliographiques

BEACCO J.-C. *et al.*, *Niveau B2 pour le français, un référentiel*, Paris, Didier, 2004a.

BEACCO J.-C. *et al.*, *Niveau B2 pour le français, un référentiel. Textes et Références*, Paris, Didier, 2004b.

BEACCO J.-C. et PORQUIER R., *Niveau A1 pour le français, un référentiel*, Paris, Didier, 2007.

BEACCO J.C. *et al*., *Niveau A2 pour le français, un référentiel*, Paris, Didier, 2008.

CONSEIL DE L'EUROPE / Division des Politiques Linguistiques (Strasbourg), *Un Cadre européen commun de référence pour les langues. Apprendre, Enseigner, Évaluer* (CECRL), trad. par S. Lieutaud, Paris, Didier, 2001.

CONSEIL DE L'EUROPE, *Cadre Européen Commun de Référence pour les Langues : volume complémentaire avec des nouveaux descripteurs*, 2018, disponible sur : <https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5> (consulté le 28/02/2020).

GOUILLER F., *Les clés du Cadre. Enjeux et actualité pour l'enseignement des langues aujourd'hui*, Paris, Didier, 2019.

LEBART L., PINCEMIN B. et POUDAT C., *Analyse de données textuelles*, Québec, Presses de l'Université du Québec, 2019.

RUGGIA S., « Le *deep learning* : un outil pour la didactique du FLE ? », *Dialettica pedagogica*, n° 1, p. 79-106, 2019, disponible sur : <hal-02274114>.
RUGGIA S., « Caractériser un texte en français : les passages-clés des niveaux A1 et A2 du CERCL », JADT 2020, in *Actes des 15èmes Journées internationales d'Analyse statistique des Données Textuelles*, Toulouse, 2021, 11p, http://lexicometrica.univ-paris3.fr/jadt/JADT2020/jadt2020_pdf/RUGGIA_JADT2020.pdf, (consulté le 28/01/2021)
VANNI L. *et al*., « Text Deconvolution Saliency (TDS) : a deep tool box for linguistic analysis », in *Actes du 56th Annual Meeting of the Association for Computational Linguistics*, jul 2018, Melbourne, vol. 1, p. 548-557, 2018a, disponible sur : <hal-01804310>.
VANNI L. *et al*., « ADT et *deep learning*, regards croisés. Phrases-clefs, motifs et nouveaux observables », JADT 2018, in *Actes des 14èmes Journées internationales d'Analyse statistique des Données Textuelles*, Rome, Italie, 2018b, p. 459-466, disponible sur : <hal-01823560>.
VANNI L. *et al*., « *Hyperdeep* : *deep learning* descriptif pour l'analyse de données textuelles », JADT 2020, in *Actes des 15èmes Journées internationales d'Analyse statistique des Données Textuelles*, Toulouse, 2021, 12p, http://lexicometrica.univ-paris3.fr/jadt/JADT2020/jadt2020_pdf/VANNI_CORNELI_LONGREE_MAYAFFRE_PRECIOSO_JADT2020.pdf, (consulté le 28/01/2021)

***Ensembles pédagogiques de FLE*[25]**

Niveau A1

ABRY D. *et al*., *Ici, méthode de français A1*, Paris, CLE International, 2007.
ALCARAZ M. *et al*., *Edito, méthode de français A1*, Paris, Didier, 2016.
BERTHET A. *et al*., *Alter Ego, méthode de français A1*, Paris, Hachette, 2006a.
CAPELLE G. et MENAND R., *Le nouveau Taxi, méthode de français A1*, Paris, Hachette, 2009.
CHAHI F. *et al*., *Défi, méthode de français A1*, Paris, Maison des langues, 2018.
DI GIURA M., BEACCO J.-C., *Alors ? A1, méthode de français fondée sur l'approche par compétences*, Paris, Didier, 2012a.
GIRARDET J. et PECHEUR J., *Echo 2e édition, méthode de français A1*, Paris, CLE International, 2013a.
GIRARDET J. *et al*., *Tendances, méthode de français A1*, Paris, CLE International, 2016a.
HIRSCHSPRUNG N. et TRICOT T., *Cosmopolite, méthode de français A1*, Paris, Hachette, 2017a.
LESCURE R. *et al*., *DELF A1 150 activités*, Paris, CLE International, 2005.
LOPES M.J. et LE BOUGNEC J.T., *Totem, méthode de français A1*, Paris, Hachette, 2014a.
PARIZET M. L. *et al*., *Activités pour le Cadre Européen Commun de Référence A1*, Paris, CLE International, 2005a.
REBOUL A. *et al*., *Mobile, méthode français A1*, Paris, Didier, 2012.

[25] Pour une meilleure lisibilité, les ouvrages sont classés en fonction des six niveaux du CECRL qui constituent les six classes de notre corpus.

Niveau A2
ABRY D. *et al*, *Ici, méthode de français A2*, Paris, CLE International, 2007.
ALEMANI L. et GIRODET C., *Mobile, méthode de français A2*, Paris, Didier, 2012.
BERTHET A. *et al.*, *Alter Ego, méthode de français A2*, Paris, Hachette, 2006b.
BERTHET A. *et al*., *Le nouveau Taxi, méthode de français A2*, Paris, Hachette, 2009.
BIRAS P. *et al.*, *Défi, méthode de français A2*, Paris, Maison des langues, 2018.
DI GIURA M., BEACCO J.-C., *Alors ? A2, méthode de français fondée sur l'approche par compétences*, Paris, Didier, 2012b.
GIRARDET J. et PECHEUR J., *Echo 2e édition, méthode de français A2*, Paris, CLE International, 2013b.
GIRARDET J. *et al.*, *Tendances, méthode de français A2*, Paris, CLE International, 2016b.
HEU E. *et al.*, *Edito, méthode de français, niveau A2*, Paris, Didier, 2016.
HIRSCHSPRUNG N., TRICOT T., *Cosmopolite, méthode de français A2*, Paris, Hachette, 2017b.
LOPES M.J. et LE BOUGNEC J.T., *Totem, méthode de français A2*, Paris, Hachette, 2014b.
PARIZET M. L. *et al.*, *Activités pour le Cadre Européen Commun de Référence A2*, Paris, CLE International, 2005b.

Niveau B1
DOLLEZ C. et PONS S., *Alter Ego, méthode de français B1*, Paris, Hachette, 2006.
DUFOUR M. *et al.*, *Edito, méthode de français, niveau B1*, Paris, Didier, 2018.
GIRARDET J. et PECHEUR J., *Echo 2e édition, méthode de français B1.1*, Paris, CLE International, 2016.
GIRARDET J. et PECHEUR J., *Echo 2e édition, méthode de français B1.2*, Paris, CLE International, 2018.
GIRARDET J. *et al.*, *Tendances, méthode de français B1*, Paris, CLE International, 2016a.
LOPES M. J., LE BOUGNEC J.T., *Totem, méthode de français B1*, Paris, Hachette, 2015.

Niveau B2
DOLLEZ C. et PONS S., *Alter Ego, méthode de français B2*, Paris, Hachette, 2007.
GIRARDET J. et GIBBE C., *Echo 2e édition, méthode de français B2*, Paris, CLE International, 2017.
GIRARDET J. *et al.*, *Tendances, méthode de français B2*, Paris, CLE International, 2017.
HEU E. et MABILAT J.J., *Edito, méthode de français, niveau B2*, Paris, Didier, 2015.
PARIZET M. L. *et al.*, *Activités pour le Cadre Européen Commun de Référence B1*, Paris, CLE International, 2006.
GRANDET E. *et al.*, *Activités pour le Cadre Européen Commun de Référence B2*, Paris, CLE International, 2007.

Niveau C1-C2
BARFETY M., *Compétences. Compréhension orale C1*, Paris, CLE International, 2017b.
BARRIERE I. et PARIZET M. L., *ABC DELF 150 exercices C1/C2*, Paris, CLE International, 2018.
CHAPIRO L. *et al.*, *Le DALF 100% réussite C1/C2*, Paris, Didier, 2018.
LESCURE R. *et al.*, *DALF C1/C2 250 activités*, Paris, CLE International, 2007.

PINSON C. (dir.), *Edito, méthode de français, niveau C1*, Paris, Didier, 2018.

Sitographie

DEEPFLE :
Plateforme pour évaluer le niveau de textes oraux en français selon les descripteurs du *Cadre européen commun de référence pour les langues*.
http://deeptext.unice.fr/FLE (consulté le 28/02/2020)

MESURE DU DISCOURS :
Discours présidentiels français de 1958 à aujourd'hui. Observatoire du discours politique français. Méthodes logométriques & deep learning.
http://mesure-du-discours.unice.fr (consulté le 28/02/2020)

DEEPTEXT :
Technical demonstration powered by hyperbase.unice.fr linguistics web tool.
http://deeptext.unice.fr (consulté le 28/02/2020)

HYPERBASE WEB :
Logiciel d'analyse de données textuelles.
http://hyperbase.unice.fr (consulté le 28/02/2020)