



HAL
open science

Caractériser un texte en français : les passages-clés des niveaux A1 et A2 du CECRL.

Simona Ruggia

► **To cite this version:**

Simona Ruggia. Caractériser un texte en français : les passages-clés des niveaux A1 et A2 du CECRL.. JADT 2020 15èmes Journées internationales d'Analyse statistique des Données Textuelles, Jun 2020, Toulouse, France. 11 p. hal-02430322

HAL Id: hal-02430322

<https://hal.science/hal-02430322>

Submitted on 17 Feb 2021

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.

Caractériser un texte en français : les passages-clés des niveau A1 et A2 du CECRL

Simona Ruggia

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

Abstract

The topic of the levels of language, as defined by the *Common European Framework of Reference for languages* (CEFR), and that of our characterization and of their mastery is a crucial topic in the didactics of French as a foreign language and it overhangs the automatic analysis of corpora. Our research is based on a sample corpus composed of six levels: namely A1, A2, B1, B2, C1 and C2. This corpus, whose extent is 595.980 occurrences, is made of several oral texts drawn from many educational packages of French. On the one hand, our study will present the statistic extraction of saliences that mark a change of level according to the CEFR thanks to Hyperdeep. This tool utilizes a deep learning model (Vanni *et al.*, 2020) able to extract the characteristics that give a unique imprint to the text. On the other hand, our research focuses on the patterns, which characterize texts corresponding to levels A1 and A2, thanks to the statistical analysis of textual data.

Keywords: deep learning, classification, prediction, description, level of language.

Résumé

La question des niveaux de langue tels que définis par le *Cadre Européen Commun de Référence pour les Langues* (CECRL) (Conseil de l'Europe, 2001), de leur caractérisation et de leur maîtrise est une question centrale de la didactique du français langue étrangère (FLE) et surplombe l'analyse automatique des corpus. Notre recherche s'est appuyée sur un corpus échantillonné qui comporte six classes de niveaux de langue : à savoir A1, A2, B1, B2, C1 et C2. Ce corpus, dont l'étendue est de 595.980 occurrences, est constitué de nombreux textes oraux extraits de plusieurs ensembles pédagogiques de français langue étrangère (FLE). Notre contribution présentera d'une part, l'extraction statistique des saillances qui marquent un changement de niveau selon le CECRL grâce à *Hyperdeep*, qui exploite un modèle de *deep learning* (Vanni *et al.*, 2020) capable d'extraire les caractéristiques qui donnent une empreinte unique du texte, et d'autre part, les passages-clés qui caractérisent les textes de niveaux A1 et A2 grâce à l'analyse des données textuelles (ADT) et plus précisément à la plateforme Hyperbase web.

Mots clés : *deep learning*, classification, prédiction, description, niveaux de langue.

1. Introduction

Depuis la publication du *Cadre Européen Commun de Référence pour les langues : apprendre, enseigner, évaluer* (CERCL) (Conseil de l'Europe, 2001)¹, la didactique des langues a développé une approche actionnelle de l'enseignement/apprentissage des langues qui est « articulée autour de descripteurs et d'échelles décrivant les compétences des apprenants selon six niveaux » (Rosen, 2007 : 7) allant de A1 à C2². Les échelles de niveaux permettent à la fois « de segmenter le processus d'apprentissage » (Conseil de l'Europe, 2001 : 20) et de lister les compétences en termes de savoir-faire dans le but de « mesurer le progrès de l'apprenant à chaque étape de l'apprentissage et à tout moment de la vie » (*Ib.* : 9). Dans ce sens, l'enseignement/apprentissage d'une langue doit s'appuyer, entre autres, sur des textes, autrement dit sur des « séquence[s] discursive[s] (écrite[s] et/ou orale[s]) » (Conseil de l'Europe, 2001 : 15) correspondant aux six niveaux. Ces textes sont prioritairement des documents authentiques mais ils peuvent être aussi fabriqués à des fins pédagogiques³. La description/évaluation du niveau d'un texte est une étape délicate qui nécessite la maîtrise des descripteurs de la part des enseignants. Voilà pourquoi, dans la lignée des travaux menés par l'équipe « Logométrie. Corpus, Traitements, Modèles » de notre laboratoire « Bases, corpus et langage » UMR 7320/CNRS sur les atouts de l'Intelligence Artificielle⁴, et plus particulièrement du modèle de *deep learning* : *Text Deconvolution Saliency* (TDS) développé par Laurent Vanni (Vanni *et al.*, 2018a) nous avons mené une recherche faisant dialoguer didactique du français langue étrangère (FLE), *deep learning* et analyse des données textuelles (ADT). Les premiers résultats obtenus à l'aide d'un corpus d'apprentissage FLE⁵ constitué de deux classes⁶ (A1 et B2) ont mis en lumière l'efficacité du TDS dans l'extraction des saillances du texte (au niveau lexical, grammatical et morphosyntaxique) qui marquent un changement de niveau selon le CECRL (Ruggia, 2019). Les saillances détectées lors de cette expérience, à partir d'un corpus FLE de test différent du corpus d'apprentissage, correspondent aux caractéristiques des textes telles que décrites dans les *Référentiels pour le français* du niveau A1 (Beacco, Porquier, 2007) et B2 (Beacco *et al.* 2004). Ces derniers décrivent sous formes d'inventaires de « mots » les contenus d'enseignement en fonction des niveaux et d'une langue donnée. Ensuite, au corpus initial d'apprentissage, nous avons intégré 4 autres classes afin de comparer les résultats pour les 6 niveaux. Ce travail a permis de réaliser la plateforme DeepFLE⁷ qui évalue le niveau de textes oraux en français selon les descripteurs du CERCL et propose la classification de textes via *deep learning* ainsi qu'une description des saillances apprises par l'Intelligence Artificielle (I.A.) qui marquent un changement de niveau. Le calcul du TDS évolue constamment permettant une analyse de plus en plus fine des observables linguistiques. Ainsi, dans cette contribution, nous nous

¹ Un *Volume complémentaire* de cet ouvrage a été publié en 2018 afin de proposer de nouveaux descripteurs des niveaux (Conseil de l'Europe).

² Ces niveaux sont d'abord organisés en 3 grandes catégories : A : utilisateur élémentaire, B : utilisateur indépendant, C : utilisateur expérimenté. Chaque catégorie est ensuite déclinée en deux sous-catégories : A1 : introductif ou découverte et A2 : intermédiaire ou de survie, B1 : niveau seuil et B2 : avancé ou indépendant, C1 : autonome et C2 : maîtrise.

³ Notamment pour les niveaux A1 et A2, car la plupart des documents authentiques correspondent aux niveaux plus élevés.

⁴ 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 UCAJEDI portant la référence n° ANR-15-IDEX-01.

⁵ Voir bibliographie : Ensembles pédagogiques de FLE.

⁶ L'utilisation du TDS nécessite la création d'un corpus d'apprentissage d'au moins deux classes. Pour la reconnaissance des niveaux du CECRL chaque classe doit comporter 100.000 occurrences au minimum.

⁷ <http://deeptext.unice.fr/FLE>. Cette plateforme a été réalisée par Laurent Vanni.

pencherons sur l'analyse des passages-clés qui caractérisent les textes de niveau A1 et A2, en considérant un passage-clé comme « une unité de surcroît textométrique ; c'est-à-dire une unité dont la pertinence est calculable et l'extraction automatique » (Vanni *et al.* 2018 : 461). L'étude des passages-clés sera illustrée grâce à *Hyperdeep*, un outil qui reprend le modèle TDS et plus précisément le dernier : le TDS pondéré et « propose une interface visuelle nouvelle afin d'en exploiter les moindres détails et d'interpréter au mieux les indices descriptifs remontés par le réseau » (Vanni *et al.* 2020). L'outil *hyperdeep* est intégré à la plateforme d'analyse des données textuelles Hyperbase web⁸.

2. Cadrage méthodologique : didactique du FLE, *deep learning* et ADT

Notre objectif est de faire dialoguer trois approches afin de mettre en lumière leur complémentarité dans l'analyse de textes en français et plus particulièrement dans la description, évaluation et classification de textes en fonction de leur niveau.

2.1. De la description des niveaux grâce aux outils du CECRL ...

Le CECRL décrit les niveaux de langue en termes de compétences. Ainsi, il propose d'abord de nombreux descripteurs organisés en fonction du macro-type d'activité, à savoir la réception : orale, écrite et audiovisuelle (Conseil de l'Europe, 2018, 58-70), la production : orale et écrite (*Ib.* : 71-84), l'interaction : orale, écrite et en ligne (*Ib.* : 85-106), et la médiation : pour médier un texte, des concepts et la communication (*Ib.* : 107-131). Chaque macro-type d'activité est ensuite déclinée en plusieurs micro-types et plus précisément en plusieurs savoir-faire. A titre d'exemple, pour la réception orale, le CECRL liste les savoir-faire en distinguant la compréhension : générale, d'une conversation entre tierces personnes, en tant qu'auditeur, d'annonces et instructions orales, d'émissions de radio et d'enregistrements. Comme nous l'avons précisé, ce sont les compétences en tant qu'utilisateur de la langue qui sont décrites, ainsi par exemple, au niveau A1 l'apprenant lors d'une conversation entre tierces personnes : « peut comprendre quelques mots et quelques expressions qui se rapportent à lui/elle, à la famille, à l'école, à ses loisirs ou à son environnement, s'ils sont prononcés lentement et clairement. Peut comprendre des mots et des expressions courtes lorsqu'il/elle écoute une conversation simple (par exemple entre un client et un vendeur dans une boutique) à condition que les gens parlent très lentement et très clairement » (*Ib.* : 59). Ensuite, d'autres descripteurs organisés aussi en cascade illustrent les échelles des « compétences communicatives langagières » sous trois intitulés : linguistique (l'étendue linguistique et du vocabulaire, la correction grammaticale, la maîtrise du vocabulaire, du système phonologique et de l'orthographe) (*Ib.* : 137-143) ; sociolinguistique (correction sociolinguistique) (*Ib.* : 143-144) ; et pragmatique (souplesse, tours de parole, développement thématique, cohérence, précision et aisance à l'oral) (*Ib.* : 145-152)⁹.

Afin de décrire davantage les spécificités des niveaux en termes cette fois-ci de « réalisations linguistiques » qui correspondent à des « énoncés ou fragments d'énoncés » et à « des descripteurs grammaticaux » les didacticiens du FLE ont rédigé des *Référentiels pour le français*¹⁰ pour chaque niveau. En ce qui concerne le niveau A1 par exemple, et plus

⁸ hyperbase.unice.fr.

⁹ Le CECRL comprend aussi des descripteurs pour la langue des signes que nous n'aborderons pas ici.

¹⁰ Les *Référentiels* sont organisés en 10 chapitres : 1) Structure du Niveau, 2) Spécifications générales du niveau de la compétence de communication au répertoire discursif, 3) Fonctions, 4) Notions générales, 5) Grammaire : morphologie et structures des énoncés et des phrases, 6) Notions spécifiques, 7) Matière sonore, 8) Matière graphique, 9) Compétences culturelles, 10) Stratégies d'apprentissage.

précisément dans le chapitre « grammaire » les inventaires illustrent « les structures syntaxiques de la phrase simple et complexe », dont nous ne reproduisons ici qu'un extrait :

5.2.1. Il V Impersonnel

IL V	<i>Il pleut</i>
IL V GN	<i>Il faut du sel</i>

5.2.2. C'est -, il y a -, voilà -

C'est GN	<i>Là, c'est la cuisine.</i>
Voilà GN	<i>Voilà le bus !</i>
Il y a GN	<i>Il y a un problème.</i>
C'est adj.	<i>C'est gentil ! (Beacco, Porquier, 2007 : 98)¹¹</i>

Certes l'étude des échelles des descripteurs du CECRL ainsi que des inventaires des *Référentiels* permet aux enseignants de FLE de s'approprier les spécificités de chaque niveau mais la tâche d'évaluation, pour laquelle des grilles criteériées¹² ont été créées par les spécialistes de la discipline, demande un haut degré d'expertise ... d'où l'intérêt de faire appel à l'Intelligence Artificielle.

2.2. ... à l'évaluation/description des niveaux grâce au deep learning et à l'ADT

En matière d'apprentissage, de classification/évaluation de textes le *deep learning* représente aujourd'hui une méthodologie particulièrement efficace, capable d'atteindre notamment des taux de reconnaissance très élevés. Dans le cadre de notre recherche, un apprentissage ou entraînement profond a été effectué avec le TDS, un modèle de *deep learning* déconvolutionnel qui ne se contente pas d'effectuer une simple prédiction mais qui est capable d'extraire les marqueurs linguistiques qui permettent l'attribution d'un texte à une classe avec « une évaluation de leur pertinence interprétative » (Vanni *et al.*, 2018b : 460). Pour ce faire, nous avons établi un corpus d'apprentissage constitué de 6 classes correspondant aux 6 niveaux de langue du CECRL. Ce corpus, constitué de textes oraux issus de nombreux ensembles pédagogiques de FLE¹³, comporte 595980 occurrences¹⁴.

Grâce à *Hyperdeep*, qui intègre le TDS, nous nous pencherons sur les résultats de la prédiction du niveau de textes en français mais également sur la description des passages-clés identifiés pour les niveaux A1 et A2 en considérant le passage-clé comme « le segment de taille fixe le plus fortement attribué à une classe » (Vanni *et al.*, 2020). L'étude des passages-clés portera également sur la comparaison de leur distribution dans les classes du corpus en faisant appel à l'ADT (Lebart, Pincemin, Poudat, 2019).

Cette dernière nous a permis également de hiérarchiser la complexité linguistique des classes de notre corpus à l'aide d'une classification ascendante projetée en forme d'arbre (Mayaffre, Luong, 2003). Comme l'illustre la figure 1, aux deux extrémités de l'arbre la classification identifie et oppose le niveau A (A1 et A2) et le niveau C (C1 et C2), en position intermédiaire au centre de l'arbre se trouve le niveau B (B1 et B2).

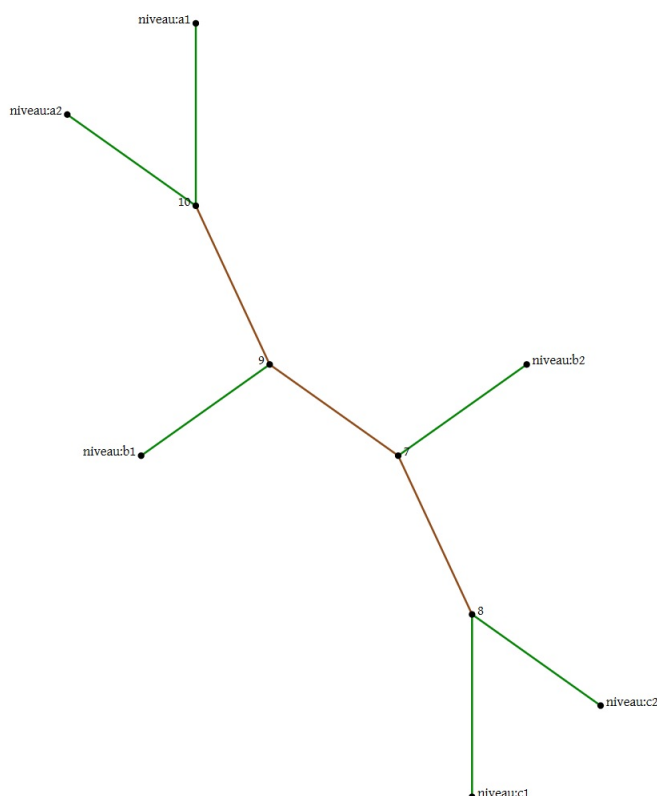
¹¹ Les inventaires sont organisés en deux colonnes : à gauche la description des structures à l'aide de codes (V. = verbe, GN = groupe nominal, etc.) et à droite les réalisations linguistiques.

¹² Notamment celles pour les diplômes du DELF A1, A2, B1 et B2 (diplôme d'études en langue française) et DALF C1 et C2 (diplôme approfondi de langue française).

¹³ Voir Bibliographie : Ensembles pédagogiques de FLE.

¹⁴ L'étendue de chaque classe du corpus est homogène, ainsi le niveau A1 comporte 101923 occurrences, le A2 : 115030, le B1 : 117949, le B2 : 97961, le C1 : 99407 et le C2 : 97175.

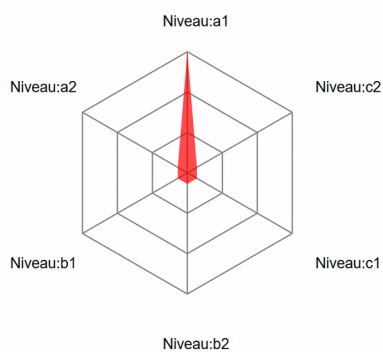
Figure 1 : Analyse arborée du corpus (distance intertextuelle selon le vocabulaire utilisé)



3. Résultats

3.1. Prédiction : du niveau d'un texte

La fonction de prédiction d'*Hyperdeep*, disponible après avoir effectué un entraînement profond, est particulièrement utile en didactique du FLE car l'utilisateur peut soumettre à la plateforme un texte nouveau (qui n'appartient pas au corpus d'apprentissage) par un simple copier/coller et afficher les résultats sous forme de diagramme type radar avec un score de reconnaissance qui indique l'appartenance à une classe :



Plutôt du a1 : 100%

Figure 2 : Visualisation de la prédiction dans *Hyperdeep* : A1 100%

Comme l'indique la figure 2, *hyperdeep* a attribué le texte soumis¹⁵ à la classe correspondant au niveau A1 avec un score de 100%. Mais la finesse de la prédiction est encore plus appréciable lorsqu'un texte, et plus précisément certains passages-clés d'un texte, appartiennent à des classes différentes :

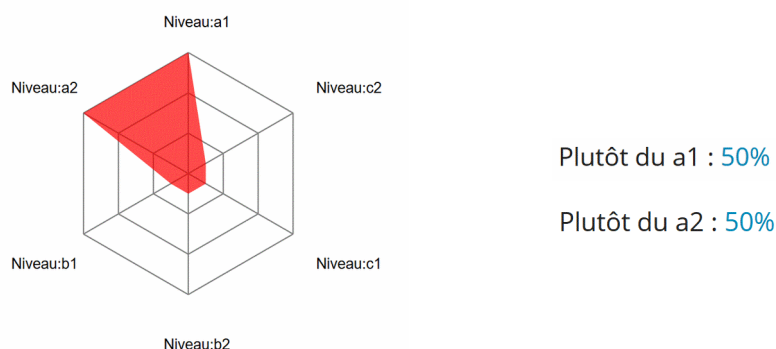


Figure 3 : Visualisation de la prédiction dans Hyperdeep : A1 50% et A2 50%

Ce qui est le cas du deuxième texte¹⁶ soumis dont 50% des passages-clés appartiennent au niveau A1 et 50% au niveau A2 (figure 3). A ce propos, ces résultats confirment que les niveaux du CECRL s'emboîtent comme l'illustre la figure 4 :

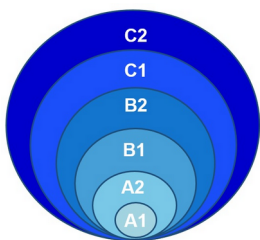


Figure 4 : Les niveaux du CECRL

et que les descriptions de ces derniers :

dessinent six ensembles différents de formes (de A1 à C2) qui sont inclus les uns dans les autres : tout A1 se trouve dans A2 qui est lui-même présent, avec tous ses éléments, dans B2. (Beacco *et al.*, dir., 2008 : 15)

¹⁵ Texte 1 : « Regarde, c'est une photo de ma famille ; il y a André, mon mari, c'est l'homme assis, il a 70 ans donc ses cheveux sont gris ; l'homme debout à gauche est mon fils, il s'appelle Arnaud, il a 46 ans ; l'autre homme, à droite, le très grand, c'est Philippe, le mari de ma fille ; j'ai deux filles, la blonde s'appelle Monique et celle qui a les cheveux bruns, c'est Béatrice. Sur cette photo, il y a mon frère, Jean-Luc, c'est l'homme petit avec une moustache ; il y a aussi Jacques, mon autre frère ; lui, il est grand, il a une moustache aussi ; ils sont avec leurs filles, il y a Mathilde, Sonia et Clara ; la femme blonde aux cheveux courts, c'est Clara ; Sonia a 42 ans, c'est la femme blonde avec les cheveux très longs et l'autre, c'est Mathilde ».

¹⁶ Texte 2 : « Salut ! enfin ça y est, je me suis installé dans mon nouveau logement ; il est super, j'ai deux chambres et un grand salon et une vraie cuisine ; j'ai aussi un petit balcon et je peux apercevoir au loin la tour Eiffel ; les chambres sont petites mais le salon est agréable ; je me sens bien dans cet appartement ; la seule chose qui manque, c'est un ascenseur et je suis au 4^e étage ; heureusement je suis sportif ! j'ai hâte de te montrer la vue, viens me voir très vite.

En ce qui concerne la prédiction, les scores peuvent atteindre des pourcentages très élevés allant jusqu’à 100% pour les niveaux A, ils sont néanmoins inférieurs pour les autres niveaux. A titre d’exemple, un texte est reconnu de niveau C1, avec un score de 60% mais avec quelques passages-clés qui correspondent à d’autres niveaux, à savoir : C2 à 20%, B1 à 10%, A2 et A1 à 5%. Ceci parce que, comme nous l’avons précisé *supra*, un niveau supérieur comporte des traits linguistiques d’un niveau inférieur (A2 →A1, B2 →B1, A2, A1 par exemple) et qu’un texte peut aussi contenir des traits linguistiques d’un niveau supérieur ce qui correspond à la matérialisation de l’intertexte, autrement dit aux passages-clés d’un corpus cible empruntés à un corpus de référence tel que décrit par Mayaffre et Vanni (2020) à travers l’étude du discours politique de Macron depuis 2017 (corpus cible) et l’ensemble du discours élyséen depuis 1958 (corpus de référence).

3.2. Description : les passages-clés de niveau A1 et A2

La description des caractéristiques des niveaux de langue peut être visualisée grâce aux passages-clés *Hyperdeep*, des segments de taille fixe, et plus précisément de 100 occurrences. A ce propos, *Hyperdeep* permet aussi de reparcourir le corpus d’entraînement en proposant un parcours interprétatif des passages-clés qui ont été les plus fortement attribués à une classe grâce à une visualisation dynamique qui prend en compte la forme graphique, le code grammatical et le lemme que l’on peut distinguer grâce aux couleurs attribuées, respectivement le bleu, l’orange et le vert¹⁷ :

“ [...] il y a un **super film sur** le nord de la France ! **rendez-vous à** 18 heures 30 devant le cinéma **PUN PRO:PER** peut venir avec Carmen bien sûr ; appelle -moi ! **à bientôt SENT NOM** , c' est Jeannette ! pour l' anniversaire de Louis **samedi** , est -ce que tu peux apporter le gâteau **SENT PRO:PER PUN** ; j' apporte les **boissons et le ballon** ; on se retrouve chez moi à 14 **heure d'accord ? PRO:PER VER:pres PRP** aller chez Louis en voiture et pas en **méto** ; **à samedi alors ! INT SENT** c' est Fabiola [...] ”

Figure 5 : Description d’un passage-clé A1 dans Hyperdeep

L’intérêt du calcul du TDS pondéré (Vanni *et al.* 2020) est de permettre à l’utilisateur de visualiser le score qu’il a attribué à chaque mot, autrement dit de distinguer les marqueurs qui ont fortement contribué à l’attribution de la classe ou à une autre classe, comme l’illustre la figure 6 :

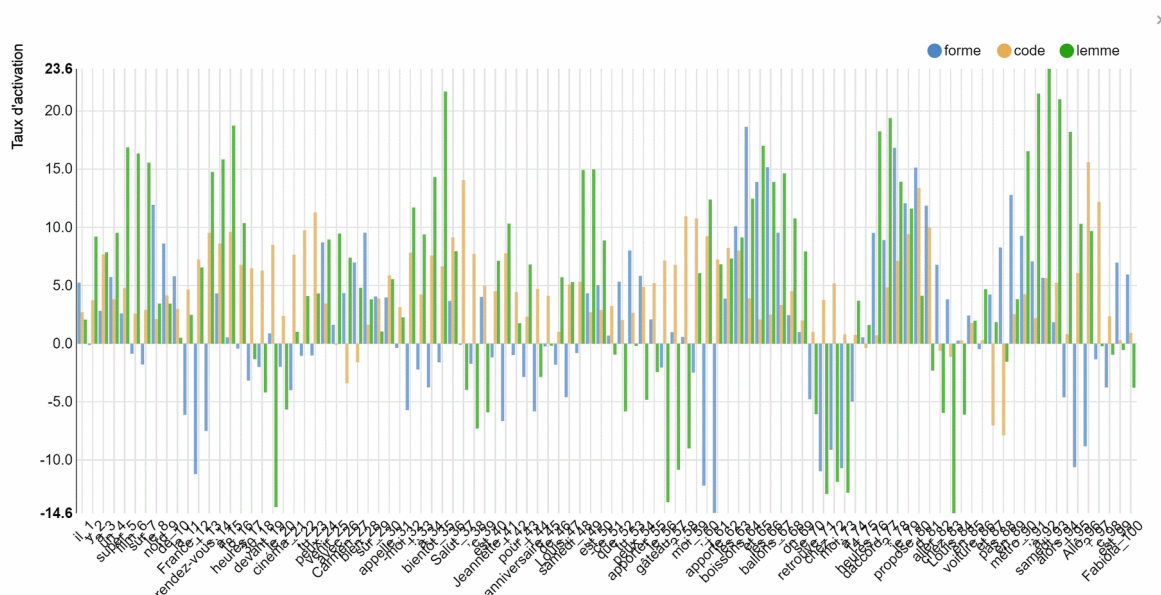
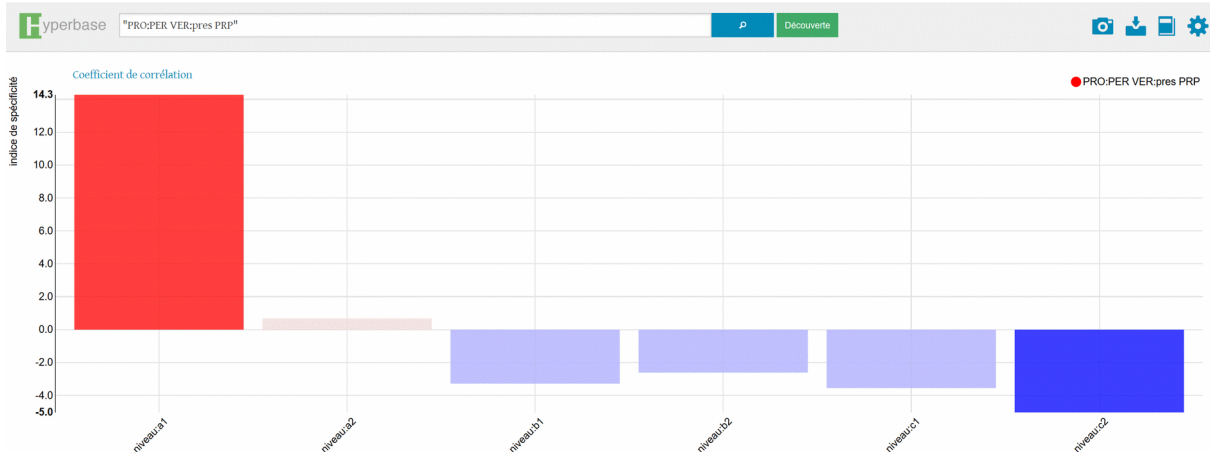


Figure 6 : Taux d’activation des marqueurs d’un des passages-clés A1 dans Hyperdeep

¹⁷ Les exemples illustrés ici appartiennent au corpus d’entraînement.

Par ailleurs, la distribution statistique des marqueurs le plus fortement attribués à une classe, et dans le cas de notre exemple au niveau A1, confirme les résultats du TDS comme par exemple la distribution de l'enchaînement syntaxique élémentaire « Pronom personnel Verbe au présent Préposition » (figure 7).

Figure 7 : Distribution statistique des marqueurs le plus fortement identifiés par le TDS pour



le niveau A1 dans Hyperbase dans tous le corpus.

En ce qui concerne le niveau A2, les TDS les plus élevés du passage suivant sont « je » et « ai » :

Figure 8 : Description d'un des passages-clés A2 dans Hyperdeep

“ [...] ils apporteront des **boisson** ; **j' ai** envie de faire **un** gâteau **au chocolat** ; si tu **vouloir** , apporte **du fruit** ; on écouterà les derniers enregistrements de notre groupe **préférer** ; **laisser moi un** message si tu es OK ; **je ne** te rappelle pas parce que je vais à la **piscine et je ne ai** plus de batterie , tchao **tchao SENT vous écouter** Radio 1 , il est 8 heures **@card@** ; !' une des principales informations , **ce être** l' incroyable douceur hivernale ; **cela fait** maintenant **presque quatre semaine** que les températures sont [...] ”

Encore une fois, l'analyse statistique de ces marqueurs, nous a permis de corroborer les résultats du TDS mais également d'attribuer des observables linguistiques : « j'ai envie de » (Pronom personnel Verbe au présent Substantif Préposition) à une classe :

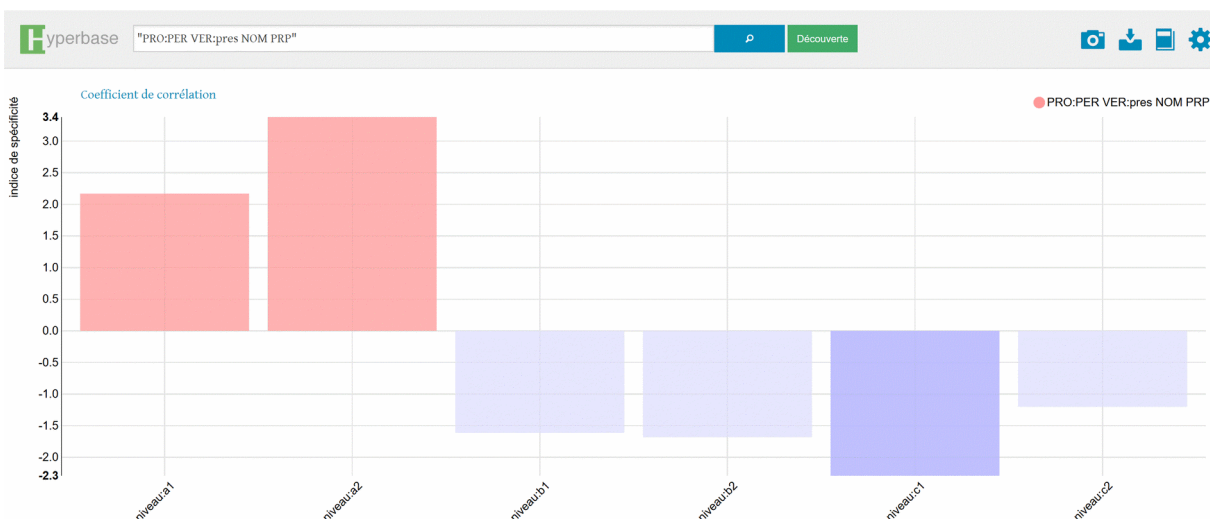


Figure 9 : Distribution statistique des marqueurs le plus fortement identifiés par le TDS pour le niveau A2 dans Hyperbase dans tous le corpus

4. Conclusion et perspectives

Cette contribution a cherché de mettre en exergue l'intérêt d'une approche qui vise à croiser didactique du FLE, *deep learning* et ADT. D'une part, la classification et la description des passages-clés *hyperdeep* détectés par le TDS pondéré représente un atout considérable pour les didacticiens et enseignants de FLE dans le cadre de l'évaluation et de l'analyse des textes en fonction de leur niveau du CECRL. D'autre part, l'analyse statistique des résultats du TDS permet de repérer les observables linguistiques typiques de chaque niveau.

Bibliographie

- Beacco J.C. *et al.* (2004). Niveau B2 pour le français, un référentiel. Didier.
- Beacco J.C. *et al.* (dir.). (2008). Niveau A2 pour le français, un référentiel. Didier.
- Beacco J.C., Porquier R. (2007). Niveau A1 pour le français, un référentiel. Didier.
- Conseil de l'Europe. (2001). Cadre Européen Commun de Référence pour les langues : apprendre, enseigner, évaluer. Didier.
- Conseil de l'Europe. (2018). Cadre Européen Commun de Référence pour les Langues : volume complémentaire avec des nouveaux descripteurs, consultable à l'adresse : <https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5>.
- Lebart L., Pincemin B. et Poudat C. (2019). Analyse des données textuelles. *Presses de l'Université du Québec*.
- Mayaffre D., Luong X. (2003). Les discours de Jacques Chirac (1995-2002), Histoire & mesure, XVIII - 3/4 | 2003, mis en ligne le 15 avril 2007, consultable à l'adresse : <http://journals.openedition.org/histoiremesure/831>.
- Mayaffre D., Vanni L. (2020 – soumis). Objectiver l'intertexte ? Emmanuel Macron, *deep learning* et statistique textuelle. JADT 2020, in *Actes des 15^{èmes} Journées internationales d'Analyse statistique des Données Textuelles*, Toulouse.
- Rosen E. (2007). Le point sur le Cadre Européen Commun de Référence pour les langues. CLE International.
- Ruggia S. (2019). Le deep learning : un outil pour la didactique du FLE ?, *Dialettica pedagogica*, n°1, p.79-106. [hal-02274114]
- Vanni L. *et al.* (2018a). *Text Deconvolution Saliency (TDS): a deep tool box for linguistic analysis*, 56th Annual Meeting of the Association for Computational Linguistics, jul 2018, Melbourne, France. [hal-01804310]
- Vanni L. *et al.* (2018b). ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables, JADT 2018, in *Actes des 14^{èmes} Journées internationales d'Analyse statistique des Données Textuelles*, Rome, Italie, p.459-466. [hal-01823560]
- Vanni L. *et al.* (2020 - soumis). Hyperdeep : deep learning descriptif pour l'analyse de données textuelles, JADT 2020, in *Actes des 15^{èmes} Journées internationales d'Analyse statistique des Données Textuelles*, Toulouse.

ENSEMBLES PEDAGOGIQUES DE FLE¹⁸**Niveau A1**

- Abry D. *et al.* (2007). Ici, méthode de français A1. CLE International.
- Alcaraz M. *et al.* (2016). Edito, méthode de français A1. Didier.
- Berthet A. *et al.* (2006a). Alter Ego, méthode de français A1. Hachette.
- Capelle G., Menand R. (2009). Le nouveau Taxi, méthode de français A1. Hachette.
- Chahi F. *et al.* (2018). Défi, méthode de français A1. Maison des langues.
- Di Giura M., Beacco J.C., (2012a). Alors ? méthode de français A1. Didier.
- Girardet J., Pécheur J. (2013a). Echo 2^e édition, méthode de français A1. CLE International.
- Girardet J. *et al.* (2016a). Tendances, méthode de français A1. CLE International.
- Hirschsprung N., Tricot T. (2017a). Cosmopolite, méthode de français A1. Hachette.
- Lescure R. *et al.* (2005). DELF A1 150 activités. CLE International.
- Lopes M.J., Le Bougnec J.T. (2014a). Totem, méthode de français A1. Hachette.
- Parizet M.L. *et al.* (2005a). Activités pour le Cadre Européen Commun de Référence A1. CLE International.
- Reboul A. *et al.* (2012). Mobile, méthode français A1. Didier.

Niveau A2

- Abry D. *et al.* (2007). Ici, méthode de français A2. CLE International.
- Alemani L., Girodet C. (2012). Mobile, méthode de français A2. Didier.
- Berthet A. *et al.* (2006b). Alter Ego, méthode de français A2. Hachette.
- Berthet A. *et al.* (2009). Le nouveau Taxi, méthode de français A2. Hachette.
- Biras P. *et al.* (2018). Défi, méthode de français A2. Maison des langues.
- Di Giura M., Beacco J.C. (2012b). Alors ? méthode de français A2. Didier.
- Girardet J., Pécheur J. (2013b). Echo 2^e édition, méthode de français A2. CLE International.
- Girardet J. *et al.* (2016b). Tendances, méthode de français A2. CLE International.
- Heu E. *et al.* (2016). Edito, méthode de français, niveau A2. Didier.
- Hirschsprung N., Tricot T. (2017b). Cosmopolite, méthode de français A2. Hachette.
- Lopes M.J., Le Bougnec J.T. (2014b). Totem, méthode de français A2. Hachette.
- Parizet M.L. *et al.* (2005b). Activités pour le Cadre Européen Commun de Référence A2. CLE International.

Niveau B1

- Dollez C., Pons S. (2006). Alter Ego, méthode de français B1. Hachette.
- Dufour M. *et al.* (2018). Edito, méthode de français, niveau B1. Didier.
- Girardet J., Pécheur J. (2016). Echo 2^e édition, méthode de français B1.1. CLE International.
- Girardet J., Pécheur J. (2018). Echo 2^e édition, méthode de français B1.2. CLE International.
- Girardet J. *et al.* (2016a). Tendances, méthode de français B1. CLE International.
- Lopes M.J., Le Bougnec J.T. (2015). Totem, méthode de français B1. Hachette.

Niveau B2

- Dollez C., Pons S. (2007). Alter Ego, méthode de français B2. Hachette.
- Girardet J., Gibbe C. (2017). Echo 2^e édition, méthode de français B2. CLE International.

¹⁸ Pour une meilleure lisibilité, les ouvrages sont classés en fonction des 6 niveaux du CECRL qui constituent les six classes de notre corpus.

- Girardet J. *et al.* (2017). Tendances, méthode de français B2. CLE International.
- Heu E., Mabilat J.J. (2015). Edito, méthode de français, niveau B2. Didier.
- Parizet M.L. *et al.* (2006). Activités pour le Cadre Européen Commun de Référence B1. CLE International.
- Grandet E. *et al.* (2007). Activités pour le Cadre Européen Commun de Référence B2. CLE International.

Niveau C1

- Barfety M. (2017b). Compétences. Compréhension orale C1. CLE International.
- Barrière I., Parizet M.L. (2018). ABC DELF 150 exercices C1/C2. CLE International.
- Chapiro L. *et al.* (2018). Le DALF 100% réussite C1/C2. Didier.
- Lescure R. *et al.* (2007). DALF C1/C2 250 activités. CLE International.
- Pinson C. (dir.) (2018). Edito, méthode de français, niveau C1. Didier.

Niveau C2

- Barrière I., Parizet M.L. (2018). ABC DELF 150 exercices C1/C2. CLE International.
- Chapiro L. *et al.* (2018). Le DALF 100% réussite C1/C2. Didier.
- Dubois I., Saintes I. (2017). DALF C2, Tests complets corrigés. La fée prépa.
- Lescure R. *et al.* (2007). DALF C1/C2 250 activités. CLE International.