



GSJ: Volume 11, Issue 5, May 2023, Online: ISSN 2320-9186
www.globalscientificjournal.com

Vers un traitement automatique des difficultés de la langue spécialisée à partir des examens du baccalauréat de la discipline physique-chimie au Maroc

hoummad chakib, faik ouahab zakaria, faik ouahab abdelmajid

Introduction

Nombreuses sont aujourd'hui les voix qui s'élèvent pour revendiquer l'appropriation des langues étrangères dans les matières scientifiques au cycle secondaire au Maroc¹. Aux yeux des professeurs ainsi des élèves, la langue étrangère fait peser un fardeau supplémentaire sur un système déjà fragile. La barrière de la langue devient plus visible dans les examens du baccalauréat (Bac) et s'ajoute aux difficultés déjà rencontrées dans les matières scientifiques.

À partir d'une extraction automatique d'un corpus des examens nationaux de la discipline physique-chimie entre 2016 -2022, nous cherchons à comprendre l'usage de la langue française dans un domaine sous un nouvel angle. Le traitement automatique, récemment actif en linguistique, offre une opportunité pour étudier quantitativement et qualitativement la langue.

Un enjeu important est alors d'être capable d'évaluer les difficultés de la langue spécialisée des examens en se basant sur des critères pertinents.

Comment sont-elles représentées les difficultés de la langue spécialisée dans les examens nationaux?

Existe-t-il des différences dans l'usage de la langue des examens nationaux entre les filières étudiées?

1. Généralités

À l'orée de ce travail réalisé en ces pages, quelques définitions s'imposent. Ce furent donc nos prémisses de départ. Nous voulons d'abord nous arrêter sur la notion de la langue spécialisée:

¹ <https://www.leconomiste.com/article/1021974-enquete-l-economiste-sunergia-langue-d-enseignement-decidement-pas-d-accord>
<https://aujourd'hui.ma/emploi/formation/reforme-du-systeme-educatif-les-dix-defis-de-lecole-de-demain-118020>

1.1 La langue spécialisée

Comme le rappelle Pierre Lerat, la langue spécialisée se caractérise par le fait de :

Utilise[r] des dénominations spécialisées (les termes), y compris des symboles non linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée. On peut donc la définir comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées. (Lerat, 1995, p.21)

Cette définition comprend trois composantes de la langue spécialisée, à savoir les dénominations spécialisées (les termes), les symboles non linguistiques et des énoncés mobilisant les sources ordinaires d'une langue naturelle. Les élèves dans les épreuves du Bac sont confrontés, dans les disciplines scientifiques enseignées en français, à ces trois niveaux de difficultés.

Le corpus que nous entamons analyser correspond à ces trois répertoires de la langue spécialisée. D'un point de vue méthodologique, il est aisé de repérer les symboles linguistiques des symboles non linguistiques, cependant la séparation des termes appartenant au domaine des mots de la langue naturelle est loin d'être résolue. Nous recourons en ce sens à la notion du spécialiste dans le domaine. Cette notion est développée par plusieurs linguistes :

1.2 La notion du spécialiste

La notion de spécialiste ou expert dans son domaine revient en force dans les recherches des langues spécialisées. On ne peut se contenter juste des traits strictement linguistiques pour désigner un terme (appartenant à la terminologie) d'un mot (appartenant à la langue commune). L'avis du spécialiste du domaine est la pierre angulaire pour la validation du caractère spécialisé ou non d'un énoncé, Pierre Lerat (Lerat, 1995, p.47) écrit en ce sens :

Le critère des critères est l'avis du spécialiste, dont on ne saurait raisonnablement tenter de faire l'économie. Ce qui est possible, linguistiquement, c'est le repérage d'un faisceau d'indices de terminologisation.

2. Bac international au Maroc

Après d'âpres controversées, ressenti parfois comme une menace à l'identité nationale², le système éducatif s'est tracé une stratégie visant à introduire le plurilinguisme dans le système éducatif marocain. La réforme de 2014 prévoit la création de filières dites « Internationales » au

² Le projet de loi-cadre prévoyant le retour de l'enseignement des matières scientifiques et techniques a provoqué une levée de boucliers du PJD et de l'Istiqlal lors de la tenue d'une commission parlementaire.

niveau de l'enseignement qualifiant où les trois matières scientifiques (mathématiques, physique-chimie et sciences de la vie et de la terre) sont enseignées en langues étrangères. Selon la note ministérielle³. Elles sont catégorisées comme suit:

- Sections internationales option langue française (BIOF).
- Sections internationales option langue anglaise.
- Sections internationales option langue espagnole.

Cependant, le français constitue la langue principalement diffusée dans l'ensemble du royaume. Les autres langues: l'espagnole est favorisée dans certains établissements au nord du Maroc et l'anglais est désormais pratiqué dans certains établissements des grandes villes au Maroc: Rabat, Casablanca, Marrakech et autres.

Cette loi-cadre prévoit en même temps l'augmentation de deux heures pour la couverture horaire hebdomadaire consacrée à la langue française dans le Tronc Commun et en première année Bac lettres et ainsi l'enseignement de toutes les matières scientifiques en langue française, auquel s'ajoute un cours de deux heures dédié aux techniques d'expression et de communication (TEC) en français dans le Tronc Commun et en première année Bac, séries Sciences Expérimentales, Sciences Mathématiques, Lettres et Sciences Humaines.

Les deux heures de la langue française consacrées aux filières susmentionnées suscitent un renforcement supplémentaire aux élèves en matière de la langue générale. C'est alors que le Ministère de tutelle a programmé deux heures par semaine de traduction pour offrir aux élèves des occasions supplémentaires de parfaire leurs connaissances en langue spécialisée. Or, ces derniers cours ne sont instaurés que dans quelques établissements.

La matière physique-chimie

Il va de soi que chaque filière a sa propre valeur et répond à une certaine politique éducative, néanmoins les choix des élèves du secondaire sont orientés par au moins un des deux aspects :

- le premier est plutôt coercitif: les filières à accès limité comme les sciences mathématiques (désormais SM) l'orientation des élèves est ... alors la décision est prise par conseil de classe avec la présence du conseiller en orientation en fonction des résultats obtenus.
- Le second revêt un caractère plutôt social lié aux écioreprésentations sociales que les élèves se donnent de la valeur et de l'importance des filières dans le système éducatif. En effet, entre les filières sciences physiques et Sciences de vie et de la terre SVT (désormais PC et SVT), par

³ Loi-cadre n° 51-17 relative au système d'éducation, de formation et de recherche scientifique.

exemple, aucun texte ne détermine une hiérarchie, mais sur le terrain la réalité est autre. Le principe de relégation n'y pas étranger. Comme le pensait Mélanie Ferrara and Nathanaël Friant (2014) « *les mécanismes de relégation ont pour conséquence de dévaloriser de facto la filière qualifiante et d'instaurer une certaine hiérarchie entre les filières* ».

Bien que les unités du programme de la matière physique-chimie sont identiques dans les trois filières, les leçons et le volume horaire des SM et PC dépasse les SVT. Les SM et Les PC comptent 198 heures par an et les SVT 132 heures seulement.

Cette portée générale sur les trois filières permet de tirer quelques conclusions: les trois filières ne sont pas sur le même pied d'égalité. Une hiérarchie s'est instaurée entre les trois filières et les faits qui se produisent sur le terrain confirment le constat avancé, la filière des SVT est la plus dévalorisée parmi les trois et ceci contribue dans une large mesure à une concentration des élèves en difficulté scolaire dans cette dernière. Par ailleurs, la filière des SM est la mieux valorisée et par conséquent elle demeure l'apanage des élèves les plus compétents. Entre ces deux filières, se positionne la filière PC qui abrite les élèves d'un niveau intermédiaire.

3. La méthode

- Le traitement des données a été réalisé en deux étapes: tout d'abord, nous avons utilisé le logiciel en ligne Voyant Tools pour effectuer des analyses automatiques de façon simple; en second lieu, nous avons entretenu les résultats de manière à ce que les imperfections du logiciel soient corrigées manuellement.
- Pour vérifier la deuxième hypothèse, nous scindons notre corpus entre les trois filières des examens nationaux à savoir les SVT, PC et SM.
 - Notre corpus soumis à l'analyse se compose de 36 documents (input) des examens nationaux du baccalauréat de la matière physique chimie des trois filières : SM, PC et SVT.
 - Tous les documents traités sont d'une extension PDF, toutefois le logiciel Voyant Tools n'a pas pu prendre en charge quelques documents. L'output est de 21 documents.
 - Les documents traités sont répartis comme suit:
SM 8 documents, PC 8 documents, SVT 5 documents

- Pour l'exploitation du corpus, nous utilisons Voyant Tools. Il s'agit en effet d'un outil informatique pour effectuer rapidement des analyses automatiques de façon simple. Un analyseur de textes gratuit et en ligne. Il a de nombreux outils intéressants facilement explorables et exportables. Toutefois, Voyant Tools n'effectue pas une lemmatisation, une opération lors de laquelle les différentes formes qu'un mot peut prendre dans un texte (par exemple déclinaisons, singulier-pluriel, etc.) sont classées sous une même entrée lexicale, appelée le lemme.

4. Résultats et discussions

Nous privilégions, d'abord, les outils d'analyse de voyant Tools qui donnent des statistiques sur le corpus étudié : nombre de mots, nombre de documents, mots les plus fréquents.

Les trois tableaux ci-dessous donnent des statistiques sur les documents traités par le logiciel Voyant Tools (tableau 1 SM; tableau 2 PC; tableau 3 SVT)

Figure 1- la filière SM

Titre	Mots	Types	Proportion	Mots/Phrase
Examen national SM 2016 session normal	2882	703	24%	51.5
Examen national SM 2016 session rattrapage	3202	755	24%	55.2
Examen national SM 2017 session normal	3005	783	26%	46.2
Examen national SM 2017 session rattrapage	2954	725	25%	50.1
Examen national SM 2018 session normal	2905	755	25%	41.5
Examen national SM 2018 session rattrapage	3187	838	26%	41.4
Examen national SM 2020 session normal	3109	849	27%	42.6
Examen national SM 2020 session rattrapage	3034	829	27%	44.6

Figure 2- la filière PC

Titre	Mots	Types	Proportion	Mots/Phrase
Examen national PC 2016 session normal	2462	779	32%	27.4
Examen national PC 2016 session rattrapage	2062	666	32.3%	26.4
Examen national PC 2018 session normal	2508	731	29.1%	30.9
Examen national PC 2018 session rattrapage	2206	678	30.7%	26.3
Examen national PC 2019 session normal	2354	676	28.7%	23.3
Examen national PC 2019 session rattrapage	2557	725	28.4%	25.8
Examen national PC 2020 session normal	2328	665	28.6%	24.8
Examen national PC 2020 session rattrapage	2428	673	27.7%	24

Figure 3- la filière SVT

Titre	Mots	Types	Proportion	Mots/Phrase
Examen national SVT 2017 session normal	2194	655	30%	21.7
Examen national SVT 2018 session normal	1984	595	30%	20
Examen national SVT 2018 session rattrapage	1954	609	31%	22.2
Examen national PC 2020 session normal	2160	678	31%	21.8
Examen national SVT 2020 session rattrapage	2024	649	32%	22.4

Dans les trois tableaux, nous réalisons que le nombre des unités lexicales dépasse les 1800. On constate que leur nombre dans les examens des filières est d'un ordre décroissant selon les filières. En effet, les filières des SM comptent le nombre des unités le plus élevé parmi les deux autres, suivi de la filière du PC et enfin viennent les SVT.

Notons que le logiciel a fourni des statistiques sur le nombre des unités lexicales. À partir desquelles est effectué le calcul dans chaque tableau de la différence du nombre minimum et le nombre maximum. Les résultats font preuve d'un état irrégulier dans les trois filières. En effet, la filière des SVT ne dépasse pas les 240 unités lexicales et les PC compte le plus grand nombre d'unités qui atteint les 495, tandis que les SM se positionnent en intermédiaire entre les deux précédentes (320 unités lexicales). C'est alors que nous pourrions dire que les épreuves des SVT montrent une constance dans le nombre des unités lexicales par rapport aux autres filières.

En ce qui concerne les proportions des unités, nous constatons que le nombre le plus élevé des hapax⁴ se réalise succinctement dans la filière des SVT, viennent ensuite les PC et enfin les SM.

4.1 Au niveau quantitatif

L'on peut légitimement se demander sur la difficulté des examens dans les trois filières :

❖ Le volume des unités lexicales

L'un des traits de la difficulté linguistique des élèves est le nombre des unités lexicales dans l'examen. Pour nous faciliter l'analyse, nous insérons, au prime abord, toutes les unités lexicales retenues sans se soucier de leur répertoire dans la langue spécialisée.

Si nous comparons le nombre de ces unités dans les trois filières, nous constatons que le nombre le plus élevé est celui des SM; tandis que la filière PC se positionne en deuxième rang et enfin les

⁴ Les hapax sont des unités à fréquence égale à un.

SVT. Au vu de ces statistiques, le volume est un critère quantitatif de la difficulté linguistique. C'est alors que les épreuves des SM sont considérées comme étant les plus difficiles.

❖ **La différence du nombre entre les unités lexicales**

Par ailleurs, comparer seulement le volume des unités dans notre cas paraît peu plausible, car il faut introduire le changement pour mieux observer les obstacles linguistiques. En effet, le premier trait mesure des entités statiques, alors que le deuxième inclut aussi la dynamique de ces mêmes entités dans la même filière. Ce trait mesure l'écart, dans le même tableau, des unités lexicales.

Nous supposons en ce sens que l'instabilité des examens constitue un autre trait de difficulté linguistique que peuvent éprouver les élèves. Sachant que l'une des sources fondamentales de la préparation pour le Bac est les examens précédents. La batterie lexicale est désormais substantielle en corrélation avec la difficulté. De sorte que le volume de la batterie lexicale soit en fonction des difficultés. En ce sens, nous pouvons en déduire à partir de ce rebondissement que la filière du PC soit la filière qui subit un changement dans les unités lexicales par un écart⁵ de 495.

❖ **Le nombre des hapax**

Le nombre des hapax est un autre trait mesurable pour supposer la difficulté des examens.

Nous distinguons dans un premier lieu deux niveaux de traitement des hapax:

- Au niveau de chaque document

Cette analyse traite le nombre des hapax au niveau du même document. C'est-à-dire la difficulté linguistique rencontrée par les élèves dans l'examen.

Les tableaux 1, 2 et 3 indiquent les proportions des hapax dans chaque document. Nous observons que les pourcentages des hapax dans les trois filières sont quasi-constants. SM (24% à 27%); PC (27.5% à 32.3); SVT (30% à 32%). À partir de ces statistiques, deux remarques se dégagent:

- ✓ la filière SVT détient le pourcentage le plus élevé des hapax parmi les autres filières.
- ✓ Les pourcentages des hapax sont quasiment réguliers. Ceci émet une hypothèse de recherche: on peut légitimement penser que ces données sont révélatrices des types des

⁵ L'écart est calculé entre la fréquence des unités lexicales la plus élevée et la fréquence des unités la plus basse dans les documents de chaque filière.

examens proposés au Bac. La lecture des pourcentages des hapax pourrait indiquer une certaine corrélation entre les examens de la même filière.

- Au niveau de tous les documents de la même filière

Le deuxième volet consiste à fusionner tous les documents de la même filière dans un seul document d'une extension PDF dans l'objectif est de traiter les hapax sur un axe diachronique (2016-2022). Cette prise en compte permettra de visualiser le nombre des hapax non seulement dans un seul examen, mais dans tous les examens de la filière. Les pourcentages des hapax obtenus sont presque réduits de la moitié:

- Le pourcentage des SM est de 10%.
- Le pourcentage des PC est de 24%.
- Le pourcentage des SVT est de 16 %.

Bien que ces résultats soient dans l'ensemble conformes à la logique des prévisions, ils sont discutables à l'égard de la filière du PC. Le pourcentage obtenu est supérieur à nos attentes. Cette valeur est due principalement à un dysfonctionnement du logiciel.

4.2 Au niveau qualitatif

Conçu à partir de la définition de Pierre Lerat, ce volet s'est basé sur cette répartition de la langue spécialisée en mots, termes et symboles. Notre procédé, donc, est de comparer les résultats dans chacune des catégories susmentionnées sur la base de la fréquence. Deux volets seront évalués: les unités lexicales les plus récurrentes et les hapax. Dans chaque volet, nous abordons une comparaison des trois répertoires de la langue spécialisée.

❖ Les unités lexicales dont la fréquence est élevée

Cette entrée mesure la fréquence des unités lexicales qui sont les plus en usage. Un usage intense se matérialise par une fréquence élevée. La répartition des unités en mots, termes et symboles montre le poids de chaque catégorie dans les examens nationaux. Les résultats obtenus sont majoritairement en faveur des mots. Ce qui implique que la langue commune est dominante. En second rang viennent les termes et en bas de la liste se positionne les symboles.

❖ Les hapax

La distribution des hapax précise une difficulté d'ordre linguistique. Il va notamment ainsi lorsqu'on sait que le nombre élevé des hapax est un indice de difficulté pour la compréhension de la langue. La répartition des hapax dans les examens traités montre que les mots sont majoritaires. Le nombre des termes et des symboles est très limité.

L'examen qualitatif de la fréquence élevée ou des hapax a permis donc de montrer le poids des mots dans le corpus étudié. Les difficultés des élèves quant à la langue commune sont fort présentes dans les examens nationaux.

Conclusion

Dans cette étude, nous avons cherché à évaluer les difficultés de la langue des examens du Bac particulièrement dans les filières SM, PC et SVT, générées par une extraction automatique. Pour y parvenir, notre procédé fait l'objet de deux approches différentes:

Une approche que l'on pourrait qualifier de quantitative, qui nous a conduits à mesurer les difficultés dans les unités lexicales en tenant en considération leur volume, les écarts et les hapax. D'autre part, une approche plutôt qualitative, basée sur la répartition de la langue spécialisée comme étant défini par Pierre Lerat, qui a concouru à l'évaluation des fréquences des mots, des termes et des symboles.

De ces évaluations menées sur notre corpus, deux résultats se dégagent:

La première approche permet de montrer que malgré le nombre des unités lexicales est conforme à nos attentes, l'écart et les hapax ne respectent pas le niveau de difficulté supposé dans l'usage de la langue entre les filières. La deuxième approche aide à montrer le poids considérable des mots par rapport aux termes et aux symboles. Les mots sont largement dominants par rapport aux termes et aux symboles. La langue des examens nationaux BIOF n'incorpore pas assez les difficultés entre les filières, et la langue commune suppose avoir un poids considérable dans la difficulté des examens.

Bibliographie

- Azizoun, M. (2019). *L'évaluation des établissements scolaires au Maroc : État actuel et perspectives*.
- Benamara, F., Grouin, C., Karoui, J., Moriceau, V., & Robba, I. (s. d.). *Analyse d'opinion et langage figuratif dans des tweets : Présentation et résultats du Défi Fouille de Textes DEFT2017*.
- Boulahcen, A. (2005). Le processus d'orientation scolaire au Maroc : Une analyse sociologique. *Revue internationale d'éducation de Sèvres*, 38, 25-34. <https://doi.org/10.4000/ries.1427>
- Chami, F., & Elbekraoui, N. (1991). L'enseignement du français au Maroc : Démarches pédagogiques - Essai d'évaluation. *L'Information Grammaticale*, 51(1), 44-47. <https://doi.org/10.3406/igram.1991.3239>
- Ferrara, M., & Friant, N. (2014). Les représentations sociales des élèves du premier et du dernier degré de l'enseignement secondaire en Belgique francophone par rapport aux différentes filières. *L'Orientation scolaire et professionnelle*, 43/4. <https://doi.org/10.4000/osp.4496>

Gregory, K., Geiger, L., & Salisbury, P. (2022). Voyant Tools and Descriptive Metadata : A Case Study in How Automation Can Compliment Expertise Knowledge. *Journal of Library Metadata*, 22(1-2), 1-16. <https://doi.org/10.1080/19386389.2022.2030635>

Hetenyi, G., Dr. Lengyel, A., & Dr. Szilasi, M. (2019). Quantitative analysis of qualitative data : Using voyant tools to investigate the sales-marketing interface. *Journal of Industrial Engineering and Management*, 12(3), 393. <https://doi.org/10.3926/jiem.2929>

Lerat Pierre. (1995). *Les langues spécialisées* (PUF).

<https://www.leconomiste.com/article/1021974-enquete-l-economiste-sunergia-langue-d-enseignement-decidement-pas-d-accord>

<https://aujourd'hui.ma/emploi/formation/reforme-du-systeme-educatif-les-dix-defis-de-lecole-de-demain-118020>

<https://www.tandfonline.com/doi/full/10.1080/19386389.2022.2030635>

<https://www.men.gov.ma/Fr/Pages/Programmes-qualifiant.aspx>

