



HAL
open science

Analyser les réponses aux questions ouvertes des questionnaires de MOOC : Considérations méthodologiques

Aurélie Bayle

► **To cite this version:**

Aurélie Bayle. Analyser les réponses aux questions ouvertes des questionnaires de MOOC : Considérations méthodologiques. [Rapport Technique] Inria. 2018, pp.1-27. hal-01735880

HAL Id: hal-01735880

<https://inria.hal.science/hal-01735880v1>

Submitted on 16 Mar 2018

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.

Analyser les réponses aux questions ouvertes des questionnaires de MOOC : Considérations méthodologiques

Aurélie Bayle – Inria Learning Lab
Janvier 2018

Résumé

Ce document a pour objectif de documenter la méthodologie d'analyse des questions ouvertes de nos questionnaires de satisfaction. Nous proposons une approche pragmatique, structurée et reproductible qui permet de prendre en compte les retours des participants à nos MOOC.

Nous prenons le MOOC "*Bioinformatique : algorithmes et génomes*" comme appui pour illustrer notre démarche.

En annexe se trouve une application de cette méthodologie à deux autres cours : *Introduction to a Web of Linked Data* et *Accessibilité numérique*.

Introduction

Inria Learning Lab conçoit des MOOC depuis 2013. Comme pour tout dispositif de formation, notre démarche inclut une étape de bilan et d'évaluation qui passe notamment par la prise en compte des retours et de la satisfaction à travers l'analyse des réponses à des questionnaires que nous leur proposons à différentes étapes du cours. Si les questions fermées font l'objet d'analyses quantitatives et de comparaisons sur l'ensemble des MOOC joués, les questions ouvertes sont plus difficilement analysées. Or, c'est bien souvent dans le discours "hors des cases" des participants que l'on entrevoit des éléments très pertinents sur leur expérience, leur ressenti et les axes d'amélioration que nous pouvons amorcer. L'objectif de ce document est de proposer une méthodologie qui facilite l'analyse manuelle des questions ouvertes de nos questionnaires de satisfaction.

Les questionnaires et la problématique des questions ouvertes

Les questionnaires

L'équipe d'Inria Learning Lab a conçu et propose aux participants des MOOC 3 questionnaires qui ont différents objectifs :

- Le questionnaire initial est intitulé "profil et attentes". Il vise à décrire le profil des apprenants (pays, ville, niveau d'études, situation professionnelle) et à cerner leurs motivations, leurs connaissances préalables et leurs attentes.
- Le questionnaire intermédiaire est intitulé "temps de travail et avis sur le MOOC". Il est diffusé au milieu du MOOC (début de 3^{ème} semaine pour un MOOC de 5 semaines)
- Le questionnaire final, diffusé en dernière semaine de MOOC, est intitulé "Suivi et avis global".

Ces questionnaires permettent de recueillir des données qui complètent celles obtenues via la plateforme FUN. Il s'agit de données provoquées (Van der Maren, 1996, p. 138) dans le sens où elles résultent d'une procédure construite pour les recueillir en fonction d'un format défini à l'avance (ici le questionnaire). On peut aussi dire qu'il s'agit de données déclaratives puisque les participants construisent ces données de manière volontaire et consciente par leur action de répondre au questionnaire. Les données que FUN met à disposition sont pour la plupart des données invoquées (ibid.), c'est à dire qui existent indépendamment de tout travail de recherche ou d'analyse.

Pour chaque MOOC, nous diffusons le même questionnaire, adapté en fonction des particularités de chaque cours. Cela nous permet de pouvoir faire le bilan de chaque MOOC sur une même base ainsi que des comparaisons entre les MOOC lorsque cela est pertinent.

Chaque question des questionnaires est facultative. Ainsi, certains questionnaires sont partiellement remplis.

Les questions ouvertes

Contrairement aux questions fermées qui peuvent faire l'objet d'un traitement statistique simple pour extraire les informations pertinentes, les questions ouvertes nécessitent un travail manuel.

En effet, il n'existe pas d'outil capable de comprendre et restituer le sens des réponses données sans intervention humaine. Des outils permettent de faire un certain nombre de comptages (nombre de mots, fréquence), de regrouper des expressions, de repérer des formes grammaticales, de faire ressortir des grandes thématiques, etc. Certains fournissent des représentations graphiques de réseaux de mots mais il ne s'agit que de manières de présenter les données, pour éventuellement faciliter l'analyse. Il s'agit d'outils d'aide à l'analyse et non pas d'outils d'analyse.

La lecture des réponses à ces questions ouvertes est utile et peut évidemment permettre de se faire une idée globale de l'appréciation du cours. Elle fait d'ailleurs partie de notre méthodologie. Cependant, chaque lecteur interprétera et retiendra des informations de manière subjective en fonction de ses expériences, de ses croyances (Berger & Luckmann, 1966 ; Schutz, 1973) ou encore de ses attentes, même inconscientes. Sans vouloir gommer totalement la subjectivité, nécessaire à l'analyse de phénomènes humains, notre objectif est de tendre vers une méthodologie d'analyse un peu plus objective, reproductible pour l'ensemble de nos MOOC.

Les questions ouvertes sont, par définition, des questions qui appellent à des réponses très diverses, sur la forme comme sur le fond. Les répondants peuvent écrire ce qu'ils souhaitent dans un espace de commentaire. Il est donc impossible d'extraire des informations facilement et rapidement de ces données au contraire de questions fermées de type vrai/faux ou seules deux variables existent.

Si l'on prend au hasard 5 réponses à la question sur les points d'amélioration, on se rend vite compte de l'hétérogénéité des réponses et de la complexité du traitement :

- *Ce mooc doit être étalé sur plusieurs cours pas celui la seulement. Alors vous devez nous programmer d'autres cours suites de celui ci. Je vous prie de le faire. Merci pour vos efforts et de nous avoir consacré tout ce temps. Le grand merci au professeur. On attend la suite.*
- *???*
- *Un petit regret de ne pas aller plus loin sur certains points*
- *-*
- *Je ne vois rien à améliorer en même temps c'est mon premier MOOC*

Certains messages contiennent beaucoup d'informations différentes, pertinentes ou non par rapport à la question. On peut noter une différence importante au niveau de la taille des messages (de 1 à 50 mots). Il faut également mentionner l'orthographe qui peut poser problème dans les traitements automatiques ultérieurs.

Le présent document analyse les réponses à deux questions ouvertes du MOOC "Bioinformatique : algorithmes et génomes" et présente différentes techniques qui peuvent aider à dégager les éléments pertinents mentionnés par les participants dans leurs réponses.

Notons bien qu'il ne s'agit pas d'un mode d'emploi à suivre à la lettre. En fonction de ce que l'on cherche et de la nature des données à traiter, une technique peut être plus intéressante qu'une autre.

Présentation des données

La présente analyse porte sur deux questions proposées dans le dernier questionnaire proposé aux participants du MOOC "Bioinformatique : algorithmes et génomes" (session 2 diffusée entre le 9 mai et le 12 juin 2016) :

- Quels sont, selon vous, les points positifs de ce cours ?
- Quels points d'amélioration nous suggèreriez-vous ?

Il faut noter qu'une proportion faible et très variable des participants à nos MOOC répondent aux questionnaires que nous proposons (de 0,5 à 20,5% des inscrits, 7% en moyenne) et que les questions ouvertes ne recueillent par une proportion de réponses importante par rapport aux questions fermées (entre 4 et 79%, 23% en moyenne).

Sur les 3270 inscrits au MOOC Bioinformatique, 117 ont répondu au questionnaire 3, soit 3,5%. Puisqu'aucune question n'est obligatoire, les participants peuvent ou non répondre aux différentes questions. Ainsi, 77 ont répondu à la question portant sur les points positifs et 63 sur les points d'amélioration.



Figure 1 : proportion de réponses par rapport au nombre d'inscrits

Préparation et traitement des données

Avant d'analyser à proprement parler le contenu de ces réponses, une préparation et un traitement s'avèrent nécessaires puisque les données doivent être organisées, nettoyées et homogénéisées.

Nous travaillons à partir des données brutes extraites de l'outil *LimeSurvey* que nous utilisons pour créer et diffuser nos questionnaires. Nous extrayons un fichier .xls pour chaque questionnaire de chaque MOOC qui présente les réponses à chaque question. Nous copions, depuis ce fichier, les réponses qui correspondent à la question à analyser (ex : les points d'amélioration) dans un fichier .txt pour faciliter la manipulation des données et leur traitement par différents outils. En fonction du type de réponses, nous pouvons également créer un fichier tableur pour coder les commentaires et les analyser.

La première étape consiste à lire l'ensemble des réponses pour se faire une idée globale du contenu, se l'approprier avant de procéder à l'analyse, et de dégager quelques grandes thématiques si le même type de réponses revient à plusieurs reprises.

Vient ensuite une étape de nettoyage des données. Il s'agit essentiellement de corriger les coquilles ("Pyton" → "Python", "interessant" → "intéressant") et les abréviations (pb pour problème, - pour moins) qui pourraient biaiser le comptage automatique des mots. Nous supprimons également les signatures et les sauts de ligne à l'intérieur des commentaires pour qu'à un paragraphe corresponde un commentaire. Une fois ce travail de nettoyage fait, nous sauvegardons le fichier pour y revenir par la suite.

Dans un autre fichier qui servira au traitement automatique, nous supprimons la ponctuation (virgule, point-virgule, deux points, point d'interrogation, point d'exclamation, parenthèses, à l'exception des points et la casse. Sur Mac, il faut également remplacer les apostrophes. Beaucoup d'outils ne reconnaissent que le caractère "'" comme une apostrophe.

Une fois que l'on a un fichier nettoyé, le traitement pour l'analyse peut commencer.

Analyse de la question sur les points positifs du cours

Comptage des mots

Nous commençons par faire un comptage du nombre d'occurrences à l'aide d'un outil d'analyse lexicale pour repérer les termes les plus employés ainsi que les éléments récurrents afin de dégager des idées-clés.

Comme pour tout champ de commentaire libre, il est possible d'avoir des commentaires en doublon, hors sujet, mal formulés, incompréhensibles, ou encore sans contenu ("RAS"). Ces types de réponses que nous appelons "hors sujet / HS" ne sont pas pris en compte dans notre analyse

Sur les 77 réponses qui ont été données à cette question, une seule était hors sujet puisqu'il s'agissait d'une suggestion d'amélioration :

"un support écrit équivalent à la vidéo. Des fois retrouver une phrase dite ou un élément dans la vidéo c'est contraignant".

Nous avons choisi de l'ôter du corpus d'analyse de cette question et de l'inclure dans l'analyse des points d'amélioration.

Outil utilisé et unités de comptage

Nous utilisons [AntConc](#), développé par Laurence Antony (Antony, 2014), un outil d'aide à l'analyse de texte qui comporte de nombreuses fonctionnalités dont un tri des mots par fréquence et un concordancier. Nous l'avons choisi parce qu'il est l'un des rares outils gratuits qui possède toutes les fonctionnalités que nous recherchions et qui fonctionne sur tous les systèmes d'exploitation, et simple d'installation comme d'utilisation.

La plupart des programmes de ce type fonctionnent de la même manière : on importe le fichier .txt à analyser et on l'explore de différentes manières.

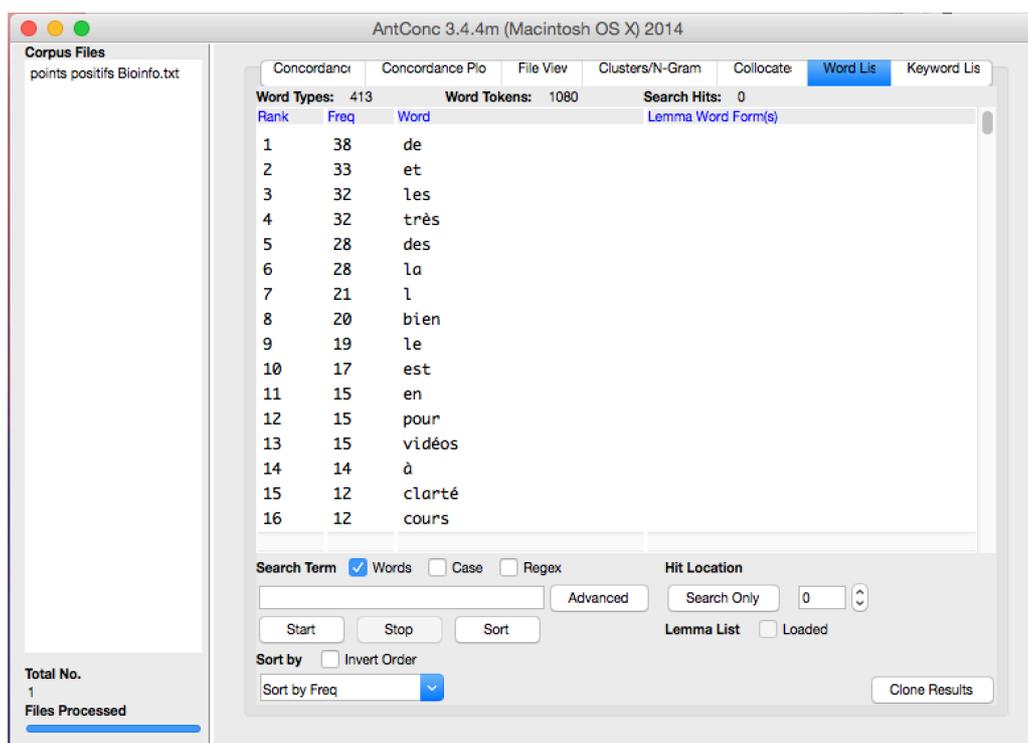


Figure 2 : capture d'écran de l'interface AntConc

Dans le logiciel AntConc (cf. figure 2), le fichier à analyser apparaît dans la colonne de gauche quand il est importé. Les onglets en haut ("concordance", "concordance plot", "file view", etc) correspondent aux différents outils de traitement. Nous nous intéressons ici à "Word List" pour lister les termes.

Nous commençons par générer la liste brute des mots. Le programme indique qu'il y a un total de 1080 mots et 413 mots différents. Les mots sont classés par fréquence.

Comme on peut le voir sur la capture d'écran, les mots les plus fréquents sont majoritairement des « stop words » aussi appelés mots vides. Il s'agit de mots très communs, caractéristiques d'une langue et qui ne sont pas significatifs comme par exemple les articles, les pronoms, les prépositions, etc. qui représentent la grande majorité des termes du texte mais n'ont pas de fonction sémantique ou lexicale.

AntConc a une fonctionnalité qui permet de mettre de côté les « stop words ». Il est même possible d'importer sa propre liste de « stop words » ce qui représente un avantage de cet outil en comparaison d'autres outils testés. Nous utilisons la [liste de « stop words »](#) du français élaborée par l'université de Neuchâtel. Cette liste contient 463 mots.

Une fois les « stop words » exclus, il ne reste plus que 513 mots dont 318 différents.

Synthèse des informations quantitatives

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points positifs à partir de AntConc.

Nombre d'inscrits au cours	3270 inscrits
Nombre de répondants au questionnaire 3	117
Nombre de réponses à la question	77
Taux de réponse à la question	64,9%
Nombre de réponses prises en compte / Nb de commentaires Hors Sujet	76 / 1
Nombre de mots	1080
Nombre de mots différents	413
Nombre de mots sans stop words	513
Nombre de mots différents sans stop words	318
Commentaire le plus long (en nombre de mots)	57
Commentaire le plus court (en nombre de mots)	1

Tableau 1 : synthèse des informations quantitatives - points positifs

Le texte de l'ensemble des réponses fait donc ressortir 318 mots potentiellement intéressants pour l'analyse sémantique. Il est alors intéressant de se concentrer sur ceux qui reviennent le plus fréquemment.

Termes les plus fréquents

Dans le tableau 2, les 10 termes les plus fréquents sont listés.

Termes	Nombre d'occurrences
vidéos	15
clarté	12
cours	12
clair	10
python	9
exercices	8
explications	6
orateur	6
accessible	5
bioinformatique	5

Tableau 2 : liste des termes les plus fréquents

Cette liste, simple à obtenir, donne déjà un certain nombre d'éléments sur les réponses les plus souvent données par les répondants au questionnaire. Les termes "vidéos", "cours", "exercices", "explications" et "orateur" indiquent les points de satisfaction des répondants. Les termes "clarté", "clair", "accessible" sont des termes positifs mais il reste à savoir ce qu'ils qualifient. Nous verrons cela avec les concordances.

Lemmatisation

Notons tout de même les limites du classement automatique brut des termes. Parmi les termes les plus fréquents, on retrouve "clair" et "claire" qui sont de la même racine et ont la même sémantique. En les regroupant, ils représentent 24 occurrences. En ajoutant les autres termes de la même famille (clarté, clairs, clairement, etc.), on arrive à un nombre encore plus élevé.

D'où l'intérêt de ce qu'on appelle la lemmatisation qui consiste à regrouper les termes de la même famille et à ne considérer que la forme canonique du mot, le lemme. Dans l'exemple ci-dessus, il s'agit de ranger derrière le mot "clair" toutes ses déclinaisons.

AntConc permet également d'importer une liste de lemmes. Nous avons utilisé Lefff (Lexique des Formes Fléchies du Français) de Sagot (2010) qui comporte 399368 entrées.

Le tableau 3 liste les termes lemmatisés :

Lemme	Fréq.	Formes					
vidéo	18	vidéo	3	vidéos	15		
clair	15	clair	10	claire	1	claires	4
clarté	12	clarté	12				
en cours	12	cours	12				
python	9	python	9				
exercice	8	exercices	8				
explication	6	explications	6				
expliquer	6	explique	1	expliqué	4	expliqués	1
orateur	6	orateur	6				
permettre	6	permet	3	permettant	1	permettent	2
accessible	5	accessible	5				
algorithme	5	algorithme	1	algorithmes	4		

Tableau 3 : liste des termes lemmatisés

Ce procédé permet d'affiner un peu la quantification des termes. Toutefois, il faudrait, pour se faire une idée plus précise de l'univers lexical employé, pouvoir aller plus loin que la lemmatisation pour rassembler automatiquement les termes de même famille (pour l'exemple précédent "clair" et "clarté").

Expressions les plus fréquentes

Nous pouvons également repérer les groupes de mots (clusters) les plus fréquents. Nous listons ici les 10 premières expressions à 3 mots et celles à 4 mots qui apparaissent au moins 2 fois :

Expressions à 3 mots	Fréq.	Expressions à 4 mots	Fréq.
Les vidéos sont	4	Cours très clair et	2
Clarté des explications	3	Des algorithmes en python	2
De la bioinformatique	3	Envie d'en savoir	2
Très clair et	3	J'ai beaucoup aimé	2
Ai beaucoup aimé	2		
Algorithmes en python	2		
Clarté des vidéos	2		
Cours est très	2		
Cours très bien	2		
Cours très clair	2		

Tableau 4 : expressions les plus fréquentes

Nous retrouvons les mêmes thématiques que celles de la liste des mots. L'intérêt ici est d'avoir un peu plus de contexte. On voit que le mot "clarté" qui apparaît en tout 12 fois est utilisé 3 fois pour parler des explications et 2 fois pour parler du cours.

Ces repérages ne constituent pas une analyse mais sont des aides qui permettent d'explorer le corpus et d'en tirer du sens.

Concordances

En allant encore un peu plus loin, il est possible, à partir de ces termes et expressions les plus fréquents, d'observer leurs contextes de production, c'est-à-dire, à l'aide d'un concordancier, d'aller voir les termes qui apparaissent avant ou après lorsque cela est pertinent.

Par exemple, il peut être intéressant de savoir ce qu'ont écrit les répondants après ""clarté" ou encore "les vidéos sont".

AntConc permet de travailler sur les concordances. On recherche un terme et la liste des occurrences apparaît avec un nombre défini de caractères du contexte à gauche et du contexte à droite.

Nous illustrons ce traitement avec les 2 occurrences les plus fréquentes. Nous avons continué de travailler à partir du fichier sans ponctuation mais il serait possible ici de reprendre un texte ponctué.

Vidéos

Le terme le plus fréquent était "vidéos". Voici l'extraction du concordancier :

sans vouloir tout dire sont en réalité. clarté des vidéos durée des e. excellent liens entre informatique et biologie e de l'intervenant. progression intéressante. les facile pour gérer son temps qu'avec des très clair et didactique un sujet passionnant des qu'ils ne sont en réalité. clarté des informatique sans être simpliste format court des la clarté de l'orateur la durée des galement. les points positifs sont la clarté des e façon didactique. les explications tant dans la avoir pu y consacré plus de temps. les des vidéos durée des vidéos bon séquençage les informatique très intéressant et très complet les exercices simples qui encouragent à continuer. les	vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos vidéos	bien calibrées pour leur durée clarté des e bon séquençage les vidéos sont claires et claires. exercices et quizz intéressants cl court précis et compréhensible la pédagog de 15 20 minutes. très positif bravo à vous de qualité je pensais que la bioinformatiqu durée des vidéos bon séquençage les extrêmement apprécié. l'orateur est excelle idéale le découpage des séances les quiz et la que les notebook python sont bien faites. t sont assez courtes pour rester concentré. l sont claires et pas trop longues. c'est sont courtes et claires. on peu donc les sont détaillées les supports complémentair
---	--	--

Tableau 5 : extrait du concordancier - terme "vidéos"

Le terme recherché se trouve au milieu avec son contexte à droite et à gauche. Nous avons mis en évidence, en gras, les éléments qui se rapportent au terme "vidéos". On voit qu'ils sont tantôt à droite (ex : "vidéos claires"), tantôt à gauche ("clarté des vidéos") et tantôt des deux côtés du mot ("durée des vidéos idéale"). Avoir d'un seul coup d'œil tous les contextes d'un terme permet de se faire une idée précise de son utilisation et de repérer des récurrences.

On peut dire ici que les points positifs liés aux vidéos sont leur durée (vidéos courtes) et leur clarté.

Il est possible de faire le même travail pour chacun des éléments pour lesquels on veut savoir ce que les participants ont dit (exercices, enseignants, rythme, forums, etc.) en choisissant les bons mots clés.

Clarté

Prenons à présent un terme qualificatif : le terme "clarté" qui apparaît 12 fois.

tivement clairement assez peu de prérequis que qu'ils ne sont en réalité. ès clairement et les schémas détaillés simplicité ormatique et initiation à la génétique. les plus n. excellente présentation. l'enthousiasme et la e vidéos claires. exercices et quizz intéressants	clarté clarté clarté clarté clarté clarté clarté	ressources complémentaires. clair et pédago des vidéos durée des vidéos bon séquençage révisions de bases approche pas que informa des explications les algorithmes et le choi des exposés. pédagogique progressif laisse cours très bien construit permettant probab
--	--	---

Nous retenons donc une grande satisfaction globale de la part des répondants, plus particulièrement, notons la clarté des exercices et des vidéos, l'organisation du cours, la durée des vidéos, la simplicité et la clarté des explications.

Venons-en à présent à une question sans doute plus intéressante à analyser dans le cadre d'un bilan de formation : les points d'amélioration suggérés.

Analyse des points d'amélioration

Dans cette section, nous traitons la question des points d'amélioration proposés par les répondants au questionnaire.

La question sur les points positifs nous a permis d'expliciter et d'illustrer la méthodologie employée d'un point de vue essentiellement quantitatif. Pour les approches similaires à celles utilisées plus haut pour les points positifs, nous nous concentrons sur les données et les résultats. Nous nous attarderons plus sur les techniques qualitatives employées du fait de la spécificité de cette question et des réponses qui ont été données.

Balayage des commentaires et traitement avant analyse

A la lecture de l'ensemble des commentaires, nous retenons une très grande diversité de réponses. Les répondants font principalement des suggestions d'amélioration mais certains commentent également le cours passé ou leur expérience du cours.

Il semble à première vue qu'une analyse presque uniquement quantitative n'apportera pas d'éléments probants. De plus, l'intérêt de ce type de question est de recueillir des témoignages et des idées pour améliorer les sessions suivantes. Ainsi, il semble pertinent de prendre en compte chacune des réponses de manière plus qualitative avec un traitement manuel. Au vu du nombre de réponses (63), ceci est tout à fait possible.

Après un nettoyage sur la forme (coquilles, sauts de ligne, etc.), nous avons extrait les réponses dans un fichier de tableur. Nous avons choisi de catégoriser les commentaires de manière inductive suite à la lecture que nous avons faite. Ainsi, nous avons codé chaque commentaire selon la ou les thématiques qu'il abordait. Nous nous sommes limités à maximum 3 catégories par commentaire. Au fil de la lecture et du codage, les catégories se sont enrichies et/ou affinées.

Ce travail a permis de dégager 9 catégories. Nous avons codé chacun des commentaires en fonction de ces catégories.

Difficulté/prérequis	16
Exercices	16
HS	9
Python	10
Supports/ressources	8
présentation/organisation	6
Approfondissement/exemples	7
point contenu précis	4
?	2

Tableau 7 : catégories définies suite au codage

Traitement avant comptage

Nous avons expliqué plus haut qu'il est nécessaire de nettoyer le texte avant de l'analyser. Pour l'analyse des points positifs, le texte ne présentait pas de problème particulier. Nous avons simplement corrigé quelques coquilles, supprimé les sauts de ligne et extrait un commentaire qui ne traitait pas de points positifs mais donnait une suggestion d'amélioration.

Les réponses à la question sur les points d'amélioration sont plus hétérogènes. Nous avons dû supprimer une signature. Deux commentaires étaient vides ("- " et "??"). Il a également fallu écarter les réponses "vides", c'est-à-dire sans contenu pertinent au regard de la question, par exemple "*rien ne me vient à*

l'esprit", "J'ai trouvé ce cours génial, pour moi rien à améliorer!" ou encore "aucun", "rien! ce cours est génial et me donne envie de reprendre mes études en bio. faites une suite :D !!".

Nous pouvons nous poser la question de la manière de considérer ces réponses qui certes n'apportent pas d'élément à la question posée mais donnent quand même des indications très intéressantes sur la grande satisfaction des répondants ou des indices sur leur perception (ex : *"Comme introduction a la bioinformatique, je trouve ce mooc très bien fait. Rien à changer ! Les notebooks amènent un outil vraiment intéressant"*).

Se pose aussi la question des commentaires que l'on pourrait qualifier d'"hybrides" qui comportent un élément de réponse à la question et un élément hors sujet. Par exemple : *"plus d'exemple illustrés d'arbres phylogénétiques. MERCI BEAUCOUP A TOUS"*. Nous avons choisi de les conserver.

Au total, nous avons écarté 9 réponses.

Éléments quantitatifs

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points d'amélioration à partir de AntConc. Les comptages se font sur les réponses non HS.

Nombre d'inscrits au cours	3270 inscrits
Nombre de répondants au questionnaire 3	117
Nombre de réponses à la question	63 (+1 de la question sur les points positifs)
Nombre de réponses prises en compte / Nb de commentaires HS	55 / 9
Nombre de mots	1362
Nombre de mots différents	488
Nombre de mots sans stop words	552
Nombre de mots différents sans stop words	369
Commentaire le plus long (en nombre de mots)	71
Commentaire le plus court (en nombre de mots)	3

Tableau 8 : synthèse des informations quantitatives - points d'amélioration

Il est intéressant de noter que pour un nombre moins important de réponses par rapport à la question sur les points positifs (55 vs. 76), le nombre de mots est plus important.

Termes les plus fréquents

Nous listons les 10 termes les plus fréquents :

Termes	Nombre d'occurrences
exercices	12
cours	10
exemple	9
questions	9
algorithmes	8
quizz	8
faire	7
python	7
mooc	6
vidéo	6

Tableau 9 : liste des termes les plus fréquents

Difficulté/prérequis

Les commentaires se répartissent entre ceux qui affirment que le cours est trop difficile et ceux qui le trouvent trop facile.

- accentuer la difficulté du contenu, des quiz. Environ 9 personnes considèrent le cours et/ou les exercices trop faciles
- préciser les prérequis (3) (connaissances en algorithmes, statistiques, programmation) et proposer des éléments pour se mettre à niveau
- commentaires difficiles à interpréter (2) ("*Prière d'améliorer le niveau des questions du Quiz*" – impossible de dire si cette personne trouve les quiz trop faciles ou trop difficiles)

Exercices

Certains trouvent les exercices trop faciles, d'autres les trouvent un peu trop difficiles.

On trouve également des commentaires qui portent sur des aspects particuliers : demande de plus de mise en pratique, réduire le nombre de questions de logique, ajouter des exercices de programmation pratique et de réflexion, préciser la différence entre quiz et exercices, préciser certaines questions, allonger ou raccourcir les quiz, rajouter des essais, "*remplacer les exercices d'écriture de pseudo-code par des logigrammes*", etc.

Python

Une personne suggère de changer de langage, une autre a trouvé que les explications sur Python allaient trop vite. Un apprenant n'a pas trouvé que les Notebooks étaient un bon moyen d'apprendre à concevoir des algorithmes et aurait préféré des astuces et techniques pour utiliser directement Python, un autre souhaiterait un tuto vidéo pour utiliser les notebooks, deux autres auraient besoin d'une vidéo ou d'un cours complet sur Python pour les débutants.

Supports/ressources

Plus de ressources, des animations 3D, des mots-clés, plus de détails dans les corrections, plus de supports écrits et une bibliographie sont demandés.

Présentation/organisation

Les commentaires portent sur la distinction quiz/exercices, les coquilles, le manque d'homogénéité du cours, l'accès au MOOC au delà de la fermeture du cours, une demande de plus de détails dans les transcriptions pour ceux qui ne peuvent pas accéder aux vidéos, une demande de vidéos plus longues.

Approfondissement/exemples

Les commentaires de cette catégorie demandent plus de détails sur la biologie ou les algorithmes. Un répondant souhaiterait des exemples avant les exercices.

Point contenu précis

3 commentaires portent sur des points de contenu très précis : un algorithme sur les chaînes de Markov, plus d'explication sur "la méthode de calcul pour remonter un arbre phylogénétique n'est pas suffisamment expliquée, à mon sens", plus d'exemples illustrés d'arbres phylogénétiques, des exemples d'accident génétique (mutation), des applications concrètes de la bioinformatique, des liens avec la bioéthique, des exemples d'expérimentation en labo.

Conclusion générale

En fonction du type de question et de ce que l'on cherche, la méthodologie de traitement et d'analyse des données ne sera pas toujours similaire.

Nous avons ici travaillé sur des questions ouvertes assez précises et quelque peu opposées puisque l'une demandait quels étaient les points positifs du cours tandis que l'autre demandait aux répondants de donner des idées pour améliorer le cours. Pour la première, des points positifs étaient attendus sur un certain

nombre d'aspects (le cours, les vidéos, les exercices, l'enseignant, l'animation, etc.). Ce type de question est plus susceptible d'obtenir un « consensus ».

Le repérage des termes les plus fréquents associé à un balayage pour repérer des éléments pertinents même si peu fréquents s'avère suffisant pour analyser ce type de questions et dresser le bilan.

Pour la question portant sur les points d'amélioration, nous avons dû adopter une méthode plus qualitative et mettre l'accent sur le codage des réponses pour les catégoriser dans de grands ensembles. Ceci nous a permis de prendre en compte les commentaires et suggestions de chacun. En effet, les réponses concernant les points d'amélioration font moins fréquemment « consensus » car elles sont plus spécifiques et plus personnelles.

Lorsque l'on a à faire à des questions plus larges, comme par exemple "que pensez-vous du forum de discussion ?", un traitement quantitatif comme nous l'avons fait peut être pertinent mais il est souvent nécessaire de catégoriser les commentaires à la main, soit en fonction de catégories prédéfinies, soit en créant les catégories au fil de l'analyse.

De nombreux outils existent pour aider à l'analyse de textes. On les retrouve souvent sous l'acronyme CAQDAS (Computer Assisted/Aided Qualitative Data Analysis Software). Comme nous l'avons déjà dit, ces outils ne permettent pas d'analyser les données, ils proposent seulement des fonctionnalités qui permettent d'aider l'humain à faire une analyse. Avant de choisir l'outil AntConc, nous en avons testé un certain nombre. En fonction de l'objectif d'analyse que l'on a, de ses compétences et du type de données à traiter, certains outils peuvent être plus pertinents et plus utiles que d'autres. Le choix est du ressort de l'analyste. Du fait des différences de calcul entre chacun des outils (comptages différents car manières de définir un mot différentes et problématiques inhérentes à l'analyse du langage naturel), il nous semble important de n'utiliser qu'un outil (ou deux à des fins de comparaison éventuellement) pour toutes les analyses réalisées, d'opérer les mêmes traitements préalables sur les données (corrections, suppression de la ponctuation, etc.) et de bien documenter la démarche méthodologique adoptée.

Références bibliographiques

Anthony, L. (2014). *AntConc* (Version 3.4.3) [logiciel]. Tokyo : Waseda University. Disponible à <http://www.laurenceanthony.net/>

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. Garden City: Anchor Books

Sagot, B. (2010). Lexique des Formes Fléchies du Français. Version 3.4. Disponible à <http://alpage.inria.fr/~sagot/leff.html>

Schutz, A. (1973). *Collected papers I : The problem of social reality*. La Hague : Martinus Hijhoff.

Van der Maren, J.-M. (1996). *Méthodes de recherche pour l'éducation*. Bruxelles : De Boeck

Wordart (2017). Wordart.com [site web]. <https://wordart.com>

Annexe 1 : Application aux questions sur les points positifs et d'amélioration du MOOC "Linked Data"

Présentation des données

La présente analyse porte sur deux questions proposées dans le dernier questionnaire proposé aux participants du MOOC "Introduction to a Web of Linked Data" (session 2 diffusée entre le 9 mai et le 4 juin 2017) :

- Quels sont, selon vous, les points positifs de ce cours ?
- Quels points d'amélioration nous suggèreriez-vous ?

Sur les 1600 inscrits (parmi lesquels 278 actifs), 47 ont répondu au questionnaire 3, soit 2,9% des inscrits et 16,9% des actifs. Puisqu'aucune question n'est obligatoire, les participants peuvent ou non répondre aux différentes questions. Ainsi, 20 ont répondu à la question portant sur les points positifs et 19 sur les points d'amélioration.

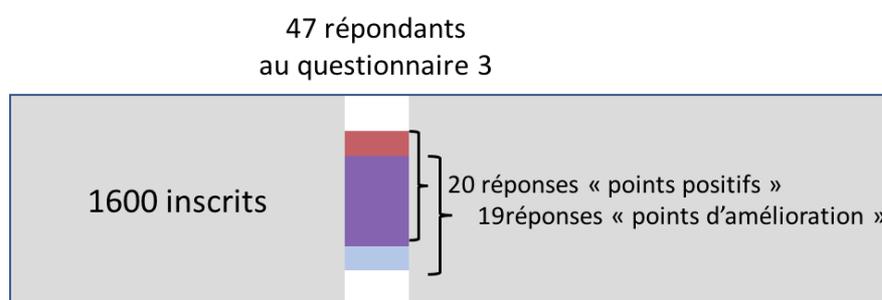


Figure 5 : : proportion de réponses par rapport au nombre d'inscrits

Analyse de la question sur les points positifs du cours

Synthèse des informations quantitatives

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points positifs à partir de AntConc.

Ce cours était en anglais. Les apprenants avaient la possibilité de répondre au questionnaire en français ou en anglais. Au vu de leur nombre, les réponses en français sont exclues du comptage mais seront prises en compte dans les catégorisations.

Nombre d'inscrits au cours	1600 inscrits
Nombre d'actifs dans le cours	278
Nombre de répondants au questionnaire 3	47
Nombre de réponses à la question	20
Nombre de réponses prises en compte / Nb de commentaires Hors Sujet	20 / 0
Nombre de réponses en anglais	18
Nombre de réponses en français	2
Nombre de mots	356
Nombre de mots différents	204
Nombre de mots sans stop words	173
Nombre de mots différents sans stop words	134

Commentaire le plus long (en nombre de mots)	70
Commentaire le plus court (en nombre de mots)	2

Tableau 11 : synthèse des informations quantitatives - points positifs

Réponses en français : "Clair, concis." Et "très pédagogique."

Termes les plus fréquents

Nous listons les 10 termes les plus fréquents avec leur lématisation:

Fréq.	Termes	Occurrences
6	understand	understand 5 understanding 1
5	data	data 5
5	material	material 5
4	clear	clear 4
4	tool	tool 1 tools 3
4	video	videos 4
3	part	part 1 parts 2
3	tag	tag 2 tagging 1
3	level	level 3
2	additional	

Tableau 12 : liste des termes les plus fréquents

Expressions les plus fréquentes

Expressions à 3 mots	Fréq.	Expressions à 4 mots	Fréq.
Help to understand	3	Help to understand better	2
A lot of	2	To understand better the	2
Some of the	2		
To understand better	2		
Understand better the	2		

Tableau 13 : expressions les plus fréquentes

Concordances

Understand

Le terme le plus fréquent était "**understand**". Voici l'extraction du concordancier :

data and data formats. Definitely improved my material is high level and help to includes practical parts where learner can test . demos: offer useful tools. quizzes: help to videos. End of week exercises: help to overview of the field. I did not	understand understand understand understand understand understand	of linked open data. The material is learning topics. There are explanations of material. Well structured, I like better the videos. End of week exercise better the subjects The course is an anything fully : it gave me the curiosi
--	--	--

Tableau 14 : extrait du concordancier - terme "understand"

Data

Prenons à présent le terme "**data**" qui apparaît 5 fois.

a wide range of Web of link the aspect of "cost" of tagging (add meta	data data	technology. Refer to multiple online tools. C) to provided data. And that designer needs
---	----------------------------	---

Vidéos

Les vidéos sont appréciées, plus particulièrement leur courte durée.

Autre

Les commentaires inclassables abordent le temps alloué pour faire les exercices, les modalités d'exercices originales et intéressantes, les ressources et activités pédagogiques proposées (demos, quiz, exercices de fin de semaine) et la pédagogie.

Analyse des points d'amélioration

Éléments quantitatifs

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points positifs à partir de AntConc. Les comptages se font sur les réponses non HS et en anglais.

Nombre d'inscrits au cours	1600
Nombre d'actifs	278
Nombre de répondants au questionnaire 3	47
Nombre de réponses à la question	19 (dont 3 en français)
Nombre de réponses prises en compte / Nb de commentaires HS	16 / 0
Nombre de mots	545
Nombre de mots différents	274
Nombre de mots sans stop words	238
Nombre de mots différents sans stop words	175
Commentaire le plus long (en nombre de mots)	125
Commentaire le plus court (en nombre de mots)	7

Tableau 18 : synthèse des informations quantitatives - points d'amélioration

Termes les plus fréquents

Nous listons les 10 termes les plus fréquents :

Termes	Nombre d'occurrences
Data	7
Mooc	5
Example	4
Link	4
Format	4
Trust	4
Time	4
Web	4
Deeper	3

Tableau 19 : liste des termes les plus fréquents

On voit, comme c'était le cas plus haut, que les termes contenus dans le titre du cours et le terme « MOOC » apparaissent le plus fréquemment. Pour une analyse d'un corpus plus dense, il serait judicieux de retirer ces termes pour ne pas parasiter l'analyse.

Annexe 2 : Application aux questions sur les points positifs et d'amélioration du MOOC "Accessibilité numérique"

Présentation des données

La présente analyse porte sur deux questions proposées dans le dernier questionnaire proposé aux participants du MOOC "Accessibilité numérique" (session 2 diffusée entre le 9 mai et le 11 juin 2017) :

- Quels sont, selon vous, les points positifs de ce cours ?
- Quels points d'amélioration nous suggèreriez-vous ?

Sur les 1925 inscrits, 109 ont répondu au questionnaire 3, soit 5,6%. Puisqu'aucune question n'est obligatoire, les participants peuvent ou non répondre aux différentes questions. Ainsi, 69 ont répondu à la question portant sur les points positifs et 50 sur les points d'amélioration.

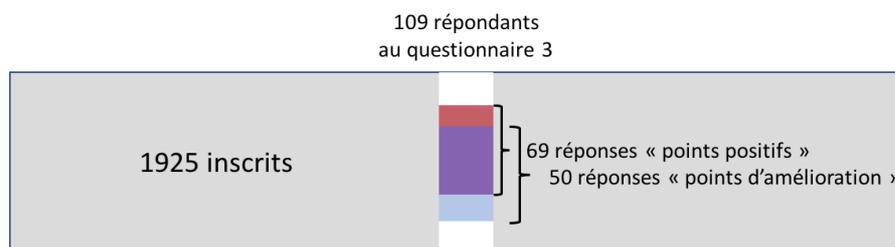


Figure 11 : proportion de réponses par rapport au nombre d'inscrits

Analyse de la question sur les points positifs du cours

Synthèse des informations quantitatives

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points positifs à partir de AntConc.

Nombre d'inscrits au cours	1925 inscrits
Nombre de répondants au questionnaire 3	109
Nombre de réponses à la question	69
Nombre de réponses prises en compte / Nb de commentaires Hors Sujet	68
Nombre de mots	1168
Nombre de mots différents	466
Nombre de mots sans stop words	561
Nombre de mots différents sans stop words	369
Commentaire le plus long (en nombre de mots)	133
Commentaire le plus court (en nombre de mots)	2

Tableau 20 : synthèse des informations quantitatives - points positifs

Termes les plus fréquents

Nous listons les 10 termes les plus fréquents :

Termes	Nombre d'occurrences
cours	14

accessibilité	9
lecteur	8
complet	7
handicap	7
permet	7
sujet	7
vidéos	7
accessible	6
numérique	6

Tableau 21 : liste des termes les plus fréquents

Expressions les plus fréquentes

Nous pouvons également repérer les groupes de mots (clusters) les plus fréquents. Nous listons ici les 10 premières expressions à 3 mots et celles à 4 mots qui apparaissent au moins 2 fois :

Expressions à 3 mots	Fréq.	Expressions à 4 mots	Fréq.
L'accessibilité numérique	4	L'accessibilité numérique et	2
La clarté des	3	La clarté des explications	2
Accessibilité numérique et	2		
Accessible à tous	2		
Aux personnes non	2		
Bien structuré complet	2		
Clarté des cours	2		
Clarté des explications	2		
Complet sur le	2		
De l'accessibilité	2		
De l'équipe			
De voir l'			
De vue sur			
Et de l'			
Le lecteur de			
Le livre numérique			
Le sous titrage			
Point de vue			
Situation de handicap			
Sur l'accessibilité			

Tableau 22 : expressions les plus fréquentes

Concordances

Cours

Le terme le plus fréquent était "**cours**". Voici l'extraction du concordancier :

l'élocution de Pascal. Les permet d'acquérir la véritable information. Ce niveaux, la réactivité de l'équipe pédagogique e se repérer facilement. Félicitations! Clarté du points sont positifs. EXCELLENT Mooc. clarté des	cours cours cours cours	sont de qualité, je ne m'attendais est une prise de conscience pour tous. très instructifs Tour complet de l'A. , le lecteur de vidéo Tous les points qualité de la formation, des intervenants e
---	--	---

Sujet / contenu

Les commentaires les plus nombreux portent sur le sujet. Les répondants ont apprécié d'avoir un cours sur l'accessibilité numérique. Sont évoqués le contexte, le grand angle de vue, les notions traitées, la sensibilisation, la réflexion.

Clarté / qualité

Le cours est jugé très clair, bien structuré, de grande qualité et accessible.

Autre

5 commentaires ne sont que des félicitations sur le cours ("très complet", "tous les points sont instructifs"). Une personne malentendante a apprécié le sous-titrage du Live. Des personnes évoquent "la facilité et l'attrait", la pédagogie, les illustrations, les exercices de fin de semaine

Vidéos / Aïana / livre numérique

Le lecteur Aïana, les vidéos et le livre numérique sont particulièrement appréciés. Plus largement, plusieurs commentaires font ressortir le souci de rendre le MOOC lui-même accessible

Intervenants

Les répondants ont apprécié les explications claires des intervenants, le fait qu'ils ne lisent pas leur diaporama, leur passion pour leur sujet. Ils sont "calmes, paisibles, agréables à écouter". Notons également la réactivité de l'équipe pédagogique

Niveau

Il revient à 5 reprises l'accessibilité du cours en termes de niveau

Forme

Sur la forme, les différents outils et fonctionnalités sont appréciés (lecteur Aïana, sous-titrage, alternance de vidéos et quiz, etc.) ainsi que l'organisation générale du cours.

Conclusion sur les points positifs

La question sur les points positifs est assez facile à analyser puisqu'il suffit d'une certaine manière de repérer les aspects que les répondants ont apprécié.

Nous retenons donc une grande satisfaction globale de la part des répondants, plus particulièrement, notons la clarté des exercices et des vidéos, l'organisation du cours, la durée des vidéos, la simplicité et la clarté des explications.

Venons-en à présent à la question des points d'amélioration suggérés.

Analyse des points d'amélioration

Balayage des commentaires et traitement avant analyse

Après un nettoyage sur la forme (coquilles, sauts de ligne, etc.), nous avons extrait les réponses dans un fichier de tableur. Nous avons choisi de catégoriser les commentaires de manière inductive suite à la lecture que nous avons faite. Ainsi, nous avons codé chaque commentaire selon la ou les thématiques qu'il abordait. Nous nous sommes limités à maximum 3 catégories par commentaire. Au fil de la lecture et du codage, les catégories se sont enrichies et/ou affinées.

Ce travail a permis de dégager 7 catégories. Nous avons codé chacun des commentaires en fonction de ces catégories.

Application / concret / exemples	24
Présentation / organisation	11
Outils / fonctionnalités	5

Contenu / exercices	5
Créer un autre cours	4
Interaction	3
HS	5

Tableau 26 : catégorisation des points d'amélioration

Éléments quantitatifs

Le tableau suivant synthétise un certain nombre d'informations quantitatives sur les réponses à la question sur les points positifs à partir de AntConc. Les comptages se font sur les réponses non HS.

Nombre d'inscrits au cours	1925 inscrits
Nombre de répondants au questionnaire 3	109
Nombre de réponses à la question	50
Nombre de réponses prises en compte / Nb de commentaires HS	45 / 5
Nombre de mots	1635
Nombre de mots différents	631
Nombre de mots sans stop words	712
Nombre de mots différents sans stop words	506
Commentaire le plus long (en nombre de mots)	127
Commentaire le plus court (en nombre de mots)	3

Tableau 27 : synthèse des informations quantitatives - points d'amélioration

Termes les plus fréquents

Nous listons les 10 termes les plus fréquents :

Termes	Nombre d'occurrences
mooc	10
numérique	8
vidéos	8
accessibilité	7
cours	7
personnes	6
lecteur	5
pratiques	5
serait	5
vidéo	5

Tableau 28 : liste des termes les plus fréquents

6 des 10 termes sont communs avec la question précédente. Il s'agit de termes liés à l'environnement MOOC les plus couramment rencontrés dans tous les questionnaires analysés (MOOC, vidéo, cours, exercice) mais aussi des termes spécifiques au sujet du MOOC : "accessibilité" et "lecteur".

Vidéos

<p>pédagogique. J'ai quelquefois trouvé les un peu trop longues et monotones. Des les 13 thématiques du RGAA (images, liens, manque la possibilité de passer à la rappelle pas ici immédiatement utiliser endroit apprécié une version ""à lire"" de chaque que j'ai appris des choses. Les (les téléchargements disponibles sous les pense qu'il serait intéressant sur des transformation de contenus accessibles. Une expliquant où sont les problèmes , ... Plus de tre du handicap , expliquant les contraintes (cf.</p>	<p>Vidéos vidéos vidéos vidéo vidéo vidéos vidéos vidéos vidéo vidéos vidéo</p>	<p>un peu trop longues et monotones. Des plus nombreuses mais de plus courtes durées etc). L'accessibilité numérique est a /quizz suivant de manière automatique à la pour traduction en langue de signes sous-titré qui soit plus structurée que les sous-titres sont longues et il faudrait plus les nécessitent trop de mise en forme pour d'illustrer vos problématiques ou solutions d'exemple avec la visite d'un de rencontre du handicap , expliquant les c sur la synthèse vocale) pour se rendre</p>
---	--	--

Tableau 31 : extrait du concordancier - terme "vidéos"

Accessibilité

<p>un réel plus pour mieux comprendre l' Ajouter des situations visibles de non situations de handicap. Parler des experts afin d'entrer davantage dans la partie ' du RGAA (images, liens, vidéos etc). L' amélioré pour un prochain Mooc sur l' sur un site présentant des difficultés d' montrer l'impacte réel que représente l'</p>	<p>accessibilité accessibilité accessibilité accessibilité accessibilité accessibilité accessibilité accessibilité</p>	<p>numérique (page web à construire). A de contenus pour une ou des situatio référents, de leur travail... plus d numérique & contributeurs (Comment c numérique est aussi l'affaire de tou numérique. Je pense qu'il serait in numérique, il aurait été intéressan numérique. Je pense en effet que de</p>
--	--	---

Tableau 32 : extrait du concordancier - terme "accessibilité"

Cours

<p>une formation beaucoup plus lourde. Support J'ai vraiment beaucoup apprécié ce Merci cependant pour votre MOOC ! Faire un construction et à la gestion de ce qu'il s'agit plus d'un que j'ai pu apprendre avec ce je n'aurais pas eues sans ce</p>	<p>Cours cours cours cours cours cours cours</p>	<p>pdf aucune suggestion J'ai vraiment beaucoup très instructifs et pour cela je remercie sur les outils CMS pour les développeurs . Comme les ressentis sont multiples et parti général que d'indications concrètes destinée général m'a finalement énormément intéressée . Peut-être prévoir un autre MOOC juste</p>
---	--	--

Tableau 33 : extrait du concordancier - terme "cours"

Synthèse des suggestions données

Si nous reprenons les catégories déterminées dans la section "balayage", nous pouvons résumer et synthétiser les différents commentaires :

Application / concret / exemples

Les commentaires de cette catégorie mettent en évidence le manque d'applications réelles, d'exemples concrets dans le cours. Les répondants auraient souhaité des aspects plus techniques, des témoignages, des mises en pratique, etc.

Présentation / organisation

Certains commentaires portent sur des aspects de forme et de présentation du cours (notamment les vidéos) et de la plateforme. Il est suggéré de réduire la taille des vidéos, d'augmenter et de baisser le rythme d'élocution des locuteurs. L'ergonomie de la plateforme est également mentionnée : problème pour localiser les quiz. Il est suggéré de permettre aux apprenants d'évaluer les questions des quiz

Outils / fonctionnalités

Deux personnes souhaiteraient avoir à disposition un support de cours pdf, notamment pour annoter. Trois remarques portent sur le lecteur Aïana : "à optimiser", sans précision et le fait qu'Aïana ne peut être utilisé qu'avec une connexion à Internet. Une suggestion porte sur l'utilisation du lecteur dans le livre numérique.

Contenu / exercices

On trouve dans cette catégorie des demandes de clarification pour quelques questions. On reproche aux quiz d'être trop faciles ou de trop porter sur des points de définitions et d'abréviations.

Créer un autre cours

4 personnes ont suggéré de créer un nouveau MOOC, soit pour approfondir, soit pour proposer un cours plus pratique ou plus technique.

Interaction

3 commentaires mentionnent les interactions entre apprenants. Une personne suggère un rendez-vous sur un réseau social. Il est suggéré d'ajouter un forum directement en dessous des vidéos. Enfin, une personne suggère de susciter plus d'échanges sur les forums.