

Représentations de la réussite au lycée : analyse exploratoire d'une question ouverte

Aurélie Seznec

Université de Bretagne Occidentale - Laboratoire LABERS Brest - France

Abstract

This presentation aims to analyse responses from high school students to an open question about their conception of success. One of the main hypotheses is that the different aspects of success that are highlighted may vary according to the context and the respondents' profile (sex, kind of schooling). We will explore and analyse our corpus (2608 responses) with the Sphinx IQ Quali software, with a particular stress on the coding protocol. Additionally, these first results will be compared to other researches about success and mainly the 1988 survey by Christian Baudelot on high school students in vocational education.

Résumé

L'objectif de cet article est d'analyser les réponses de lycéens et lycéennes à une question ouverte portant sur leur conception de la réussite. Une des principales hypothèses est que les différentes facettes de la réussite mises en avant évoluent suivant le contexte et le profil des enquêtés (sexe, filière de scolarisation). Nous tenterons donc d'explorer et d'analyser notre corpus de 2608 réponses avec le logiciel Sphinx IQ Quali en revenant plus particulièrement sur le protocole de codification. En parallèle, ces premiers résultats seront mis en perspective avec des recherches déjà effectuées autour de la réussite, principalement celles de Christian Baudelot, enquête réalisée auprès des lycéens de filière professionnelle (1988).

Mots-clés : population lycéenne ; question ouverte ; analyse exploratoire ; représentations de la réussite

1. Introduction

1.1. *Qu'est-ce que la réussite ?*

« Selon vous, qu'est-ce que la réussite ? » A cette question, chaque personne peut apporter sa propre réponse, il serait difficile de tenter de faire consensus autour d'une définition unique. Les mots que l'on peut mettre derrière le terme « réussite » dépendent en partie de notre trajectoire, de notre histoire, de notre parcours mais aussi de nos valeurs ou encore de nos projections dans l'avenir.

1.2. *Une question ouverte s'inscrivant dans une recherche sociologique*

Avant d'aller plus loin, il nous faut préciser que cette question de la réussite s'inscrit plus largement dans un travail de thèse portant sur les pratiques et stratégies éducatives familiales mises en place en dehors de l'école. Il s'agit d'une enquête quantitative menée dans onze établissements scolaires en région Bretagne. Chaque élève de la classe de seconde à la classe de terminale a été invité à répondre à un questionnaire en ligne. 2 608 élèves ont apporté une réponse à cette question ouverte (soit plus de 98% de l'échantillon), le faible taux de non-réponse à cette question signifie que celle-ci semble faire sens pour eux. Comment prendre en compte et analyser cette pluralité de réponses? Notre présentation est basée sur l'analyse exploratoire des réponses apportées à cette question ouverte et a débuté par un travail d'harmonisation et de correction orthographique des réponses. Le lexique brut compte 1 598

formes différentes pour un corpus total de 25 475 occurrences. La longueur médiane des réponses est de 8 mots. En utilisant le texte lemmatisé¹ et en ignorant les mots de moins de deux lettres, le lexique compte 1 442 mots et 16 432 occurrences.

2. Choix et posture d'une méthodologie de l' « entre-deux »

2.1. Mises en perspectives avec des recherches antérieures

L'intérêt pour l'analyse textuelle a été stimulé à la lecture des textes de Christian Baudelot (1988) et de Georges Felouzis (1993). Le travail et les analyses qu'ils ont effectués sur leurs corpus dans des contextes différents, en mobilisant des méthodologies et échantillons semblables à notre corpus, nous ont poussés à expérimenter l'analyse de données textuelles à notre tour. En quelques mots, Christian Baudelot s'appuyait sur un questionnaire adressé exclusivement à des lycéens professionnels (LEP), là où nos enquêtés sont inscrits dans des filières générales et technologiques (section suivie par 21% de notre échantillon). La recherche de Georges Felouzis, quant à elle, mobilisait les réponses de 276 lycéens et lycéennes de filières générales, technologiques et de LEP, qui ont été amenés à définir la réussite au cours d'un entretien.

2.2. Expérimentation de l'ADT avec Sphinx IQ Quali

Expérimenter, plonger dans son corpus, en s'appuyant et en remobilisant des recherches précédentes (Baudelot, Felouzis, Zazzo) permet de baliser davantage les procédés à mettre en œuvre. Nous avons donc décidé de suivre la procédure mise en place par Christian Baudelot, en créant une variable avec les formes graphiques les plus utilisées, reprenant les principaux mots. L'objectif est d'arriver à une centaine de mots pour se rapprocher de l'exploitation des réponses des lycéens de LEP². Nous avons décidé de simplifier davantage le lexique en utilisant le dictionnaire des mots-outils de Sphinx auquel nous avons par contre retiré toutes les formes de l'auxiliaire être et avoir, afin qu'ils soient pris en compte dans l'analyse, « leurs conceptions de la réussite dépendent fortement de leur position et de leur passé scolaire. Notamment entre ceux qui veulent « avoir » parce qu'ils n'ont rien ou peu de chose, et ceux qui veulent « être » parce qu'ils ont déjà un peu plus » (Felouzis, 1993).

Le travail présenté ici s'est effectué grâce au logiciel Sphinx IQ Quali, il est délicat de décrire un tel outil en quelques lignes, précisons simplement que nous nous sommes appuyés largement sur les travaux de Stéphane Ganassali (2014) pour expérimenter le logiciel.

3. Exploration et codification du corpus

3.1. Présentation du corpus à l'aide des verbatims

Dans l'objectif de se familiariser et de décrire le type de réponses données par les lycéens et lycéennes, nous avons décidé non pas de tirer au hasard, de façon aléatoire, des réponses dans le corpus, mais d'utiliser le protocole des verbatims, c'est-à-dire d'extraire les conceptions de la réussite les plus représentatifs d'une catégorie de répondants. La variable de contexte choisie ici est le sexe.

¹ Lemmatiser : ramener chaque mot à son lemme, à sa racine (singulier, masculin pour les noms et adjectifs, infinitif pour les verbes)

² Les analyses statistiques effectuées par Christian Baudelot s'appuie sur une variable composée des 102 mots distincts de plus de deux lettres apparaissant plus de 5 fois dans son corpus.

« Avoir une vie qui nous rend heureux, dans notre vie personnelle comme dans notre vie professionnelle » (Générale – Seconde - Une femme) / « Atteindre nos objectifs et exercer un métier qui nous plaît réellement, et qui nous permet d'avoir une bonne situation » (Générale - Terminale - Une femme) / « Avant tout faire quelque chose qui nous plaît et qui nous aide à nous épanouir » (Générale - Première - Une femme) / « De la motivation, du travail, prendre du plaisir, aimer, savoir gérer son temps, un bon futur » (Générale - Première - Un homme) / « Un bon diplôme, un bon travail, avec un salaire convenable, une maison, une famille » (Générale - Terminale - Un homme) / « Avoir un bon travail, un bon salaire et surtout que le travail en question ne soit pas une corvée » (Technologique - Première - Un homme)

A partir de ces extraits, nous pouvons remarquer que leurs conceptions de la réussite sont multiples. La réussite prend différentes formes, et peut être perçue de façon plurielle selon les individus. Nous observons des interprétations différentes de la question ; certains ont davantage donné une définition de la réussite en mettant en avant des facteurs et qualités comme l'ambition, l'assiduité, le courage. Plus nombreux sont ceux qui ont mis en avant leur propre conception de la réussite, en donnant une dimension plus personnelle de celle-ci. Au niveau de la forme, nous avons également noté que la longueur des réponses est variable, sur le corpus de 16 432 mots, la longueur médiane des réponses est de 5 mots. Des enquêtés ont par exemple répondu sous forme de liste, en écrivant des mots en face des tirets, certains ont construit des phrases relativement longues jusqu'à 35 mots (226 observations comptent simplement 1 mot et 3 observations comptent plus de 30 mots). La longueur moyenne des réponses et l'indice de richesse³ calculé par Sphinx indiquent des différences suivant la variable de contexte prise en compte, par exemple les réponses des élèves de filière technologique comptent en moyenne 5 mots et l'indice de richesse est de 0,93 contre 6 mots pour les filières générales et un indice de 1,07. Suivant la classe, il est intéressant de noter que la longueur moyenne est la même, mais l'indice de richesse est corrélé avec l'avancée dans la scolarité (de 0,97 en classe de seconde à 1,03 en terminale). Le postulat de départ est bien que les mots employés ont un sens, ils ne s'emploient pas au hasard.

3.2. Protocole de codification et grille d'analyse

De multiples aspects composent la réussite et peuvent être perçus comme indissociables pour certains lycéens. Ces conceptions sont davantage tournées vers la réussite scolaire, professionnelle pour les uns, ou vers une réussite personnelle, sociale pour les autres. La place centrale du champ scolaire redéfinit-elle les conceptions de la réussite des lycéens et lycéennes ? Quelles dimensions sont mises en avant par les élèves ? Dans le contexte actuel, de plus en plus compétitif, marqué par le chômage, les réponses apportées par ces jeunes vont-elles mettre en avant des conceptions moins hédonistes et plus stratégiques et utilitaristes de la réussite ? A la fin des années 1980, Christian Baudelot et Roger Establet (1989) avaient souligné que « l'immense majorité des réponses, refusant le rêve ou l'utopie, semble exprimer les souhaits et les vœux que les jeunes interrogés forment eux-mêmes pour leur propre avenir ». Dans l'ensemble des réponses de notre enquête, le terme de *rêve* apparaît seulement dans 19 observations et ce mot est très souvent accolé au terme métier, cela relativisant le caractère idéaliste de leur conception.

Différents types de réussite sont souvent exprimés, souhaités dans une même phrase. Nous pouvons nous interroger sur l'importance donnée à l'une ou l'autre de ces facettes de la réussite. Pour cela, nous avons codifié les réponses, en adoptant une grille d'analyse des

³ Si l'indice de richesse est supérieur à 1, alors les réponses de cette catégorie sont plus riches que la moyenne, et inversement quand l'indice est inférieur à 1.

différents critères donnés à la réussite, en nous s'appuyant sur l'article de Bianka Zazzo publié au début des années 1960. Cette auteure identifie trois grandes rubriques : réussite sociale, réussite sentimentale et la réalisation de soi-même. Avec le logiciel Sphinx IQ Quali, il est possible d'identifier les thématiques présentes dans les réponses et les principaux concepts avec différents niveaux de détails (Ganassali, 2014, Boughzala et al. 2014). Coder le corpus à l'aide d'une grille de codification peut être effectuée de façon automatisée. La distribution des différents thèmes de la grille produite par le logiciel Sphinx⁴ ne semble pas être à première vue inadapté à notre corpus, mais quelque soit le niveau du thésaurus⁵ choisi, cette nomenclature arborescente ne fait pas sens pour nous. Dans l'idée d'actualiser des grilles thématiques déjà pensées comme celle de Bianka Zazzo, il nous a semblé plus pertinent de construire notre propre terminologie (composée de thèmes et sous-thèmes), afin de faciliter la mise en perspective avec d'autres recherches. Afin d'être dans nos mondes lexicaux, il nous a semblé difficile de nous adapter à une grille thématique lorsque nous avons déjà pensé et identifié en amont nos propres catégories de codification.

Le codage manuel a le grand avantage de nous rendre plus sensible à notre corpus par rapport à une simple lecture exploratoire des réponses, mais l'objectivité recherchée et attractive en laissant le logiciel établir une grille thématique ne semble pas se prêter à notre corpus.

Lors de la phase de codification, il s'agit de mettre en avant les interprétations différentes de la question, les élèves ayant par exemple défini la réussite se retrouveront dans la grille de codification, sous le thème « définition de la réussite ». Nous avons également décidé de différencier la réussite scolaire et professionnelle. Certains lycéens et lycéennes ont une conception ancrée dans le présent ou le futur proche, utilisant des mots tels que *notes*, *trimestre*, *bac*. En reprenant la typologie utilisée par Bianka Zazzo, il s'agit de percevoir les différences, les évolutions par rapport à notre propre corpus. Il faut préciser que le traitement des résultats, la méthodologie employée n'ayant pas été identiques, une comparaison rigoureuse n'était pas possible. L'auteure a mis en avant que la réussite sociale et notamment le critère de l'argent sont souvent invoqués par les garçons, la réussite sentimentale et la réalisation de soi plus souvent soulevées par les filles. Ces quelques éléments nous permettent de relever les évolutions avec les résultats de notre corpus. Quelles sont les conceptions dominantes pour tel ou tel groupe d'individus enquêtés ? La variable du sexe croisée aux différentes thématiques, nous montre que les différences soulevées dans la recherche de Bianka Zazzo, ne se réitèrent pas forcément. La relation est peu significative (PS), mais il faut noter la sur-représentation de la réussite personnelle chez les lycéennes (tableau 1). En parallèle, dans une approche lexicométrique, *épanouissement* et *épanouir* font partie des termes les plus représentatifs⁶ dans les réponses féminines. Lorsque nous nous penchons sur les sous-thèmes relatifs à cette catégorisation, des différences apparaissent plus nettement. Pour ce qui est de la réussite matérielle, on note par exemple une sur-représentation du besoin d'argent du côté des lycéens. En ce qui concerne la réussite sentimentale, les sous- thèmes de la mise en couple et de l'envie d'avoir des enfants, sont largement sous présentés chez les lycéennes.

⁴ Les thèmes générés par le logiciel Sphinx pour notre corpus sont : activité professionnelle, action et réaction, causalité, intention, emploi et salaire, instruction recherche, motivation, degré de satisfaction, genre littéraire et degré d'aisance.

⁵ Le thésaurus est un ensemble de significations, idées, concepts organisé suivant une nomenclature arborescente allant du général au particulier.

⁶ En explorant le verbatim sous Sphinx l'onglet *mots spécifiques* liste les mots présents dans les observations d'un contexte donné, ici le sexe.

	Réussite scolaire	Réussite professionnelle	Réussite personnelle	Réussite matérielle	Réussite sentimentale	Définitions de la réussite	Non classé
Une femme	9,3%	29,1%	<u>29,6%</u>	7,2%	3,8%	20,4%	0,5%
Un homme	10,5%	27,8%	<u>25,2%</u>	8,0%	5,2%	22,6%	0,7%

p = 0,08 ; Khi2 = 11,42 ; ddl = 6 (PS)

Tableau 1 : répartition des thèmes de la réussite en fonction du sexe (en %)

Des hypothèses ont été formulées lors de la première lecture exploratoire des réponses données par les uns et les autres. L'importance de l'emploi, du travail, la centralité de l'école et des diplômes pour y parvenir semblent avoir un fort impact sur la façon de percevoir la réussite. Ces postulats de départ, au regard des résultats du tableau croisé sont confirmés, effectivement la réussite professionnelle occupe une place principale dans leur conception.

4. Les lycéens et leurs conceptions de la réussite

4.1. Explorer avec les outils de data mining : graphe de relations et key view

Les graphes de relations permettent de voir les liens, les relations entre différentes variables. La variable des principaux mots a été croisée avec un ensemble de variables, afin de déterminer quelles sont celles qui l'influencent la plus fortement. Les thématiques abordées dans le questionnaire adressé aux lycéens et lycéennes étant assez larges (75 questions), nous avons choisi d'expliquer la variable des mots en la croisant avec plusieurs variables. Le seuil de significativité est fixé à 5%. Nous pouvons voir sur la figure 1 que la variable clé au centre du graphe est influencée fortement par les variables qualifiées avec un symbole TS (relations très significatives) telles que la série, la classe ou encore le sexe et a une relation significative (symbole S) avec le fait d'avoir redoublé ou non dans l'enseignement secondaire.

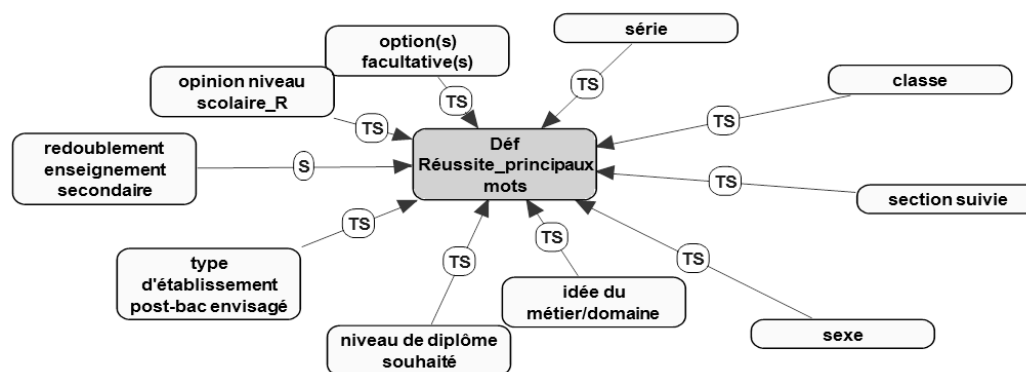


Figure 1 : graphe de relations établi autour de la variable des principaux mots de la réussite

Les Key views (correspondant aux tableaux de caractéristiques) élaborés par les équipes de Sphinx permettent de visualiser les résultats les plus significatifs d'un ensemble d'analyses croisées. La variable centrale utilisée est celle des principaux mots ressortant des conceptions de la réussite croisée à différentes variables comme le sexe, la classe, la filière. Cette analyse revient à réaliser plusieurs tris croisés pour identifier les couples de modalités dont la correspondance est significative. Le tableau étant trop volumineux, nous préférons insérer la représentation graphique de celui-ci (Figure 2). Il s'agit ici de revenir sur quelques modalités sur-représentées. Le tableau de caractéristiques indique par exemple que les termes *métier* et *épanouir* sont utilisés davantage chez les femmes, le mot *salaire* se retrouve plus chez les lycéens, le verbe *atteindre* est davantage employé par les élèves de classe de terminale et en

des objectifs) ou plus hédoniste et expressive par d'autres, dépendent plus fortement de la socialisation familiale ou de la socialisation scolaire ? Ces premières analyses devront être élargies à d'autres variables (milieu d'origine, CSP et diplômes des parents) afin de mesurer l'impact des conditions sociales et ses effets sur leur conception de la réussite.

Il semble important de pouvoir passer du lexique au sens, mais la reconnaissance des significations sous Sphinx soulève quelques interrogations. Lorsque nous avons en tête un monde lexical plus ou moins arrêté, il semble plus pertinent de construire sa propre grille thématique et de procéder manuellement à la codification des réponses même si un des inconvénients est bien de voir disparaître les spécificités de certaines réponses. Il semble alors difficile de représenter réellement le contenu, sans perdre la qualité du discours, ou même de devoir écarter certaines réponses qui ne rentrent pas dans les cases pensées. Cette grille a été construite par rapport au corpus, et reflète donc les conceptions de lycéens. Les thèmes et sous-thèmes développés sont-ils adaptable à une autre catégorie de population, peut-on la transposer facilement ? Une dizaine d'entretiens semi-directifs auprès de parents de lycéens ont été effectués, nous avons réitéré cette question de la réussite avec eux et ces premières réponses nous laissent penser que le grille peut s'accorder avec les représentations d'individus en dehors de l'univers scolaire, mais doit évidemment s'adapter à la situation, à l'âge des enquêtés.

Références

- Baudelot C. (1988). « Confiance dans l'avenir et vie réussie », *Mélanges économiques. Essais en l'honneur d' Edmond Malinvaud*, Paris, Economica, EHESS.
- Baudelot C. et Establet R. (1989). *Le niveau monte. Réfutation d'une vieille idée concernant la prétendue décadence de nos écoles*, Paris, Seuil.
- Boughzala Y., Moscarola J. et Hervé M. (2014). *Sphinx quali : un nouvel outil d'analyses textuelles et sémantiques*, 12ème journées des analyses de données textuelles.
- Felouzis G. (1993). Conceptions de la réussite et socialisation scolaire: Le cas des lycéens des filières générales, technologique et de LEP. *Revue française de pédagogie*. Volume 105, N°1.
- Ganassali S. (2014). *Enquêtes et analyse de données avec Sphinx*, Collection LM, Person.
- Ganassali S. et Roederer C. (2014). *L'analyse des courriers au Père Noël comme expression des styles de consommation familiaux*, 12ème journées des analyses de données textuelles.
- Lebart L. et Salem A. (1994). *Statistique textuelle*, Paris, Dunod.
- Zazzo B. (1962). La représentation de la réussite chez les adolescents. *Enfance*. Tome 15, n°3.