

De l'automatisation de l'affectation dans l'enseignement supérieur

Judicaël Courant, Julien Gossa

▶ **To cite this version:**

Judicaël Courant, Julien Gossa. De l'automatisation de l'affectation dans l'enseignement supérieur. La sélection dans l'enseignement supérieur : Politiques, instruments, pratiques, Feb 2019, Paris, France. hal-02967579

HAL Id: hal-02967579
https://hal.science/hal-02967579

Submitted on 15 Oct 2020

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.

De l'automatisation de l'affectation dans l'enseignement supérieur

Judicaël Courant et Julien Gossa*

5 janvier 2019

1 Introduction

Aborder les enjeux conceptuels relatifs à l'affectation des bacheliers dans l'enseignement supérieur sous l'angle informatique est essentiel : des choix d'apparence technique peuvent avoir d'importantes répercussions sur les plans éthique et sociologique. L'objectif de ce texte est donc de permettre aux non-informaticiens de se saisir des questions soulevées par ces choix techniques.

Rappelons tout d'abord deux points essentiels :

1. Un algorithme ou un système informatique ne fait qu'exécuter une procédure. Il ne peut inventer des places s'il n'y en a pas assez, ni même satisfaire tous les bacheliers à partir du moment où le nombre de places offertes dans certaines formations est inférieur à la demande.
2. L'algorithme n'est qu'une partie de la question technique. Car si l'algorithmique est une grande partie de l'informatique théorique, [1] l'autre consiste essentiellement à s'assurer qu'un programme répond à la spécification attendue. [2] En effet, un ordinateur appliquant les ordres à la lettre, la moindre imprécision dans l'expression d'un programme peut conduire à la catastrophe. L'apprenti sorcier [3] en est une métaphore éclairante : s'il est aisé d'oublier de préciser qu'il faut cesser de remplir le bassin lorsqu'il est plein, il est à plus forte raison normal qu'un système informatique complexe comporte des erreurs.

Les citoyens sont donc légitimement fondés à connaitre les programmes qui décident de leur sort, non seulement pour comprendre les procédures qui les concernent, mais aussi pour obtenir des garanties sur leur bon fonctionnement.

*Judicael.Courant@free.fr, gossa@unistra.fr

1. Jan van Leeuwen. *Handbook of Theoretical Computer Science, Vol. A : Algorithms and Complexity*. The MIT Press, 1994. isbn : 0262720140.
2. Jan van Leeuwen. *Handbook of Theoretical Computer Science, Vol. B : Formal Models and Semantics*. The MIT Press, 1994. isbn : 0262720159.
3. Johann Wolfgang von Goethe. *Der Zauberlehrling*. 1797.

Nous présentons section 2 un cadre conceptuel utilisé pour décrire les algorithmes d'affectation, qui englobe Admission Post-Bac (APB) comme ParcourSup. Nous expliquons quelles grandes propriétés on est en droit d'exiger de ces algorithmes.

Section 3, nous expliquons en quoi, pour répondre au problème tel que nous l'avons posé, il existe un algorithme incontournable s'appuyant sur une notion intuitive, la notion de barres d'admissions.

Section 4, nous nous intéressons à la mise en œuvre de ces procédures d'affectation et étudions un certain nombre de conditions pour que les citoyens — les bacheliers en premier lieu — puissent avoir confiance en elles.

Section 5, nous synthétisons nos remarques dans une proposition de procédure tentant de corriger les problèmes les plus flagrants de ParcourSup et d'APB, en adoptant une approche de *transparence par conception*.

2 Cadre conceptuel

Le cadre conceptuel de l'affectation de candidats dans le supérieur s'articule en trois parties : la modélisation du problème, les propriétés de la solution elle-même, et les propriétés de la mise en œuvre du système qui la calcule.

2.1 Modélisation du problème

Sur le plan mathématique, l'affectation des bacheliers dans le supérieur peut se modéliser comme un *problème* auquel il s'agit de trouver des solutions.

Pour une même modélisation du problème, plusieurs solutions sont parfois possibles et il s'agit de comprendre lesquelles sont les meilleures.

Différentes modélisations du problème sont également possibles et induisent des solutions différentes mais on s'accorde généralement sur le cadre suivant :

- des établissements d'enseignement supérieur proposent des formations ;
- chaque formation est dotée d'un nombre maximum de places (appelé capacité d'accueil de la formation) ;
- les candidats à chaque formation sont classés, suivant des critères de sélection qui peuvent être propres à la formation considérée ;
- chaque candidat est intéressé par une ou plusieurs formations (on dira que ces formations forment ses *vœux*) ;
- on supposera de plus que le candidat a ordonné ses vœux : son premier vœu est celui qu'il préfère, et son dernier vœu est son dernier choix avant de renoncer à obtenir une formation.

Dans cette modélisation, la sélection peut prendre bien d'autres formes que celle de la sélection sur le dossier scolaire du candidat :

- sélection sur la date de candidature (premier arrivé, premier servi) ;
- sélection par l'argent (des frais d'inscription élevés éliminent des candidatures) ;
- sélection par tirage au sort ;
- sélection par défaut de publicité…

En ce sens, l'accès aux universités dans les années 80, ainsi qu'avec APB et ParcourSup rentrent tous trois dans le cadre ci-dessus.

2.2 Spécification du problème

Avant de donner un algorithme, c'est-à-dire une méthode pour résoudre un problème, il convient de préciser ce que serait une solution à ce problème : en informatique, on parle de cahier des charges ou de *spécification* des propriétés attendues du système.

Le cahier des charges d'APB n'a jamais été publié ; en ce qui concerne ParcourSup, le ministère de l'enseignement supérieur a publié, *a posteriori*, un faux cahier des charges, [4] au mépris de ses obligations légales. Celui-ci semble d'ailleurs se focaliser sur les détails et non sur les objectifs généraux.

Une spécification doit permettre de lever les implicites inhérents à toute modélisation. En l'espèce, la résolution du problème que nous avons modélisé repose sur un implicite important : les vœux de chaque acteur sont indépendants de ceux de chaque autre acteur. Par exemple :

1. Un couple de candidats ne peut pas émettre un vœu du type « aller dans n'importe quelle formation, pourvu que nous soyons dans la même ville ».
2. Lorsqu'elle examine une candidature, une formation ne se vexe pas de ne pas être le premier choix.

Ces implicites peuvent avoir une grande importance. Par exemple, dans le cas des internats de médecine aux États-Unis, [5] le premier point a conduit à une refonte du système utilisé. [6]

Le second point était un des fondements d'APB. Malheureusement, au cours du temps, des « bonus premier vœu » ont été introduits : de rares formations ont eu accès à l'ordre des vœux, et ce dernier était utilisé lors des rares tirages au sort. Malgré cette rareté, cette introduction a jeté le doute sur le rôle de l'ordre vœux dans la procédure et des candidats ont élaboré des stratégies au lieu d'exprimer honnêtement leurs préférences. Nous discutons de de cela plus loin.

Un des avantages de la transparence est précisément d'éviter ces errements, ce qu'aurait permis la publication de véritable spécifications.

4. Guillaume Ouattara. *Parcoursup : le ministère publie le vrai/faux cahier des charges*. Le Figaro.fr étudiant `https://etudiant.lefigaro.fr/article/parcoursup-le-ministere-publie-le-vrai-faux-cahier-des-charges_fea3df26-03aa-11e9-a935-933e550b769f/`. Déc. 2018.

5. A.E. Roth. *Who Gets What — and Why : The New Economics of Matchmaking and Market Design*. Houghton Mifflin Harcourt, 2015. isbn : 9780544288393.

6. Alvin E Roth et Elliott Peranson. *The Redesign of the Matching Market for American Physicians : Some Engineering Aspects of Economic Design*. Working Paper 6963. National Bureau of Economic Research, fév. 1999. doi : `10.3386/w6963`. url : `http://www.nber.org/papers/w6963`.

2.3 Propriétés de la solution

On veut trouver, entre toutes les solutions possibles, celle qui respecte au mieux l'ordre des vœux des candidats et les classements des candidats par les formations. La solution que nous construisons devrait vérifier les critères suivants :

Réalisme une solution est *irréaliste* si, pour au moins une formation, le nombre de bacheliers affectés sur celle-ci excède sa capacité.

Absence de gabegie il y a *gabegie* s'il reste au moins une place libre dans une formation qu'au moins un candidat préférerait à celle qu'on lui assigne.

Optimalité pour les bacheliers une solution réaliste est dite optimale au sens de Pareto pour les candidats si toute autre solution réaliste améliorant le sort d'au moins un candidat dégrade celui d'au moins un autre candidat.

Optimalité pour les formations de la même façon, une solution est optimale au sens de Pareto pour les formations si toute autre solution réaliste qui améliore le recrutement d'une formation dégrade celui d'au moins une autre formation.

Absence de jalousie légitime si un candidat découvre qu'une formation qu'il préfère à celle qu'on lui a attribuée a pris un candidat moins bien classé que lui sur cette formation, il peut légitimement être jaloux.

2.3.1 Réalisme et gabegie

Le réalisme d'une solution n'est pas nécessairement une notion *rigide* avec un nombre de places absolu et prédéfini. En effet, il est tout à fait possible que les formations adaptent leur capacité d'accueil au cours de la procédure d'affectation. Cela peut servir, par exemple, à s'assurer de l'absence de gabegie face aux défections de dernière minute, ou encore à prendre le dernier candidat sur la liste d'attente.

2.3.2 Objectifs d'optimalité

On peut montrer que les objectifs d'optimalité pour les candidats et d'optimalité pour les formations sont parfois incompatibles. Un choix doit donc être fait. On peut privilégier les candidats pour une raison éthique : privilégier des humains plutôt que des institutions ; et pragmatique : les étudiants sont plus motivés dans une formation choisie. Mais on peut aussi privilégier les formations pour une raison politique : affecter les meilleurs étudiants aux meilleures formations favorise la concurrence, la concentration des élites, et la concentration des moyens sur ces élites.

2.3.3 Jalousie légitime

La jalousie légitime survient lorsqu'un étudiant a intérêt à se dédire de son engagement vis-à-vis de son affectation pour aller dans une formation qu'il préfère, cette formation ayant elle-même intérêt à se dédire de son engagement vis-à-vis d'un autre étudiant pour pourvoir l'accueillir. Ces solutions d'affectation sont donc instables, au sens où la tentation est forte pour certains acteurs de ne pas respecter les affectations données. D'où le terme également employé de « mariages stables » pour qualifier les solutions exemptes de toute jalousie légitime.

Le critère d'absence de jalousie légitime n'est donc pas seulement issu de considérations morales mais surtout de considérations pratiques venant de l'étude concrète de l'affectation des étudiants de médecine dans les internats de médecine dans les années 40. [7]

Malheureusement, les objectifs d'optimalité et d'absence de jalousie légitime sont parfois contradictoires, ce qui conduit à rechercher des solutions optimales au risque de susciter de la jalousie, ou bien à se contenter des solutions optimales *parmi* les solutions sans jalousie légitime. Le premier choix conduit à des algorithmes relativement délicats à expliquer et justifier (et probablement insuffisamment compris à l'heure actuelle) alors que le second conduit sur des chemins mieux établis. Il est considéré comme le mécanisme adapté dans le cas de l'affectation dans le supérieur. Dans toute la suite, nous ne nous intéresserons donc qu'à des solutions sans jalousie légitime.

2.4 Propriétés de la mise en œuvre

Le système que nous voulons mettre en œuvre devra également respecter un certain nombre de propriétés supplémentaires, qui dépassent le cadre purement algorithmique :

Limiter l'anxiété. Le système doit éviter de mettre les candidats dans une situation de stress répété sur une longue durée. Le système empirique précédant APB en France était assez mauvais : un candidat pouvait attendre tout l'été une réponse qui ne venait pas. ParcourSup est pire encore : l'attente interminable se double de l'obligation de vérifier quotidiennement sa situation.

Favoriser une expression honnête des préférences. Si les acteurs du système mentent, il est difficile de savoir quelles sont leurs véritables préférences, ce qui est problématique pour adapter l'offre de formation à leurs désirs réels. De plus, les stratégies basées sur le mensonge profitent le plus souvent aux bacheliers les mieux informés sur le système, déjà les plus favorisés socialement. Enfin, ces stratégies ne font pas qu'augmenter l'anxiété des candidats, mais diminuent en plus la qualité de l'affecta-

7. Alvin E. Roth et Marilda A. Oliveira Sotomayor. *Two-Sided Matching : A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monographs. Cambridge University Press, 1990. doi : 10.1017/CCOL052139015X.

tion,[8] sans réel bénéfice pour les candidats dans la plupart des cas. On veut donc un système aussi insensible que possible aux choix stratégiques d'expression des vœux : révéler honnêtement ses préférences doit être la meilleure stratégie.

Être digne de la confiance du public. Non seulement le système doit être fiable, mais sa mise en œuvre doit être suffisamment transparente pour que les différents acteurs soient convaincus qu'il fonctionne correctement et équitablement.

2.4.1 L'exigence de transparence

Sur ce dernier point, il convient d'attirer l'attention sur les enjeux de la transparence, qui apparait comme une difficulté majeure dans ce domaine. En effet, la transparence oblige les formations à affirmer leurs stratégies de recrutement, ce qui peut leur poser problème dans une logique de concurrence. On sait que certains critères ne seraient pas utilisés s'ils devaient être publiés, et on sait aussi que l'examen des dossiers serait moins confortable s'il était public. C'est d'ailleurs pour ces raisons qu'une exception à la Loi pour une république numérique a été introduite dans la Loi pour l'orientation et la réussite des étudiants.

En effet, la transparence ouvre la porte à la contestation des candidats, tant sur les critères de sélection que sur leur mise en œuvre. La transparence ne peut donc être vue comme simplement souhaitable sans prendre le risque qu'elle soit affaiblie par les acteurs. Si on la décide, il faut donc la considérer comme une exigence et concevoir une procédure qui s'y conforme. C'est ce que nous appelons un *algorithme transparent par conception.*

3 Une solution connue et incontournable

La solution connue et incontournable est d'utiliser un algorithme du type « mariage stable », tel que proposé par Gale et Shapley.[9] Nous préférons à la présentation traditionnelle une présentation par barres d'admission, plus aisée à expliquer et à comprendre.

Nous présentons tout d'abord la méthode empirique utilisée pour l'affectation dans les universités en France dans les années 80, qui conduit assez naturellement à mettre en place une procédure centralisée où les candidats expriment leurs vœux de façon hiérarchique.

Nous introduisons dans un second temps la notion de barres d'admissions et montrons comment elle conduit aux algorithmes de Gale et Shapley.

8. Róbert F. Veszteg, Ágnes Pintér et Joana Pais. *College admissions and the role of information : an experimental study.* UC3M Working papers. Economics we080302. Universidad Carlos III de Madrid. Departamento de Economía, jan. 2008.

9. Gale et Shapley. « College Admissions and the Stability of Marriage ». In : *The American Mathematical Monthly* 69.1 (1962), p. 9–15. issn : 00029890, 19300972. url : `http://www.jstor.org/stable/2312726`.

3.1 Une méthode empirique

L'affectation des bacheliers en université dans les années 80 en France relève du cadre présenté plus haut, bien que ne faisant pas appel à l'outil informatique. Chaque candidat pouvait s'inscrire sur plusieurs formations et, une fois accepté sur l'une d'elles, se désistait de celles qui l'intéressaient moins. Les formations en surcapacité mettaient en liste d'attente une partie des candidats (souvent les derniers inscrits) et prenaient les candidats dans cette liste au fur et à mesure des désistements.

La situation peut se modéliser ainsi :

1. les formations classent leurs candidats et les candidats classent (mentalement) les formations ;
2. chaque formation décide d'un *rang d'admission*. Initialement, ce rang est égal à la capacité de la formation ;
3. chaque formation admet (potentiellement) chaque candidat dont le rang dans le classement est inférieur ou égal au rang d'admission et informe chaque nouveau candidat ainsi admis ;
4. chaque candidat admis sur plusieurs formations démissionne de celles qui lui plaisent le moins pour ne garder que celle qui lui plaît le plus (sur la base de son classement) ;
5. chaque formation qui enregistre une démission augmente son rang d'admission de façon à admettre un nouveau candidat. Le processus reprend alors, à l'étape 3, tant qu'il y a des démissions.

3.1.1 Centralisé vs. Distribué

Pour un informaticien, cette procédure est un algorithme exécuté non pas en un seul endroit, mais par les différents acteurs (bacheliers et universités). On dit qu'il est distribué.

Il s'exécute très lentement, les démissions s'effectuant en chaînes, dont chaque étape prend plusieurs jours. Ce problème a été perçu dès les années 40 aux États-Unis pour l'affectation dans les internats de médecine et a posé problème jusqu'à l'introduction en 1951 de la procédure centralisée NIMP. Basée sur le volontariat, elle a remporté l'adhésion de toutes les parties pendant plusieurs dizaines d'années. [10]

Une procédure centralisée permet en effet d'avoir une vision globale. C'est impossible dans la procédure empirique où aucun acteur ne peut savoir si des démissions vont encore se produire.

3.1.2 Hiérarchisation des vœux

En réalité, il ne s'agit pas seulement de centraliser la procédure mais aussi les informations utiles à son exécution. NIMP est rapide parce qu'elle demande une

10. Roth et Sotomayor, *Two-Sided Matching : A Study in Game-Theoretic Modeling and Analysis*.

seule fois ses préférences à chaque candidat, avant l'exécution de l'algorithme, sous forme d'une liste de vœux hiérarchisée.

Il s'agit d'ailleurs de la différence majeure entre APB et ParcourSup : les vœux, hiérarchisés dans APB, ne le sont plus dans ParcourSup. Dès lors, de nombreux candidats ont attendu tout l'été une réponse qui n'est jamais venue ; d'autres ont accepté définitivement un vœu qui n'était pas leur préféré, par manque d'espoir ou parce qu'ils n'avaient pas les moyens de s'adapter à un changement d'affectation à la veille de la rentrée. Il y a donc eu objectivement une prime offerte à ceux qui étaient en mesure d'attendre jusqu'au bout.

La hiérarchisation des vœux est le seul moyen raisonnable d'améliorer significativement la rapidité de la procédure.

En outre, seule la hiérarchisation des vœux permet d'éviter de demander à de multiples reprises leurs préférences aux candidats. Elle semble donc nécessaire pour limiter leur anxiété.

3.2 Notion de barre d'admission

3.2.1 La méthode empirique comme calcul de barres d'admissions

Dès que les candidats sont classés par les formations et que les bacheliers ont exprimé et classé leurs vœux, il est possible de simuler informatiquement la procédure empirique. En quelques minutes au plus, on peut ainsi calculer les affectations des candidats : c'est ce que faisait la procédure APB.

Une façon de comprendre cette procédure est de se dire que son rôle est de calculer un rang d'admission pour chaque formation, ce qui est équivalent à choisir une affectation des candidats sans jalousie.

Dans certains systèmes d'affectation dans le supérieur, comme ceux utilisés en Espagne, en Hongrie, en Turquie et en Irlande, [11] au lieu de donner un rang aux candidats, on leur attribue une note.

Dans les deux cas, on parle de *barres d'admissions* (référence métaphorique au saut en hauteur) : on dira qu'une barre d'admission est haute si le rang d'admission est petit/la note d'admission élevée.

La méthode empirique, APB et ParcourSup consistent à prendre initialement un choix de barres d'admissions suffisamment hautes pour être réalistes et à baisser ces barres jusqu'à ce qu'elles soient sans gabegie. Cette procédure est équivalente à la version de l'algorithme des mariages stables de Gale-Shapley connue sous le nom *College Optimal Student Matching* (COSM).

3.2.2 SOSM : une procédure de calcul de barres optimales

Il peut y avoir plusieurs choix de barres d'admissions à la fois réalistes et sans gabegie. COSM donne en réalité les barres les plus hautes c'est-à-dire les plus défavorables aux étudiants parmi tous ces choix.

11. Péter Biró et Sofya Kiselgof. « College Admissions With Stable Score-Limits ». In : *Central European Journal of Operations Research* 23.4 (2013), p. 727–741. doi : 10.1007/s10100-013-0320-9. url : https://doi.org/10.1007/s10100-013-0320-9.

Une autre version de l'algorithme de Gale-Shapley, appelée *Student Optimal Student Matching* (SOSM), choisit plutôt des barres les plus basses possibles, ce qui donne une affectation sans gabegie, et les fait monter jusqu'à obtenir une affectation réaliste.

Les barres fixées par SOSM sont les plus basses conduisant à une affectation réaliste : elles donnent donc une affectation optimale sans jalousie, réaliste et sans gabegie, plus favorable aux étudiants que COSM.

SOSM est utilisé en Espagne et en Hongrie,[12] ainsi que pour l'affectation des jeunes médecins en Écosse et pour les internats de médecine aux États-Unis. En ce qui concerne la France, la différence entre COSM et SOSM sur les affectations d'APB en 2017 serait, semble-t-il, relativement marginale (de l'ordre de 400 bacheliers sur plus de 800000[13]). Ceci plaide pour SOSM : il est beaucoup plus facile de faire connaitre cette absence de différence aux formations qu'aux bacheliers, qui changent naturellement chaque année.

Enfin, COSM peut parfois inciter les candidats à mentir sur leurs préférences, alors que SOSM les incite à être honnêtes.[14] Notons toutefois que les formations peuvent parfois avoir dans ce cas un intérêt à mentir sur leurs préférences ; en outre il est important, pour que cette propriété soit vérifiée, qu'il n'y ait pas d'ex-æquo dans les classements des candidats par les formations.[15]

4 Mise en œuvre

4.1 Publicité du système

Un système affectant les étudiants dans l'enseignement supérieur devrait voir ses principes clairement spécifiés, ce qui n'a été le cas que très partiellement pour APB et ParcourSup (via la loi ORE ou les décrets d'application).

Le public est en droit d'attendre aussi que les algorithmes et le code utilisé soient publiés largement avant leur mise en mise en œuvre pour permettre d'éventuelles corrections. La description de l'algorithme et le code l'implantant devraient être suffisamment courts et clairs pour être compris par un informaticien moyen.

Dans le cas d'APB, le ministère s'est contenté de publier une infographie d'une page et le code du tirage au sort des candidats. Ni l'un ni l'autre n'étaient d'une qualité conforme à ce qu'on attend d'un système gérant chaque année l'avenir de 800000 personnes.

12. Biró et Kiselgof, « College Admissions With Stable Score-Limits ».
13. Claire Mathieu. Communication privée. 2018.
14. L. E. Dubins et D. A. Freedman. « Machiavelli and the Gale-Shapley Algorithm ». In : *The American Mathematical Monthly* 88.7 (1981), p. 485–494. issn : 00029890, 19300972. url : http://www.jstor.org/stable/2321753.
15. Atila Abdulkadiroğlu, Parag A. Pathak et Alvin E. Roth. « Strategy-Proofness versus Efficiency in Matching with Indifferences : Redesigning the NYC High School Match ». In : *American Economic Review* 99.5 (déc. 2009), p. 1954–78. doi : 10.1257/aer.99.5.1954. url : http://www.aeaweb.org/articles?id=10.1257/aer.99.5.1954.

L'algorithme de Parcoursup a été publié, [16] bien que tardivement, ce qui est un progrès.

Le code du cœur algorithmique de ParcourSup a été publié, ce qui est aussi un progrès. Mais c'est insuffisant pour avoir confiance dans les résultats car la moindre erreur à la périphérie de cet algorithme, par exemple sur le code gérant les données des candidats ou les classements retournés par les formations induit des erreurs sur l'affectation des étudiants.

La composante logicielle d'un futur système d'affectation des étudiants doit donc être rendue publique — spécifications, algorithmes, code — en intégralité (y compris le code de la plate-forme sur laquelle elle est exécutée).

Enfin, à l'échelle de l'individu, la partie affectation n'est pas la partie la plus importante : la partie classement des candidatures par les formations, souvent désignée par « algorithmes locaux », a sans doute plus de poids dans son affection finale.

4.2 Caractère vérifiable de l'exécution

Le caractère public du logiciel d'affectation ne suffit pas à assurer que le résultat de la procédure est correct. Il est possible que :

- une erreur liée au matériel ou au système d'exploitation utilisé affecte le résultat d'un programme ;
- le logiciel exécuté ne soit pas le logiciel publié. Cela peut provenir d'une action délibérée, par exemple d'un administrateur du serveur calculant le classement ou d'un « pirate » modifiant furtivement le logiciel présent sur le serveur depuis l'extérieur, mais aussi d'une simple erreur, par exemple celle qui consiste à exécuter par erreur une version expérimentale du programme au lieu de la version publiée ;
- les données fournies au système soient incorrectes en raison d'erreurs dans les données collectées, ou d'altérations délibérées de ces données avant leur entrée dans le système ;
- les résultats publiés ne soient pas les résultats calculés par le système.

En raison de son importance, la question de la qualité des données fait l'objet de la partie suivante.

Pour s'assurer que le logiciel exécuté est bien le logiciel publié, il n'existe en pratique qu'un seul moyen : reproduire le calcul officiel. Ce critère de reproductibilité est d'ailleurs traditionnellement celui de la vérité scientifique. C'est pourquoi il convient de publier les données en plus du logiciel.

4.3 La qualité des données : un impensé majeur

Les données utilisées en entrée d'APB et ParcourSup sont les vœux des (futurs) bacheliers mais aussi leurs dossiers, remplis en partie par leur établissement d'origine, et par le candidat lui-même.

16. Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation. *Document de présentation des algorithmes de Parcoursup*. `https://framagit.org/parcoursup/algorithmes-de-parcoursup`. Mai 2018.

Bien que les situations des formations soient extrêmement diverses, on peut cependant mentionner les problèmes qui peuvent y survenir.

4.3.1 Des données difficiles à exploiter

Les dossiers sont généralement d'abord classés à partir d'informations quantitatives chiffrées (notes), puis un examen manuel plus ou moins superficiel permet de monter ou descendre certains candidats dans le classement.

La notation différant suivant les lycées, les commissions d'examen des dossiers tentent de corriger le problème en incorporant dans leurs classements automatisés le rang de l'élève dans sa classe, l'effectif de sa classe ainsi que des informations sur son établissement d'origine (comme le taux de réussite ou de mention au baccalauréat). Aucun processus rigoureux officiel n'encadre ces ajustements qui tiennent de l'emplâtre sur une jambe de bois.

4.3.2 Un gaspillage

Lorsqu'un étudiant postule à dix endroits différents, son dossier est étudié dix fois : cette multiplication d'effort constitue un formidable gaspillage du temps des enseignants, qu'éviterait une mutualisation de l'examen des candidatures.

Cet examen local des dossiers a deux conséquences :

— en raison du temps nécessaire à l'examen des dossiers, l'envoi des dossiers a lieu dès la fin du second trimestre et le dernier trimestre de terminale ne peut être évalué ;

— les enseignants de lycée ne peuvent noter leurs élèves en toute indépendance, les notes de l'année ayant un impact direct sur la poursuite d'études de leurs élèves.

4.3.3 Gestion des erreurs : *garbage in, garbage out*

Si les données sont mal renseignées, par exemple si la note de mathématiques d'un étudiant postulant sur une filière scientifique est absente ou erronée, le résultat de l'algorithme risque fort de n'être pas le bon. Or sur le nombre de candidatures, les dossiers erronés sont inévitables.

Ce cas n'est pas prévu dans les procédures : il revient à l'examinateur du dossier d'improviser une correction.

Le cas des données erronées est le plus préoccupant : si une absence de saisie est toujours détectable, les établissements classant les candidatures ne peuvent pas savoir si les données qu'on leur donne sont erronées, en dehors de situations aberrantes (qui nécessitent un examen approfondi pour être détectées).

Ils n'ont donc pratiquement aucun moyen de découvrir et faire rectifier d'éventuelles erreurs.

Ce problème est connu, puisque la loi dite *Informatique et libertés* consacre en France depuis 1978 le droit à la rectification des erreurs dans les fichiers.

4.3.4 Contrôle des commissions d'examen des vœux

Même avec des dossiers corrects et complets, il peut y avoir des erreurs lors de leur classement par les formations. Classer les dossiers est une combinaison de tâches manuelles et automatisées (« algorithmes locaux »).

L'automatisation est préoccupante. Si le code des « moulinettes » utilisées pour effectuer un premier tri des dossiers est gardé secret par les établissements, alors aucun moyen sérieux ne peut en assurer la qualité.

La partie manuelle est tout aussi préoccupante. Une erreur d'examen ou de report du classement des candidats est toujours possible. Or, si aucun processus de vérification clair n'est prévu, des candidats seront refusés ou admis par erreur, ce qui les lèse ou en lèse d'autres.

4.3.5 Anonymat

La question de l'anonymat se pose sous différentes formes.

En ce qui concerne un algorithme transparent par conception, l'anonymat est indispensable pour protéger la vie privée des citoyens. On peut par exemple noter le cas d'un candidat ne souhaitant pas que sa famille connaisse ses vœux de formation, suite à un différend à ce sujet.

En revanche, l'anonymat n'est pas souhaitable lorsqu'on en vient à l'examen des dossiers. En effet, bien qu'on puisse instinctivement penser que cet anonymat est une mesure antidiscriminatoire, on constate en réalité sur l'étude des CV anonymes qu'il n'en est rien, et qu'en plus cela diminue la qualité du recrutement. [17]

En ce qui concerne les concours de la fonction publique, l'anonymat est la règle pour les épreuves écrites mais pas pour les dossiers et les épreuves orales. Et les noms et le rang de ceux qui sont admis sont publiés lors des résultats.

4.3.6 État actuel du système français

Pour ce qui est de ParcourSup, chaque candidat peut vérifier son dossier avant qu'il soit transmis aux formations, puis peut connaitre son rang, une fois l'examen fait. Même si l'appréciation de ce rang est difficile sans connaitre les dossiers des autres candidats, il peut éventuellement se rendre compte d'une grave aberration. Toutefois, si cette erreur est en sa faveur, il devra être très altruiste pour le signaler. Si elle est en sa défaveur, aucun mécanisme officiel n'est prévu pour corriger le problème *a priori*. En pratique, le bachelier contacte la formation et, s'il est entendu, sera mis premier en liste d'attente et pris au premier désistement… s'il s'en produit un !

Aucune règle ne semble prévue pour les commissions de classement. En particulier, la personne qui envoie à APB/ParcourSup les résultats du classement

17. Bruno Crépon† et Thomas Le Barbanchon Luc Behaghel. *Evaluation de l'impact du CV anonyme*. Rapp. tech. Center for Research in Economics et Statistics (CREST), mar. 2011. URL : `http://www.crest.fr/images/CVanonyme/rapport.pdf`.

peut le faire seule, sans aucun contrôle. Rien n'empêche non plus un membre d'une commission d'instruire le dossier de son propre enfant.

4.3.7 Alternative

Plutôt que de procéder à un classement sur dossier scolaire, il est possible de procéder à un classement sur une épreuve commune. C'est notamment l'approche adoptée en Turquie, Espagne et Chine. En France, c'est ce que l'on retrouve par exemple dans les concours communs, après les CPGE. L'Angleterre (BMAT ou UKCAT en médecine par exemple) et les Etats-Unis (SAT) ont également ce type de dispositif pour l'entrée dans les universités.

Ce type de classement des candidatures présente l'avantage d'être moins sujet aux erreurs et manipulations. En un sens, il est plus équitable puisque tous les candidats se confrontent à la même épreuve. Cependant, cela fait retomber cette équité sur l'équité du système scolaire : les candidats ne sont égaux face aux épreuves qu'autant que les conditions de leur préparation sont égales.

5 Synthèse : une proposition de procédure

De ce qui précède, nous pouvons dessiner les grandes lignes d'une nouvelle procédure transparente par conception. Nous synthétisons cette proposition sous la forme d'un calendrier.

Octobre Publication de la procédure. La spécification, l'algorithme et le code de la procédure sont publiés, sous une licence libre. Le code repose uniquement sur une infrastructure à base de logiciel libre (base de données, système d'exploitation) afin de pouvoir être exécuté par tous.

Novembre Publication des formations. En novembre, les formations publient les capacités qu'elles proposent via la procédure.

Nous proposons que les formations participent uniquement sur la base du volontariat. Cela permet de mesurer de façon claire l'acceptation du système : une formation peut décider de ne pas participer au système, de participer totalement, ou de participer en partie, par exemple en proposant l'essentiel des places disponibles par le système et en réservant quelques places pour gérer des cas particuliers.

Les seuls critères de sélection admis sont les suivants :

— les notes du baccalauréat ;
— les notes d'*épreuves spécifiques*, ces épreuves pouvant être organisées par chaque formation sous la forme qu'elle désire (écrit, oral) ou mutualisées entre formations ;
— à titre transitoire, une *note de dossier* basée sur l'examen des dossiers des étudiants par les formations.

Chaque formation choisit et publie une *formule* qui sera utilisée pour combiner ces différents critères et obtenir une note de classement pour chaque candidat sur la formation.

Janvier Attribution d'un numéro d'anonymat à chaque candidat. Chaque candidat se voit attribuer un numéro unique, qu'il est le seul à connaître.

Janvier à mi-mars Formulation des vœux par les candidats. Les candidats formulent leurs vœux sur une plateforme web.

Mi-mars à début avril Rectification des erreurs. Mi mars, les vœux de tous les candidats sont publiés, de façon anonyme : pour chaque candidat, on publie son numéro d'anonymat et les vœux qu'il a formulés. On ne publie en revanche pas son nom.

Les candidats disposent alors de quelques jours pour faire rectifier les éventuelles erreurs les concernant.

Début avril, les vœux corrigés sont de nouveau publiés.

Avril-mai Examen des dossiers par les formations. Les formations dont la formule prend en compte une note de dossier examinent les dossiers des candidats, sans anonymat.

Avril à mi-juin Hiérarchisation des vœux et vœux additionnels. D'avril à mi-juin, les candidats hiérarchisent leurs vœux. En outre, les candidats peuvent formuler de nouveaux vœux sur les formations qui ne prennent pas en compte la note de dossier. Ils peuvent également formuler de nouveaux vœux sur les formations qui prennent en compte la note de dossier mais leur note de dossier pour ces formations sera alors zéro.

Mi-juin, pour chaque candidat, on publie son numéro d'anonymat, la liste de ses vœux ainsi que l'ordre de ces vœux.

Avant les premières épreuves du baccalauréat Les épreuves spécifiques sont organisées et leurs notes remontées dans le système informatique au plus tard mi-juin. Pour chaque candidat à ces épreuves, on publie le numéro d'anonymat du candidat ainsi que ses notes. Les candidats peuvent alors faire rectifier les erreurs éventuelles. Les rectifications ont lieu avant fin juin, où les résultats corrigés sont de nouveau publiés.

Dès les résultats du baccalauréat Pour chaque formation, les notes de classements des bacheliers et le classement qui en découle sont calculés. En dernier recours, pour départager les éventuels ex-æquo, on utilise un tirage au sort transparent. On publie alors, pour chaque candidat, son numéro d'anonymat, ses notes (baccalauréat, épreuves spécifiques et dossier) et ses notes de classement.

Le système informatique calcule les barres d'admissions et l'affectation des candidats qu'elles induisent en utilisant l'algorithme SOSM. Ces barres et l'affectation des candidats sont alors publiées à titre provisoire.

Chacun peut vérifier qu'il n'a pas à être jaloux des candidats pris sur les formations où il a été refusé. De plus, chacun peut vérifier que les notes de classement, les barres d'admission et l'affectation ont été correctement calculées pour tous les candidats en rejouant l'algorithme d'affectation. Même si ceci n'est pas à la portée de tout le monde, la vérification publique pourra être menée par les acteurs intéressés.

Les candidats peuvent alors demander la correction d'erreurs éventuelles.

Mi-juillet Les erreurs sont corrigées et les barres d'admission définitives calculées et publiées.

En parallèle, les candidats ont pu refuser ou accepter la formation qui leur était proposée, ainsi que refuser ou accepter de continuer à participer à la procédure.

Les candidats pris sur des offres de formations hors procédure doivent démissionner de la procédure.

Au fur et à mesure des démissions, les admissions sont recalculées pour les candidats qui ont décidé de continuer à participer. Les résultats ne sont pas affichés chaque jour : les candidats à qui une nouvelle proposition est faite sont contactés directement.

Mi-août les résultats définitifs sont publiés et la procédure est arrêtée. On lève alors l'anonymat de tous les bacheliers ayant accepté une affectation : chacun peut ainsi vérifier que les candidats qu'il connaît n'ont pas eu des notes exagérément élevées.

Notons que nous avons quelque peu simplifié ce que serait une procédure d'affectation dans le système français. En pratique, plusieurs autres points sont à prendre en compte :

— Les quotas sur des critères géographiques ou sociaux ;
— La gestion des places d'hébergement pour les formations disposant d'un internat.
— Les admissions sous réserve de mise à niveau des candidats.

Pour les deux premiers points, on peut utiliser les mêmes moyens que ParcourSup, ou d'autres systèmes classiques de quotas. [18] Pour le troisième, il s'agirait de rajouter des critères décidant si le candidat doit ou non suivre une remise à niveau.

Notons enfin qu'il serait possible de permettre aux bacheliers de changer l'ordre de leurs vœux ou même de formuler de nouveaux vœux après les premiers résultats.

6 Conclusion

Nous avons présenté le cadre conceptuel dans lequel s'inscrivent les algorithmes d'affectation. Si ceux-ci ne peuvent évidemment pas remédier au manque de places, il peuvent tout du moins être choisis et mis en œuvre de façon à les utiliser au mieux.

L'implantation d'un système automatisé requiert une grande attention à des points périphériques à l'algorithme, qui touchent très rapidement à l'humain. Ici, il s'agit du choix des critères de classement locaux, de la détection et de

18. Atila Abdulkadiroğlu et Tayfun Sönmez. « School Choice : A Mechanism Design Approach ». In : *American Economic Review* 93.3 (juin 2003), p. 729–747. DOI : 10.1257/000282803322157061. URL : http://www.aeaweb.org/articles?id=10.1257/000282803322157061.

la correction des erreurs, qu'elles proviennent des données d'entrée ou de son fonctionnement.

La procédure transparente par conception dont nous avons dressé les grandes lignes présente de nets progrès par rapport à APB et ParcourSup, en regard des spécifications que nous avons présentées : réalisme, absence de gabegie, optimalité de l'affectation, absence de jalousie, limitation de l'anxiété, récompense de l'honnêteté et transparence la rendant digne de la confiance du public. Cependant, elle représente un changement substantiel pour les formations, qui devront alors adapter leurs pratiques à l'exigence de transparence.