

Une nouvelle évaluation : construction et expérimentation d'un barème graduel pour l'exercice de la dictée

par O. Barbarant
Igen groupe des Lettres

L'apprentissage de la norme orthographique constitue l'un des sujets les plus épineux et sensibles dans le système éducatif français, et plus encore dans le champ social, qui accorde à une complexité, d'ailleurs assez particulière à notre langue, un attachement symbolique considérable. Toute ouverture de réflexion à ce propos se trouve noyée dans un halo de nostalgie pour une ancienne école mythifiée, dont on veut croire qu'elle tirait sa grandeur d'un enseignement dont l'excellence est en grande partie imaginaire. Elle renvoie aussi aux incontestables difficultés dans la maîtrise de la langue de certains élèves les plus fragiles, et actuellement les plus malmenés par un système éducatif dont toutes les évaluations, nationales et internationales, se rejoignent pour dire qu'il ne sait pas répondre aux inégalités. Dans tous les cas, la dictée représente davantage qu'un exercice, dont il serait aisé de contester les limites. Elle apparaît comme l'emblème de tout souci de la correction graphique, et semble de ce fait difficile à réformer ailleurs que dans ses marges. Il ne s'agit donc pas ici d'en refaire l'histoire, d'évaluer ses limites, les évolutions de ses pratiques, non plus que de considérer une éventuelle suppression qui susciterait des débats dépassant largement le cadre pédagogique.

Outre qu'elle demeure largement pratiquée dans les classes, et qu'elle constitue aussi un pont entre l'enseignement primaire et le secondaire, la dictée, qui est loin d'être le seul exercice d'appropriation de l'orthographe, présente d'ailleurs, en dépit de son ancienneté qui la rend quelquefois suspecte aux yeux de l'innovation pédagogique, quelques atouts. Certes, elle ne doit surtout pas apparaître comme le seul vecteur de l'enseignement de la maîtrise orthographique, selon une pédagogie par l'exercice dont on sait les limites, ou pis encore selon une logique d'enseignement des règles qui croirait fonder magiquement leur application. Il reste qu'à côté de l'orthographe de production (dans les écrits des élèves) elle offre une concentration sur le seul objet orthographique, tandis que les sciences cognitives nous apprennent que des défaillances rédactionnelles naissent aussi quand la concentration doit régler à la fois l'invention d'une parole et le respect des normes. Un élève ayant à gérer simultanément la réflexion, la recherche de l'expression, la graphie et la mise en page de son écrit, et le souci de la correction de la langue, se trouve ainsi confronté à des tâches mentales multiples, dont on sait qu'elles ne peuvent solliciter également son attention. Si, à terme, l'enseignement de l'orthographe vise bien à libérer la conscience d'une vigilance coûteuse pour favoriser l'expression, il est donc évident que cet objectif final ne contredit pas l'intérêt pédagogique de cloisonner les tâches. Par conséquent, l'orthographe de production ne contredit donc pas l'utilité d'avoir recours à des exercices de réplique des textes – lesquels peuvent d'ailleurs être mieux articulés aux activités de lecture, et favoriser l'appropriation de textes littéraires avant leur étude.

A l'occasion de la réforme des épreuves de Français aux DNB¹ (série générale comme série professionnelle), qui se traduit par un allongement de la dictée proposée aux élèves (désormais de 600 à 800 signes) à partir de la session 2013, le groupe des Lettres de l'inspection générale avait émis le souhait que s'engageât une réflexion sur l'évaluation, de manière à répondre aux exigences des nouvelles modalités de l'examen. Cette perspective de travail donna lieu, lors de l'arrivée, en septembre 2012, d'un nouvel inspecteur général de Lettres prenant en charge l'Académie de Poitiers

¹ Note de service n°2012-029 du 24-2-2012.

(conceptrice des sujets nationaux jusqu'en 2015) à la construction d'une autre modalité de correction. Il s'agissait de proposer, non une simple sanction par la note, mais une discrimination des acquis et des difficultés des élèves, susceptible de faire apparaître des profils de compétences, dans le droit fil des ambitions désormais portées par l'institution pour une approche renouvelée de l'évaluation.

Sur une telle question, la dictée offre une entrée certes paradoxale, mais particulièrement efficace par ce qu'elle offre de concentration exclusive, de réflexion logique et de vigilance orthographique. Exercice canonique en dépit des inflexions et modulations pédagogiques qui l'enrichissent quelquefois dans les classes (dictée préparée, dictée négociée...), elle est l'emblème même de l'évaluation descendante. Loin d'apprécier, voire de valoriser, les compétences des élèves, elle propose d'abord un idéal de perfection - le texte-source - à partir duquel l'enseignant décompte, en négatif, les erreurs commises (qui portent encore, dans la langue commune, le nom révélateur de « fautes »). On n'y part pas des capacités d'un élève moyen, mais d'un modèle d'écriture accomplie, vers lequel l'élève est censé tendre, mais dont on lui montre qu'il est presque nécessairement indigne. Avec un élève en difficulté, une telle pratique ne manque pas d'être décourageante. Même à faire preuve d'application et de vigilance, il est confronté à nombre d'obstacles divers qui font que, s'il tente de consacrer son attention aux accords, il ne manquera pas d'être pris en défaut concernant l'orthographe lexicale. Le plus souvent, une notation disqualifiante et répétitive ne lui offre aucune perspective de progrès. Aussi le verdict tombe-t-il, estimant une « nullité » singulièrement répandue, que l'enseignant ne peut que constater, et le corps social déplorer à sa suite. Le souci de l'orthographe, et la difficulté à faire progresser les élèves, produisent alors autant de niches pour des initiatives privées, proposant des certifications (le plus souvent payantes) qui ne manquent pas de rencontrer de l'intérêt, aussi bien chez les parents d'élèves que chez certains employeurs. Tant pour mettre en œuvre le projet d'inventer des évaluations au service des élèves et de leurs progrès, que pour répondre aux demandes du corps social risquant autrement de ne plus se fier à l'école pour ce domaine d'enseignement, il paraît utile, et peut-être même urgent, que l'Education nationale propose, par de nouvelles modalités d'évaluation, de nouvelles démarches et méthodes. Le projet ici exposé s'inscrit donc dans une triple actualité :

- concernant le traitement de la difficulté scolaire ;
- concernant l'évolution de l'évaluation ;
- concernant l'articulation entre les pratiques du primaire et celles du secondaire, articulation favorisée par la mise en place du cycle CM1-CM2-6^{ème}.

Sur proposition de Mme Anxionnaz, IEN Lettres-Histoire ayant en responsabilité dans l'Académie de Poitiers la commission de sujets nationaux de français pour le DNB série professionnelle, il fut donc décidé de concevoir un barème graduel, prenant en compte les réussites comme les difficultés des élèves. Elaboré d'abord dans un dialogue entre l'IEN et l'inspecteur général, l'instrument a été ensuite évalué avec l'aide des membres de la commission de sujets de la série professionnelle, puis présenté à la commission de sujets de la série générale, où siègent notamment les IPR-IA de Lettres de l'Académie (dont l'une en charge des sujets de la série générale, Mme Catherine Mottet ; l'autre, M. Eric Barjolle, CARDIE). Ce barème a donné lieu à des évaluations et des aménagements au fil de quelques allers-retours avec quelques classes de collège, puis, après sa présentation à M. Delahaye, à une expérimentation de grande ampleur, effectuée sur huit centres de correction lors de la session de juin 2013 du Diplôme national du brevet. L'analyse des résultats, sur les retours de relevés de notes exploitables, a porté sur 1148 copies en série générale et 415 en série professionnelle².

Le cahier des charges que nous nous étions fixé devait répondre à trois exigences : proposer une évaluation ascendante capable de repérer les réussites comme les erreurs (d'où l'enjeu d'un décompte graduel) ; distinguer les difficultés que la notation traditionnelle traite globalement (d'où l'intérêt de proposer des catégories pertinentes) ; mais aussi et peut-être surtout mettre à la disposition des professeurs un outil simple et maniable (d'où l'utilité d'automatiser les décomptes, et

² L'on trouvera, en annexe à ce texte, le document rendant compte de cette expérimentation et de l'analyse de ses résultats par la DGESCO, tel que présenté à M. Delahaye, directeur général de l'enseignement scolaire, le 27 novembre 2013.

de construire des seuils). Le résultat de ce qui fut à la fois une réflexion théorique et une mise en pratique de quelques mois en a permis l'expérimentation Dgesco de juin 2013. Sans formation particulière, des correcteurs ont montré, dans le temps contraint de l'examen, que l'instrument était exploitable.

Les documents, que l'on joint à ce dossier (diaporama et tableur agrémenté de consignes explicatives), permettent de faire l'économie ici de longues explications. A partir d'un texte donné, introduit dans le tableur, le logiciel répartit en colonnes l'ensemble des mots du texte, et propose les entrées : « accords du groupe nominal », « accords du groupe verbal », « orthographe lexicale » (éventuellement « majuscules ») qui permettent de couvrir le champ des erreurs, sans sombrer dans un étiquetage des difficultés orthographiques dont les aspirations à la finesse contredisent l'utilisation pédagogique. Un même terme relève (par exemple « chantait » dans le groupe « il chantait ») de plusieurs catégories : de l'accord du groupe verbal pour la désinence (-« ait »), de l'orthographe lexicale pour la graphie (« chant -»). Ces catégories couvrent le champ des réussites et des erreurs. Le renseignement du tableau permet donc au professeur de considérer l'étendue des difficultés proposées par chaque texte, et le cas échéant de choisir de porter l'accent sur telle ou telle dimension selon ses besoins pédagogiques ou sa progression. Il garde la main pour appliquer une pénalité en fonction d'autres erreurs afin de ne pas laisser une part des compétences orthographiques dans l'ombre. Réciproquement, l'on peut choisir de tenir compte ou non de certaines données ; ainsi l'expérience a-t-elle montré qu'il était préférable, pour ne pas alourdir à l'excès la catégorie de l'orthographe lexicale, d'extraire des décomptes les déterminants, considérés comme relevant d'un degré zéro de l'orthographe, quand l'élève défaillant sur un article fera aussi des erreurs sur d'autres formes plus accomplies (par exemple l'accord dans le groupe nominal). A l'issue de ces choix, le logiciel calcule de lui-même les occurrences, et propose un tableau de proportionnalité, qui attribue des points par des seuils (eux-mêmes mobiles, et donc modifiables selon les besoins). Le professeur relève ainsi les erreurs par catégories, et obtient dès lors un bilan chiffré par catégories d'erreurs (x formes bien orthographiées sur y formes : z points).

La note finale est l'addition des points obtenus dans les trois catégories, qui prennent donc en compte le rapport entre formes réussies et formes erronées. Lors de l'expérimentation de juin 2013, ce travail avait été réalisé en amont, pour ne proposer aux correcteurs que la grille qui en résultait. L'enjeu principal du barème n'est cependant pas d'abord certificatif : l'un des intérêts de la feuille de calcul, notamment pour d'éventuelles formations, est qu'à se pencher sur les catégories et les valeurs, les professeurs prennent en considération les difficultés diverses proposées par l'exercice (ce qui relève de la mémoire lexicale, ce qui relève de la marque des pluriels du nom, ce qui relève des désinences verbales). A l'issue de la correction se dessinent ainsi des profils d'élèves, que les expérimentations réalisées lors de la construction de l'outil mettent clairement en lumière.

En effet, lors de nos travaux, nous avons fait réaliser des dictées dans des classes de 3^{ème} à partir des sujets des annales zéro. Celle extraite d'*Une vie* de Maupassant dans les conditions de l'examen s'est traduite selon le décompte traditionnel par 19 copies sur 25 notées zéro et 6 copies notées de 0,5 à 2,5, dans une 3^{ème} de la voie générale d'un collège ordinaire. Parmi ces productions d'élèves, il est aisé de voir des profils de compétences très distincts, ainsi d'un extrait de la copie de Thibault (« *Nous avons remis notre réponse à tantôt* ». *Elle alleter, étrangler par l'émotion...*) au regard de celle de Gaylord (« *Nous avons remis notre réponse a tanto* ». *Elle alletait, étranglée par l'émotion...*). Dans un cas, les chaînes d'accords paraissent mal maîtrisées ; dans l'autre, c'est la part lexicale qui concentre les difficultés (*tanto, *alleter), mais la marque de l'imparfait (« -ait ») comme l'accord du participe passé adjectival (étranglée) sont acquis. Or les deux copies avaient obtenu 0/20, sans que la correction fasse apparaître des hiérarchies ou à tout le moins des distinctions entre les difficultés et les compétences manifestées par l'une et par l'autre.

Un tel écrasement des profils, décourageant pour les élèves et inefficace pour l'enseignement, apparaît plus nettement encore dans la comparaison de deux copies évaluées selon les deux barèmes, dans une classe de 3^{ème} prépa-pro, à partir d'un autre exercice proposé par les annales zéro. La copie de Corentin, sur un texte extrait d'Annie Ernaux, *La Place*, montrait les formes suivantes : « *Mon père s'ennervé de me voir à longueure de journée dans les livres, mettant sur l'heure conte leurs contre ...et ma movese umere. La lumière sous la porte de ma chambre le soir lui fessais dire que je mussais la santé [...]. Il grennier que l'on me prene pour une pareseuse et luiir pour un graneur* ». L'élève cumule ici nombre de difficultés, touchant à la fois la syntaxe, le lexique, mais aussi régulièrement la correspondance entre le son et la graphie, et pour finir la lisibilité même d'un

texte régulièrement suspendu, troué, dont la restitution ne permet par l'intelligence. Le travail réalisé par Jordann montre un tout autre degré de compétence : « Mon père s'enervait de me voir à longueur de journée dans les livres, mettant sur leur compte mon visage fermé et ma mauvaise humeur. La lumière sous la porte de ma chambre le soir lui faisait dire que je musé la santé. [...] Ils craignait que l'on ne me prenne pour une paresseuse, et lui pour un crâneur ».

Une telle copie montre des irrégularités traduisant une maîtrise de la langue certes encore vacillante : de réelles réussites y voisinent avec des erreurs surprenantes (ainsi de la maîtrise de l'orthographe lexicale, de certaines chaînes d'accords, avec cependant des relations irrégulières entre sujets pourtant proches et désinences verbales). Toutefois, ces manques ne mettent pas en crise la lisibilité d'un texte dans nombre de ses parties restitué. Or, avec une notation traditionnelle, la copie de Corentin obtenait 0/20, et celle de Jordann 2/20, compte tenu d'erreurs de même nature dans d'autres passages. Avec le barème graduel, l'écart se trouve singulièrement plus marqué : 2/20 pour Corentin, contre 8/20 pour Jordann. L'outil ne conduit donc pas à distribuer des points par indulgence excessive ; il permet de distinguer les compétences et de hiérarchiser les difficultés. Une évaluation de cette nature permet dès lors d'envisager des remédiations personnalisées, adaptées à des capacités et des situations distinctes. Ce barème donne ainsi consistance à l'idée, souvent mal comprise, d'une évaluation « bienveillante », c'est-à-dire qui ne renonce pas aux exigences sans lesquelles il n'est pas d'enseignement, mais qui soit mise au service d'une progression des élèves, et pour ce faire capable d'indiquer leurs acquis, leurs éventuelles réussites, et d'aider dès lors à construire le parcours au sein duquel construire leurs apprentissages.

L'intérêt accordé à cette expérience a conduit la Dgesco à réaliser l'expérimentation sur une échelle plus vaste en juin 2013. Les ambitions de cette expérience étaient les suivantes :

- vérifier la maniabilité de l'instrument pour des professeurs qui, n'ayant pas été associés à la recherche, devaient le découvrir et se l'approprier ;
- évaluer (grâce à des fiches permettant de dire l'intérêt ou les limites de l'instrument aux yeux des correcteurs) son degré de recevabilité et de pertinence, pour d'éventuels aménagements ;
- étudier, statistiquement, les effets produits sur la notation.

Les résultats de l'étude dont on joint en annexe les données statistiques, ont fait apparaître dans l'ordre les points suivants :

- de la part de correcteurs contraints à corriger selon les modalités certificatives en vigueur leur lot de copies, puis de revenir sur la notation de la dictée selon le barème expérimenté, dans une seule journée et sur le lieu d'examen, les retours sont largement positifs. Les quelques remarques concernant le caractère fastidieux de la répartition en trois catégories selon la nature des erreurs paraissent même, aux yeux des concepteurs, en grande partie explicables par les conditions d'urgence et de fatigue liées à la situation de l'examen.
- Les correcteurs ont marqué majoritairement un intérêt pour les conséquences pédagogiques de l'outil. Ils se sont interrogés sur sa pertinence dans une démarche certificative, attendu que ce sont surtout les profils d'élèves et donc les remédiations possibles qui ont, à juste titre, retenu leur attention. Certains se sont exprimés sur le « laxisme » (item délibérément proposé sur les fiches recueillant les avis) d'une nouvelle évaluation. Nombreux ont déclaré le barème « injuste », en ce que les seuils proposés, sur une échelle réduite à 6 ou 5 points selon les filières du DNB, alors que nos expérimentations portaient sur une échelle de 20 points, au dégradé plus subtil, produisaient des effets désavantageux pour les élèves compétents.

Cette remarque appelle deux analyses. D'une part, les professeurs ont pointé ici un défaut qu'il est aisé de corriger, les seuils étant mobiles : reconnaissons que, soucieux surtout des élèves en difficulté et de la production d'écart entre les deux systèmes de notation, nous avons concentré les seuils autour de ceux concernant les élèves moyens ou en difficulté. En relevant le niveau d'exigence du premier seuil, les élèves compétents retrouveraient leur notation initiale. Encore une fois, le barème est entièrement mobile, et dépend de choix dont il est important que les professeurs, souvent

peu au fait des enjeux d'une orthographe plus évaluée qu'enseignée, puissent les faire et les appréhender. D'autre part, l'insistance sur les effets produits sur les « bonnes » copies nous paraît révélatrice d'une constante de l'enseignement français, et de ce qui est encore la majorité des mentalités professorales : tandis que les concepteurs de l'outil portaient leurs regards – presque exclusivement, comme nous venons de le reconnaître – sur les difficultés, c'est d'abord à l'excellence pénalisée que les professeurs ont accordé leur attention. Les remontées de l'expérimentation éclairent ainsi sur le travail qu'il faut continuer de mener pour que le système éducatif prenne en considération, comme une priorité, le sort fait à la difficulté, voire à la grande difficulté, dont on sait combien un traitement défaillant pénalise aujourd'hui l'ambition éducative.

Du côté de la Dgesco, l'expérimentation a conduit aux conclusions suivantes, citées depuis le compte rendu adressé par le bureau des collègues à la direction :

- de l'analyse des résultats saisis il ressort la nécessité d'ajuster les seuils et les occurrences du barème - ce que permet l'instrument ;
- à l'heure où prolifèrent les projets en tous genres pour « redresser » les performances orthographiques, il s'impose que l'Éducation nationale prenne la main et l'initiative pour impulser une dynamique de l'enseignement de l'orthographe en particulier, de la maîtrise de la langue française en général.

Compte tenu de ces éléments, une réflexion réunissant M. Jean-Paul Delahaye (Direction générale de l'enseignement scolaire), Mme Brigitte Réauté (bureau des collèges) et M. O. Barbarant, inspecteur général, a conclu que le barème devait faire l'objet d'une diffusion, à travers des informations et formations, afin qu'il puisse contribuer à sa mesure à la rénovation des modalités d'évaluation. Il a été conclu que l'enjeu pédagogique dépassait l'utilisation certificative, si bien qu'il serait, stratégiquement, préférable d'inviter les professeurs à se l'approprier plutôt que de le faire connaître à travers le DNB, selon une logique de pilotage par l'examen dont les effets sur la réception pourraient être néfastes. Pour l'inspection générale, le barème offre aussi un outil aidant à considérer l'enseignement de la maîtrise de la langue, dont on sait qu'il est constamment mis en avant comme une priorité, mais assez mal réalisé dans les classes. Le barème prolonge ainsi dans la mesure de ses moyens le bilan, et propositions formulées, lors du travail collaboratif mené par le groupe des Lettres et le groupe Premier degré au cours de l'année 2012-2013 (« comité de convergences sur l'étude de la langue ») qui a fait l'objet d'un rapport transmis via le Doyen de l'inspection générale, puis d'une note de synthèse. Cette analyse rejoint celle formulée notamment par les inspecteurs Lettres-Histoire de la voie professionnelle, lors d'une présentation de ce projet faite à l'ensemble du corps en décembre 2012 pour recueil d'avis et mise en discussion.

Du côté de l'inspection générale, il s'agit dès lors de poursuivre le travail de diffusion, notamment dans l'Académie de Poitiers qui en fut l'initiatrice, à travers les réseaux des IPR-IA et notamment du CARDIE. Mme Sophie Anxionnaz travaille à cette heure en collaboration avec une inspectrice du premier degré, qui, découvrant l'outil, veut envisager ses applications et modulations pour les formations des professeurs des écoles. Une présentation de l'outil aura lieu aussi pour l'ensemble des IPR de Lettres. Nous avons mis à disposition du bureau des collègues de la Dgesco l'outil numérique, une brève vidéo de présentation et de son mode d'emploi, ainsi que le présent texte, s'efforçant de situer les enjeux méthodologiques et pédagogiques de ce travail. Une communication des inspecteurs concepteurs à l'occasion des assises de l'évaluation (programmées désormais pour l'automne 2014) assurera aussi la diffusion de l'information, de manière à ce qu'il se répande dans les classes.