

Répondre aux besoins de formation des adultes francophones en partant des erreurs qu'ils produisent

Camille Martinez, Groupe EDUCLEVER

Résumé. Face au manque de maîtrise des adultes francophones dans la compétence écrite, les formations en orthographe se sont installées parmi les formations professionnelles. Comment recueillir leurs besoins de formation et y répondre ? Pour construire un programme de formation adapté, on évite d'aborder exhaustivement toutes les difficultés de la langue, pour se focaliser plutôt sur les erreurs courantes et leurs conditions d'apparition. On explore pour cela de grands corpus d'erreurs (WiCoPaCo, Ortholexies) et des corpus d'écrits ordinaires (Grand Débat) et professionnels. Recourir à plusieurs corpus permet de recouper leurs données pour gommer leurs éventuels biais respectifs. Les erreurs relevées sont organisées dans une typologie d'erreurs, qui repose sur la théorie du système graphique de N. Catach. Des briques de formation pourront ensuite être construites à partir de cette typologie.

« Savoir communiquer en français » est l'un des piliers du *Socle de connaissances et de compétences professionnelles* (COC 2014). Avec le large développement des outils numériques au travail, l'expression écrite est devenue une compétence de base, dont la maîtrise est attendue pour s'insérer sur le marché de l'emploi. Or, plusieurs études relèvent un manque de maîtrise des adultes francophones natifs dans cette compétence (Blondel *et al.*, 2016 ; Cogis, 2001 ; Martin-Lacroux, 2015).

La première conséquence de ce manque de maîtrise concerne les individus. Dans le pire des cas, le lecteur leur prête un manque de connaissances en français et éventuellement dans d'autres domaines ; il estime qu'ils n'ont pas de légitimité pour s'exprimer, qu'ils manquent de sérieux, qu'ils ne sont pas respectueux (Martinez, 2013). La productivité de l'individu à l'écrit est entravée, tout comme sa carrière.

L'autre conséquence, c'est l'émergence d'un besoin de formation en communication écrite. Longtemps, des livres ont rempli ce rôle (dictionnaires de difficultés, manuels d'orthographe ; mais aussi Gaillard, 2017 ; Maillet, 2020, etc.), mais depuis la fin des années 2000 des formations en e-learning ont investi ce créneau (*Orthodidacte, Projet Voltaire, Ecri+, Frantastique, Mon coach Bescherelle, Formation Le Robert*, etc.). Elles s'adressent principalement aux adultes francophones, actifs ou étudiants, désirant consolider leurs acquis. Certes, de par leur format numérique, les formations en e-learning innovent. Mais sur le plan du contenu enseigné, y a-t-il du nouveau ? Pas vraiment. Dans presque toutes ces formations, les contenus enseignés recourent ceux des livres déjà mentionnés : collection de difficultés déconnectées des pratiques réelles, description exhaustive de règles et d'exceptions. Ces traditions ont la vie dure.

Pourtant, dans ce domaine, une autre approche est possible, en repartant du public visé. Quel est le besoin des adultes francophones, pour devenir autonomes à l'écrit dans un cadre professionnel ? Pas de réviser toutes les difficultés de la langue française, pas de devenir des grammairiens, mais plutôt de faire face à des questions pratiques qui les bloquent quand ils rédigent. Que proposer pour répondre à ce besoin ? Pas une transposition des méthodes employées dans un cadre scolaire, mais plutôt des contenus clairs et directs, dénués de métalangage, qui s'appuient sur l'expérience des adultes. D'où l'idée de repartir des erreurs produites par les adultes en situation de rédaction. Quels mécanismes sont impliqués dans les

erreurs produites ? Y a-t-il des erreurs fréquentes, des schémas d'erreur qui se répètent ? À défaut de disposer de recherches ou d'enquêtes sur les besoins en formation FLM exprimés par les adultes, on se propose donc d'étudier leurs productions écrites.

La première étape consiste à collecter et à créer des corpus d'erreurs linguistiques. Le corpus WiCoPaCo (Wisniewski *et al.*, 2010) est un recueil de 400 000 modifications apportées à l'encyclopédie Wikipedia en 2009-2010, dont une majorité de corrections d'erreurs linguistiques. Après une étape de tri, ses données deviennent utilisables. Cependant, ce corpus n'est pas optimal pour notre besoin : il n'est pas produit par des adultes en situation de rédaction professionnelle et il comporte des biais thématiques, lexicaux et grammaticaux (rareté de la deuxième personne, surutilisation du passé simple). Le corpus Ortholexies (Manguin, 2016) a été récolté à partir des requêtes envoyées par les utilisateurs sur le dictionnaire de synonymes du Crisco. Il ne peut pas être téléchargé, mais consulté pour connaître les formes erronées associées à un lemme correct. Ce corpus comporte un artéfact lié à l'instrument d'écriture des utilisateurs, le clavier : les graphies sans signes diacritiques semblent surreprésentées.

Pour atténuer les biais, il est bon de recourir à un plus grand nombre de corpus. Pour cela, on se propose de constituer d'autres collections d'erreurs. On part de grands corpus existants (Grand débat 2019) et de corpus d'écrits constitués pour l'occasion (écrits professionnels, 450 000 mots, non publié), sur lesquels on passe un outil de détection des erreurs linguistiques, afin de les cartographier.

Encore faut-il adopter une typologie d'erreurs adéquate pour classer le matériau variationnel détecté. Plusieurs auteurs (Gey, 1983 ; Mitaine, 1982-1983 ; Karabetian, 1983 ; Millet, 1989 ; Aubergé *et al.*, 1999 ; Rey et Sabater, 2008 ; Wisniewski *et al.*, 2010 ; Yvon et Segal, 2012) ont proposé des typologies, appliquées dans divers contextes (scolaire ou non), reposant sur divers modèles théoriques (en particulier la théorie du système graphique, Catach, 1980) et plus ou moins englobantes (erreurs phonographiques seules, composante morphologique, composante syntaxique, etc.). Plusieurs d'entre elles alimentent la typologie d'erreurs retenue, qui se décompose en trois volets : une arborescence qui s'applique aux opérations sur la chaîne graphique, un pan qui nécessite un taggage morphosyntaxique préalable, enfin une partie qui décrit les erreurs de type « mot mis pour un autre » (homophonie, paronymie). Cette typologie doit être interprétable par la machine, afin de faire fonctionner l'outil de détection d'erreurs.

L'étude des erreurs codées à l'aide de la typologie révèle des schémas d'erreur : des patterns dans lesquels les chances d'erreurs paraissent plus élevées. Par exemple, sur la question des accords, plusieurs facteurs jouent un rôle dans l'apparition d'erreurs : le fait que le donneur d'accord ait une forme non canonique, la présence d'autres mots entre le donneur et le receveur, le fait que les constituants soient dans un ordre non canonique, etc. Autre exemple, qui touche l'homophonie : on observe que les mots *parti* et *partie* sont surtout confondus dans l'expression *faire partie de*, et plus rarement dans d'autres contextes. Troisième exemple, qui concerne la phonographie : des mots comme *notamment*, *suffisamment*, *britannique*, *appellation*, *occurrence*, *débarrasser*, *fréquemment*, sont des mots dans lesquels une consonne double a de très grandes chances d'être absente. De tous ces relevés, on déduit les difficultés qui se présentent aux scripteurs.

Il reste à construire des briques de formation à partir de ces données et à les ordonner dans un parcours d'apprentissage. Ces briques de formation répondront à des difficultés réelles rencontrées par les scripteurs adultes, en passant sous silence certaines difficultés théoriques (jusque-là considérées comme incontournables dans les approches classiques exhaustives). Dans le parcours d'apprentissage, elles seront dosées proportionnellement au volume des erreurs rencontrées en corpus. Comme les contenus de formation pour adultes, dans ce domaine, ne reposent sur aucun programme officiel, on se laisse guider par les données.

Références bibliographiques

- Aubergé, V., Ghneim, N. et Belrhali, R. (1999). Analyse du corpus ORTHOTEL : apport du traitement automatique à la classification des déviations orthographiques. *Langue française*, 124, 90-103. https://www.persee.fr/doc/AsPDF/lfr_0023-8368_1999_num_124_1_6308.pdf.
- Blondel, C., Brissaud, C. et Rinck, F. (2016). Description des pratiques orthographiques de scripteurs ordinaires et de scripteurs en difficulté avec l'écrit à partir de l'analyse d'une dictée. *Actes du Congrès mondial de linguistique française – CMLF 2016*. DOI: 10.1051/shsconf/20162707.
- Catach, N. (1980). *L'orthographe française. Traité théorique et pratique*. Paris : Nathan.
- Comité observatoire et certifications [COC] (2014). *Socle commun de connaissances et de compétences professionnelles*.
- Cogis, D. (2001). Difficultés en orthographe : un indispensable réexamen. *Revue française de linguistique appliquée*, VI, 47-61. <https://doi.org/10.3917/rfla.061.0047>.
- Gaillard, B. (2017). *100 jours pour ne plus faire de fautes*. Paris : Les éditions de l'Opportun.
- Gey, M. (1983). Typologie des erreurs orthographiques : résultats de tests. *Liaisons HESO*, 9, 43-55.
- Karabetian, É. (1983). Utilisation de la typologie des erreurs et intégration du plurisystème graphique. Remarques théoriques et exploitation pédagogique. *Liaisons HESO*, 10, 45-49.
- Maillet, J. (2020). *5 minutes par jour pas plus ! Pour ne plus faire de fautes*. Paris : L'étudiant.
- Manguin, J.-L. (2016). « Ortholaxies », une base de données publique pour l'orthographe lexicale. *Actes du Congrès mondial de linguistique française – CMLF 2016*. https://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_11006.pdf.
- Martinez, C. (2013). L'orthographe du français : représentations et comportements des scripteurs. Linguistisches Kolloquium Romanistik Universität Freiburg im Breisgau, non publié.
- Martin-Lacroux, C. (2015). *L'appréciation des compétences orthographiques en phase de présélection des dossiers de candidature : pratiques, perceptions et implications pour la GRH*. Thèse de doctorat, Université de Toulon. <https://tel.archives-ouvertes.fr/tel-01295234>.
- Millet, A. (1989). Essai de typologie des variations graphiques. *Lidil*, 1.
- Mitaine, G. (1982-1983). Expérimentation des recherches HESO appliquées au cycle primaire (C.E.2). Compte-rendu. *Liaisons HESO*, 7(10), 54-66 ; 49-53.
- Pach, C. et Jacquemin, D. (1994). Le quotidien de l'orthographe ordinaire. Usages et normes dans les écrits ordinaires. *L'orthographe de tous les jours. Enquête sur les pratiques orthographiques des Français*. Paris : Champion.
- Rey, V. et Sabater, C. (2008). *Écriture, orthographe, dysorthographe*. Presses universitaires de Provence.
- Wisniewski, G., Max, A. et Yvon, F. (2010). Recueil et analyse d'un corpus écologique de corrections orthographiques extrait des révisions de Wikipédia. *TALN 2010*, <https://wicopaco.limsi.fr/pub/taln10.pdf>.
- Yvon, F. et Segal, N. (2012). Des corpus d'erreurs pour TRACE. Rapport technique, <https://anrtrace.limsi.fr/SpellingCorpus.pdf>.