

## Orthographe du français (particularités)

Le mot **orthographe** (mot latin repris du grec, *ortho* « correct » et *graphein* « écriture ») fait référence à une norme pour la représentation écrite d'une langue, de sa **graphie** (« manière dont on écrit un mot »). Les langues parlées en Europe ont adapté un système d'écriture appelé 'alphabet', constitué d'un ensemble de symboles appelés 'lettres' et que les linguistes (selon le modèle terminologique usuel en linguistique - 'phonème', 'morphème', 'sémème', etc.) appellent parfois **graphèmes** (« unité graphique minimale »). Selon le type d'écriture (phonétique, syllabique ou idéographique) un graphème peut représenter un phonème, une syllabe (comme le système kana du japonais) ou un concept (les idéogrammes du chinois, les hiéroglyphes égyptiens). Dans l'écriture alphabétique, les graphèmes représentent, en principe, les phonèmes d'une langue.

L'alphabet grec est le premier alphabet qui a noté aussi les voyelles, à la différence de l'alphabet phénicien, dont il provient. L'alphabet grec se trouve à l'origine de deux autres alphabets européens importants, l'alphabet latin (adapté pour les langues de l'Europe de l'ouest, de l'Europe du nord et de l'Europe centrale) et l'alphabet cyrillique, employé surtout par les langues slaves, des peuples slaves orthodoxes (bulgare, macédonien, serbe, ukrainien, russe, biélorusse, ...) mais, jusqu'à la moitié du 19<sup>e</sup> siècle, utilisé aussi par une langue romane d'un peuple orthodoxe, le roumain.

Pour la grande majorité des langues qui emploient l'alphabet, en règle générale chaque lettre (chaque graphème) représente un phonème. Pour s'adapter aux particularités phonologiques, certains graphèmes ont reçu des diacritiques : accents (é, â, û), cédille (ç, è), tréma (ë, î), une 'brève' (la lettre ă en roumain), etc. Certains groupes de lettres peuvent représenter un ou deux phonèmes (par exemple, en italien et en roumain le groupe *che, chi* sont lus [ke, ki], en italien le groupe *sc* + voyelle note le phonème /ʃ/, qui, en roumain, est écrit avec un *s* portant une virgule diacritique, ș, en français par le groupe *ch*). On dit que les langues qui appliquent le principe général un graphème ↔ un phonème sont des langues à orthographe phonétique.

Pour des raisons historiques et culturelles, certaines langues ont conservé un nombre important de graphies qui rappellent l'étymon ou la forme du mot dans un état de langue antérieur. Les deux langues qui présentent un nombre important de mots à écriture étymologique sont le français et l'anglais.

Le premier texte important en ancien français qui nous est parvenu est le manuscrit des *Serments de Strasbourg* (842). Tout au cours du Moyen Âge, on enregistre une riche production de textes, surtout poétiques, dans les deux dialectes (au nord les dialectes d'oïl, au sud - les dialectes d'oc), tandis que la langue des documents administratifs et judiciaires du royaume reste le latin. En absence d'une autorité normative, chaque auteur, chaque copiste choisissait sa propre manière d'écrire, donc la même parole 'jouissait' d'un nombre important de variantes graphiques. Vu que les personnes apprenaient à lire et à écrire d'abord en latin, beaucoup de ces formes graphiques étaient 'ennoblies' par une graphie proche à leur étymon latin, rappelant le lien génétique, prestigieux, entre les mots français et les mots latins correspondants. Selon Catach (2001), le français écrit commence à se détacher du latin seulement à partir du 13<sup>e</sup> siècle, mais, dans l'ensemble, les textes écrits en français restent minoritaires par rapport au latin.

La situation du français change entièrement en 1539, quand le roi François Ier (1515-1547) signe l'ordonnance de Villers-Cotterêts qui impose l'emploi du français dans les actes juridiques et administratifs, à la place du latin. Cet acte s'inscrit dans un courant européen qui, au cours de la Renaissance, a affirmé la noblesse des 'nouvelles' langues (italien, espagnol, français), considérées jusqu'alors 'vulgaires' (Dante Alighieri en 1303 pour l'italien, Antonio de Nebrija en 1492 pour l'espagnol, Joachim du Bellay en 1549 pour le français). Pour le français et l'espagnol, une langue unitaire était nécessaire, en tant qu'expression linguistique d'un État monarchique centralisé (v. Costăchescu 2020).

L'unification normative de l'orthographe du français commence en 1635, avec la création de l'Académie française, dont la principale mission était d'établir le 'bon usage' du français par la rédaction d'une grammaire et, surtout, d'un dictionnaire (première édition en 1694). Jusqu'à nos jours, l'Académie française et son dictionnaire font actes d'autorité et les formes recommandées sont obligatoires en France tant dans les documents de l'administration et de la justice que dans l'enseignement. L'Académie française a repris et autorisé certaines graphies à caractère étymologique qui étaient utilisées depuis des siècles. La norme n'a pas changé la manière d'écrire de certains mots, même quand leur prononciation a été modifiée par l'évolution de la langue au cours des siècles, preuve que l'orthographe change moins vite que la langue parlée, surtout si la norme (comme celle du français) est très conservatrice. Pour cette raison, l'orthographe française, tel qu'elle a été fixée aux 16<sup>e</sup>-17<sup>e</sup> siècles, est « étymologisante » (Dubois *et al.* 2002).

Nous prenons comme exemple l'évolution phonétique du diphtongue *oi* qui, en français moderne, est prononcé [wa] dans des mots comme *poids, voix, avoine*, etc. Au 12<sup>e</sup> siècle cette diphtongue était prononcée

[ɔj] et la graphie illustre cette manière de la prononcer. Il existe une preuve indirecte de cette situation, offerte par des mots d'origine française contenant cette diphtongue qui sont entrés en anglais. On sait qu'après la Conquête normande de Guillaume le Conquérant (1066) il y avait, en l'Angleterre, deux communautés linguistiques : une communauté petite (environ 10.000 personnes) d'aristocrates, qui parlait l'anglo-normand (un dialecte du français médiéval de Normandie) et une communauté majoritaire (1,5 million d'habitants) qui parlait une langue germanique (l'anglo-saxon, appelé aussi 'vieux anglais'), communauté qui vivait sous le 'joug normand'. Ces deux groupes ont fusionné, du point de vue linguistique, par un processus graduel considéré achevé au 14<sup>e</sup> siècle, processus qui a conduit à la présence d'un nombre important d'emprunts du français à l'anglo-saxon, durant la période du moyen-anglais, nombre estimé à environ un tiers du lexique (Iliescu 2013, Trotter 2013). Une bonne partie de ces mots empruntés ont conservé la forme de l'ancien français. Le mot anglais *voice* [vɔis] « voix » existait dans l'anglo-normand (depuis 1121 cf. AND s. v.) ; le mot est attesté en ancien français aussi, à la fin du 11<sup>e</sup> siècle (provenant de *vocem* l'accusatif de *vox*, *vocis* « voix » TLFi. s. v.) ; en anglais, la diphtongue continue à être prononcée comme dans l'ancien français *oi* [ɔi]. Dans le français parlé sur le continent, cette diphtongue a continué son évolution : comme montré par les rimes du *Roman de la Rose*, au 13<sup>e</sup> siècle elle a été prononcée [wɛ] et, au début du 19<sup>e</sup> siècle, [wa], prononciation autorisée par l'édition de 1835 de *Dictionnaire de l'Académie française* (Pasques 1975) et qui s'est conservé jusqu'à présent.

On dit que le français a une orthographe étymologisante parce que la graphie de beaucoup de mots n'a pas tenu compte des modifications qui se sont manifestées dans la phonétique/ phonologie historique de cette langue, la norme a maintenu une forme qui correspond à la prononciation dans une période antérieure, souvent médiévale.

Le caractère 'étymologique' de l'orthographe du français est relatif. Un très grand nombre de mots ont une orthographe phonétique : *amical* [amikal], *café* [kafɛ], *malaga* [malaga], *nodal* [nɔdal], *spirituel* [spiritɥɛl], *tact* [takt], etc. À l'autre extrémité se situe un mot comme *oiseau*, prononcé [wazo], où aucune lettre ne correspond à un phonème, mais c'est un cas rare. Les éléments étymologiques apparaissent quand il s'agit de mots contenant les groupes *oi*, *ai*, *eau*, *au*, etc. prononcés, en français moderne, [wa], [ɛ], [o], et dont la graphie reflète une prononciation antérieure, car, pour ces mots, les normes d'orthographe n'ont pas suivi les modifications phonologiques historiques.

## Bibliographie

- AND (2024) = *Anglo-Norman Dictionary*, Aberystwyth University, <<https://anglo-norman.net/>>  
 Catach, Nina (2001), *Histoire de l'orthographe française*, Paris, Honoré Champion.  
 Costăchescu, Adriana (2020) 'Le roumain, une langue hospitalière - parce que (plus) libre', in Zoltán Rostás/ Theodora-Eliza Văcărescu (éds.), *In honorem Sanda Golopenția*, București, Spandugino, 413-437.  
 Dubois, Jean *ed alii* (1994/ 2002), *Dictionnaire de linguistique*, Paris, Larousse.  
 Iliescu, Maria (2013), 'À la recherche des universaux de contact : Anglo-saxon et normand vs roumain et slave', in Maria Iliescu, *Varia Romanica*, Berlin, Frank & Timme, 33-44).  
 Pasques, Liselotte (1975), 'L'ancienne diphtongue *oi*. Son évolution phonique et graphique en français moderne', in *Romania* 318, 67-82; <[https://www.persee.fr/doc/roma\\_0035-8029\\_1975\\_num\\_96\\_381\\_2453](https://www.persee.fr/doc/roma_0035-8029_1975_num_96_381_2453)>  
 Riegel, Martin / Jean-Christophe Pellat / René Rioul (1994/ 2009), *Grammaire Méthodique du Français*, Paris, Presses Universitaires de France ; quatrième édition revue et augmentée 2009  
 TLFi (2002) = *Trésor de la langue française informatisé*, <<http://www.atilf.fr/tlfi>>  
 Trotter, David (2013), 'Where does anglo-norman begin and end?', in *Romance Philology*, 67, 139-177  
 <<https://www.jstor.org/stable/44742003>>