

Grammaire et histoire de la langue

Code du cours : L1LFB011

Horaire : mardi, 13h00 à 14h00

Salle : Grand Amphi, Centre Universitaire Malesherbes

CM : André Thibault

TD : Élodie de Oliveira, Hélène Biu, Gilles Couffignal, Camille Bellenger

Semaine 4 : Lexicologie historique ; les composantes historiques du lexique français. – Concepts de substrat, superstrat, adstrat ; l'emprunt. – Mots héréditaires, mots savants, doublets. – Innovations (formelles) internes. – Bref aperçu de la lexicographie historique du français.

1. Les concepts de substrat, superstrat et adstrat ; l'emprunt

On distingue dans la longue histoire d'une langue comme le français un certain nombre de « strates » historiques qui ont contribué à façonner son vocabulaire (entre autres). Il est usuel en lexicologie historique de distinguer les concepts suivants : substrat ; superstrat ; adstrat.

En outre, il convient de préciser que ces concepts sont intimement liés à celui de l'emprunt. Nous utiliserons ici le mot *emprunt* pour nous conformer à un usage établi (et indéradicable) mais il faut rappeler qu'il repose sur une métaphore trompeuse, qui réifie les mots en en faisant des objets extra-linguistiques que les locuteurs se passeraient d'une langue à une autre comme on le fait d'un livre ou d'un ballon. Cela opacifie complètement le fait que les mots appelés à tort *emprunts* ne sont en rien des emprunts, mais bien des créations nouvelles, différentes de leur modèle à tous les points de vue (phonologique, morphologique, sémantique). Il serait infiniment préférable de parler d'**imitations** (approximatives et encadrées par les structures de la langue d'accueil), d'**adaptations**, ou d'**imitations adaptées**.

1.1. Substrat.

On appelle « langue de substrat » une langue qui a été abandonnée par sa communauté linguistique d'origine, tout en ayant laissé des traces dans la nouvelle langue adoptée par cette communauté. Dans la préhistoire de la langue française, le gaulois est considéré comme une langue de substrat pour le latin vulgaire de la fin de l'Antiquité en Gaule, celui qui allait devenir le galloroman, puis éventuellement l'ancien français (ainsi que le francoprovençal et l'occitan)¹. En effet, les Gaulois ont abandonné leur langue celtique ancestrale, le gaulois, pour adopter le latin, la langue de l'Empire Romain, auquel ils ont été annexés. Le latin était donc, à la base, une langue étrangère pour eux, qu'ils ont dû assimiler tant bien que mal. Dans ce genre de situation, on constate toujours que la langue ancestrale laisse des traces dans la nouvelle langue adoptée (cf. par exemple les mots d'origine alsacienne dans le français parlé par les Alsaciens aujourd'hui). On trouve de même dans le latin de Gaule certaines unités lexicales (ainsi que certaines habitudes articulatoires) qui représentent des « emprunts » du latin au gaulois ; il s'agit, si l'on veut, de mots gaulois latinisés, ou pour le dire plus justement, de néologismes latins créés en imitant/adaptant un mot gaulois.

¹ La plupart des auteurs disent simplement que le gaulois est un substrat du français, mais il nous semble que cela consiste à brûler les étapes ; lorsque le gaulois disparaît (au plus tard au V^e siècle), le français n'existe pas encore. Entre le gaulois et l'ancien français, il y a d'abord l'étape intermédiaire du roman, plus précisément du galloroman, qui n'était plus vraiment du latin mais qui n'était pas encore de l'ancien français.

Voici quelques exemples de mots français remontant à un étymon du latin de Gaule résultant d'un « emprunt » au gaulois ; il faut noter que le nombre total de mots français remontant au gaulois et dont l'origine est sûre n'est pas très élevé.

- *arpent* n. m. « ancienne mesure agraire » (du lat. *arepennis*, accusatif *arepennem*, empr. du gaulois).
- *cervoise* n. f. « bière d'orge, de blé, etc. en usage chez les Anciens et au moyen âge » (du lat. impérial *cerevisia*, d'orig. gauloise).
- *chêne* n. m. « (esp. d'arbre) » (du bas lat. **cassanus*, mot d'origine gauloise).
- *crème* n. f. « matière grasse et onctueuse du lait » (du lat. pop. *crama* d'orig. gauloise).
- *dru*, *-ue* adj. « épais, serré, touffu » (d'un mot du lat. pop. de Gaule remontant au gaulois **drûto* qui signifiait « fort, vigoureux »).
- *if* n. m. « (esp. d'arbre) » (d'un mot du lat. pop. de Gaule remontant au gaulois **ivos*).
- *lieue* n. f. « ancienne mesure itinéraire » (du lat. *leuca*, d'un mot gaulois, selon les auteurs latins).
- *marne* n. f. « mélange naturel d'argile et de calcaire (glaise) » (d'un lat. pop. **margila*, mot gaulois).
- *pièce* n. f. « partie séparée d'un tout » (d'un mot du lat. pop. de Gaule remontant au gaulois *pettia*).

1.2. Superstrat

1.2.1. Définition

On appelle « langue de superstrat » une langue apportée par des envahisseurs qui n'arrivent pas à l'imposer aux populations qu'ils ont conquise, mais ayant tout de même laissé des traces dans la langue de celles-ci. Dans l'histoire de la langue française, les langues germaniques que sont le francique et le norrois, apportées respectivement par les envahisseurs francs et scandinaves, sont considérées comme des langues de superstrat car elles sont venues se superposer au galloroman en l'influençant mais sans toutefois réussir à le supplanter (car si c'était le cas la France parlerait aujourd'hui une langue germanique et non une langue romane !).

1.2.2. Les superstrats du français

1.2.2.1. Le francique

Vers la fin de l'Antiquité et le début du Moyen Âge, l'Empire romain est envahi par des hordes de guerriers germaniques. Parmi ceux-ci, les plus importants pour la préhistoire de la langue française sont les Francs. Ils déferlent sur le nord de la Gaule et s'installent en seigneurs de guerre au sein des populations galloromanes sur place ; comme nous l'avons déjà mentionné il y a quelques semaines lorsque nous avons commenté les Serments de Strasbourg, une période de deux ou trois siècles de bilinguisme roman / germanique a régné sur le Royaume des Francs. La langue germanique des Francs, le francique, n'a pas réussi à s'imposer et a fini par disparaître, mais non sans avoir d'abord légué au galloroman (et donc à l'ancien français naissant) un assez bon nombre de types lexicaux qui caractérisent le français au sein des langues romanes.

Quelques exemples de mots d'origine francique en français :

- *adouber* v. tr. « accorder (à quelqu'un) le titre de chevalier, en (le) frappant rituellement du plat de l'épée » (probablement du francique *dubban* « frapper », parce que le futur chevalier recevait de son parrain un coup sur la nuque).
- *bannir* v. tr. « condamner » (du francique **bannjan* « proclamer, convoquer des troupes »).
- *baron* n. m. « grand seigneur du royaume » (du francique **baro* « homme libre »).
- *bleu* adj. (du francique **blao*, de même sens).
- *bois* n. m. (du francique **bosk* « buisson »).
- *fauteuil* n. m. « siège à dossier et à bras, à une seule place » (du francique **faldistól* « siège pliant »).
- *framboise* n. f. (du francique **brambasia* « mûre », devenu *frambeise*, *framboise* sous l'influence de l'initiale de *fraise*).
- *hache* n. f. « outil servant à fendre le bois » (du francique **happia*, de même sens).
- *hareng* n. m. « (esp. de poisson de mer) » (du francique *haring*).
- *heaume* n. m. « casque d'une armure médiévale » (du francique **helm*).
- *héberger* v. tr. (du francique **heribergôn* « loger »).
- *honnir* v. tr. « vouer au mépris public, conspuer, mépriser » (du francique **haunjan*, cf. all. *höhn*).
- *marcher* v. intr. (du francique **markôn* « marquer, imprimer le pas »).
- *rang* n. m. (du francique **hring* « anneau, cercle », cf. all. *Ring*, angl. *ring*).
- *troupe* n. f. (du francique *throp* « village, troupeau, tas »).

1.2.2.2. Le norrois

Au 9^e siècle, des invasions en provenance de Scandinavie commencèrent à dévaster les côtes septentrionales de la France. Ces envahisseurs finirent par s'implanter dans le pays, dans la région qu'on appella désormais *Normandie*, du nom de ces « hommes du nord » (*Normanni*). Ils s'assimilèrent très rapidement, et ne légèrent à la langue française que quelques mots de leur langue ancestrale (le *norrois*, ou ancien scandinave), la plupart tirés du vocabulaire maritime (*crique*, *vague*, *tillac*, *étrave*, *étambot*, etc.), ainsi que quelques dizaines de toponymes.

1.3. Adstrat

1.3.1. Définition

On parle de « langue d'adstrat » lorsque deux communautés linguistiques « s'empruntent » des mots sans que la langue de l'une de ces deux communautés linguistiques cède la place à l'autre (comme c'est le cas avec la langue de substrat, celle d'un peuple envahi qui l'abandonne pour passer à celle de l'envahisseur ; et comme c'est le cas aussi avec la langue de superstrat, celle d'un peuple envahisseur mais qui l'abandonne aussi pour passer à celle du peuple qu'il a conquis).

À vrai dire, les langues de substrat et de superstrat commencent d'abord par être des langues d'adstrat, elles aussi ; ce n'est qu'à partir du moment où elles sont abandonnées qu'on peut dire qu'elles deviennent des langues de substrat ou de superstrat (ces concepts impliquent des langues ayant été abandonnées par leurs communautés linguistiques respectives, la langue des conquis dans le cas du substrat, la langue du conquérant dans le cas du superstrat). Les

langues d'adstrat, en revanche, ne sont pas des langues ayant été abandonnées par leurs locuteurs, mais simplement des langues considérées du point de vue de l'influence qu'elles ont exercé sur leurs voisins.

1.3.2. Principales langues d'adstrat du français

Tout au long de son histoire pluriséculaire, le français a côtoyé (et continue de le faire) un très grand nombre de langues, auxquelles il doit l'existence de nombreux mots. Nous allons énumérer et exemplifier ci-dessous les plus importantes d'entre elles, mais cette liste est loin d'être exhaustive.

1.3.2.1. Le latin écrit et le grec ancien ; le concept de « doublets »

Les mots des langues de l'Antiquité ont souvent servi de modèle pour la formation de néologismes savants, en particulier à partir de la Renaissance. On doit donc légitimement considérer le latin et le grec ancien comme des langues d'adstrat pour le français, dans la mesure où de nombreux mots français doivent leur existence à un processus d'imitation/adaptation de mots appartenant à ces langues.

Vous allez peut-être vous dire : « mais comme le français vient du latin, ne devrait-on pas dire qu'il a *hérité* les mots de cette langue, plutôt qu'*emprunté* ou *imité/adapté* ? ». En fait, il faut distinguer ces deux cas de figure :

a) Certains mots se sont transmis de façon ininterrompue, sans rupture, d'une génération à l'autre, de l'Antiquité jusqu'à nos jours. De tels mots ont subi d'innombrables phénomènes d'évolution phonétique, au point de devenir souvent méconnaissables. On peut légitimement considérer que ces mots ont été *hérités* du latin. Ces mots sont dits *héréditaires*.

b) En revanche, d'autres mots ont été réinjectés artificiellement dans le vocabulaire de la langue française, avec un minimum d'adaptations phonétiques et morphologiques, le plus souvent dans le cadre de traductions d'ouvrages scientifiques ou liturgiques du latin au français. De tels mots sont dits *savants*.

Il arrive souvent que de tels mots, jusqu'à un certain point artificiels dans le parcours historique de la langue, remontent au même étymon qu'un mot héréditaire. Lorsqu'un mot savant et un mot héréditaire remontent au même étymon latin, on parle de « doublets ». Voici une série de doublets :

Étymon latin	Mot savant	Mot héréditaire
AUSCULTĀRE	<i>ausculter</i>	<i>écouter</i>
CAPTĪVUS	<i>captif</i>	<i>chétif</i>
CARBŌNE	<i>carbone</i>	<i>charbon</i>
CAUSA	<i>cause</i>	<i>chose</i>
COAGULĀRE	<i>coaguler</i>	<i>cailler</i>
CUBĪTUS	<i>cubitus</i>	<i>coude</i>
FRAGĪLE	<i>fragile</i>	<i>frêle</i>
FRIGĪDU	<i>frigide</i>	<i>froid</i>
GRACĪLE	<i>gracile</i>	<i>grêle</i> (adj.)
LEGĀLIS	<i>légal</i>	<i>loyal</i>

MUTĀRE	<i>muter</i>	<i>muer</i>
NATĪVU	<i>natif</i>	<i>naïf</i>
NAVEGĀRE	<i>naviguer</i>	<i>nager</i>
PALMA	<i>palme</i>	<i>paume</i>
RIGĪDU	<i>rigide</i>	<i>raide / roide</i>
SIMULĀRE	<i>simuler</i>	<i>sembler</i>

Il peut aussi arriver que deux mots remontent à un même étymon, sans nécessairement que l'un soit savant et l'autre héréditaire. Les deux peuvent être héréditaires, mais avoir connu un cheminement différent. Cf. l'exemple suivant :

Étymon latin	Mot français	Emprunt à l'italien
CABALLARIŪS	<i>chevalier</i>	<i>cavalier</i> (de l'it. <i>cavaliere</i>)
CAPPA	<i>chape</i>	<i>cape</i> (de l'ital. <i>cappa</i>)

Le latin et (le plus souvent à travers lui) le grec ancien représentaient un énorme réservoir de mots à adapter/imiter en français, en particulier dans les nomenclatures juridiques, techniques et scientifiques ainsi que dans le champ sémantique du christianisme, mots à partir desquels les langues modernes se sont enrichies en créant d'innombrables néologismes à leur imitation. Quelques exemples :

– du latin :

- *aquatique* adj. « relatif à l'eau » (emprunt savant au latin *aquaticus*).
- *auriculaire* adj., n. m. (du bas latin *auricularius*, dérivé de *auricula* « oreille »).
- *ausculter* v. tr. (emprunt savant au latin *auscultare*, qui a donné *écouter* par voie héréditaire).
- *cause* n. f. « ce qui produit un effet » (emprunt savant au latin *causa* ; par évolution phonétique régulière, ce mot latin a donné *chose*).
- *digital* adj. (du latin *digitalis*, dérivé de *digitus* « doigt »).
- *pellicule* n. f. « petite peau ; fine membrane » (du latin *pellicula* « petite peau », diminutif de *pellis* « peau »).
- *potion* n. f. « médicament liquide » (emprunt savant au latin *potione* ; par évolution phonétique régulière, ce mot latin a donné *poison*).

– du grec, le plus souvent à travers le latin (en particulier le latin ecclésiastique) :

- *antipathie* n. f. (du latin *antipathia*, lui-même du grec *antipathia*, de *anti-* et *pathos* « passion »).
- *aphrodisiaque* adj. (du grec *aphrodisiakos*, de *Aphrodité* « Aphrodite »).
- *apocalypse* n. f. (du latin ecclésiastique *apocalypsis*, lui-même du grec *apokalupsis* « révélation », du verbe *apokaluptein* « découvrir, dévoiler »).
- *borborygme* n. m. « bruit produit par le déplacement des gaz dans l'intestin ou l'estomac » (du grec *borborugmos*).
- *eucharistie* n. f. (du latin ecclésiastique *eucharistia*, lui-même du grec *eukharistia* « action de grâce »).
- *hippopotame* n. m. (du latin *hippopotamus*, lui-même du grec *hippos* « cheval » et *potamos* « de fleuve »).

- *hypocrite* n., adj. (du bas latin *hypocrita*, lui-même du grec *hupocritês* « acteur, mime, imitateur », du verbe *hupokrinesthai* « feindre, jouer un rôle »).

1.3.2.2. Les autres parlers galloromans (les dialectes d'oïl, les dialectes francoprovençaux, les dialectes d'oc)

Une autre source non négligeable d'enrichissement lexical pour la langue française ont été les emprunts à d'autres parlers galloromans, c'est-à-dire à d'autres langues ou dialectes issus de l'évolution du latin vulgaire de Gaule. Quelques exemples :

– de parlers d'oïl :

- *calumet* n. m. « pipe à long tuyau » (forme normanno-picarde de *chalumeau*, avec substitution de suffixe).
- *campagne* n. f. (forme normanno-picarde de *champagne*, lequel ne s'est conservé en français que comme toponyme).
- *gambette* n. f. « (pop.) jambe » (forme picarde de *jambette*).

– de parlers occitans :

- *aigue-marine* n. f. « pierre précieuse » (de *aigue*, mot provençal signifiant « eau »).
- *aïoli* n. m. « mayonnaise à l'ail » (du provençal *aioli*, de *ai* « ail » et *oli* « huile »).
- *bastide* n. f. « ferme ; petite maison de campagne » (de l'ancien provençal *bastida*, participe passé du verbe *bastir*, équivalent du français *bâtir*, et remontant en dernière analyse au francique **bastjan*).
- *brandade* n. f. « morue émietée (etc.) » (du provençal *brandado*, participe passé d'un verbe signifiant « remuer »).

1.3.2.3. Les autres langues romanes

Le français a évidemment « emprunté » tout au long de son histoire de très nombreux mots à d'autres langues romanes ; nous ne mentionnerons rapidement ci-dessous que les deux plus importantes d'entre elles, l'italien et l'espagnol.

1.3.2.3.1. L'italien

À l'époque de la Renaissance, un vent d'italianomanie, si l'on peut dire, déferle sur la France, en particulier à la Cour. Il était à l'époque aussi chic d'utiliser des mots italianisés que cela l'est devenu aujourd'hui, dans certains milieux, d'utiliser des mots anglo-américains. Ce ne sont pas tous les mots italiens attestés à l'époque qui se sont définitivement enracinés dans l'usage, mais plusieurs d'entre eux sont devenus entièrement français et ne sont plus du tout reconnus de nos jours comme des mots d'origine italienne. D'autres sont venus s'ajouter à une époque plus récente. Quelques exemples :

- *ambassade* n. f. « députation auprès d'un souverain ou d'un gouvernement étranger » (de l'italien *ambasciata*).
- *aquarelle* n. f. (de l'italien *acquarella*, dérivé de *acqua* « eau »).
- *ballerine* n. f. (de l'italien *ballerina*, dérivé du verbe *ballare* « danser »).
- *fugue* n. f. « composition musicale » (de l'italien *fuga*, de même sens).

- *piano* n. m. « instrument de musique » (troncation de *pianoforte*, qui signifie littéralement « doucement / fort », cet instrument de musique permettant, à la différence du clavecin, de jouer à volonté « doucement » ou « fort »).
- *pittoresque* adj. (de l'italien *pittresco*, dérivé du substantif *pittore* « peintre »).

1.3.2.3.2. L'espagnol

La langue espagnole a aussi fourni de nombreux modèles lexicaux au français, et a joué un grand rôle d'intermédiaire pour la transmission des mots issus de l'arabe ainsi que des langues autochtones d'Amérique centrale et d'Amérique du sud.

- *ar(r)obase* n. f. « symbole typographique » (de l'espagnol *arroba*, mot très ancien d'origine arabe exprimant une mesure de poids et dont le symbole traditionnel est @).
- *azur* n. m. « (nom de couleur) » (de l'espagnol *azul*, lui-même d'origine arabe, le mot arabe étant à son tour d'origine perse).
- *caïman* n. m. « reptile crocodylien » (de l'espagnol *caimán*, lui-même probablement d'origine caraïbe).
- *condor* n. m. « (esp. d'oiseau) » (de l'espagnol *condor*, lui-même d'origine quichua).
- *moustique* n. m. (d'abord attesté en français sous la forme *mousquite*, adaptation de l'espagnol *mosquito*, diminutif de *mosca* « mouche »).
- *tchatche* n. f. « disposition à s'exprimer facilement, à parler beaucoup » (de l'espagnol *cháchara* n. f. « abondance de paroles inutiles ; conversation frivole »).
- *tomate* n. f. (de l'espagnol *tomate*, lui-même emprunté au nahuatl *tomatl*).

1.3.2.4. Les langues germaniques

Le groupe des langues germaniques est évidemment très important aussi dans le panorama des langues d'adstrat qui ont fourni des types lexicaux au français ; nous n'allons nous pencher rapidement que sur les deux plus importantes d'entre elles, l'allemand et l'anglais.

1.3.2.4.1. L'allemand

L'allemand joue un grand rôle depuis au moins le début du 19^e siècle dans l'établissement de nomenclatures scientifiques qui sont éventuellement reprises et adaptées par les autres langues. Quelques exemples :

- *aspirine* n. f. (emprunt de l'allemand *Aspirin*, nom déposé par Bayer en 1899).
- *allergie* n. f. (emprunt de l'allemand *Allergie*, lui-même formé à partir des éléments d'origine grecque *allos* « autre, différent » et *-ergeia* « action, efficacité »).
- *diesel* n. m. (du nom de l'inventeur du moteur diesel, l'ingénieur allemand Diesel, 1858-1913).
- *ego* n. m. « (terme de psychanalyse et de philosophie) » (il s'agit évidemment d'un mot latin, le pronom personnel sujet de première personne du singulier, mais son utilisation dans les sciences humaines provient des travaux de nombreux intellectuels et savants germanophones, en particulier le philosophe Kant).
- *morphologie* n. f. « étude de la configuration et de la structure externe » (employé pour la première fois en français en 1822 par Blainville, mais repris d'un mot allemand créé par Goethe en 1790 à partir du grec *morpho-* « forme » et de l'élément de formation *-logie* « étude »).

- *surmoi* n. m. « (terme de psychanalyse) » (calque de l'allemand *Über-ich*, employé d'abord par Freud en 1923, de *über* « au-dessus de » et *ich* « moi »).

L'alsacien, qui est un dialecte germanique, a aussi transmis des mots au français :

- *chope* n. f. « récipient cylindrique destiné à boire la bière » (de l'alsacien *Schoppe*).
- *choucroute* n. f. « spécialité culinaire préparée avec du chou » (de l'alsacien *Surkrut*, correspondant à l'allemand standard *Sauerkraut*, et signifiant littéralement « aigre » et « chou » ; le mot a toutefois subi une attraction par étymologie populaire, et **soucroute* est devenu *choucroute* à cause de l'influence des mots *chou* et *croûte*).

1.3.2.4.2. L'anglais

Le rôle de l'anglais, assez secondaire du moyen âge jusqu'à l'époque classique, est devenu aux 18^e et 19^e siècles comparable à celui de l'allemand, en particulier dans les nomenclatures techniques, scientifiques et politiques ; au 20^e siècle, son importance a explosé et il est devenu la principale langue exportatrice de mots et de concepts à l'échelle mondiale.

Certains de ces mots ont une forme qui trahit leur origine anglaise (on pense à tous les noms en *-ing*) ; toutefois, de nombreux anglicismes passent totalement inaperçus, soit parce que dans leur forme ils ont été profondément francisés (*paquebot*, *boulingrin*, *redingote*), soit parce qu'ils ont été forgés à l'imitation de mots anglais eux-mêmes formés de racines grecques et latines dont les équivalents existent aussi en français (*téléphone*, *dinosaure*). Enfin, la catégorie des calques sémantiques passe aussi inaperçue, car il s'agit de mots dont la forme est française mais dont le sens a été emprunté à un mot anglais de forme et d'origine apparentées (un exemple parmi tant d'autres : l'emploi du verbe *réaliser* avec le sens de l'anglais *to realize* « se rendre compte ») ; les calques de traduction passent aussi inaperçus, comme la *souris* de l'ordinateur, de l'anglais *mouse* ; le *gratte-ciel*, de l'anglais *sky-scrapers*). Signalons pour terminer la catégorie des faux-anglicismes ; c'est le cas de *footing*, formé de façon arbitraire par des francophones à partir de *foot* « pied » et de *-ing* pour désigner une certaine forme de course à pied, alors qu'en anglais le mot *footing* signifie tout autre chose (l'*Oxford English Dictionary* propose 17 définitions pour ce mot, mais aucune d'entre elles n'a le moindre rapport sémantique avec le mot français).

1.3.2.5. Les langues slaves

Parmi les grands groupes de langues indo-européennes, outre les langues romanes et germaniques, on compte également la grande famille des langues slaves, au premier rang desquelles on retrouve le russe. Quelques exemples :

- *bélouga* n. m. « mammifère cétacé carnivore » (du russe *bieluga*, de *biely* « blanc »).
- *blinis* n. m. « petite crêpe » (du russe *blini*).
- *cosmonaute* n. (de *cosmo-*, du russe *kosmonavt*, de même sens, lui-même formé à partir d'éléments de formation d'origine grecque).
- *mazout* n. m. « résidu de la distillation du pétrole » (du russe *mazut*, de même sens).
- *taïga* n. f. « forêt de conifères » (du russe *tajga*, de même sens).

1.3.2.6. Les langues du Proche-Orient (arabe, turc)

De nombreuses langues parlées dans l'est du bassin méditerranéen ont fourni au français des modèles lexicaux pour la création de néologismes ; leur histoire est souvent assez complexe, car il arrive fréquemment que l'on assiste à de véritables « chaînes » d'emprunts, les langues se copiant les unes les autres en faisant subir aux mots d'innombrables modifications qui souvent les rendent méconnaissables. Les deux principales langues dans ce contexte sont l'arabe et le turc. Les arabismes ont souvent transité par l'espagnol (v. ci-dessus) ; les types lexicaux d'origine turque peuvent aussi avoir connu de nombreux intermédiaires avant d'aboutir en français. Quelques exemples :

– de mots dont l'étymon est arabe

- *azimut* n. m. « (terme d'astronomie) » (de l'arabe *'āz-sāmt*, proprement « le chemin »).
- *clebs* n. m. « (pop.) chien » (de l'arabe maghrébin *klāb*, arabe classique *kīlāb* « chiens »).
- *goudron* n. m. (de l'arabe d'Égypte *qātrām*).
- *nénufar* n. m. « (nom de plante) » (de l'arabe *nīnūfār*).
- *safran* n. m. (de l'arabe *za'frān*).

– de mots dont l'étymon est turc

- *café* n. m. (du turc *kahve*, peut-être par l'intermédiaire de l'italien de Venise *caffé*).
- *minaret* n. m. (du turc *menaret*, de même sens, lui-même d'un mot arabe, *mānārāh*, signifiant « tour qui éclaire, phare »).
- *tulipe* n. f. (du turc *tülbend* « (plante) turban », avec adaptation formelle).
- *turban* n. m. (du turc *tülbend*, de même sens, avec une adaptation formelle différente).
- *yaourt*, *yogourt* n. m. (emprunté au bulgare *yugürt*, *yaürt*, lui-même emprunté au turc *yogürt*).

1.3.2.7. Les langues d'Extrême-Orient (hindi, chinois, japonais)

Les grandes explorations, à partir de la fin du Moyen Âge et du début des Temps Modernes, ont mis en contact les Européens avec de nouveaux mondes, de nouveaux peuples, et de nouveaux mots et concepts. De très nombreuses langues pourraient être ici mentionnées ; nous allons nous limiter aux plus importantes d'entre elles, le hindi, le chinois et le japonais.

– le hindi

- *basmati* n. m. « type de riz indien » (de l'hindi *bāsmatī*, de *bās* « parfum » et *matī*, suffixe exprimant la possession).
- *gourou* n. m. « maître spirituel » (de l'hindi *gurū* « vénérable »).
- *sari* n. m. « longue étoffe drapée » (de l'hindi *sāri*).

– le chinois

- *dazibao* n. m. « journal mural » (emprunt à un mot chinois, de même sens).
- *ginseng* n. m. « (nom de plante) » (du chinois *jēn shēn* signifiant littéralement « plante homme », la racine de ginseng rappelant une silhouette humaine).

- *wok* n. m. « grande poêle » (mot du chinois cantonais, probablement diffusé en Occident d’abord à travers l’anglais).

– le japonais

- *bonsai* n. m. « arbre nanifié » (d’un mot japonais signifiant « arbre (*sai*) en pot (*bon*) »).
- *futon* n. m. « matelas pliable » (d’un mot japonais de même sens).
- *sumo* n. m. « lutte japonaise » (d’un mot japonais signifiant « lutte »).

Et, pour clore cette catégorie, un mot qui réunit ces trois langues :

- *zen* n. m. « forme du bouddhisme », adj. « calme, impassible » (mot japonais, lui-même adapté du chinois *chan* signifiant « quiétude », à son tour originaire du sanskrit *dhyāna* signifiant « méditation » ; rappelons ici que le sanskrit est l’ancêtre du hindi).

1.3.2.8. Les langues amérindiennes

Si l’on se tourne maintenant vers l’ouest, on retrouve les langues indigènes du continent américain, qui ont très tôt commencé à léguer des types lexicaux aux langues d’Europe (dès le premier voyage de Christophe Colomb en Amérique). Il peut s’agir de langues amérindiennes du grand nord (les langues des Esquimaux), de l’Amérique du Nord, de l’Amérique centrale, des Antilles ou du continent sud-américain. Dans la plupart des cas, une autre langue européenne (l’espagnol, le portugais ou l’anglais) a servi d’intermédiaire.

- *alpaga* n. m. « mammifère ruminant voisin du lama » (de l’espagnol *alpaca*, adapté d’un mot quichua ou aymara).
- *canot* n. m. (de l’espagnol *canoas*, lui-même adapté d’un mot arawak, langue indienne des Caraïbes).
- *caribou* n. m. « renne du Canada » (adapté à partir d’un mot algonquin).
- *guacamole* n. m. « préparation à base de purée d’avocat » (emprunt à l’espagnol, qui le tient lui-même du nahuatl *ahuacamulli* ; le nahuatl est la langue des Aztèques).
- *iglou* n. m. (de l’inuit, par l’intermédiaire de l’anglais).
- *maringouin* n. m. « (mot désignant le moustique au Canada et dans les Antilles) » (du tupi-guarani, langue indigène du continent sud-américain).

2. Les innovations (formelles) internes

Abordons maintenant un aspect totalement différent de l’histoire du vocabulaire, mais qui a également contribué dans une très large mesure à l’enrichissement lexical du français. Il s’agit des innovations internes, dues au dynamisme propre de la langue. Toutes les langues disposent de la possibilité de créer de nouveaux mots à partir des ressources lexicogéniques à leur disposition. Nous allons voir ci-dessous les nombreux processus par lesquels le français s’y prend pour former de nouvelles unités lexicales.

2.1. La dérivation zéro

On peut former un mot nouveau sans changer quoi que ce soit à sa forme, mais juste en le faisant fonctionner comme s’il appartenait à une autre catégorie grammaticale. On parle dans ce cas de « dérivation zéro » ou de « dérivation impropre » ; d’autres auteurs parleront d’hypostase, de translation ou de conversion. Un exemple :

- *blessé* part. passé (« je me suis blessé ») > adj. (« un homme blessé ») > substantif (« les nombreux blessés »)

2.2. La dérivation régressive

La dérivation régressive (aussi appelée « rétroformation » par certains auteurs) consiste à former un mot nouveau en lui retranchant un morphème (ce qui change aussi le sens du mot résultant ; à ne pas confondre, donc, avec les phénomènes de troncation dont on parlera ci-dessous 2.7., et qui n'impliquent aucun changement de sens). Quelques exemples :

- *conte* n. m., déverbal de *conter* (historiquement, ce n'est pas le verbe *conter* qui dérive du substantif *conter*, mais bien l'inverse).
- *marche* n. f. (en parlant d'une marche d'escalier, ou du fait de marcher), déverbal de *marcher* (historiquement, ce n'est pas le verbe *marcher* qui dérive du substantif *marche*, mais bien l'inverse).
- français d'Afrique *alphabète* n. « personne qui sait lire et écrire », formé à partir de *analphabète* par retranchement du préfixe *an-*.

2.3. L'affixation

Jusqu'ici, nous avons vu deux procédés de formation de mots nouveaux qui consistent à créer un néologisme sans changer la forme du mot de base, ou alors en lui retranchant quelque chose. Toutefois, la plupart des néologismes reposent plutôt sur l'ajout de nouveaux morphèmes dérivationnels. On appelle ces morphèmes des « affixes », et le procédé de formation de mots nouveaux grâce à des affixes s'appelle « affixation ». Les affixes peuvent être ajoutés au début du mot ou à la fin ; lorsqu'ils sont ajoutés au début du mot, on parle de « préfixes » ; lorsqu'ils sont ajoutés à la fin du mot, on parle de « suffixes ». Un affixe est dépourvu d'autonomie ; il n'existe qu'en combinaison avec une base.

2.3.1. La préfixation

Le procédé de formation de mots nouveaux qui consiste à ajouter un préfixe devant une base est appelé « préfixation ». Quelques exemples :

- *anti-* + *gel* > *antigel*
- *auto-* + *financement* > *autofinancement*
- *co-* + *participant* > *coparticipant*
- *dés-* + *enchanter* > *désenchanter*
- *in-* + *capable* > *incapable*
- *pré-* + *disposer* > *prédisposer*

2.3.2. La suffixation

Le procédé de formation de mots nouveaux qui consiste à ajouter un suffixe à la suite d'une base est appelé « suffixation ». Quelques exemples :

- *abord(er)* + *-able* > *abordable*
- *désert* + *-ifier* > *désertifier*

- *écriv(ain) + -aillon > écrivain*
- *marginal + -ité > marginalité*
- *Mitterrand + -isme > mitterrandisme*
- *réalis(er) + -ateur > réalisateur*

2.3.3. La dérivation parasynthétique

Le procédé de formation de mots nouveaux qui consiste à ajouter *conjointement* un préfixe et un suffixe à une base est appelé « dérivation parasynthétique ». Exemple :

- *dé- + vital + -iser > dévitaliser* (il ne s'agit pas d'un dérivé préfixal formé sur un hypothétique verbe **vitaliser*, qui n'existe pas ; il ne s'agit pas non plus d'un dérivé suffixal formé sur un hypothétique adjectif **dévital*, qui n'existe pas davantage ; la préfixation et la suffixation ont eu lieu en même temps, et c'est bien cela qui est définitoire de la dérivation parasynthétique telle qu'on l'entend de nos jours)

2.4. La composition

L'affixation a lieu, comme on vient de le voir, entre une base (un mot doté d'autonomie) et un affixe (un morphème dépourvu de vie autonome). Il est toutefois possible de créer des mots nouveaux en réunissant deux mots qui existent déjà de façon autonome. On parle dans ce cas de « composition », et les unités lexicales nouvelles ainsi formées sont appelées « composés » ou « mots composés ». Quelques exemples :

- *homme + grenouille > homme-grenouille* (subst. + subst.)
- *petit + déjeuner > petit-déjeuner* (adj. + subst.)
- *tire + bouchon > tire-bouchon* (verbe + subst.)
- *arc + en + ciel > arc-en-ciel* (subst. + complément prépositionnel)
- *pomme + de + terre > pomme de terre* (subst. + complément prépositionnel)

2.5. Les composés savants (ou « confixés »)

La langue scientifique fait un très grand usage d'un certain procédé qui consiste à créer des mots nouveaux à partir d'éléments lexicaux d'origine gréco-latine qui n'existent pas en tant que tels comme mots individuels mais auxquels on reconnaît un sens lexical « plein », ce qui n'est pas le cas des affixes, dont le sens est plus abstrait. On parle habituellement de « composition savante » pour se référer à ce processus lexical, mais le linguiste André Martinet a proposé le terme de « confixés » pour désigner les mots ainsi formés. Quelques exemples :

- *biblio- + -phile > bibliophile*
- *herbi- + -vore > herbivore*
- *phono- + -logie > phonologie*
- *radio- + -graphie > radiographie*
- *xylo- + -phone > xylophone*

2.6. La délocutivité

Le processus appelé « délocutivité » consiste à extraire une locution du discours et à en faire une unité lexicale figée. Attention de ne pas confondre ce phénomène avec la dérivation zéro

(v. ci-dessus 2.1.). Le phénomène de la délocutivité ne se définit pas comme un changement de catégorie grammaticale ; il est caractérisé par le fait qu'une locution de discours, qui appartient à l'axe syntagmatique et qui se concrétise dans des actes énonciatifs, devient un élément cristallisé du répertoire lexical de la communauté linguistique, et s'insère dans des ensembles paradigmatiques. Le phénomène peut toucher un mot isolé (« pour un *oui* ou pour un *non* »), ou une phrase entière :

- *m'as-tu-vu* n. inv. « personne vaniteuse » (« allusion à la question que se posent entre eux les acteurs évoquant leurs succès » Nouveau Petit Robert 2007).
- *je-ne-sais-quoi* n. m. inv. « chose qu'on ne peut définir ou exprimer, bien qu'on en sente nettement l'existence ou les effets » Nouveau Petit Robert 2007).
- *je-m'en-fichisme* n. m. « attitude d'indifférence envers ce qui devrait intéresser ou préoccuper » (« de *je m'en fiche* » Nouveau Petit Robert 2007). Ici, la base dérivationnelle s'est en outre vue adjoindre le suffixe *-isme*.

2.7. La troncation

La troncation est un phénomène purement formel, qui n'affecte pas le sens du mot. Il consiste à retrancher une partie du mot pour l'abrèger. Cette troncation peut avoir lieu au début du mot (aphérèse) ou à la fin du mot (apocope).

2.7.1. L'aphérèse

- *(omni)bus* > *bus*

2.7.2. L'apocope

- *métro(politain)* > *métro*

Attention de ne pas confondre la troncation avec l'ellipse ; dans le cas de l'ellipse, ce ne sont pas des parties de mots qui tombent, mais bien des mots entiers :

- *(chemin de fer) métropolitain* > *métropolitain*

2.8. Les mots-valises

Des formes aphérésées et apocopées peuvent se combiner pour former ce qu'on appelle des mots-valises, ou des formes télescopées ; en anglais, on dit *portmanteau word* depuis Lewis Carroll. La définition du Nouveau Petit Robert 2007 de *mot-valise* spécifie qu'on a affaire à un « mot composé de morceaux non signifiants de deux ou de plusieurs mots », ce qui est en effet une précision nécessaire pour les différencier d'un mot comme *télévore*, dans lequel *télé-* est un véritable élément de formation. Quelques exemples :

- *velcro* n. m., de *vel(ours)* et *cro(chet)* (*vel-* et *-cro* ne sont pas des éléments de formation, comme ce serait le cas de *télé-* « relatif à la télévision » dans *télévore*).
- *progiciel* n. m., de *pro(duit)* et de *(lo)giciel* ; *pro-* et *-giciel* ne sont pas des éléments de formation, ils résultent de la troncation arbitraire des mots correspondants, et non d'une analyse en formants).

2.9. Sigles, acronymes et abréviations

Ils abondent, comme vous le savez, dans la société moderne. La plupart d'entre eux restent cantonnés au statut de noms propres.

- Sigle : se prononce en épelant, en prononçant le nom de chaque lettre (*B.C.B.G.* [besebeze]).
- Acronyme : se prononce comme s'il s'agissait d'un mot et non d'une suite de lettres (CNED « Centre National d'Éducation à Distance » prononcé [knɛd] plutôt que [seɛnɔde]).
- Abréviation : n'est pas destiné à se prononcer, ou alors se prononce comme le mot entier (« n. f. » ne se prononce pas [ɛnɛf], mais bien « nom féminin », si tant est qu'on ait besoin de le prononcer). Évitez d'utiliser *abréviation* au sens de « troncation », cela sème la confusion.

3. Bref aperçu de la lexicographie historique du français

Voici une très brève bibliographie des principaux grands dictionnaires historiques de la langue française :

DEAF : BALDINGER, Kurt (*et al.*), *Dictionnaire étymologique de l'ancien français*, Québec, Les Presses de l'Université Laval / Tübingen, Niemeyer. (En cours de publication depuis 1974).

FEW = WARTBURG, Walther von. *Französisches Etymologisches Wörterbuch. Eine Darstellung des galloromanischen sprachschatzes*, Bonn / Leipzig / Bâle : Teubner / Klopp / Zbinden, 1922-2002, 25 vol.

Gdf = GODEFROY, Frédéric. *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, 1880-1902, Paris : Vieweg, puis Bouillon, 10 vol.

GdfC = Complément du *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, (vol. 8-10 du Gdf).

GLLF = *Grand Larousse de la langue française en sept volumes*, Paris : Libr. Larousse, 1971-1978.

Huguet = HUGUET, Edmond. *Dictionnaire de la langue française du seizième siècle*, t. 1, Paris : Champion, 1925 ; t. 2, Champion, 1932 ; t. 3-7, Didier, 1946-1967.

Rob 1985 = *Le Grand Robert de la langue française. Dictionnaire alphabétique et analogique de la langue française* de Paul Robert, 2^e éd. entièrement revue et enrichie par Alain Rey, Paris : Le Robert, 1985, 9 vol.

TL : TOBLER, Adolf ; LOMMATZSCH, Erhard. *Altfranzösisches Wörterbuch*, Adolf Toblers nachgelassene Materialien bearbeitet und hrsg. von Erhard Lommatzsch. Berlin, Preußische Akademie der Wissenschaften (vol. I et II) ; Mainz, Akademie der Wissenschaften und Literatur (depuis le vol. III).

TLF = *Trésor de la langue française. Dictionnaire de la langue du XIX^e et du XX^e siècle*, sous la dir. de P. IMBS (vol. 1-7) puis de B. QUEMADA (vol. 8-16), Paris : Gallimard, 1971-1994.