



HAL
open science

Les mots de bouche à oreilles : le cas de Polymots

Véronique Rey, Núria Gala

► **To cite this version:**

Véronique Rey, Núria Gala. Les mots de bouche à oreilles : le cas de Polymots. Approches sociolinguistiques et didactiques : études offerte à Alain Giacomi., L'Harmattan, 2011, 2-296-13991-4. hal-03198400

HAL Id: hal-03198400

<https://hal.science/hal-03198400>

Submitted on 14 Apr 2021

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.

Les mots de bouche à oreilles : le cas de Polymots¹

V. Rey & N. Gala

Résumé

Dans cette étude, la comparaison phonologique est employée dans le but de rendre compte de la construction des mots en synchronie sur un corpus de 20 000 mots. Partant des correspondances entre les formes et les sens des mots, on cherche à dégager l'ensemble des familles de mots et à dresser un inventaire fermé des radicaux. Ces radicaux donnent à voir des unités de désignation qui permettent de retrouver des primitives du vocabulaire contemporain.

1. Introduction

Il est toujours étonnant de lire dans un dictionnaire que tel mot est un mot familier, tel autre, un mot populaire et tel autre encore un mot soutenu. Ces informations anonymes (hormis le nom du dictionnaire) peuvent parfois laisser le lecteur surpris ou agacé d'une telle classification présentée comme une information non discutable. L'histoire de la langue française écrite peut en parti rendre compte de ces normes sociales. Mais le droit d'inventaire se poursuit sur l'histoire des mots. Une distinction apparaît entre étymologie populaire (comme entendre « choux » dans le mot « choucroute ») et étymologie savante (comme « sel » dans le mot « salaire »). Enfin, la définition des mots est donnée suivant une longue tradition de lexicographe mais la définition est toujours objet à controverse : les fameuses définitions circulaires (blanc : couleur du lait; lait, liquide de couleur blanche) sont toujours bien représentées et certaines définitions peuvent surprendre comme anarchiste « ». Il n'existe pas à ce jour de corpus de référence permettant de valider ces définitions.

Oui décidément aborder les mots français est un exercice sous haute surveillance et ceci depuis François 1^{er} avec le décret imposant la langue française et ouvrant ainsi la voie aux grammairiens et aux lexicographes. L'ordonnance de Villers-Cotterêts, signée par François Ier en août 1539, a donné une assise juridique à ce processus :

Article 110 : Afin qu'il n'y ait cause de douter sur l'intelligence des arrêts de justice, nous voulons et ordonnons qu'ils soient faits et écrits si clairement, qu'il n'y ait, ni puisse avoir, aucune ambiguïté ou incertitude, ni lieu à demander interprétation.

Article 111 : Et pour ce que telles choses sont souvent advenues sur l'intelligence des mots latins contenus dans lesdits arrêts, nous voulons dorénavant que tous arrêts, ensemble toutes autres procédures, soit de nos cours souveraines et autres subalternes et inférieures, soit de registres, enquêtes, contrats, commissions, sentences, testaments, et autres quelconques actes et exploits de justice, soient prononcés, enregistrés et délivrés aux parties, en langage maternel français et non autrement. (source Académie française, site internet)

Richelieu en créant l'Académie française, donnera les lettres de noblesse à cette langue écrite qui conquiert progressivement les espaces linguistiques dévolus au latin et au grec.

Mais tout ceci concerne la langue écrite. Imaginons un instant le français, une langue à tradition orale (sans écriture) dont il faudrait décrire le lexique. Dans cette tradition, nous nous éloignons définitivement de la définition graphique du mot. Les mots relèvent à la fois de pratiques sociales et d'initiatives singulières. Chaque mot est donc unique mais traversé par deux types de continuité :

¹ Ce travail a été commencé en 2000poursuivi malgré les grands bouleversements universitaires. Juste pour te rassurer, Alain, le travail artisanal aura toujours quelques défenseurs !

- 1) Une continuité phonologique. Les travaux de Aronoff (1976) et de Kiparsky (1982), à la suite des travaux précurseurs de Troubetzkoy (1939) mettent en relief la dimension morphophonologique des constructions des mots. Le principe est que le radical d'un mot connaît des alternances consonantiques et vocaliques et présente donc plusieurs formes phonologiques comme dans les séquences suivantes : *naïf*, *naïveté*, *pain*, *panification*, *panier*, *roc*, *rocaillieux*, *rocher*.
- 2) Une continuité sémantique. La reconnaissance d'un lien entre deux objets peut conduire au maintien d'un mot avec une construction affixale, évitant ainsi la création d'un mot nouveau. Par exemple, le mouvement de l'eau sur la mer peut être comparé aux mouvements des tissus qui encadrent une fenêtre : les rides de la mer deviennent alors des rideaux. Le mouvement des manifestants dans la rue est un mouvement continu comparable au mouvement de la bobine de fil entraîné de se dévider : défiler évoque ces deux mouvements. Enfin, une personne défendant un lieu est un « lieu-tenant ».

Ces deux continuités conduisent à générer des familles de mots. Cependant, le concept de famille de mots est peu discuté. Il est même présenté comme une évidence. Lehmann & Martin-Berthet (2008) décrivent la famille de mots comme « une expression courante » rendant compte des mots qui partagent un élément radical commun. Les auteurs précisent qu'il s'agit de familles morpho-sémantiques synchroniques. En morphologie lexicale, Apothéloz (2002) et Huot (2005) font un état de l'art de la construction des mots du français, sans aborder le concept de famille de mots. Dans la même veine, Rey-Debove (2004) dans le Robert Brio présentant les mots et leurs définitions par famille. L'objet final est un dictionnaire, c'est-à-dire un métalangage sur les mots. Enfin, Fradin & al. (2009) rendent compte des processus de constructions des mots mais n'appréhendent pas non plus ce concept. D'autres auteurs comme Gruaz (2004) présentent des regroupements sur un principe phonologique et sémantique. La famille de mots constitue un groupe de mots associés entre eux en raison d'un élément commun qui est le radical, la racine ou encore l'étymon. Cependant « des ruptures sémantiques » interdisent des rapprochements entre des mots comme « aventure » et « convention ». De même Apothéloz (2002) ne reconnaît pas comme légitime la possibilité de rapprocher X et Y car les locuteurs n'ont plus conscience du lien sémantique entre ces mots.

Néanmoins, en dehors de la finalité lexicographique, le concept de « familles de mots » pourrait être envisagé comme une pratique sociale et temporelle qui renvoie à une capacité des locuteurs à réaliser des analogies phonologiques et sémantiques. Ces derniers construisent des mots nouveaux à partir de mots connus. Cette pratique obéit au principe de réduction de l'arbitraire du signe. En effet, construire un mot nouveau comme *avalanche*, à partir du mot *val*, permet de construire un lien phonologique et sémantique, économisant la construction d'un nouvel arbitraire à apprendre.

Nous voudrions pour notre part approfondir le concept de famille phonologique. L'approche phonologique ne rend pas compte de la prononciation des mots qui, elle, relèverait de l'étude phonétique des mots. La représentation phonologique ne renvoie pas à des variations phonétiques ni sociolinguistique. Dans une zone géographique donnée, quelque soit les âges des locuteurs, leur sexe, leur appartenance sociale, ces locuteurs accèdent aux mêmes mots oraux par l'intermédiaire des représentations phonologiques partagées.

Il s'agit pour un ensemble de mots de dégager la structure phonologique sous-jacente dérivée à partir de la structure de surface. Pour ce faire, il faut faire un état des lieux des différentes attestations d'un mot en construction et en proposer la représentation phonologique complète (abstraite par essence car elle ne prend pas en compte la réalisation phonétique individuelle). Par exemple, les mots souris, souricière, souriceau ont comme radical phonologique commun /suris-/ . Le phonème /s/ final est une consonne latente qui apparaît lors de la construction d'autres mots. Autre exemple, les mots faveur, favorable, défavorable. La voyelle « eu » alterne avec la voyelle « o » (comme dans la série fleur,

floraison) : on a donc deux radicaux phonologiques possibles faveur/favor. Dans la suite de cet exposé, pour des raisons pratiques, la présentation graphique reprend l'orthographe des mots.

Cette approche phonologique est essentielle dans notre propos car dans la perspective d'une langue à tradition orale, le maintien phonologique d'un mot à un autre est une aide mnésique. Selon Gathercole et Baddeley (1990), la boucle phonologique assure non seulement la coordination entre la perception (et la compréhension) et la production (et l'énonciation) mais permet également le maintien en mémoire. C'est ce principe qui régit aussi les règles de poésie : les allitérations, les rimes, ..., assurent la transmission de textes oraux de génération en génération. L'Illiade et l'Odyssée en sont une des illustrations. Le fait que le phénomène des règles poétiques existe dans toutes les langues atteste également de l'importance du phonologique.

Cependant, dans les exemples que nous venons de voir, la question sémantique est triviale : les mots relèvent bien de la même famille sémantique (fleur, floraison). Nous dirons alors qu'il y a isomorphisme entre la famille sémantique des mots et la famille phonologique. Mais ceci n'est pas toujours le cas. Il n'y a plus alors isomorphisme entre ces deux types d'organisation. Par exemple, si *table* et *tablette* appartiennent à la même famille de mots, à l'inverse *fin*, *finesse* et *finance* peuvent relever d'une même famille phonologique (radical phonologique fin/fine) et ne pas relever immédiatement de la même famille sémantique. Or si l'on se tourne vers l'approche étymologique, la *finance* est bien un élément relevant de quelque chose de terminer. Donc la famille phonologique contemporaine peut être validée par la présence d'un étymon commun tandis que la famille sémantique présente une rupture. Nous obtenons des familles phonologiques, validées par l'étymologie mais traitées par les locuteurs en sous partie sur un principe sémantique. L'approche étymologique valide donc les rapprochements phonologiques.

On pourrait reprocher à cette démarche d'avoir recours à une dimension graphique (la recherche d'étymon) dans une approche phonologique contemporaine. Il s'agit en fait d'une méthodologie mise en place en linguistique comparée pour reconstruire les langues et proposer des Proto-langues (Meussen, Greenberg). Sans document écrit, la construction d'un proto-bantu a nécessité l'élaboration d'étymons à partir de formes phonologiques contemporaines. Pour les langues indo-européennes, les linguistes ont eu la chance de posséder des documents écrits en latins et en grecs, permettant de valider la démarche. La validation par la présence d'un étymon s'inscrit donc dans une démarche comparative et non dans un rapport à une langue écrite.

Les familles de mots sont donc élaborées par deux approches complémentaires : une approche sémantique et une approche phonologique sous contrôle d'un étymon commun.

Actuellement l'approche sémantique est dominante lors des rapprochements lexicaux. On peut même dire que l'accès au sens bloque la reconnaissance phonologique. Entendre « bis-cuit » dans un *biscuit* ou « re-cul-er » dans *reculer* relève d'un exercice peu commun. Et pourtant! Le geste phonologique et la langue orale demeurent premiers dans le maniement des mots. Nous nous sommes proposés d'appréhender ces familles phonologiques afin de rendre compte de la pertinence ou non de l'organisation lexicale qui en découlerait.

Le principe est de regarder tous les mots présentant un radical phonologique commun, puis de vérifier si ces mots partagent un étymon commun (étymologiquement motivé).

Cette recherche s'intéresse exclusivement aux comparaisons phonologiques des familles de mots. Nous avons élaboré un corpus de 20 000 mots (20 007 exactement) POLYMOTS, organisés en famille phonologique de mots. Le propos de cet article est donc :

- 1) de rendre compte les types de familles de mot obtenues et d'illustrer le propos par quelques exemples
- 2) de dégager des types de radicaux sur la base de leur capacité observée à construire des mots nouveaux. Aucune information catégorielle de type « substantif, adjectif, verbe » n'est employée pour rendre compte de l'organisation de cette base. Les structures syllabiques des radicaux de POLYMOTS sont été dégagées ainsi que l'ensemble des

alternances vocaliques et consonantiques et leur fréquence dans ce corpus. Les alternances consonantiques et vocaliques apparaissant lors de la construction des mots sont-elles en nombre limité ? Par exemple, l'alternance f/v apparaît dans la construction des séries comme actif, activité, activement; passif, passivité, naïf, naïveté. Nous proposons de dresser l'inventaire des alternances observées.

3) De rendre compte des familles de mots les plus productives.

Il s'agit donc d'étudier l'organisation des mots communs dans un corpus extrait d'un dictionnaire monolingue français contemporain en familles phonologiques de mots. Nous tentons, par la comparaison, de faire apparaître la structure phonologique des familles de mots et de tester la pertinence de ce concept.

2. Les familles phonologiques des mots français

2.1. Type de famille

A partir d'un dictionnaire Larousse (2000), 20 000 mots communs ont été traités manuellement. Le principe est d'appliquer la comparaison des unités et de dégager un radical commun entre plusieurs unités.

La construction de la base POLYMOTS a obéi à plusieurs principes. Tout d'abord, les mots communs extraits du dictionnaire Larousse ont été saisis dans un tableur. Nous avons choisi d'écrire les mots selon l'orthographe en vigueur et non selon une transcription phonologique. En effet, le propos n'est pas de s'interroger sur les différentes réalisations phoniques d'un même mot, mais des mots conçus à travers une norme, nous permettant de parler de « la » langue française partagée. Ceci est une des limites à ce travail mais travaillant sur un corpus de 20 000 mots, ce choix permet de donner un accès habituel aux mots.

Nous avons donc élaboré des familles phonologiques. Pour chaque famille, nous avons cherché à connaître les mots partageant le même radical. Pour les constructions immédiates du type « cigare, cigarette », cela ne soulevait pas de difficultés particulières : la commutation sur la base de la comparaison (une cigarette est un petit cigare) permet de dégager le radical commun « cigare ». Avait-on le droit de réaliser la même opération pour les mots « lune et lunette ». Deux raisons autorisent le rapprochement de ces deux mots : tout d'abord, il y a bien une continuité phonologique entre les deux mots (« lun »); ensuite, la présence d'un étymon commun attesté valide le rapprochement. En dégageant le radical commun, nous obtenons aussi une analyse en affixe.

Pour chaque mot, la continuité sémantique, phonologique et la présence d'un étymon ont constitué les éléments autorisant les rapprochements. Les étymons ont été recherchés dans un des deux dictionnaires historiques de la langue française (Bloch & VonWarburg (2004) et Rey (1998)). Ainsi, le mot « confiture » peut se rapprocher de « fit » du verbe « faire » et donne donc la segmentation suivante « con-fit-ure ». Si la segmentation se trouvait contredite par un des deux dictionnaires historiques, alors le rapprochement était rejeté. Ceci nous a permis de séparer sans trop de difficultés les homonymes du type « mine, mineur » et « mine, minois ».

De plus nous avons rapproché des mots dans lesquels le lien sémantique n'était pas transparent. Par exemple, « val, avalanche, avaler » ou « établi, table ». La démarche se distingue de la démarche étymologique stricte par le fait de ne pas proposer le radical avant la segmentation : c'est la comparaison phonologique des mots contemporains qui a conduit à questionner l'étymon commun et non l'attestation de la forme ancienne qui génère la recherche. C'est ainsi que la forme ancienne « bure » donnée en explication historique du mot « bureau » n'a pas été retenue car la comparaison des mots (*bureau, burautique, buraliste, bureaucratie*) permet de dégager les radicaux bureau/bural et non de faire émerger la forme « bure ». A l'inverse, la comparaison des formes « bouleverser, divers,

devers, pervers, envers, renverser,... » a dégagé la forme « vers », que l'on retrouve dans le dictionnaire historique.

Ces comparaisons ont fait émerger au moins deux types de problème. Tout d'abord, les alternances phoniques ne devaient pas être un frein aux rapprochements des mots en famille. Par exemple, il n'y avait pas de raisons de ne pas ranger ensemble, les mots « naïf, naïveté ». De même, pour la séquence « croc, crocher, croquer ». Là aussi, l'alternance c~ch est bien connue des historiens. Elle demeure active aujourd'hui et nous autorise à ranger les mots ensemble car l'alternance est attestée dans d'autres familles phonologiques et que ces mots partagent un étymon commun. Cette démarche est également validée pour les alternances vocaliques. Ainsi, les mots « manoeuvre, oeuvre, ouvrier, ouvrage » relèvent de la même famille avec une alternance vocalique (eu ~ ou).

Le deuxième type de problème relève de la présence de radicaux qui apparaissent par la comparaison mais qui s'avère vide de sens. C'est le cas par exemple de « pâat » dans « pâatir, pâaturation, pâature », de « duct » dans « aduction, production, conduction ». Le sens contemporain du radical n'est plus accessible. Cependant, là aussi, le dictionnaire historique nous autorise la validation du radical phonologique. Ces radicaux peuvent eux aussi subir des alternances phoniques comme le radical « dou/dol » dans la série « douleur, douloureux, indolore... ». Ces radicaux peuvent apparaître avec des mots nouveaux comme le mot « bikini » qui donne en construction « monokini » faisant émerger un radical « kini » sans signification. Plus récemment, le nom d'un bateau, le « catamaran » a donné un autre nom de bateau, le « trimaran » donnant « maran » comme radical. Ces constructions s'avèrent très intéressantes car elles montrent, comme l'a déjà mentionné Fradin (2003), que si la segmentation d'un mot en phonèmes ou d'une phrase en syntagmes sont totales, la segmentation d'un mot en morphèmes comme le mot « monokini » pose la question du statut de la forme « kini ».

La construction avec des radicaux sans signification est donc un usage possible dans le lexique français.

Ce procédé a fait apparaître deux grandes classes de radicaux dans notre base POLYMOTS. Tout d'abord des radicaux étant eux-mêmes des mots, c'est-à-dire des unités de désignation. Nous les avons appelé « mots transparents » (TSP). Puis des radicaux n'ayant aucun sens en isolation. Il s'agit soit de radicaux ayant eu un sens (étymologie latine ou grecque) soit de radicaux contemporains comme « kini » dans les mots bikini et monokini, ou « panta » dans « pantalon » et « pantacourt ». Ces mots sont des « mots opaques » (OPAK).

Une des difficultés rencontrées est le critère de divergence sémantique pour distinguer homonyme et polysème. Dans bien des cas, il est impossible d'établir des critères rigoureux car la substance sémantique forme un continuum. Picoche (1995) propose que le polysème ait un signifié sous-jacent unique, réalisé dans des effets de sens en contexte. Nous avons donc privilégié la polysémie aux homophones.

Le tableau suivant donne la répartition des mots selon cette grille.

Type de radicaux	Nombre de radicaux	Nombre de mots construits	Pourcentage
TSP	1541	15 241	76,32
OPAK	478	4 766	23,68
TOTAL	2019	20 007	100

Tableau 1 : Répartition des mots de POLYMOTS selon la valeur lexicale des radicaux

Sur un corpus de 20 007 mots, nous obtenons donc 2019 radicaux. Ceci atteste de la puissance constructive du lexique français contemporain. Cependant, tous les radicaux ne construisent pas autant de mot. Le nombre de construction est très hétérogène d'un mot à un autre. Nous en reparlerons plus loin.

Par ailleurs, la répartition entre les types de mot montre une nette dominance pour les mots TSP. Ceci est en faveur de l'hypothèse que la construction des mots réduit l'arbitraire du signe. Dans la majorité des cas, le radical est un mot.

Néanmoins, les affixes construits avec les radicaux OPAK montrent aussi que le sens d'un mot ne dépend pas exclusivement du radical : les affixes peuvent suppléer comme dans les cas de « accident, acide, décider » où le radical -cid- a perdu sa valeur sémantique. Ce radical n'est plus un morphème mais peut connaître une lexicalisation. Par exemple, le cas de « burger » en anglais. « hamburger » est dans un premier temps « les sandwiches de Hambourg ». Comme « am » signifie « jambon » en anglais, les locuteurs ont créé des « sheesburger ». « burger » est devenu le terme générique pour les deux types de sandwich.

Dans la suite de ce travail, nous ne rendons compte que de la structure des radicaux et non de l'ensemble des mots qui feront l'objet d'autres travaux.

2.2 Exemple de familles de mots

Ces familles phonologiques donnent à entendre des mots dont nous avons perdu ou oublié le rapprochement sémantique. Voici quelques exemples afin de faire entendre de nouveau cette langue orale.

ail ail, ailler, aillade, ailloli

On entend bien le mot ail dans ailloli, et il faut tout de même un peu d' « oli », d'huile, pour que la sauce prenne.

Ban ban, banal, banalement, banalisation, banaliser, banalité, banneret, bannière, bannir, bannissement, forban, abandonner, abandon, abandonnateur, abandonnataire, banlieue, banlieusard

Bien sûr, on publie encore les bans du mariage. Un ban pour l'orateur, qu'on le félicite. Loin de la proclamation du maire, là où on ne se s'entend plus, se situe la banlieue. On peut faire don du ban, on l'abandonne. Mais si cela n'a plus de sens, cela devient banal.

Batte, abat, abattage, abattant, abattement, abatteur, abattoir, abattre, bataille, batailler, batailleur, bataillon, batée, battage, battant, battante, batte, battement, batterie, batteur, batteuse, battoir, battre, battue, combat, combatif, combative, combativité, combattante, combattre, débattement, ébats, ébattre, imbattable, rabat, rabattre

Une porte comme une personne peuvent être battantes. Abattre de nouveau conduit à se rabattre, se replier. Un pli sur lui-même et voici le rabat. Enfin, un débat a parfois toutes les allures d'un combat.

Cadence cadence, cadencer, décadence, décadent

Ne pas marquer la cadence peut générer la décadence!

Cale caler décalage, décaler, décalotter, recalage, recalier

Enlever une cale peut produire un décalage. Caler de nouveau, recalier peut être bien pour un meuble, mais non pour une voiture ou pour un étudiant à un examen.

Cane/chen- : canal, canalisable, canalisateur, canalisation, canaliser, cane, canebière, canette, canneler, cannelle, canon, canonnade, canonnier, canule, canulard, canuler, chenal, caniveau

La forme du roseau, la canne, génère le mot canelle et aussi bien sûr, les mots canule et canal. Nous ne pouvions pas ne pas citer *La Canebière!*

Cochon, cochonaille, cochonner, cochonnerie, cochonnet

Le cochon fait des cochonneries et il n'est pas le seul!

Coq : acoquiner, cocotte, cocotter, coquard, coquarde, coquelet, coquet, coqueter, coquette, coquetterie, coquin, coquinerie

Un coquin se prend pour un coq. Une coquette aussi. Tout cela cocotte allègrement et n'empêche pas de s'acoquiner au premier venu.

Droit : adroit, adroitement, droite, droitière, droiture, endroit

Ce qui est droit est sans détour. On peut aussi être adroit. Aller de travers, on deviendra maladroit ou gaucher. Un lieu bien choisi est un endroit.

Masse : amas, amasser, masse, masselotte, massette, massière, massif, massification, massifier, massique, massive, massivement, massue, mastoc, mastodonte, ramassage, ramasser, ramassette, ramassis.

Quand un volume constitue une masse, cela permet d'identifier un massif montagneux ou un massif de plantes. L'objet est une massue ou une masselotte. Conserver beaucoup d'argent revient à l'amasser. Ramasser indique un mouvement et conduit au ramassage scolaire. Mais il n'y a aucun mouvement dans un ramassis d'objet.

Note : annotation, annoter, connotation, connoter, dénotation, dénoter, notabilité, notable, notablement, notaire, notamment, notarial, notariat, notarié, notation, note, noter, notice, notification, notifier, notion, notionnel, notoire, notoirement, notoriété, notule

Une note, pour faire des marques afin de se souvenir, il fallait bien prendre des notes. D'où les notes de musique. Mais un bloc-notes permet aussi de faire des annotations. Cela ne dénote pas d'une mauvaise mémoire. Il est notoirement connu, d'une façon bien notée, que les notaires étaient d'abord des scribes des chancelleries. Une notion dénote de la signification d'un mot.

Ras : arasement, araser, déraser, rasade, rasage, raser, raseuse, rasibus, rasoir

Verser de l'eau à ras bord évoque le sens d'une mesure exacte et il vaut mieux être précis lorsqu'on manie le rasoir ou lorsqu'on arase les saillies d'un toit.

Solide : désolidariser, solidaire, solidariser, solidarité, solide, solidement, solidification, solidifier, solidité

Une relation ferme, solide crée des solidarités qu'il sera difficile de désolidariser

Verbe : adverbe, adverbial, adverbialement, déverbal, préverbe, proverbiale, proverbialement, verbal, verbalisation, verbaliser, verbalisme, verbe, verbeux, verbeusement, verbiage

Le verbe dans les textes religieux signifie la parole...et cela peut donner des proverbes qui peuvent amuser davantage les enfants que les adverbes.

Vol : antivol, convoler, envoler, revoler, survol, survoler, volage, volaille, volant, volante, volatile, volatilisable, volatilisatation, volatiliser, voler, volerie, volet, voletant, voletante, voleter, voleuse, volièrè

La rapidité et le déplacement sont contenus dans le mot voler : l'antivol permet de lutter contre la rapidité du voleur. Mais pour admirer les oiseaux, nous ne craignons pas de les enfermer dans des volièrès afin qu'ils ne se volatilisent pas!

3. Analyse des radicaux

L'analyse porte successivement sur le nombre de mots construits, les structures syllabiques et les alternances phoniques.

3.1. Les radicaux et le nombre de mots construits

Nous nous sommes intéressés au nombre de mots produits avec un radical. Cela ne concerne absolument pas le concept de productivité proposé par des linguistes comme Dal (2003) qui évaluent l'aptitude d'un procédé à former de nouvelles unités lexicales mais simplement la richesse de construction à partir des radicaux. Le tableau 2 donne la répartition par nombre de radicaux.

Nbre de mots produits	Nbre de radicaux	Pourcentage
entre [20; 117]	220	10,9
entre [10; 19]	440	21,79
entre [5; 8]	696	34,47
entre [2; 7]	572	28,33
1 mot	91	4,51
Total	2019	

Tableau 2: Nombre de mots construits à partir d'un radical. 220 radicaux construisent entre 20 et 117 mots, et cela représente 10,9% de la totalité du corpus.

Les 220 radicaux les plus productifs comprennent 136 radicaux TSP et 44 OPAK. Nous retrouvons la répartition entre TSP/OPAK.

Seuls 91 radicaux ne donnent aucun mot nouveau : le radical est un mot et ne donne aucune construction nouvelle. Il s'agit donc de 91 mots (comme lèvre, adulte, acerbe, agrume, aisselle, bar, bocal, bouée, calendrier, chaise, falaise, hangar, métier, nuque, paupière...), sur un corpus de 20 007 mots.

Ce premier résultat concernant la productivité de construction est un premier argument en faveur du concept de famille phonologique de mots. Si les radicaux ne donnaient que 4 à 5 mots, nous aurions alors à rendre compte de procédures morphologiques mais l'organisation en famille phonologique serait moins nécessaire. A l'inverse, le résultat obtenu atteste que le lexique a une organisation en mots apparentés phonologiquement.

3.2. Les structures syllabiques des radicaux

Les radicaux présentent de nombreuses structures syllabiques. Nous avons relevé 105 structures syllabiques différentes. La structure la plus fréquente est de type CVCV et représente 20,58% des structures syllabiques des radicaux de notre corpus. Nous présentons ci-dessous les 10 structures syllabiques les plus fréquentes de ce corpus. Il est à noter que la structure CV n'apparaît pas dans le tableau : sa fréquence est de 19 et cette structure représente 1,23% du corpus, à la différence de la structure CV(C) qui représente 7,08% des radicaux.

Fréquence	Pourcentage	Structure syllabique du radical
317	20,58	CVCV
143	9,29	CVCCV
122	7,92	CCVCV
109	7,08	CV(C)
95	6,17	CVCV(C)
87	5,65	CVCVCV
56	3,64	CCV(C)
55	3,57	CVC
53	3,44	VCVCV
36	2,34	VCCV

Tableau 3 : *Combinaisons syllabiques les plus fréquentes des radicaux du corpus. La consonne marquée entre parenthèse est une consonne latente qui apparaît au moment des constructions.*

Ce tableau atteste de la richesse de la combinatoire syllabique des radicaux : la structure segmentale avec ses 105 combinaisons est une garantie de source de créativité lexicale.

Par ailleurs, il existe 41 structures syllabiques isolées. Cette structure du radical est présente une seule fois comme radical possible mais génère une famille de mots comprenant de nombreux mots. Ainsi, même si un radical a une structure syllabique unique, cette dernière sera néanmoins employée dans les nombreuses procédures de construction, assurant ainsi la pérennité de cette combinaison. On est donc à la fois devant des structures syllabiques dominantes et des structures syllabiques particulières, qui, par le jeu de la construction des mots, ne sont pas isolées.

3.3. Les alternances consonantiques et vocaliques

Le principe des alternances a été décrit par Troubetzkoy (1933) et a été largement utilisé pour rendre compte de l'organisation lexicale des langues à tradition orale comme les langues Niger-Congo (Creissels, 1989).

Ce principe permet d'établir des familles de mots en considérant que cette variation phonique rend compte des allomorphes.

A notre connaissance, il n'existe pas d'inventaire de ces alternances sur un corpus en langue française. Tout comme pour la structure syllabique, nous avons des alternances isolées, attestées donc une seule fois dans le corpus, mais qui seront en fait réemployées lors du maniement des affixes eux-mêmes.

Le cas d'alternance le plus marqué est le phénomène de la consonne latente. Nous traitons dans la catégorie consonne la semi-consonne comme [j].

Consonne latente du radical	nombre d'attestation sur les radicaux	exemple
Voyelle nasale > voyelle orale suivie de consonne nasale	101	brun, brune pension, pensionnat
t	136	Court, courte, courtisan
s	37	Pays, paysage
g	6	Joug, conjuguer
d	22	Nord, nordique
j	10	Paie, payable; joie, joyeuse; boue, barbouiller

Tableau 4 : *Quelques exemples de consonnes latentes les plus fréquentes*

D'autres consonnes alternent avec une autre consonne. Par exemple, /f/ alterne avec /v/ (comme dans naïf, naïveté) et cela pour 8 occurrences; /k/ alterne avec /ch/ comme dans roc et roche et cela pour 15 occurrences.

Les voyelles ne sont jamais latentes : elles n'alternent jamais, dans notre corpus, avec zéro. Elles connaissent de nombreuses alternances. Nous avons déjà parlé des voyelles nasales. La voyelle /e/ alterne avec /a/ pour 19 radicaux comme dans vaisseau et vasculaire; la voyelle /o/ alterne avec /el/ pour 15 radicaux comme dans veau et vèler; la voyelle /eu/ alterne avec /o/ pour 18 radicaux comme dans faveur et favorable.

Les autres alternances sont en nombre plus réduits et dans de nombreux cas (124), ne sont attestées qu'une seule fois dans notre corpus.

Enfin, les clusters consonantiques ont plusieurs configurations. La plus régulière est le cluster Cl du type {bl, pl, scl, cl, gl}. Lors de la construction, un /y/ formatif apparaît : table, tabulaire; peuple,

populaire; article, articulaire; règle, régulateur; muscle, musculaire. Dans d'autres cas, un formatif en /o/ apparaît: arbre, arborescent (mais aussi arbuste, rien n'est parfait!) ; diable, diabolique. Dans d'autres cas encore, un élément du cluster disparaît : démonstration et démontrable avec un radical monstr-. Seuls les clusters en position finale du radical connaissent ce type de construction. Dans les autres positions du radical, les clusters se maintiennent. Enfin, nous avons un cas qui pourrait être une métathèse des éléments phoniques « bre »/ « ber », dans la construction libre et liberté.

4. Les 100 radicaux phonologiques les plus productifs

Les radicaux phonologiques les plus productifs des mots TSP se répartissent entre les mots désignant une action (29/100), les mots désignant des objets (53/100) et les mots désignant une qualité (18/100). Ces mots produisent chacun plus de 25 mots.

clos clus
fond fonc fus
jet ject
pli plé
signe seign
acte ag
don dote
écrit script scribe écriv
état stat
fait fact fi
forme
lien lie li lig
mue mu
ordre ord
passe pas
peine pen pun
point ponct
pose pô
presse
prise prend
quête qué quis
sens sent
terme
test
tordre tors tort
tour tourn
trait tract
vers vert
vue voi vis

batte
bord
boule
bout
brique bris
fil
fleur flor
fleuve flu
gaz

graphe
lecte lec lig li
leg
lieu loc lou
lime
mode
or
pic
raie rail ray
règle régul
sel sal
son sonne
tribut tribus
cape chape
carne cham chair
centre
chant cant
chrone
corde
coup
course cour cur
croc croche
cul
cycle
fiche fix fict fig
gel
lait lact lay
mer mar
mets mett mess miss
meuble mobil
mont mon
note
orgue
part port
père patr
pied pode ped
port
roue roul
sol sou
temps
terre
us
val
voie voy v

doux douce
baisse bass
dur durc
mou molle
vive
faux fal

libre liber
menu minu
force fort
un
égal équi équat
quatre quadr car
tendre tente tens
fin fign
mal maux

dans la perspective de Benveniste (1966), pourraient donner à voir les éléments culturels fondateurs et les relations sémantiques réalisées soit par association (*pain* dans *copain*) soit par métaphore (*ride* dans *rideau*). Un traitement systématique pourra être envisagé par la suite.

5. Conclusion

Au terme de cette analyse, il s'avère que l'approche phonologique a permis de dégager des familles comprenant un grand nombre de mots, démontrant ainsi une grande productivité de construction, à partir d'un radical phonologique commun. Ce fait est en faveur de l'hypothèse de Gathercol et Badeley sur la boucle phonologique, à savoir l'importance du maintien des patrons phonologiques. De plus cette approche appréhende la synchronie et la diachronie de manière complémentaire : les rapprochements synchroniques sont systématiquement validés par l'attestation d'un étymon commun partageant les mêmes informations phonologiques.

Les structures syllabiques dominantes sont de type CVCV, CVCCV, CCVCV et CV(C). Elles donnent à voir des régularités dans l'organisation phonologique mais n'empêche pas un nombre important de combinaison syllabique singulière.

Enfin, les alternances révèlent deux alternances majoritaires : voyelle nasale / voyelle orale suivie de consonne nasale et consonne « t » latente au point de créer par analogie de nouvelle construction comme « bijou, bijoutier » par analogie phonologique.

Ces différents éléments sont en faveur d'une organisation lexicale en famille phonologiques. En effet, par le jeu phonologique, la variation des constructions syllabiques et le maniement des alternances donnent une souplesse pour traduire des analogies sémantiques.

Au terme de ce travail, une famille phonologique de mots pourrait se définir non seulement par un radical commun à un ensemble de mots, mais aussi par un inventaire des radicaux pour un certain nombre de mots construits. La perspective est comme inversée : il ne s'agit pas de donner les régularités de familles de mots, mais de donner l'ensemble des radicaux phonologiques qui structurent les familles de mots.

En ce qui concerne notre propos, les familles phonologiques dégagent à la fois des éléments dominants et des singularités créant une perméabilité du système et donc une adaptabilité aux besoins des locuteurs.

Les applications tant dans une perspective TAL (création d'une base de donnée en libre accès est en cours de réalisation) que dans une perspective de comparaison des langues d'une même famille (les langues romanes ont-elles les mêmes familles de mots?) sont les travaux à venir à partir de cette étude.

Références bibliographiques

- Aronoff M. (1976). *Word formation in Generative Grammar, Linguistic Inquiry Monograph 1*. MIT Press ed. Cambridge : Massachusetts.
- Benveniste E. (1966). *Problèmes de linguistique générale, T.1 & 2*. Paris: Gallimard.
- Bloch O. & VonWarburg W., eds. (2004[1932]). *Dictionnaire étymologique de la langue française*. Paris: PUF.
- Bloomfield L. (1970 [1933]). *Le langage*. Paris: Payot.
- Boissière, P. (1862). *Dictionnaire analogique de la langue française : répertoire complet des mots par les idées et des idées par les mots*. Paris: Aug. Boyer.
- Brun, C., Jacquemin, B., Segond F. (2001) Exploitation de dictionnaires électroniques pour la désambiguïsation sémantique, *Traitement Automatique des Langues*, 42(3), pages 667-691.
- Corbin D. (1987). *Morphologie dérivationnelle et structuration du lexique*, 2 vol. Tübingen: Verlag.
- Creissels D. (1989). *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: Ellug.
- Dal G. (2003). La productivité morphologique en questions et en expérimentations. *Langue française*, 140.
- Doneux J.L. (2001). *L'écriture du français : Prédicibilité et Aléa*. Aix-en-Provence: PUP.
- Fradin B. (2003). *Nouvelles approches en morphologie*. Paris: P.U.F.
- Fradin, B., Hathout, N., & Meunier, F., 2003, "La suffixation en –ET et la question de la productivité", *Langue Française*, 140, 56-78.
- Fradin B. & Kerleroux F. (2003). Quoi de neuf en morphologie? *Langages*, 152.
- Fradin B., Kerleroux F. & Plénat M., eds. (2009). *Aperçus de morphologie du français*. Paris: Presses universitaires de Vincennes.
- Gruaz Cl. Ed., (2004), *Dictionnaire synchronique des familles dérivationnelles de mots français (DISFA)*. Mise en ligne en novembre 2004.
- Harris Z.S. (1955). From phoneme to morpheme. *Language*, 31(2), 190-222.
- Haspelmath M. (2002). *Understanding Morphology*. London: Oxford University Press.
- Huot H. La morphologie. (2005 (2nd édition)). *Forme et sens des mots du français*. Paris: Armand Colin.
- Kiparsky P. (1982). Phonology to lexical Phonology. In Van Der Hulst H. & Smith N.S., eds. *The structure of Phonological Representations (1)*. New York: Foris Dordrecht, 131-175.
- Langlais P. (2009). Etude quantitative de liens entre l'analogie formelle et la morphologie constructionnelle. *TALN 2009*. Senlis
- Lehmann A, Martin-Berthet F. (2008). *Introduction à la lexicologie*. Paris: Armand Colin.
- L'Homme M. C. (2003) Acquisition de liens conceptuels entre termes à partir de leur définition. *Cahiers de lexicographie* 83 (2), 19-34.
- Matthews PH. (1991 [1974]). *Morphology*. Cambridge: Cambridge University Press.
- Picoche J. (1995[1986]). *Structures sémantiques du lexique français*. Paris: Nathan.
- Rey A, éd. (1998). *Dictionnaire historique de la langue française*. Paris: Le Robert.
- Rey-Debove J. (1998). *La linguistique du signe*. Paris: Armand Colin.
- Rey-Debove J. (2004). La morphologie lexicale dans le dictionnaire. *Euralex Proceedings*, 707-716.
- Saussure Fd. (1985 [1915]). *Cours de linguistique générale*. Paris: Payot.
- Trubetzkoy N. (1964 [1939]). *Principes de Phonologie, traduit de l'allemand par J.Cantineau*. Paris: Klincksieck.