

De l'attestation textuelle à l'attestation lexicographique :

*Evolution épistémologiques et enjeux
méthodologiques*

From textual attestation to lexicographical attestation :

*Epistemological evolutions and
methodological challenges*

Samir HAMAMDIA *,
Université 8 mai 1945 Guelma (Algérie),
samir.hamamdia@gmail.com

Date de soumission : 22.10.2025

Date d'acceptation : 09.04.2026

Date de publication : 20.04.2026

**Ex
PROFESSO**

Volume 11 / Numéro 04 / Année 2026

* - Auteur correspondant.

Résumé

Cet article se propose d'analyser la transition fondamentale entre l'attestation textuelle, preuve brute de l'existence d'un mot dans un corpus, et l'attestation lexicographique, construction savante visant à en documenter l'usage. Nous examinerons cette évolution sous l'angle épistémologique, en montrant comment le statut de l'attestation a changé avec l'émergence de la linguistique moderne et des humanités numériques. D'un simple exemple illustratif, l'attestation est devenue la pierre angulaire de la description objective de la langue. L'article abordera également les défis méthodologiques posés par les grands corpus numériques, où la profusion des attestations nécessite de nouvelles stratégies de sélection, d'analyse et de représentation pour une lexicographie véritablement scientifique.

Mots-clés : Lexicographie, Attestation textuelle, Corpus, Épistémologie, Sémantique.

Abstract

This article aims to analyze the fundamental transition between textual attestation—raw evidence of a word's existence in a corpus—and lexicographical attestation, a scholarly construction aimed at documenting its usage. We will examine this evolution from an epistemological perspective, demonstrating how the status of attestation has changed with the emergence of modern linguistics and digital humanities. From a mere illustrative example, attestation has become the cornerstone of the objective description of language. The article will also address the methodological challenges posed by large digital corpora, where the abundance of attestations requires new strategies for selection, analysis, and representation to achieve a truly scientific lexicography.

Keywords: Lexicography, Textual attestation, Corpus, Epistemology, Semantics.

Url de la revue :

<https://www.asjp.cerist.dz/en/PresentationRevue/484>

INTRODUCTION

La lexicographie, en tant que discipline visant à inventorier et à définir le lexique d'une langue, ne s'est pas constituée ex nihilo ; ses racines plongent dans un long processus d'analyse textuelle, débutant par les modestes gloses médiévales et s'épanouissant avec les grands dictionnaires plurilingues de la Renaissance.

La composition d'un dictionnaire est un acte de sélection et d'analyse. Avant de pouvoir prescrire ou décrire, le lexicographe doit collecter. Les premiers artisans de la langue française ont donc naturellement tourné leur regard vers les textes fondateurs qui incarnaient l'autorité, le savoir et la littérature de leur temps. L'analyse de ces textes, motivée par des impératifs pédagogiques, religieux et culturels, a fourni la matière première des premiers recueils lexicographiques.

L'attestation constitue alors, le fondement même de la lexicographie. Pendant des siècles, elle fut principalement comprise comme une citation textuelle, un exemple prélevé dans un texte jugé « autorisé » (littéraire, juridique, scientifique) pour justifier l'inclusion d'un mot dans un dictionnaire. Cependant, le XXe et le XXIe siècles ont opéré une mutation profonde dans la conception et l'usage de cette attestation. Le passage d'une approche prescriptive et littéraire à une approche descriptive et scientifique de la langue a transformé l'attestation d'une illustration en une donnée probante.

Notre étude vise à cartographier cette transition, en définissant d'abord les concepts clés "attestation textuelle" et "attestation lexicographique". Nous retracerons ensuite le contexte historique et épistémologique de cette évolution, avant d'analyser les implications méthodologiques de l'ère numérique. Enfin, nous discuterons des enjeux contemporains liés à la sélection et à l'interprétation des attestations dans les ressources lexicographiques modernes.

I- L'ATTESTATION TEXTUELLE VERSUS L'ATTESTATION LEXICOGRAPHIQUE

I.1. L'attestation textuelle

C'est la forme la plus ancienne d'attestation. Elle consiste en un extrait textuel, souvent court, dont la fonction principale est de prouver l'existence d'un mot ou d'un sens à une époque donnée. Sa valeur est souvent qualitative et subjective : elle est choisie pour son origine prestigieuse (un « bon auteur ») ou son élégance stylistique. Dans les dictionnaires historiques comme le *Oxford English Dictionary* (OED) dans ses premières éditions, l'attestation servait souvent à légitimer un usage en l'ancrant dans une œuvre canonique. Elle est donc sélective et peut donner une vision partielle de la langue.

Les trois principales strates du corpus d'attestations textuelles

Le corpus d'attestations textuel s'organise autour de trois strates fondamentales : les textes sacrés et classiques hérités du Moyen Âge, les œuvres littéraires de la Pléiade et de ses contemporains, et enfin les "trésors" linguistiques provenant de l'Antiquité et de l'Italie renaissante.

Le Corpus Médiéval : Les Glisses, Acte Lexicographique Fondateur

Avant l'avènement de l'imprimerie, le travail lexicographique se manifestait principalement sous une forme manuscrite et utilitaire, prenant l'apparence de gloses.

Ces annotations manuscrites, insérées en marge ou entre les lignes de textes considérés comme difficiles ou prestigieux, répondaient à un besoin immédiat de compréhension. Les textes sources faisant l'objet de ces gloses étaient principalement de trois ordres. Les manuscrits de la Vulgate, version latine de la Bible, figuraient parmi les plus fréquemment annotés, avec des traductions ou explications en ancien français ajoutées pour aider les clercs et étudiants. Les textes de droit et de liturgie, essentiels au fonctionnement institutionnel, étaient également glossés pour en garantir une compréhension et une application correctes. Enfin, les œuvres des auteurs classiques comme Virgile, Ovide ou Cicéron, copiées dans les scriptoria monastiques, faisaient l'objet d'annotations visant à en éclaircir le sens.

De la Glose au Recueil : Les Glossaires

L'accumulation progressive de ces gloses dispersées a conduit à la création des glossaires, qui se présentaient comme des listes alphabétiques ou thématiques de mots latins accompagnés de leur équivalent en ancien français. Des ouvrages comme le Glossaire de Reichenau, datant des VIII^e-IX^e siècles, témoignent de cet effort systématique de traduction et d'explication. Il est important de souligner que l'analyse de ces textes sacrés et classiques ne poursuivait pas un objectif de description de la langue vernaculaire, mais visait essentiellement à faciliter l'accès au latin, qui demeurait la langue du savoir et du pouvoir. Bien que le corpus fût alors relativement limité, ces premières tentatives contenaient en germe l'acte lexicographique fondamental consistant à associer un signifiant à un signifié.

La Renaissance : L'Élargissement du Corpus

Le XVI^e siècle marque un tournant décisif dans l'histoire de la lexicographie, sous l'influence conjuguée de l'humanisme, de l'imprimerie et de l'émergence des États-nations. Les lexicographes de la Renaissance, à l'image de Robert Estienne, ne se contentent plus des gloses éparses mais entreprennent une analyse systématique et exhaustive des textes antiques pour établir le "bon usage" du latin. Dans son *Thesaurus linguae latinae* publié en 1531, Estienne procède à un dépouillement méthodique de l'ensemble de la littérature latine disponible, de Cicéron à Plaute, afin d'en extraire le vocabulaire, d'en illustrer l'usage par des citations précises et d'en retracer l'étymologie. Ce travail philologique rigoureux établit un nouveau standard pour la lexicographie.

Parallèlement, la défense et l'illustration de la langue française, prônée par Du Bellay, incite les lexicographes à élargir considérablement leur corpus en intégrant les textes en français. Les œuvres littéraires contemporaines, notamment celles des poètes de la Pléiade comme Ronsard et Du Bellay, ainsi que des auteurs comme Rabelais ou Montaigne, deviennent des sources privilégiées pour attester de l'usage de mots nouveaux ou rares. De plus, les textes techniques et artisanaux commencent à retenir l'attention des lexicographes désireux de capturer toute la richesse de la langue. Ils s'intéressent ainsi aux jargons des métiers, qu'il s'agisse de la marine, de la chasse ou de la fauconnerie, consignait méticuleusement des termes jusqu'alors absents de la littérature classique.

La légitimité d'un mot ou d'une expression reposait entièrement sur son attestation dans des documents textuels considérés comme faisant autorité. Cette pratique lexicographique s'articulait autour de trois piliers fondamentaux. Le premier et le plus prestigieux était constitué par les œuvres littéraires et les classiques. Les dictionnaires puisaient leur légitimité dans les citations d'auteurs consacrés du panthéon culturel – Rabelais, Montaigne, Corneille ou Hugo pour le français, Shakespeare et Milton pour l'anglais. Leur langue était érigée en modèle de pureté et d'élégance, si bien qu'un terme ne trouvait pleinement droit de cité que s'il pouvait s'abriter sous l'autorité d'un grand écrivain. Cette approche créait une lexicographie essentiellement prescriptive, qui valorisait la langue écrite et littéraire au détriment des usages courants et oraux.

Le deuxième pilier de l'attestation résidait dans les textes juridiques, administratifs et les traités savants. Ces documents incarnaient la langue du pouvoir, de la loi et du savoir institutionnalisé. Les anciennes chartes, les édits royaux, les codes de lois ou les ouvrages de philosophes et de scientifiques comme Descartes ou Buffon offraient un vocabulaire précis et stable. Leur autorité émanait de leur fonction même : ils devaient être sans ambiguïté et leur langage technique ou spécialisé était perçu comme particulièrement fiable. Ils fournissaient ainsi au lexicographe les attestations nécessaires pour légitimer le vocabulaire abstrait de la vie civique et intellectuelle, depuis les concepts juridiques comme la "souveraineté" jusqu'aux termes techniques des sciences naissantes.

Enfin, dans des sociétés où la religion occupait une place centrale, les textes sacrés et les ouvrages théologiques formaient le troisième fondement de l'autorité linguistique. La Bible dans ses différentes versions, le Coran ou les écrits des Pères de l'Église jouissaient d'un statut quasi transcendant. Leur langue était souvent perçue comme inspirée, et leur influence culturelle profonde – through la lecture quotidienne, la récitation et le commentaire – avait imprégné le lexique de millions de locuteurs. Ces textes fournissaient le vocabulaire fondamental de la métaphysique, de la morale et de la foi, et beaucoup des premiers travaux lexicographiques furent motivés par l'exégèse de ces écrits sacrés.

Ce système d'attestation, bien que rigoureux en apparence, présentait des limites considérables. Il était profondément élitiste, ignorant superbement la langue vivante du peuple, des artisans, des femmes et de tous ceux qui n'appartenaient pas à l'élite lettrée. En se focalisant presque exclusivement sur des modèles prestigieux du passé, il engendrait un décalage croissant avec l'usage contemporain et prescrivait la langue plus qu'il ne la décrivait. Le grand bouleversement apporté par le structuralisme fut précisément de remplacer cette "attestation d'autorité" par une "attestation par le système", considérant que la légitimité d'un élément linguistique réside dans ses relations avec les autres termes du système, indépendamment du prestige de sa source littéraire ou institutionnelle.

I.2 L'attestation lexicographique

C'est le résultat d'un processus méthodique et réflexif. Elle ne se contente pas d'attester l'existence d'une unité lexicale ; elle vise à en documenter l'usage de manière systématique.

Documenter l'usage d'une unité lexicale revient à mener une enquête minutieuse visant à cerner son identité et son comportement dans son environnement naturel, le texte. Cette investigation commence par une observation de sa forme et de sa catégorie grammaticale : on relève scrupuleusement sa graphie exacte, sa catégorie (est-elle un nom, un verbe, un adjectif ?), et sa forme morphologique spécifique (est-elle conjuguée, déclinée ?). Cette première étape permet d'établir la carte d'identité formelle du mot.

L'enquête se poursuit par l'analyse de son contexte linguistique immédiat et de ses habitudes syntaxiques. On examine les mots qui l'entourent, pour identifier ses collocations, ses partenaires fréquents qui révèlent une partie de son sens. On décortique également la construction syntaxique dans laquelle il s'insère : régit-il un complément, requiert-il une préposition particulière ? Ceci permet de comprendre ses modalités d'action dans la phrase.

Le cœur de l'enquête réside dans l'élucidation de sa signification et de sa valeur sémantique. Au-delà d'une simple définition, on cherche à saisir les nuances de son sens en contexte, ses connotations (est-il péjoratif, mélioratif, technique ?) et le champ sémantique auquel il appartient. On explore également ses relations avec d'autres mots, comme ses synonymes ou antonymes dans cette occurrence précise, pour mieux le situer sur la carte lexicale de la langue.

Pour compléter le portrait, il est essentiel de reconstituer le cadre énonciatif et le contexte de communication. On détermine le type de texte où l'attestation figure (un discours politique, un roman, un article scientifique), son registre de langue (soutenu, familier) et, le cas échéant, sa variété géographique. On s'interroge aussi sur l'acte de parole accompli, car le mot peut être au service d'une argumentation, d'une ironie ou d'une menace.

Enfin, une documentation rigoureuse exige de situer le mot dans le temps et d'attester sa provenance. La date de l'attestation est cruciale pour le situer dans l'histoire de la langue et juger de son ancienneté, qu'il s'agisse d'un néologisme, d'un terme courant ou d'un archaïsme. En parallèle, on cite toujours sa source avec une précision bibliographique irréprochable, en évaluant la fiabilité et la nature du corpus, qu'il soit littéraire fermé ou issu du web, ce qui garantit la solidité et la vérifiabilité de l'analyse.

Ainsi, l'attestation lexicographique se définit par trois caractéristiques fondamentales. Elle est d'abord représentative, puisant ses exemples dans des corpus équilibrés qui reflètent la diversité des usages, qu'ils soient écrits ou oraux, formels ou informels, dépassant ainsi le cadre traditionnel des seuls textes littéraires. Elle est ensuite descriptive, ayant pour vocation de rendre compte de manière objective du comportement du mot, en documentant son sens, ses constructions syntaxiques, ses collocations et ses connotations telles qu'elles apparaissent en contexte. Enfin, elle est **contextualisante**, en veillant à ce que l'extrait choisi offre un environnement

linguistique assez éclairant pour que le lecteur puisse en déduire le sens et l'emploi du mot sans aucune ambiguïté.

II. Le tournant épistémologique : du prescriptif au descriptif

Cette transition conceptuelle s'inscrit profondément dans l'évolution de la transition entre les attestations textuelles et les attestations lexicographiques. Trois mouvements majeurs en ont été les catalyseurs :

II.1 L'émergence des approches structurales et descriptives

Avec des linguistes comme Ferdinand de Saussure et Leonard Bloomfield, Les approches structurales ont fondamentalement changé le regard porté sur la langue. Celle-ci est désormais appréhendée comme un système dont l'objectif premier est de décrire le fonctionnement concret, plutôt que d'édicter des règles normatives. Dans cette perspective, l'attestation textuelle acquiert un nouveau statut : elle sert de témoignage concret de ce système linguistique en activité.

L'impact des approches linguistiques structurales sur la lexicographie a été une révolution méthodologique, transformant profondément la conception et la fabrication des dictionnaires. Avant le structuralisme, la lexicographie était principalement philologique et historique, traitant le mot comme une entité isolée dont on retraçait l'histoire et les emplois via des citations d'auteurs. Le structuralisme a imposé l'idée que la langue est un système où chaque élément n'a de valeur que par ses relations avec les autres. Ce changement de paradigme a conduit les lexicographes à ne plus considérer le mot comme une île, mais comme un nœud au sein d'un vaste réseau lexical. Dès lors, la définition d'un terme ne se contente plus de pointer vers un référent dans le monde ; elle doit décrire sa place dans le système en le situant par rapport à ses synonymes, antonymes, hyperonymes et hyponymes. Cette nouvelle logique a favorisé le développement de l'étude des champs lexicaux et sémantiques et a directement inspiré la création d'outils spécialisés comme les dictionnaires analogiques ou les thésaurus, qui organisent le vocabulaire par idées et par relations plutôt que par la simple liste alphabétique.

Cette vision nouvelle a également remodelé la microstructure même de l'article de dictionnaire. Le traitement de la polysémie en a été bouleversé : l'ordre de présentation des sens a évolué d'une logique historique ou intuitive vers une organisation logique et systémique. On cherche à identifier un sens fondamental à partir duquel les autres sens se dérivent par des processus métaphoriques ou métonymiques, créant une arborescence qui reflète la structure sous-jacente du lexique. Parallèlement, l'attention portée aux rapports syntagmatiques a mis en lumière l'importance de la combinatoire des mots. Les dictionnaires ont commencé à systématiser l'indication des collocations, des verbes supports et des restrictions de sélection, enrichissant ainsi la description de l'usage concret des mots au-delà de leur seule définition.

Sur le plan méthodologique, le structuralisme a légitimé l'approche onomasiologique, qui consiste à partir d'un concept pour trouver les mots qui

l'expriment. Cette démarche, inverse de la méthode sémasiologique traditionnelle, est au cœur des dictionnaires de synonymes, des planches thématiques et, de manière ultime, des bases de données lexicographiques électroniques interrogées par concepts. L'incarnation la plus aboutie de ces principes est le dictionnaire de Paul Robert, dont chaque article fonctionne comme une cartographie d'un fragment du système lexical, guidant l'utilisateur des définitions vers un réseau de mots analogiquement liés.

Cependant, l'approche structuraliste classique a aussi révélé ses limites. En privilégiant une vision de la langue comme système synchronique et fermé, elle a eu tendance à sous-estimer la variation sociolinguistique, le changement linguistique en temps réel et la vitalité des usages. La lexicographie moderne a donc dû intégrer d'autres courants, comme l'analyse de corpus et la sémantique du prototype, pour mieux rendre compte de la fréquence d'usage, des registres de langue et de la nature plus floue des catégories lexicales. Malgré ces dépassements nécessaires, l'héritage structuraliste reste constitutif de la discipline. La présence systématique des synonymes, l'organisation raisonnée des sens et la conception du lexique comme un réseau relationnel demeurent les piliers sur lesquels repose la lexicographie contemporaine, qui est passée, grâce à cette révolution, du statut de compilation érudite à celui de cartographie scientifique du système de la langue.

II.2 L'essor de la lexicographie de corpus

Matérialisé par des entreprises telles que le *Trésor de la Langue Française* (TLF) ou le *Collins COBUILD*, la lexicographie de corpus a introduit une rupture méthodologique. Le recours à de vastes corpus informatisés a permis une collecte d'attestations à la fois massive et systématique. Ce changement d'échelle a entraîné une évolution des critères de sélection : la valeur d'une attestation ne se juge plus principalement à son « autorité » (l'auteur ou l'œuvre), mais à sa « fréquence » d'usage et à sa « typicité ».

L'avènement de la lexicographie de corpus a représenté une révolution méthodologique aussi profonde que le structuralisme, mais de nature différente. Alors que les lexicographes dépendaient auparavant de leur intuition, de citations sélectives et des dictionnaires antérieurs, cette nouvelle approche a fondé la pratique sur l'observation massive et systématique de la langue réelle, grâce à l'analyse d'immenses collections de textes électroniques.

Ce passage de l'intuition aux données a eu des conséquences concrètes et majeures. La fréquence des mots et des sens, désormais mesurable objectivement, est devenue un pilier central pour organiser les articles, donnant souvent la priorité aux usages les plus courants plutôt qu'aux sens historiques.

Les exemples, autrefois fabriqués ou soigneusement choisis, sont maintenant extraits directement du corpus, garantissant leur authenticité et offrant une couverture plus complète de tous les registres de la langue. Surtout, l'analyse assistée par ordinateur a permis une cartographie d'une précision inédite des modèles de collocation et de la grammaire des mots, révélant les partenaires privilégiés de chaque terme avec une exactitude scientifique.

Plutôt que de remplacer l'approche structurale, la lexicographie de corpus l'a renforcée et rendue empirique. Le structuralisme postulait que les mots forment un système de relations ; le corpus fournit les preuves tangibles de ces relations. Par exemple, l'étude quantitative des collocations permet de vérifier et de préciser les liens de synonymie en montrant comment des mots comme "acheter" et "acquérir" se distribuent dans des contextes différents, cartographiant ainsi les champs sémantiques avec une rigueur nouvelle. Cette synergie a fait émerger un structuralisme empirique, où la théorie qui pose les bonnes questions sur les relations entre les mots est alimentée par la méthode qui y répond avec des données massives.

Cependant, cette approche n'est pas sans limites. La question de la représentativité du corpus est cruciale, car tout biais dans la sélection des textes se répercute dans le dictionnaire. Le lexicographe doit aussi exercer son jugement pour interpréter les données brutes, distinguer les usages significatifs du bruit, et filtrer les textes de qualité médiocre.

Malgré ces défis, la lexicographie de corpus a transformé le métier de lexicographe, qui est passé d'un artisan de la langue guidé par son érudition à un analyste de données linguistiques. Aujourd'hui, la dialogue entre la vision structurale du langage et la méthode du corpus définit l'état de l'art de la lexicographie moderne.

II.3 L'avènement des humanités numériques

Cet avènement a bouleversé l'ordre de grandeur des données disponibles. L'accès à des corpus contenant des milliards de mots (tels que Google Books, Google Ngram Viewer ou Frantext) a transformé la nature même de la collecte. Découvrir une attestation n'est plus une chasse au trésor visant une « perle rare » ; c'est désormais identifier un point de données parmi une multitude, ce qui requiert le déploiement d'outils statistiques et des méthodes comme la « lecture distante » (*distant reading*) pour en extraire une connaissance significative.

L'avènement des humanités numériques représente une évolution décisive qui dépasse et englobe la révolution méthodologique de la lexicographie de corpus. Si cette dernière avait déjà introduit l'analyse quantitative des données textuelles, les humanités numériques élargissent considérablement le paysage en intégrant pleinement la dimension computationnelle et technologique à la recherche sur la langue. Cette transformation ne se limite plus à l'utilisation d'outils d'analyse, mais refonde profondément l'épistémologie même des études lexicales.

La lexicographie entre ainsi dans une ère nouvelle où l'échelle des données accessibles devient vertigineuse. Les méga-corpus et les archives numérisées permettent désormais des analyses diachroniques et synchroniques d'une ampleur inédite, tandis que le traitement automatique des langues ouvre des perspectives jusqu'alors inimaginables. On assiste à un véritable changement d'échelle dans la recherche lexicographique, où la modélisation des phénomènes linguistiques peut s'appuyer sur des masses de données qui dépassent largement les capacités de traitement humain. Cette évolution transforme le statut même du dictionnaire, qui tend à devenir une base de données relationnelle et dynamique, constamment actualisable et interrogeable selon de multiples axes.

Cette transformation s'accompagne d'une interdisciplinarité renouvelée, la lexicographie dialoguant désormais avec l'informatique, la science des données et la visualisation de l'information. Les projets lexicographiques deviennent des entreprises collaboratives où se croisent des compétences variées, et les résultats prennent des formes de plus en plus diversifiées : plateformes interactives, visualisations de réseaux lexicaux, ou interfaces de requêtes complexes. Cette ouverture s'accompagne d'un impératif de partage et de réutilisabilité des données, les humanités numériques prônant l'accès ouvert et l'interopérabilité des ressources.

Néanmoins, cette évolution ne va pas sans soulever des questions fondamentales. Le risque d'une fascination pour la quantité au détriment de la qualité et de l'interprétation existe bel et bien. La massification des données ne dispense pas de la réflexion théorique sur ce qui constitue un fait linguistique pertinent. Par ailleurs, la nécessaire médiation technologique introduit une nouvelle couche d'abstraction entre le lexicographe et son objet d'étude, ce qui exige une vigilance épistémologique constante. La tentation pourrait être grande de s'en remettre aveuglément aux algorithmes, au risque de méconnaître la complexité subtile des phénomènes linguistiques.

En définitive, les humanités numériques ne se contentent pas de fournir de nouveaux outils à la lexicographie : elles transforment en profondeur ses questions, ses méthodes et ses objets. Loin de rendre obsolète l'héritage structuraliste, elles lui offrent au contraire de nouveaux terrains d'expression et de validation. La cartographie du système lexical, rêvée par les structuralistes, trouve avec le tournant numérique les moyens techniques de sa réalisation à une échelle et avec une précision jusque-là inimaginable. La lexicographie devient ainsi pleinement une science des données linguistiques, tout en devant préserver l'intelligence interprétative qui fait la spécificité des sciences humaines.

III. implications méthodologiques à l'ère numérique

La technologie numérique a profondément bouleversé les méthodes lexicographiques, imposant une refonte complète du processus de travail, de la collecte des données à la diffusion des résultats. La première implication majeure réside dans le changement d'échelle et de nature des données disponibles.

Le lexicographe n'arpente plus les mêmes territoires : aux corpus soigneusement construits mais limités de la lexicographie sur corpus ont succédé des méga-corpus, le web comme corpus, et des flux de données massifs et hétérogènes. Ce nouveau paysage oblige à une délégation croissante du travail d'analyse à des algorithmes et à des outils de traitement automatique des langues (TAL).

Le lexicographe doit désormais maîtriser, ou du moins comprendre, des techniques comme l'extraction de termes, l'analyse distributionnelle, l'apprentissage automatique pour le regroupement des sens, et la visualisation de données. Son rôle évolue ainsi de manière significative : il devient un ingénieur et un interprète, qui conçoit des pipelines de traitement des données, formule des requêtes

computationnelles et, surtout, donne un sens aux patterns que les machines font émerger de masses de textes trop vastes pour être lues par un humain.

Cette évolution technique s'accompagne d'une transformation épistémologique profonde. La méthode ne consiste plus seulement à observer des exemples pour en inférer des règles, mais à construire des modèles computationnels du lexique. Le dictionnaire lui-même change de statut : il n'est plus conçu comme un artefact imprimé et figé, mais comme une base de données relationnelle et dynamique, constamment actualisable et interrogeable selon de multiples facettes.

Cette nouvelle nature ouvre la voie à une lexicographie différentielle, capable de modéliser finement les variations (diachroniques, diatopiques, diastratiques) et de produire des ressources adaptatives, voire personnalisées. La méthode devient donc plus hypothético-déductive : on formule des hypothèses sur le comportement des mots que l'on peut tester à grande échelle grâce aux données et aux algorithmes.

Enfin, le travail lexicographique est désormais fondamentalement collaboratif et interdisciplinaire. La complexité des outils et l'ampleur des données rendent obsolète la figure du lexicographe solitaire. Les projets associent nécessairement des linguistes, des informaticiens, des spécialistes de la visualisation de données et des documentalistes. Cette collaboration s'étend également à la collecte des données, avec l'émergence de méthodes participatives qui intègrent les contributions des utilisateurs, et s'inscrit dans une exigence nouvelle de science ouverte.

Les méthodologies doivent donc garantir l'interopérabilité, la réutilisabilité et la pérennité des ressources numériques produites, faisant de la lexicographie non plus seulement une discipline de description de la langue, mais aussi une discipline de gestion et de curation de données linguistiques complexes.

IV. Enjeux et perspectives

L'avènement du numérique et des humanités numériques place la lexicographie devant des enjeux fondamentaux qui en redéfinissent simultanément les risques et les promesses. Le premier défi est d'ordre épistémologique : face à la masse des données disponibles et à la puissance des algorithmes, comment préserver la place du jugement expert et de l'interprétation qualitative ?

Le risque est réel de voir s'imposer un positivisme numérique où la quantification des occurrences l'emporterait sur la compréhension fine des nuances sémantiques et des usages créatifs. Cette tension s'accroît avec la question cruciale de la représentativité des corpus, qui véhiculent inévitablement des biais – de genre, social, générationnel ou géographique – risquant de perpétuer une vision partielle et partielle de la langue. À ces défis scientifiques s'ajoutent des enjeux de souveraineté, alors que le coût des infrastructures numériques pourrait accorder une influence démesurée aux acteurs privés, au détriment d'une lexicographie indépendante et de bien commun.

Pourtant, les perspectives qui s'ouvrent sont considérables et laissent entrevoir un renouveau profond de la discipline. Le numérique permet d'imaginer une

lexicographie « différentielle » et hypercontextualisée, capable de générer des descriptions adaptées à des variétés régionales, à des publics spécifiques ou à des registres particuliers, mettant ainsi fin à l'idée d'un dictionnaire universel et moyen.

Cette évolution s'inscrit dans le projet plus large d'un « dictionnaire total », une plateforme intégrée où les données lexicales dialogueraient avec des ressources encyclopédiques, des archives multimédias et des atlas linguistiques. L'intelligence artificielle, au-delà de son rôle analytique, pourrait devenir une véritable assistante de création, suggérant des ébauches de définitions, générant des exemples contextualisés ou identifiant des lacunes dans la description du lexique.

Enfin, cette transformation s'accompagne d'une démocratisation des pratiques, où le lexicographe, devenu curateur et valideur, saura intégrer les contributions des communautés de locuteurs pour une description plus vivante et représentative de la langue telle qu'elle se parle et se vit.

Ainsi, entre les écueils d'une technicisation excessive et les promesses d'un enrichissement sans précédent, l'avenir de la lexicographie semble devoir passer par une alliance renouvelée entre l'intelligence humaine et la puissance computationnelle. Loin de rendre le lexicographe obsolète, le numérique lui offre l'opportunité de réaliser pleinement l'ambition structurale de cartographier le système de la langue avec une précision et une ampleur jusqu'alors inimaginables, tout en préservant la finesse interprétative qui reste au cœur du travail sur le sens.

Le passage aux attestations lexicographique n'est alors pas sans soulever des questions pertinentes :

La perte de l'autorité : En se détachant des « grands auteurs », la lexicographie gagne en objectivité mais perd-elle en légitimité aux yeux du grand public ? Un équilibre doit être trouvé.

La surinterprétation des données : La facilité d'accès aux attestations peut-elle conduire à des interprétations hâtives ? Le rôle du lexicographe comme expert interprétant le contexte reste plus que jamais nécessaire.

L'attestation multimodale : À l'avenir, l'attestation devra-t-elle intégrer des données provenant de discours oral (transcriptions alignées avec l'audio) ou même d'interactions multimodales (gestes, images), élargissant encore le concept d'« usage en contexte » ?

Le parcours de la lexicographie, du structuralisme à l'ère numérique, peut se lire comme une transformation profonde du principe même d'attestation. Nous sommes passés d'une attestation textuelle, de nature essentiellement qualitative et illustrative, à une attestation lexicographique quantitative et systémique.

L'ère pré-structuraliste et philologique fondait son autorité sur l'attestation par la citation, un fragment d'autorité puisé chez un « bon auteur ». Le lexicographe était un collectionneur de ces preuves textuelles, et le dictionnaire, une galerie de citations. Le structuralisme a opéré un premier basculement en déplaçant l'enjeu de la preuve : il ne s'agissait plus seulement d'attester qu'un mot existe dans un texte prestigieux, mais de démontrer sa valeur et sa place au sein du système linguistique. L'attestation devenait relationnelle ; elle se prouvait par les liens de synonymie, d'antonymie et par l'organisation logique des sens.

La révolution décisive est intervenue avec la lexicographie de corpus, qui a transformé l'attestation en une pratique empirique et statistique. La preuve n'est plus dans une citation unique, mais dans la récurrence massive des patterns d'usage observables dans des millions de textes. L'autorité ne réside plus dans le prestige de la source, mais dans la fréquence et la régularité des phénomènes linguistiques. On ne dit plus « ce mot existe parce que Victor Hugo l'a employé », mais « ce sens est le plus central parce qu'il est statistiquement le plus fréquent, et ces collocations sont attestées par des milliers d'occurrences ».

Actuellement, les humanités numériques et l'intelligence artificielle poussent cette logique à son paroxysme en faisant de l'attestation un processus algorithmique et dynamique. Le dictionnaire n'est plus le réceptacle statique d'attestations compilées, mais une base de données interactive où l'attestation se génère à la demande, modélisée par des réseaux de neurones et vérifiable en temps réel contre des corpus toujours actualisés. La preuve lexicale devient computationnelle.

Ainsi, la transition est complète : de l'attestation philologique (la preuve par l'exemple d'auteur), on est passé à l'attestation structuraliste (la preuve par la relation), puis à l'attestation statistique (la preuve par la fréquence), pour aboutir à l'attestation algorithmique (la preuve par le modèle computationnel). Cette évolution n'a pas annulé les étapes précédentes, mais les a intégrées et complexifiées. Le défi contemporain du lexicographe est de maîtriser cette chaîne de l'attestation, de la donnée brute à l'interprétation experte, pour produire une description de la langue qui soit à la fois scientifiquement fondée sur les données et humainement intelligible, préservant la richesse du sens au-delà de la simple quantification.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. **Bray, L.** (1986). César-Pierre Richelet (1626-1698) : Biographie et œuvre lexicographique. Niemeyer.
2. **Quemada, B.** (1968). Les Dictionnaires du français moderne, 1539-1863. Didier.
3. **Matoré, G.** (1968). *Histoire des dictionnaires français*. Larousse.
4. **Académie française** (1694). *Le Dictionnaire de l'Académie française*. [Première édition].
5. **Wooldridge, T. R.** (1977). *Les Débuts de la lexicographie française*. University of Toronto Press.

POUR CITER L'AUTEUR :

HAMAMDIA Samir, (2025), « De l'attestation textuelle à l'attestation lexicographique : évolution épistémologique et méthodologique », Ex Professo, V 11, N 01, pp. 37- 48, Url : <https://www.asjp.cerist.dz/en/PresentationRevue/484>