

VERS UNE ANALYSE AUTOMATIQUE DU DISCOURS EN HISTOIRE DE L'ARCHITECTURE

This contribution examines the methodological potential of automated discourse analysis in architectural history. Language, a fundamental component of architectural practice, can now be approached through digital methods derived from corpus linguistics and natural language processing. Starting with a historiographical overview of the relationship between architecture and language since the 1960s, the paper shows how figures such as Summerson, Zevi, and Jencks conceptualized architecture as a linguistic system. Yet, despite these early insights, few studies have applied computational methods to the analysis of architectural texts. The pioneering work of Alexander Tzonis in the 1970s on “conceptual systems” in French architectural texts laid the groundwork for the computerized study of architectural discourse. Today, the widespread digitization of sources and major technological advances make it possible to assemble large textual corpora suitable for automated discourse analysis. Approaches developed in the digital humanities – such as textometry, topic modeling, argument analysis, and the use of large language models – have proven particularly fruitful. They have opened up a promising interdisciplinary field, offering new ways to trace the evolution of deontic discourse and theoretical conceptions of architecture, as well as the formulation of aesthetic judgments and the documentation of the reception of architectural works.

La question de la littérature artistique a retenu de longue date l'attention des historiens d'art¹ et plus récemment des historiens de l'architecture², mais peu de travaux ont jusqu'à présent eu recours à l'analyse de discours. Au-delà de la théorie et de ses contenus déictiques³, le discours occupe pourtant une place fondamentale dans la pratique architecturale. Qu'il s'agisse de la présentation du projet, de la délibération sur l'opportunité des solutions apportées par les architectes, ou encore de la réception critique d'un édifice, le langage est sans cesse mobilisé dans la pratique (fig. 1)⁴. Ainsi, une large part des sources de l'historien sont le résultat d'actes de langage. Acquérir une meilleure connaissance sur le discours architectural, ses modalités et ses évolutions est donc déterminant tant pour pouvoir mener une réelle critique des sources que pour en produire l'analyse.

Dans cette contribution, nous nous proposons de montrer comment les approches informatiques développées ces dernières années, au croisement des domaines disciplinaires de la linguistique computationnelle⁵ et de la textométrie, présentent un potentiel déterminant pour l'histoire de l'architecture. À partir d'une historiographie des relations entre langage et architecture et du recours à l'informatique pour le traitement des textes architecturaux, il s'agira d'aborder les enjeux épistémologiques soulevés par l'application des méthodes d'analyse du discours en architecture. Au-delà de la stricte dichotomie entre lecture distante et rapprochée fréquemment signa-

lée par le champ des humanités numériques⁶, ou de problématiques de fouille et de restructuration de corpus grâce à la reconnaissance d'entités nommées ou de *topics*⁷, il convient de déterminer comment arriver à saisir des notions architecturales pertinentes dans le contexte d'une recherche historique. L'apparition récente des grands modèles de langue (LLM) nous paraît, en effet, rendre désormais possible l'étude des conceptualisations architecturales et leur saisie diachronique⁸.

L'architecture et la langue

Il n'y a pas d'architecture sans discours. L'omniprésence du langage dans le processus de création architecturale est patente. Que ce soit pour exprimer un programme ou présenter le projet, ou encore pour communiquer sur la réalisation, chacune des étapes du processus de production architecturale mobilise le langage. Il n'est pas surprenant que la théorie, en tant qu'entreprise scripturale, ait toujours fait grand cas du langage⁹. L'analogie entre l'architecture et la langue, qui apparaît dès le traité de Vitruve, est reprise à la Renaissance par Leon Battista Alberti. Elle devint, par la suite, un lieu commun du discours théorique, tout particulièrement sur les ordres. Plusieurs auteurs ont exploré ces dernières années les rapports entre la rhétorique et l'architecture dans les traités¹⁰. Toutefois, ces travaux n'abordent pas le matériau textuel proprement dit, mais des questions esthétiques concernant l'art de bâtir et ses effets.

Parler d'architecture et de langage renvoie immédiatement à la littérature. D'autres recherches, engagées dans le courant des années quatre-vingt du vingtième siècle, se sont plus particulièrement penchées sur la représentation de l'architecture dans la littérature. Dès 1981, Philippe Hamon organisait à l'Université de Rennes un colloque consacré au texte et à l'architecture. Cette veine a, depuis, nourri de nombreuses publications en France¹¹. Toutefois, ce n'est pas le point de vue littéraire qui nous intéresse ici, mais plus spécifiquement la dimension linguistique de l'architecture.

Trois ouvrages majeurs dans l'historiographie présentent des titres symptomatiques d'une conception langagière de l'architecture. En 1964, John Summerson publiait *Le langage classique de l'architecture*¹². Bruno Zevi faisait paraître une contrepartie contemporaine avec *Le langage de l'architecture moderne* en 1973¹³. Enfin, la même année, Charles Jencks publiait *Le langage de l'architecture postmoderne*¹⁴. À des niveaux différents, ces trois livres dénotent un intérêt particulier pour la langue et peuvent être mis en rapport avec le tournant linguistique observé à partir des années soixante dans les sciences humaines¹⁵. Si l'influence de Ferdinand de Saussure était sensible, c'était principalement dans la volonté d'identifier une syntaxe et une grammaire architecturale. Issu d'une série de conférences sur la BBC, l'*opus* de John Summerson était avant tout un texte d'introduction sur l'architecture moderne.



page 53

Fig. 1 J.D. Steingruber, *Architectonisches Alphabet...*, Schwabach 1773. Frontispice (ETH-Bibliothek Zürich, Rar 10420, domaine public).

¹ J. VON SCHLOSSER, *La littérature artistique : manuel des sources de l'histoire de l'art moderne*, Paris 1984.

² M. CARPO, *Architecture in the Age of Printing: Orality, Writing, Typography, and Printed Images in the History of Architectural Theory*, Cambridge 2001 ; *La construction savante : les avatars de la littérature technique*, actes du colloque (Paris, mars 2005), sous la direction de J.P. Garric, V. Negre, A. Thomine-Berrada, Paris 2008 ; *Le Livre et l'architecte*, actes du colloque (Paris, 31 janvier-2 février 2008), sous la direction de J.P. Garric, E. d'Orgeix, E. Thibault, Wavre 2011 ; M. HVATTUM, A. HULTZSCH, *The Printed and the Built: Architecture, Print Culture, and Public Debate in the Nineteenth Century*, New York 2018 ; H. JANNIÈRE, *Critique et architecture : un état des lieux contemporain*, Paris 2019 ; R. WITTMAN, *Architecture, Print Culture, and the Public Sphere in Eighteenth-Century France*, New York-London 2007.

³ A. TZONIS et al., *Les systèmes conceptuels de l'architecture en France de 1650 à 1800*, Paris 1975.

⁴ C.G. CRYSLER, *Writing Spaces: Discourses of Architecture, Urbanism and the Built Environment 1960-2000*, London 2003 ; T.A. MARKUS, D. CAMERON, *The Words between the Spaces: Buildings and Language*, London 2002.

⁵ F. RASTIER, *La mesure et le grain : sémantique de corpus*, Paris 2011, p. 20.

⁶ F. MORETTI, *Distant Reading*, London 2013 ; M. PIOTROWSKI, *Natural Language Processing for Historical Texts*, San Rafael 2012.

⁷ *Doing Digital Humanities: Practice, Training, Research: Open Approaches to Creation, Growth, and Development*, edited by C. Crompton, R. Lane, R. Siemens, London 2016 ; P.J. ORTIZ SUÁREZ et al., *Establishing a New State-of-the-Art for French Named Entity*, in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, conference proceedings (Marseille, 11-16 may 2020), edited by N. Calzolari et al., Marseille 2020, pp. 4631-4638.

⁸ *L'intelligence artificielle des textes : des algorithmes à l'interprétation*, sous la direction de D. Mayaffre, L. Vanni, Paris 2021 ; C. ZIEMS et al., *Can Large Language Models Transform Computational Social Science?*, "Computational Linguistics", 50, 2024, 1, pp. 237-291, <https://aclanthology.org/2024.cl-1.8.pdf> (consulté le 30 août 2025).

⁹ W. SZAMBIEN, *Symétrie, goût, caractère : théorie et terminologie de l'architecture à l'âge classique 1550-1800*, Paris 1986.

¹⁰ *Architecture and Language: Constructing Identity in European Architecture, c. 1000 -c. 1650*, edited by G. Clarke, P. Crossley, Cambridge-New York 2000 ; A.A. PAYNE, *Ut Poesis Architectura: Tectonics and Poetics in Architectural Criticism circa 1570*, in *Antiquity and Its Interpreters*, edited by Ead., A. Kuttner, R. Smick, Cambridge 2000, pp. 145-158 ; C. VAN ECK, *Classical Rhetoric and the Visual Arts in Early Modern Europe*, Cambridge 2007 ; A.A. PAYNE, *The Architectural Treatise in the Italian Renaissance: Architectural Invention, Ornament and Literary Culture*, New York 2011.

¹¹ P. HAMON, *Expositions : littérature et architecture au XIX^e siècle*, Paris 1989 ; *Architecture et littérature : une interaction en question, XX^e-XXI^e siècles*, acte du colloque (Cerisy, Centre Culturel International, 1-8 Septembre 2009), sous la direction de P. Hyppolite, A. Leygonie, A. Verlet, Aix-en-Provence 2014 ; F. MOULIN, *Embellir, bâtir, demeurer : l'architecture dans la littérature des Lumières*, Paris 2017 ; D. ROUILLARD, *Les monuments de la langue : architecture, mémoire, écriture*, Genève 2021.

¹² J.N. SUMMERSON, *The Classical Language of Architecture*, London 1964.

Néanmoins, il proposait une stylistique de l'architecture classique et en définissait les emplois et les variations au cours des époques. Si toute trace de linguistique y est absente, il témoigne néanmoins d'un intérêt particulier pour la langue. Bruno Zevi cherchait, quant à lui, à mettre à jour un "code anticlassique" en faisant plus directement appel à la linguistique et à la sémiotique. Il s'agissait pour lui d'une tentative d'énonciation des principes non-écrits qui sous-tendent la conception moderne de l'architecture. Cet objectif était principalement pédagogique, mais il confinait l'architecture à la forme ; une idée que reprit directement Charles Jencks pour son architecture postmoderne.

Au cours des années soixante et soixante-dix, plusieurs chercheurs en histoire de l'architecture embrassèrent plus directement la linguistique et la sémiologie (ou la sémiotique) comme une nouvelle manière d'accéder et de révéler des niveaux de signification dans leurs objets d'étude. Toutefois, la transposition de ces approches aux arts visuels posait des difficultés particulières. Les tentatives initiales pour articuler l'architecture à la sémiologie remontent au milieu des années trente. Mais ce fut surtout en Italie, dans l'immédiate après-guerre, que s'élaborèrent les premiers concepts d'une sémiologie architecturale, notamment autour de l'École d'architecture de Florence¹⁶.

La thèse de doctorat de Robert G. Hershberger, soutenue à l'Université de Pennsylvanie en 1969, *A Study of Meaning and Architecture*, constitua visiblement un jalon important dans l'application de l'informatique à des corpus architecturaux¹⁷. Menée sous la direction de Holmes Perkins, assisté par le spécialiste de sémantique différentielle Percy Tannenbaum, l'approche déployée était principalement sémiotique.

Dans l'ouvrage collectif de 1970, *Meaning in Architecture*, Charles Jencks et George Baird, se proposaient de démontrer comment la sémiolo-

gie, la linguistique, l'anthropologie structurelle et la cybernétique ou la théorie de l'information pouvaient être appliquées à l'analyse et l'interprétation de l'architecture¹⁸. L'objet central de la sémiologie est "l'étude des systèmes de signes institués par la société dans des contextes donnés destinés à comprendre les conventions culturelles sur lesquelles le processus de signification est basé"¹⁹. Une telle approche suppose donc de considérer l'architecture comme un 'langage' de formes avec l'objectif de capturer sa dimension de communication et de signification sociale. En tant que discipline, la sémiologie s'intéresse aux signes et symboles verbaux et non verbaux, et elle intègre les théories de la communication et la linguistique comme des sous-divisions. Mais, en dépit de tentatives destinées à fonder une sémiotique proprement visuelle, ou architecturale, cette approche reste fondamentalement centrée sur les mots.

Le domaine fut d'abord dominé, on l'a vu plus haut, par l'influence du *Cours de linguistique générale* de Ferdinand de Saussure, traduit en anglais en 1959, qui établissait une théorie générale du signe qu'il appela *sémiologie*. Dans ses écrits, il relevait une distinction entre des phénomènes diachroniques et synchroniques qui déboucha sur le paradigme intellectuel du structuralisme. Au cours des années soixante, Donald Preziosi, qui avait étudié la linguistique avec Roman Jakobson, développa une approche strictement visuelle des espaces bâtis qui découlait des réductions élémentaires développées par la linguistique²⁰. D'autres auteurs s'inspirèrent de la sémiotique de l'américain Charles Sanders Pierce comme Geoffrey Broadbent ou Robert Venturi.

Cette veine donna lieu à plusieurs colloques et publications en France. Les actes du colloque d'Albi *Espace et représentation : sémiotique de l'architecture* de 1981 accueillent une contribution du célèbre linguiste et sémioticien Algirdas Julien

Greimas²¹. Cette approche se retrouve également dans la *space syntax* de Bill Hillier²². Sa manière de considérer les formes de l'architecture et de distinguer le bâti et le non-bâti peut apparaître comme le prolongement des postulats saussuriens sur le rapport entre signifiants et signifiés. On peut aussi en retrouver la trace dans le concept de *pattern language* de Christopher Alexander²³ ou même les recherches de George Stiny²⁴.

Les premières applications de l'informatique aux textes architecturaux

Il n'est guère étonnant que ce soit la théorie architecturale qui ait donné lieu aux premières applications de l'informatique à l'architecture. Le texte se prête en effet facilement à des manipulations computationnelles. À cet égard, l'approche développée sous la direction d'Alexander Tzonis dans *Les systèmes conceptuels de l'architecture en France de 1650 à 1800*, un ouvrage publié en 1975, fut tout à fait pionnière²⁵. Enseignant à la Graduate School of Design d'Harvard depuis 1968 au sein d'un Institut des sciences de la planification, il y avait développé des recherches dans le domaine des méthodes de conception avec l'aide de Walter Isard et Ovidia Salama. En 1972, le chercheur fut invité par le ministère de la Culture en France pour une année d'étude à Strasbourg. Il mena ce projet collectif multidisciplinaire au ministère de l'Équipement en 1974 et 1975, bénéficiant du soutien du Comité pour la recherche et le développement en architecture (CORDA) (fig. 2).

La recherche de Tzonis et de ses collègues se distingue des nombreux autres travaux publiés à l'époque sur la théorie architecturale par l'emploi d'une méthodologie formelle utilisant le traitement informatique. Les auteurs formulaient l'hypothèse que les textes étaient porteurs de "systèmes conceptuels" qui servent de "modèles décisionnels" pour les constructeurs dans des situations réelles. L'étude rassemble donc

une investigation épistémologique des modèles décisionnels de design ainsi qu'une analyse historique. Puisque le matériel examiné est constitué par des textes directement et indirectement liés aux décisions de design, les techniques de la linguistique moderne, de l'analyse du discours et de l'analyse du contenu sont mobilisées conjointement.

Au sein de cette recherche collaborative, Michael Freeman de l'université Harvard avait étudié l'interface entre la linguistique et les modèles décisionnels. Liane Lefavre avait travaillé les problèmes liés à la mise en rapport de l'analyse du discours et du matériel historique. Ovidia Salama de l'université Columbia avait contribué à la méthodologie générale du projet et à la théorie des modèles décisionnels. Robert Berwick, qui était alors l'étudiant de Tzonis, s'était penché sur les rapports entre l'intelligence artificielle et la méthode d'analyse. De nombreux chercheurs avaient, par ailleurs, contribué au projet conçu, semble-t-il, avec René Tabouret à Strasbourg²⁶. À partir d'une analyse serrée des textes, il s'agissait de pouvoir mettre à jour et identifier des châssis décisionnels qui permettaient en retour de classifier les textes les uns par rapport aux autres. Cette recherche s'inscrivait dans un cadre plus large visant à développer des outils de prise de décision en matière de conception architecturale. Étonnamment, les propositions de Tzonis eurent peu de suite. L'auteur, qui se consacra par la suite plus spécifiquement aux systèmes de décisions, ne prolongea pas cette approche textuelle qui apparaissait peut-être trop coûteuse à l'époque, ou qui ne présentait pas aux yeux des spécialistes des résultats suffisamment convaincants.

Il fallut ensuite attendre la fin des années 90 et le début des années 2000, pour que plusieurs publications qui s'inscrivaient dans le courant de l'histoire intellectuelle s'emparent à nouveau du discours architectural comme objet

¹³ B. ZEVI, *Le Langage moderne de l'architecture*, Paris 1981.

¹⁴ C. JENCKS, *The Language of Post-modern Architecture*, New York 1977.

¹⁵ C. GAILLARD, *Du paradigme linguistique en architecture : Illusion, désillusion, abandon*, article en ligne, section 'Philosophie générale (anciens articles)', 2019, pp. 1-19, <https://www.clementgaillard.com/articles/philosophie/duparadigmelin-guistiqueenarchitecture> (consulté le 28 mars 2025).

¹⁶ I. GAMBERINI, *Per una analisi degli elementi dell'architettura: introduzione ai corsi propedeutici di architettura nella Facoltà di Firenze*, Firenze 1953 ; ID., G.K. KOENIG, *Introduzione al primo corso di elementi di architettura e rilievo dei monumenti: gli elementi di architettura come parole del linguaggio architettonico*, Firenze 1959.

¹⁷ R.C. CLEMENTS, R.G. HERSHBERGER, *A Study of Meaning and Architecture [Abstract and Review]*, "Review of Research in Visual and Environmental Education", I, 1973, 1, pp. 75-82 ; R.G. HERSHBERGER, *Architecture and Meaning*, "Journal of Aesthetic Education", IV, 1970, 4, pp. 37-55. Malheureusement, nous n'avons pas pu consulter cette thèse.

¹⁸ *Meaning in Architecture*, edited by C. Jencks, G. Baird, New York 1970.

¹⁹ *Signs, Symbols, and Architecture*, edited by G. Broadbent, R. Bunt, C. Jencks, Chichester 1980.

²⁰ D. PREZIOSI, *Architecture, Language and Meaning: The Origins of the Built World and its Semiotic Organization*, Berlin-Boston 1979.

²¹ *Espace et représentation : sémiotique de l'architecture*, sous la direction de A. Renier, Paris 1982.

²² B. HILLIER, J. HANSON, *The Social Logic of Space*, Cambridge 1984.

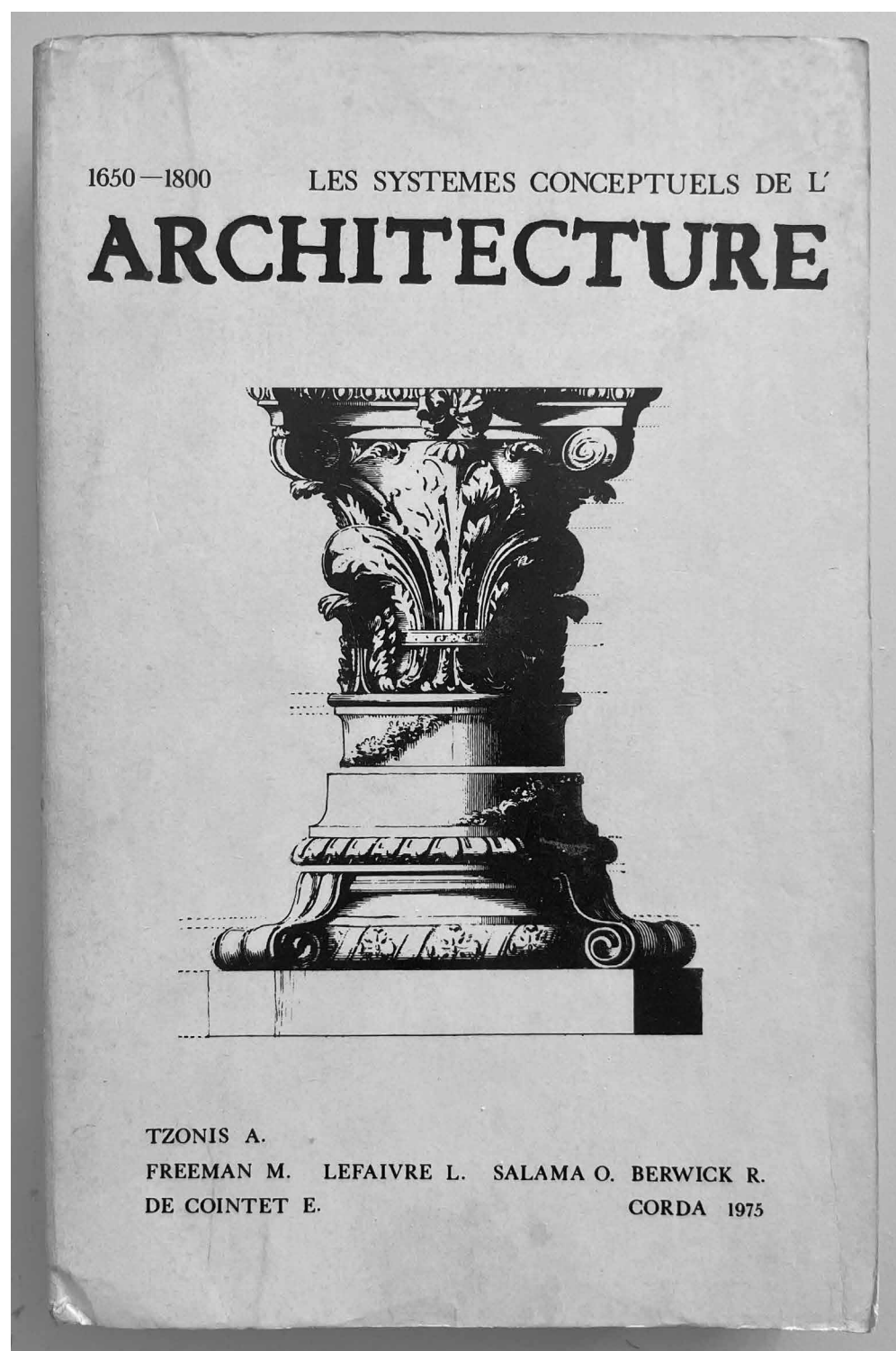
²³ *A Pattern Language: Towns, Buildings, Construction*, edited by C. Alexander, S. Ishikawa, M. Silverstein, New York 1977.

²⁴ G. STINY, W.J. MITCHELL, *The Palladian Grammar*, "Environment and Planning B: Planning and Design", V, 1978, 1, pp. 5-18.

²⁵ Tzonis et al., *Les systèmes conceptuels de l'architecture...* cit.

²⁶ *Ibidem*.

Fig. 2 Tzonis et al., *Les systèmes conceptuels de l'architecture...* cit., 1975 (photo E. Château-Dutier).



d'étude mais, cette fois-ci, sans recours à l'outil informatique. En s'intéressant aux changements de signification de termes clefs du modernisme, Adrian Forty identifiait, dans *Words and Buildings*, leur rôle structurant tant en dedans qu'en dehors de la profession²⁷. Thomas Markus et Deborah Cameron, dans *The Words between the Spaces*, montraient également que les édifices n'étaient pas de simples représentations autonomes en ayant extensivement recours à l'analyse du discours²⁸. La lecture rappro-

chée des textes révélait à quel point notre compréhension de l'environnement est déterminée par le langage. De même, Greig Crysler, en se consacrant plus particulièrement à la théorie de l'environnement bâti dans les revues d'architecture au cours de la seconde moitié du vingtième siècle, montrait l'importance des discours et l'effet déterminant des textes sur la manière dont on comprend, imagine et agit sur le monde²⁹. Ces approches peuvent donc être rattachées au *linguistic turn* et à l'analyse critique du discours, un

²⁷ A. FORTY, *Words and Buildings: A Vocabulary of Modern Architecture*, London 2000.

²⁸ MARKUS, CAMERON, *The Words between the Spaces...* cit.

²⁹ CRYSLER, *Writing Spaces...* cit.

“courant qui assigne aux langages un rôle décisif dans la construction de la réalité sociale, en particulier des mouvements collectifs”, et qui eut un impact considérable dans la culture anglo-saxonne³⁰. Mais plutôt que d’une exploration distributionnelle des textes, il s’agissait avant tout d’une herméneutique³¹.

Juan Pablo et Sylvia Bonta furent parmi les rares chercheurs à appliquer des traitements informatisés aux textes architecturaux dans *American Architects and Texts* publié en 1996. L’ouvrage était issu d’une analyse assistée par ordinateur de près de quatre cents textes sur l’architecture américaine et s’accompagnait d’un complément numérique sous la forme d’une base de données³². La préface de William J. Mitchell (1944-2010) soulignait à quel point cette étude méticuleuse permettait de dégager des mécanismes de construction de la notoriété, ainsi que l’intérêt de l’utilisation de graphiques et de statistiques. La mise à disposition des données et du logiciel développé pour les besoins du projet, permettant de rejouer soi-même les résultats, ce qui annonçait des changements dans les pratiques de lecture. En réalité, les auteurs s’étaient contentés de produire une analyse des index des ouvrages pour des raisons d’accessibilité au matériau textuel.

Néanmoins, les deux auteurs identifiaient plusieurs enjeux relatifs à l’application des approches automatiques d’analyse du discours aux corpus architecturaux. La plupart des termes utilisés dans le discours architectural classique, tels que ‘colonne’ ou ‘entablement’, disposaient d’un référent formel univoque. Il était donc possible d’étudier la fréquence de leur emploi en recherchant les occurrences des termes associés et en les mettant en rapport avec la connaissance architecturale. Mais plusieurs mots employés par les architectes comme la logique, la vérité, ou la justice n’avaient pas d’équivalents formels. Même si la fréquence et l’évolution de

termes esthétiques tels que la beauté, le fonctionnalisme et certaines idées connexes, pouvaient être explorées, les auteurs pointaient un certain nombre de limitations technologiques, mais également méthodologiques. Comment rendre compte de la fréquence d’un terme en tenant compte des termes associés, et aussi de la diachronie ? S’ils identifiaient bien l’efficacité de l’analyse automatique pour mesurer la répétition, l’application de ces méthodes réclamait l’examen d’un volume important de publications, ce qui exigeait une numérisation à grande échelle du patrimoine écrit qui n’était encore que balbutiante.

Lorsqu’en 2012, Diane Zorich constatait l’ambivalence de l’histoire de l’art à l’égard du numérique dans son rapport sur l’histoire de l’art numérique pour Ithaka S+R, la situation n’était guère différente en histoire de l’architecture³³. Malgré l’engagement de longue date des architectes avec l’ordinateur, ce champ de recherche avait peu investi les humanités numériques et un nombre restreint de projets avaient jusqu’à porté sur les textes architecturaux ou la production de corpus textuels³⁴. Or, la disponibilité de grands ensembles textuels était déterminante pour rendre possible des approches automatiques de la langue. Ainsi, en 2007, Stephen Ramsay pouvait expliquer que la révolution numérique n’avait pas encore pénétré le cœur des études littéraires faute de l’existence de tels corpus. Mais dix ans plus tard, Matthew L. Jockers et Ted Underwood suggéraient que l’effet conjoint de la création d’archives numériques massives et de nouveaux outils prometteurs changeait radicalement la donne³⁵.

Aujourd’hui, l’industrialisation de la numérisation des collections patrimoniales, de même que les progrès réalisés du côté de la reconnaissance automatique de textes lèvent en grande partie le verrou de la disponibilité des corpus. En effet, des avancées extraordinaires ont été faites dans

³⁰ S. CERUTTI, *Le linguistic turn en Angleterre. Notes sur un débat et ses censures*, “Enquête, Anthropologie, Histoire, Sociologie”, V, 1997, pp. 125-140.

³¹ J.E. TOEWS, *Intellectual History after the Linguistic Turn: The Autonomy of Meaning and the Irreducibility of Experience*, “The American Historical Review”, 92, 1987, 4, pp. 879-907.

³² J.P. BONTA, S. BONTA, *American Architects and Texts: A Computer-Aided Analysis of the Literature*, Cambridge 1996.

³³ D.M. ZORICH, *Transitioning to a Digital World. Art History, Its Research Centers, and Digital Scholarship*, Report to the The Samuel H. Kress Foundation et The Roy Rosenzweig Center for History and New Media, George Mason University, juin 2012, <https://www.kressfoundation.org/Resources/Sponsored-Research/Research-Items/Transitioning-to-a-Digital-World> (consulté le 25 mars 2025).

³⁴ E. CHÂTEAU-DUTIER, *Des éditions numériques pour l’histoire de l’art : deux décennies d’expérimentations*, “Histoire de l’Art”, 87, 2021, 1, pp. 77-90.

³⁵ M.L. JOCKERS, T. UNDERWOOD, *Text Mining the Humanities*, in *A New Companion to Digital Humanities*, edited by S. Schreibman, R. Siemens, J. Unsworth, Chichester 2016, pp. 291-306.

³⁶ B. GATOS *et al.*, *Ground-Truth Production in the Transcriptorium Project*, in 2014 11th IAPR International Workshop on Document Analysis Systems, conference proceedings (Tours, 7-10 April 2014), Tours 2014, pp. 237-241.

³⁷ F. MORETTI, *Conjectures on World Literature*, "New Left Review", 2000, 1, pp. 54-68.

³⁸ M. BOILLET *et al.*, *HORAE: an Annotated Dataset of Books of Hours*, in *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing*, conference proceedings (Sydney, 20-21 September 2019), New York 2019, pp. 7-12 ; A. CHAGUÉ *et al.*, *Deux siècles de sources disparates sur l'industrie textile en France : comment automatiser les traitements d'un corpus non-uniforme ?*, dans *DHNord* (2019) "Corpus et archives numériques", présentation de colloque (Lille, 16-18 octobre 2019), <https://inria.hal.science/hal-02448921> (consulté le 25 mars 2025) ; A. CHAGUÉ, L. TERRIEL, L. ROMANY, *Des images au texte : LECTAUREP, un projet de reconnaissance automatique d'écriture*, dans *DHNord2020: The Measurement of Images: Computational Approaches in the History and Theory of the Arts*, présentation de colloque (Lille, novembre 2020), <https://hal.science/hal-03008579v1> (consulté le 25 mars 2025) ; M. GOHIER, *Nouvelle-France numérique : l'Intelligence artificielle au service de l'analyse du patrimoine documentaire*, Midiconférence de l'Association of Critical Heritage Studies, Université du Québec en Outaouais, 1^{er} février 2023, <https://nouvellefrance-numerique.info> (consulté le 25 mars 2025) ; P. KAHLE *et al.*, *Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents*, in 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), conference proceedings (Kyoto, 9-15 November 2017), IV, Kyoto 2017, pp. 19-24 ; D. STUTZMANN, J.F. MOUFFLET, S. HAMEL, *La recherche en plein texte dans les sources manuscrites médiévales : enjeux et perspectives du projet HIMANIS pour l'édition électronique*, "Médiévales. Langues, Textes, Histoire", 73, 2017, pp. 67-96.

³⁹ KAHLE *et al.*, *Transkribus*... cit. ; B. KIESSLING *et al.*, *eScripta: A New Digital Platform for the Study of Historical Texts and Writing*, in *Digital Humanities 2019*, conference paper (Utrecht, July 2019), <https://ephe.hal.science/hal-02310781v1> (consulté le 25 mars 2025).

⁴⁰ F. MORETTI, *Graphes, cartes et arbres : modèles abstraits pour une autre histoire de la littérature*, Paris 2008 ; *Id.*, *Distant Reading*... cit. ; *La littérature au laboratoire*, sous la direction de F. Moretti, Paris 2016.

⁴¹ S. RAMSAY, *In Praise of Pattern*, "TEXT Technology", 14, 2005, 2, pp. 177-190 ; *Id.*, *Reading Machines: Toward an Algorithmic Criticism*, Champaign 2011.

⁴² M.L. JOCKERS, *Macroanalysis: Digital Methods and Literary History*, Champaign 2013 ; JOCKERS, UNDERWOOD, *Text Mining the Humanities*... cit.

⁴³ G. ROCKWELL, *What is Text Analysis, Really?*, "Literary and Linguistic Computing", 18, 2003, 2, pp. 209-219 ; *Id.*, S. SINCLAIR, *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, Cambridge-London 2016.

⁴⁴ PIOTROWSKI, *Natural Language*... cit.

⁴⁵ T. POIBEAU, *Traitement automatique du contenu textuel*, Paris 2011 ; *Id.*, *Le traitement automatique des langues pour les sciences sociales. Quelques éléments de réflexion à partir d'expériences récentes*, "Réseaux", 188, 2014, 6, pp. 25-51 ; RASTIER, *La mesure et le grain*... cit.

le domaine de la numérisation des textes tant imprimés que manuscrits. La reconnaissance automatique de textes (*Automatic Text Recognition-ATR*, en anglais) consiste en l'acquisition d'un texte numérique plus ou moins structuré à partir d'images numérisées. Il s'agit d'un processus automatisé qui permet de traiter de larges masses documentaires en réduisant le temps nécessaire à leur traitement et les coûts humains relatifs à la transcription³⁶. Ceux-ci permettent l'indexation d'un fonds et la création d'un accès plein-texte, la mobilisation de fouille de texte et de lectures distantes³⁷, ou encore plus simplement l'édition et la publication de textes.

Ces dernières années, plusieurs initiatives de grande ampleur ont ainsi fait sortir la reconnaissance des écritures manuscrites du strict domaine expérimental de la recherche sur la vision par ordinateur pour prendre une place plus affirmée au sein des programmes de numérisation industriels conduits par les institutions patrimoniales ou dans le cadre de projets en humanités numériques³⁸. Ces différentes réalisations ont notamment permis de réunir d'importantes données d'annotation et la constitution de modèles de reconnaissance d'écritures manuscrites très efficaces pour les écritures anciennes. Parallèlement, les modèles et les outils de segmentation des pages pour les documents écrits se sont considérablement perfectionnés ainsi que les chaînes de traitement documentaire destinées à passer d'un format à l'autre.

La création de plusieurs logiciels interfacés a beaucoup facilité la mobilisation de ces différentes techniques pour le traitement des corpus patrimoniaux. Le projet européen *Transkribus* ou le projet français *eScriptorium* ont, par exemple, permis la mise au point d'outils qui rendent la prise en main et l'application de ces techniques relativement aisées sur de grands corpus manuscrits³⁹. En outre, l'initiative *HTR United* permet le partage de modèles de recon-

naissance d'écritures manuscrites anciennes. Ce n'est donc plus le seul corpus théorique imprimé, mais potentiellement l'ensemble des écrits architecturaux auquel il est désormais possible d'appliquer des lectures distantes : critiques architecturales, correspondance, délibérations administratives, etc.

Plusieurs types de traitements automatiques peuvent être aujourd'hui mobilisés pour analyser les évolutions du discours architectural. De très nombreuses questions deviennent explorables à travers les textes, qu'il s'agisse – cette liste n'est pas limitative – de la définition des types architecturaux ou de la convocation des précédents, de la formulation des énoncés stylistiques et des références historiques, en passant par la formation des énoncés déontiques et de l'argumentation architecturale. Chaque fois, l'historien de l'architecture dispose de méthodes et d'outils bien éprouvés dans d'autres domaines d'étude.

Des méthodes d'analyse automatique du discours en architecture

Les travaux conduits dans le champ des humanités numériques au cours des deux dernières décennies ont clairement démontré l'intérêt de l'analyse automatique du discours en littérature et dans les sciences historiques. Franco Moretti⁴⁰, Stephen Ramsay⁴¹, Matthew L. Jockers et Ted Underwood⁴², Geoffrey Rockwell et Stefan Sinclair⁴³ et Michel Piotrowski⁴⁴ ont contribué à fixer plusieurs méthodologies qui relèvent à la fois de la critique littéraire et de la linguistique de corpus⁴⁵. Souvent quantitatives, ces approches ne se limitent pas à la description de la langue dans des corpus donnés. Elles permettent aussi de travailler sur la sémantique ou encore l'argumentation. En effet, l'objet de l'histoire de l'architecture ne concerne pas à proprement parler les phénomènes textuels ou langagiers. Si ceux-ci ont une place dans l'analyse, il

s'agit plutôt de comprendre comment le langage peut nous informer sur la pratique architecturale. De ce point de vue, utiliser ces techniques ne fait pas disparaître l'interprétation, il s'agit plutôt d'une méthode complémentaire qui permet d'accéder aux contenus des textes et de réunir des faits qui sont soumis dans un second temps à la sagacité de l'historien.

Parmi les techniques à notre disposition, plusieurs nous semblent particulièrement pertinentes dans notre discipline. Il peut s'agir de méthodes supervisées ou non supervisées, par exemple le *topic modeling*, la textométrie, l'analyse de l'argumentation et l'utilisation de grands modèles de langue. De nombreux algorithmes sont d'ores et déjà disponibles pour les chercheurs et les chercheuses et sont intégrés dans des logiciels comme Weka ou MALLET, TXM, ou encore disponibles par l'intermédiaire de bibliothèques de code informatique. Cependant, ces outils ne suffisent pas pour parvenir à des résultats, puisqu'il est souvent nécessaire de préparer et standardiser les données, de rassembler des métadonnées et des corpus, et enfin de disposer de compétences statistiques.

Depuis les années soixante-dix, l'introduction de méthodes et de techniques issues de la linguistique a véritablement renouvelé l'approche historique⁴⁶. Cependant, celles-ci ont largement pâti de la remise en question du structuralisme quantitatif dans le domaine des sciences humaines par rapport au domaine des sciences sociales. Pourtant, l'approche textométrique a notamment su se faire une place en histoire dans le domaine de l'analyse des discours politiques. C'est une méthode de comparaison des textes, réunis en corpus, sur la base d'un comptage des unités de sens : la forme, le lemme, le groupe de formes. La lexicométrie constitue donc avant tout un ensemble de méthodes de description. Différentes échelles permettent de multiplier les prismes d'analyse et ces indicateurs ne sont ni in-

terprétatifs ni explicatifs ; c'est le chercheur qui fait l'interprétation et procède à la constitution des catégories. Il s'agit en somme d'étudier les textes du corpus non pas avec des classifications prédéfinies, mais avec des classifications créées spécifiquement pour le corpus à partir du corpus lui-même.

L'intérêt de l'approche textométrique est donc qu'elle contribue à se défaire de l'idée que les textes auraient un sens qu'il faudrait dévoiler. Elle ne remplace pas l'historien et l'interprétation mais fournit une aide et des outils à l'historien pour l'interprétation. Par exemple, en architecture, elle peut permettre d'identifier l'évolution de certaines occurrences dans le discours ou de contraster les énoncés théoriques de tel ou tel auteur. Cette approche n'est pas toujours pertinente selon la nature des corpus. Elle est appropriée pour les contenus qui varient les uns des autres selon les critères appliqués dans l'analyse factorielle de correspondance. En revanche, lorsque le champ lexical général est relativement similaire et les contenus explorés emploient des formules stéréotypées, l'analyse peut apparaître triviale.

C'est la raison pour laquelle il peut être utile de s'orienter vers d'autres méthodes d'analyse automatisées. À cet égard, le domaine de l'analyse d'arguments (*argument mining*) a récemment connu un certain essor avec l'émergence de techniques d'analyse de textes plus puissantes fondées sur l'utilisation des réseaux neuronaux. L'argumentation est une activité sociale dans laquelle des agents cherchent à convaincre des interlocuteurs qu'une déclaration est vraie ou fausse en mettant en avant des arguments. Là où le raisonnement concerne la recherche de la vérité d'une déclaration, l'argumentation vise à persuader des agents⁴⁷. Ce domaine de recherche interdisciplinaire avait déjà été profondément renouvelé dans la seconde moitié du vingtième siècle par les travaux d'Hamblin, de Rescher, de

⁴⁶ *L'analyse textuelle des idées, du discours et des pratiques politiques*, sous la direction P.M. Daigneault, F. Pétry, Paris-Québec 2017 ; *Histoire de mots. Saisir le passé grâce aux données textuelles*, sous la direction de L. Dumont, O. Julien, S. Lamassé, Paris 2023 ; L. LEBART, A. SALEM, *Statistique textuelle*, Paris 1994 ; D. MAYAFFRE, *Analyses logométriques et rhétorique du discours*, dans *Introduction à la recherche en SIC*, sous la direction de S. Olivesi et al., Grenoble 2007, pp. 153-180 ; ROCKWELL, SINCLAIR, *Hermeneutica...* cit.

⁴⁷ F.H. VAN EEMEREN et al., *Handbook of Argumentation Theory*, Dordrecht 2014, p. 7.

Chaïm Perelman et Lucie Olbrechts-Tyteca⁴⁸ ou de Stephen Toulmin⁴⁹. Un système d'analyse d'arguments a pour objectif la génération automatique d'un graphe d'arguments à partir de textes non structurés, et peut généralement être divisé en une séquence d'étapes comportant notamment la détection d'arguments et la modélisation des liens unissant ces derniers⁵⁰.

Plusieurs sources qui intéressent l'historien, même si elles ne sont pas fondamentalement dialogiques (au sens où elles ne mettent pas directement en rapport des avis contradictoires), peuvent être assimilables dans le champ rhétorique au domaine du discours délibératif, au sens aristotélicien du terme. En prolongeant l'analyse classique des *topoi* de la rhétorique antique, Chaïm Perelman et Lucie Olbrechts-Tyteca ont notamment mis en avant l'idée de schèmes d'argumentation qui jouent un rôle important dans l'analyse et l'évaluation des discours argumentatifs⁵¹. Une telle approche débouche sur de fines classifications argumentaires qui intéressent évidemment l'historien d'architecture que ce soit pour l'analyse du discours déontique que pour celle de la critique d'art, ou des délibérations des jury de concours.

Nous souhaitons enfin évoquer diverses technologies d'analyse automatique des textes qui reposent sur des stratégies de vectorisation plus ou moins complexes⁵². D'abord, l'utilisation d'extraction de contenus thématiques (*Topic maps*, ISO/IEC 13250 2003) constitue une méthode classique. Ces méthodes fondées sur la similarité sémantique peuvent s'avérer utiles pour classer les textes. D'autres solutions d'analyse automatisées peuvent être pertinentes lorsque les champs lexicaux sont trop similaires. L'utilisation de trigrammes peut, par exemple, aider à effectuer des alignements de séquences et identifier des passages similaires. Ces dernières années, les méthodes de traitement naturel de la langue ont considérablement évolué avec l'apparition

des prolongements lexicaux (*word embeddings*) et des transformateurs (*transformers*) ou encore les grands modèles de langue⁵³. Qu'il s'agisse de Word2Vec ou de BERT, ces représentations des documents permettent notamment de retrouver des passages qui expriment les mêmes idées de façons complètement différentes et potentiellement de repérer les paraphrases et les allusions⁵⁴. Ils ont donc un potentiel énorme pour étudier la transformation des idées ou les idées analogues à travers des grands corpus de textes.

Conclusion

En conclusion, bien que le paradigme langagier ait été abondamment mobilisé pour l'étude des objets architecturaux, peu de travaux se sont réellement intéressés à la matière même du texte architectural. Plus rares encore sont les recherches qui ont eu recours à des moyens informatiques pour aborder le discours. À cet égard, les recherches conduites au début des années soixante-dix par Alexander Tzonis apparaissent tout à fait singulières. Cela s'explique, en partie seulement, par un accès limité aux corpus. Le même phénomène peut d'ailleurs s'observer en histoire de l'art. C'est plutôt l'approche sémiologique qui mobilise les chercheurs puisqu'elle paraît plus à même de rendre compte de la spécificité de leurs objets d'étude. Si l'on observe une influence marquée du tournant linguistique en histoire de l'architecture, c'est principalement dans les approches critiques.

Étonnamment, le domaine de l'analyse automatique du discours reste très peu investi en architecture, alors que ce domaine de recherche pourrait fortement bénéficier de l'apport de la linguistique de corpus. Les campagnes de numérisation massive engagées depuis le milieu des années quatre-vingt-dix du vingtième siècle ainsi que le perfectionnement des techniques d'acquisition automatisée des textes mettent aujourd'hui à disposition des historiens de l'archi-

⁴⁸ C. PERELMAN, L. OLBRECHTS-TYTECA, *Traité de l'argumentation : la nouvelle rhétorique*, Bruxelles-Paris 2009 ; C. PERELMAN, *Rhétoriques*, Bruxelles 2012.

⁴⁹ S.E. TOULMIN, *Les usages de l'argumentation*, Paris 1993.

⁵⁰ J.C. MENSONIDES et al., *Analyse automatique d'arguments et apprentissage multi-tâches : un cas d'étude*, "Revue Ouverte d'Intelligence Artificielle", 3, 2022, 3-4, pp. 201-222, <https://imt-mines-ales.hal.science/hal-03638222/document> (consulté le 11 mars 2024).

⁵¹ VAN EEMEREN et al., *Handbook of Argumentation Theory*... cit., p. 20.

⁵² C.C. AGGARWAL, *Machine Learning for Text*, Cham 2018 ; PIOTROWSKI, *Natural Language Processing*... cit.

⁵³ L. MARTIN et al., *CamemBERT: a Tasty French Language Model*, in *ACL 2020. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, conference proceedings (Seattle, 5-10 July 2020), pp. 7203-7219, <https://inria.hal.science/hal-02889805v1> (consulté le 25 mars 2025) ; R. MILLIÈRE, *Language Models as Models of Language*, in *The Oxford Handbook of the Philosophy of Linguistics*, edited by R. Nefdt, G. Dupre, K. Stanton, Oxford, à paraître, <https://arxiv.org/abs/2408.07144> (consulté le 25 mars 2025).

⁵⁴ J. EISENSTEIN, *Introduction to Natural Language Processing*, Cambridge 2019 ; J. GRIMMER, M.E. ROBERTS, B.M. STEWART, *Text as Data: A New Framework for Machine Learning and the Social Sciences*, Princeton 2022.

tecture des corpus étendus pour l'étude du discours architectural. Il devient donc désormais possible de tirer parti des approches développées dans d'autres disciplines en analyse automatique du discours, que ce soit en littérature ou en histoire, dans le contexte des humanités numériques.

Sur la question de l'interdisciplinarité en histoire de l'architecture sur laquelle porte ce volume, précisons que les acteurs des humanités numériques ont le plus souvent évité de se qualifier de manière disciplinaire, parlant plus volontiers de la constitution d'un champ voire d'une méta-discipline. L'adoption de cette bannière avait un intérêt stratégique ; c'était en quelque sorte un slogan pour ouvrir les disciplines traditionnelles aux transformations engendrées par l'irruption des outils numériques. En fait, l'utilisation des moyens informatiques dans les sciences historiques ou dans les disciplines littéraires existait déjà de longue date sous l'appellation *computing humanities*. Dès l'immédiate après-guerre, les premiers travaux du père jésuite Roberto Busa avaient mobilisé le traitement automatique du texte, tout comme ceux de Jean-Claude Gardin quelques années plus tard. Toutefois, investir l'analyse automatique du discours en architecture c'est bel et bien pratiquer une double interdisciplinarité, puisqu'il s'agit de mobiliser tout à la fois l'apport de la linguistique de corpus et celui du traitement automatique de la langue qu'autorise l'informatique.

Cette interdisciplinarité ne constitue pas seulement une commodité méthodologique, elle est porteuse de conséquences épistémologiques importantes. D'abord, elle implique de s'intéresser au discours et à la langue plutôt qu'aux objets dessinés ou construits qui retiennent habituellement l'intérêt des historiens d'architecture. Ensuite, le transfert de ces technologies en histoire de l'architecture suppose des approches quantitatives et descriptives qui ne constituent pas seu-

lement des résultats, mais aussi des points de départ pour interroger les textes à partir d'un questionnement architectural. Si les textes ont des effets qu'il s'agit de révéler et de mettre à jour, il convient de ne pas oublier non plus le caractère construit des corpus sur lesquels on travaille ou des techniques utilisées. Enfin, en abordant la matière langagière des discours architecturaux, ce n'est pas seulement la langue de l'architecture qui doit nous intéresser. Comme l'ont montré de nombreux travaux sur le discours architectural, le sens d'un texte est toujours en négociation dynamique avec une pragmatique qu'on ne peut ignorer dans l'analyse.