AFSCET

Res-Systemica

Revue Française de Systémique

Fondée par Evelyne Andreewsky

Volume 25, Rochebrune 2022

Systèmes complexes ; théorie et pratiques

Res-Systemica, volume 25, article 11

Une modélisation catégorielle du débat numérique

Antsa Nasandratra Nirina Avo, Dominique Luzeaux, Jean Sallantin

19 pages

UNE MODÉLISATION CATÉGORIELLE DU DÉBAT NUMÉRIQUE

Antsa Nasandratra Nirina Avo[1], Dominique Luzeaux[2], & Jean Sallantin[3]

1. IST Ambositra, Madagascar, nirhina_avo@yahoo.fr
2. Agence Numérique de Défense, dominique.luzeaux@polytechnique.org
3. DR émérite LIRMM Montpellier, jean.sallantin@gmail.com

Résumé

Dans un débat numérique, l'indexation du contenu de chaque propos se fait par des termes du français, déterminés par une intelligence artificielle. Cette indexation est flexible et réflexive car renouvelée sous la supervision de tous les débattants. Certaines relations entre deux termes de cette indexation, dévoilent des propriétés caractéristiques du débat en cours qui suscitent une réaction des débattants, et relancent ainsi le débat de manière ciblée. Nous montrons comment un topos de Grothendieck modélise le débat numérique en définissant de manière cohérente ses éléments constitutifs, leurs traitements algorithmiques et les interventions des débattants.

Abstract
In a digital debate, the indexation of the content of each statement is done by French terms determined by an artificial intelligence. This indexation is flexible and reflexive because it is updated under supervision of all the debaters. Specific relations between two terms of this indexation reveal paradigms of the debate in progress, which generate a reaction of the debaters and thus relaunch the debate in a targeted way. We show how a Grothendieck topos models the digital debate by defining in a coherent way its constitutive elements, their algorithmic treatments, and the debaters' interventions.

Mots clés : Grothendieck topos, digital debate, artificial intelligence

Introduction

De nombreuses civilisations pratiquent une forme normée de débat pour fonder leur gouvernance. Par exemple, dans la civilisation grecque, le principe de l'isègoria, désignait l'égalité de droit à la parole des citoyens dans le débat public.

Nous avons des techniques numériques pour rendre effective une égalité de droit à la parole dans le débat public. Aussi la question est de faire du débat numérique le support d'un exercice d'intelligence collective.

Il y a des conditions préalables pour qu'un débat puisse servir à penser ensemble :

1) Un débat est indépendant des langues, du statut social des participants.
2) Tous les propos d'un débat – anciens comme récents – restent discutables.

3) Tout débattant apporte librement des opinions et des arguments.
4) L'indexation des propos est automatique et supervisée par tous les débattants.
5) La structuration du débat se ramène à celle des sous-ensembles de propos.
6) La progression du débat vient de l'apport de nouveaux arguments.

En français, il existe plus de trente millions de termes susceptibles d'indexer un propos tenu (Lafourcade 2007). La supervision de l'indexation se fera en observant comment l'indexation des propos structure par inclusion les sous-ensembles de propos et dégage du débat des conditions favorables à la formulation de propos originaux.

L'ensemble des calculs à faire pour indexer les propos est important et doit se faire automatiquement par une intelligence artificielle (IA). Mais la supervision par les débattants est indispensable pour améliorer la pertinence de l'indexation. Encore faut-il que cette supervision aboutisse à dégager quelques paradigmes intéressants pour relancer le débat.

Le rôle de la modélisation mathématique, comme celle que nous allons introduire, est ici de fonder la supervision du débat numérique sur une indexation de ses propos.

Plus précisément cette modélisation doit rendre compte de :
A) la modélisation de l'indexation de chaque propos par des termes ;
B) la modélisation de la structuration d'un débat dont les propos sont indexés ;
C) la modélisation de la supervision de l'indexation et de la structuration des propos.

Les modélisations des outils réalisant A) et B) posent que les termes servant à indexer sont ordonnés, et qu'il existe une correspondance appelée correspondance de Galois entre les sous-ensembles ordonnés de termes et les sous-ensembles ordonnés de propos.

La modélisation de la supervision de l'indexation C) se fait en supposant qu'il existe un objet final, noté **1**, qui est associé à tous les termes indexant un propos. Cela revient ici à permettre à tout débattant d'établir si un terme proposé pour indexer un propos est « à propos », « hors de propos » ou « manquant ».

La modélisation catégorielle que nous allons décrire est faite avec un topos de Grothendieck. Elle rend compte des modélisations A), B) et C).

Un premier essai de modélisation de l'argumentation numérique a été réalisé avec des topos élémentaires (Nirina Avo, 2020). Cette modélisation suppose que les opinions et les arguments sont des propos portant sur des propos. En effet le nouveau propos apporte une justification à un jugement sur une reformulation partielle d'un propos. Pour justifier d'un jugement, un débattant peut se servir de dispositifs de preuve. S'il ne s'en sert pas, il profère une opinion. S'il s'en sert, il profère un argument. Notons qu'un argument est pas prouvé et pas réfuté (remarque : pour éviter toute ambiguïté, nous préférons être plus précis sur le plan logique et moins sur celui de la langue française, *hic* et *infra*). Si la preuve ou la réfutation d'un argument se font par un seul dispositif de preuve, cet argument est qualifié d'incertain et sa justification portera sur l'estimation de son incertitude. Si le débattant se sert de plusieurs dispositifs de preuve, par exemple s'il est pas prouvé par un dispositif de preuve formelle et pas réfuté par un dispositif de preuve empirique, alors une hypostase désignera la nature attendue à la justification de l'argument (Sallantin, Nirina Avo, et al., 2019).

Par exemple, considérons le propos « planter un arbre a un intérêt écologique », qui utilise le terme T1 « planter un arbre » et T2« impact écologique », ainsi que la relation « T1=>T2 ». Des débattants peuvent en faire un argument du débat se justifiant comme une hypothèse, un problème, une définition, une théorie. Chaque fois, l'hypostase désigne la nature de la justification attendue de l'argument (Sallantin, Pinet, et al., 2019).

Ainsi dans le débat numérique, des techniques d'IA produisent une indexation de l'ensemble des propos, une structuration des sous-ensembles de propos, la mise en évidence de certains termes qui caractérisent des sous-ensembles indivisibles de propos, des relations d'implication entre certains termes. Ces dernières suggèrent parfois aux débattants des arguments nouveaux, car ils ne sont pas tirés d'un propos précédent, mais rendent compte de relations entre des sous-ensembles de propos.

La notion d'argumentation sera modélisée ici par un topos élémentaire, alors que celle du débat numérique sera faite par un topos de Grothendieck. La mise en correspondance de ces deux formalismes catégoriels se pose donc : on montre que tout topos de Grothendieck est un topos élémentaire (P.T. Johnstone, 1977), et que la topologie de Lawvere-Tierney généralise celle de topologie de Grothendieck à tout topos, mettant en relation les logiques correspondantes (P.T. Johnstone, 1977 ; Prouté, 2019).

Modéliser avec des catégories

L'objectif est de donner une formalisation des différentes notions introduites précédemment, sans rentrer dans les détails mathématiques, mais en introduisant de la manière la plus intuitive possible les définitions avec le vocabulaire technique approprié. Le lecteur curieux pourra ainsi se référer aux ouvrages de référence comme *Sheaves in geometry and logic: a first introduction to topos theory* de Saunders MacLane et Ieke Moerdijk, paru en 1992 aux éditions Springer (1992), ainsi que *Topos theory* de P.T. Johnstone, paru en 1977 aux éditions

Academic Press pour les détails. Pour réaliser une modélisation avec la théorie de catégorie, l'identification des objets et les relations ou morphismes sont des étapes essentielles. Connaître le sens de ces morphismes permet une meilleure compréhension de chaque objet de la catégorie, surtout ses objets particuliers (objet final, initial, exponentiel, etc.). Une formalisation catégorielle de l'argumentation numérique permet donc de comprendre les unités qui composent une argumentation, leur relation et l'importance de chaque élément caractéristique.

Des formalisations catégorielles de l'argumentation et de débat ont été mentionnées. Ces deux concernent une catégorie en particulier, qui est un **topos**.

La capacité de la théorie des catégories à faciliter la généralisation d'un système et fournir un langage commun, un ensemble d'unification de concepts est ici essentielle (Luzeaux, 2015). Deux modélisations sont utilisées ici, la modélisation par topos élémentaire et la modélisation par topos de Grothendieck.

La formalisation avec le **topos élémentaire** est longue, séquentielle, pour parvenir à définir la différence entre argument et opinion grâce aux hypostases (Nirina Avo, 2020). Mais elle est facile à comprendre dans la mesure où les correspondances entre les éléments du topos et l'argumentation dans un débat se font de manière relativement intuitive.

La **reformulation** d'un propos comme étant un passage nécessaire pour formuler une opinion ou un argument se modélise par l'opération d'exponentiation dans une catégorie : cette opération classique en informatique revient à identifier une fonction à *n* paramètres avec un ensemble de fonctions à *n-p* paramètres indexées par les autres *p* paramètres. La reformulation est ainsi une fonction paramétrée par la justification d'un jugement sur un propos.

En revanche, la formalisation avec un **topos de Grothendieck** est plus souple (on pourrait dire qu'elle est « *bottom-up* » alors que l'autre modélisation est « *top-down* »), et se construit avec une telle finesse que les éléments du débat se placent de façon évidente dans le formalisme. La modélisation en catégorie de faisceaux d'ensembles des termes dans le débat démontre de manière subtile l'importance de la place de termes et permet une indexation et une structuration des propos par des termes.

La modélisation catégorielle du débat

---

les termes

On se donne un ensemble d'éléments, appelés *termes*. On va munir cet ensemble d'une structure de catégorie **T**, c'est-à-dire que l'on peut définir des associations (appelées *morphismes de* **T**) entre termes (appelée *objets de* **T**), de telle sorte que chaque terme soit associé avec lui-même et que si un terme est associé avec un deuxième terme qui lui-même est associé avec un troisième, alors le premier terme est associé avec le troisième (dit autrement, l'association est une opération *associative* !).

Pour avoir une représentation intuitive de la catégorie **T**, il suffit d'imaginer un graphe quelconque dont les sommets seraient des termes et les arêtes des associations.

---

les propos

Informellement, un propos est un fragment de texte, et il est indexé par une liste finie de termes : on peut voir cela comme une fonction multivoque qui associe soit des termes à un propos, soit des propos à un terme. Ce qui nous intéresse donc est cette association entre un terme et un ensemble d'éléments, qui ontologiquement ne nous intéressent pas si ce n'est par l'existence de cette association. C'est donc ceci qui va définir ce que nous appellerons un propos.

Un propos est une application de **T** dans **Set**, la catégorie des ensembles (ses objets sont des ensembles, ses morphismes sont des fonctions entre ensembles). Ceci permet alors de définir **P**, la catégorie des propos, comme étant **Set**^(**T**^*op*), où ^ est l'opérateur d'exponentiation, et *op* est une subtilité mathématique pour faciliter la suite de l'exposé (mathématiquement, on a inversé le sens des morphismes entre les termes, ce qui ne change rien épistémologiquement, comme la relation d'association n'est pas fondamentalement dirigée).

On définit un morphisme entre deux propos q et p, quand le sous-ensemble de termes associé au propos p est inclus dans celui associé à q : on dira que le propos q est un successeur au propos p. Cela permet de définir une relation d'ordre entre p et q, $p<q$, si q est un successeur de p.

La catégorie des propos est **Set**^(**T**^op) souvent appelée catégorie des *préfaisceaux d'ensembles* construite sur **T**. C'est une structure mathématique bien connue qui a de très bonnes propriétés mathématiques : c'est un *topos de Grothendieck*. On peut donc en particulier définir des intersections de propos, mais aussi des réunions de propos ; on peut aussi définir ce que l'on appelle l'exponentielle de propos, ce qui revient à être capable de définir un propos qui représente une application associant un propos à un autre (l'idée est la même qu'en informatique où une fonction à deux variables peut être vue comme une fonction à une variable et paramétrée par l'autre : $f(x,y)\equiv f_x(y)$). On peut déjà remarquer les propriétés remarquables de la catégorie des propos, telle que définie, qui permet d'internaliser les opérations d'intersection, de réunion, d'exponentiation, y compris pour un nombre quelconque d'opérandes.

Il est alors possible d'utiliser un résultat classique de théorie des catégories, le *lemme de Yoneda*, qui permet de plonger pleinement (on a bien accès à toute l'information) et fidèlement (on ne perd pas de l'information) toute catégorie dans la catégorie des préfaisceaux d'ensembles construite sur elle-même.

Par rapport à ce qui précède, cela revient à pouvoir identifier un propos à tous ses successeurs possibles, tout en gardant toutes les bonnes priorités mathématiques qui vont nous intéresser dans la suite (mathématiquement on identifie un objet *p* à tous les objets pouvant être mis en relation avec *p*, à savoir *Hom(_,p)* : c'est pour cela que les morphismes ont été définis à partir de la notion de successeur). Ceci n'est pas choquant dans la vision synthétique adoptée ici, puisque la catégorie des propos rassemble globalement tous les propos potentiels pouvant être indicés à partir des termes : ce ne sera qu'en lui rajoutant des propriétés particulières qu'il sera possible de distinguer des structures parmi tous ces propos potentiels.

Un topos de Grothendieck peut aussi être vu comme un treillis de Heyting (en fait c'est le treillis des sous-objets de l'objet terminal du topos), ce qui permet d'interpréter alors l'intersection comme une conjonction, la réunion comme une disjonction, l'exponentiation comme une implication.

Les notions d'intersection (respectivement réunion, exponentiation) de propos s'interprètent alors comme la conjonction (respectivement disjonction, implication), ce qui donne un sens formel aux acceptions habituelles.

---

les agents intervenant dans le débat

Précédemment, on a défini une structure globale incluant tous les propos potentiels. Maintenant on va identifier des sous-structures : cela va permettre de définir pour un agent les propos qu'il est susceptible de tenir potentiellement, de les distinguer de ceux d'un autre agent, et également de définir des logiques associées à chaque agent permettant de passer des propos aux arguments, et formaliser la notion d'hypostase.

On va définir des *topologies de Grothendieck* sur la catégorie des propos, ce qui revient à choisir pour chaque propos *p* un sous-ensemble de ses successeurs *J(p)* (*J* est appelé une famille couvrante ou recouvrement, et *J(p)* un *crible*), qui vérifie certaines propriétés. Sans rentrer dans les détails mathématiques, les idées sous-jacentes sont qu'un recouvrement se doit d'être recouvert par lui-même, qu'un recouvrement se doit de recouvrir également les recouvrements plus petits, et enfin que pour être un recouvrement, il suffit d'en être un sur chaque morceau d'un autre recouvrement (ces deux dernières propriétés correspondent mieux à l'intuition véhiculée par la traduction anglo-saxonne « *sieve* » d'un crible, qui signifie aussi « tamis »).

L'intérêt de la topologie de Grothendieck est que si on considère ensuite les faisceaux (qui sont des préfaisceaux recollés localement là où ils se superposent) construits sur une catégorie munie d'une topologie de Grothendieck (appelée alors *site*), on obtient un topos de Grothendieck qui est immergé dans la catégorie des préfaisceaux construite sur la catégorie (Grothendieck & Dieudonné, 1971).

Dit autrement, si un agent A_1 a une topologie de Grothendieck T_1 qui lui permet de définir « ses » successeurs d'un propos (ce sont les propos qu'il est susceptible de tenir), l'ensemble de tous ces propos a encore de très bonnes propriétés mathématiques comme précédemment, permettant notamment de définir la conjonction et la réunion de propos pour A_1 (attention ce ne sont pas forcément les mêmes conjonctions et réunions que dans **P**, car A_1 n'a pas accès à tous les propos potentiels mais seulement aux siens).

la reformulation de propos

Si on se donne maintenant un autre agent A_2 avec une topologie de Grothendieck T_2, il se peut que le propos p formulé par A_1 ne soit pas immédiatement compréhensible par A_2 (car pas dans la sous-structure associée à l'agent A_2). Il est alors nécessaire à A_1 de *reformuler* son propos.

Or, on peut montrer que l'ensemble des topologies de Grothendieck peut être muni d'une structure de treillis, ce qui permet de définir en particulier l'intersection de deux topologies données. Il suffit donc de prendre dans les successeurs de p le plus petit qui est dans l'intersection de T_1 et de T_2.

De même, si q est un propos formulé par A_2 suite à un propos p formulé par A_1, q peut être reformulé pour A_1 comme le plus petit des successeurs de p (au sens de T_1) tel que son intersection avec q soit toujours celle de p et q (au sens de l'intersection de T_1 et T_2). Si l'on n'avait pas les restrictions liées aux topologies propres à chaque agent, donc si on raisonnait simplement dans la catégorie **P**, alors cette reformulation serait tout simplement l'implication *q implique p*, c'est-à-dire l'exponentiation au sens catégoriel.

---

les arguments

La formalisation présentée va trouver son intérêt dans le fait qu'elle permet de faire de l'algèbre, de la topologie et de la logique en même temps : au prix d'une complication initiale, elle offre en fait du coup le champ permettant d'unifier toutes ces problématiques de manière naturelle ! L'algèbre et la topologie ont été rapidement évoquées précédemment. En ce qui concerne la logique, il convient de rappeler que l'on peut associer à tout topos une logique interne intuitionniste du premier ordre, et réciproquement on peut associer canoniquement un topos à toute logique intuitionniste du premier ordre.

En effet, il y a bijection entre une topologie de Grothendieck sur **T** et ce que l'on appelle une *topologie de Lawvere-Tierney* (ou aussi opérateur local ou *modalité locale*) sur **Set**^(**T**^*op*), qui est un opérateur tel que : $x<y$ implique $j(x)<j(y)$; $x<j(x)$; $j(j(x))=j(x)$; $j(x\cap y)=j(x)\cap j(y)$. L'appellation de modalité locale n'est pas fortuite, car d'une part les propriétés rappelées ci-dessous sont en fait analogues à celles qu'a un opérateur de possibilité ◊ dans une logique modale, d'autre part tout topos a un objet particulier, noté classiquement **Ω**, qui correspond aux valeurs de vérité quand on munit un topos de sa logique interne, et l'opérateur j est défini de **Ω** dans **Ω** et correspond donc à un opérateur logique.

Via cette bijection, on peut donc définir une modalité $\Diamond_1$ (respectivement $\Diamond_2$) pour l'agent A_1 (respectivement A_2). Ceci permet donc pour tout agent d'appliquer cet opérateur modal sur les propos qu'il formule, ce qui permet de passer d'un propos à un *argument.* Par ailleurs, cela permet, quand on a un propos p et 2 agents A_1 et A_2 de définir les formules du type : $\Diamond_1 p$ et $\Diamond_2$ non p, c'est-à-dire les *hypostases*. On retrouve aussi l'incertitude qui a été évoquée dans l'introduction.

Conclusion

Le débat public est posé par les Athéniens comme étant un fondement de la démocratie. Le débat numérique ici consiste à se servir des développements actuels du numérique pour donner à tous citoyens un droit égal à la parole publique. L'utilisation de l'intelligence artificielle est nécessaire pour indexer le contenu de chaque propos tout en laissant au citoyen la faculté de superviser cette indexation et de s'en inspirer pour relancer l'argumentation du débat.

La modélisation catégorielle du débat est indispensable pour donner une base solide à cette pratique qui met en jeu différents algorithmes d'IA, et qui doit donner, aux débattants observant l'ontologie en cours de construction, un regard sur le déroulement du débat, ces zones stériles, fécondes et confuses, et une incitation à y intervenir par ses propres propos.

Références

Grothendieck, A., & Dieudonné, J. (1971). *Éléments de géométrie algébrique* (Vol. 166). Springer Verlag.

Lafourcade, M. (2007). Making people play for lexical acquisition. Symposium on Natural Language Processing.

Luzeaux, D. (2015). A Formal Foundation of Systems Engineering. *In* F. Boulanger, D. Krob, G. Morel, & J.-C. Roussel, *Complex Systems Design & Management*, Springer Verlag.

MacLane, S., & Moerdijk, I. (1992). Sheaves in geometry and logic: A first introduction to topos theory. Springer Science & Business Media.

Nirina Avo, A. N. (2020). Modélisation mathématique de l'argumentation numérique. Thèse Université de Fianarantsoa, Madagascar.

Johnstone, P.T. (1977). *Topos theory*, Academic Press.

Prouté, A. (2019). *Introduction à la logique catégorique*, cours http://163.172.10.123:8080/cours_2010.pdf