Tribune *(-La-tribune-des-mathematiciens-.html)*

Mathématiques et langages : le feuilleton de l'automne

LA LANGUE COMMUNE AU CŒUR DU RAISONNEMENT, NOTAMMENT MATHÉMATIQUE

Le *27 octobre 2017* - Ecrit par Christian Retoré *(_Retore-Christian_.html)*

La connexion entre mathématiques et langage est en plein renouveau, notamment en sciences cognitives, en didactique des mathématiques ou encore dans certaines applications de traitement automatique des langues, comme l'analyse automatique de l'argumentation ou l'inférence textuelle — une phrase est-elle conséquence d'un texte ? Cette interaction entre mathématiques et langage naturel, très liée à la logique, bénéficie depuis l'Antiquité aux deux domaines qu'elle met en relation. D'une part, on raisonne et on fait des mathématiques dans la langue commune, en classe comme entre mathématiciens. D'autre part les phrases ont une structure qui peut être décrite mathématiquement : arbres représentant la structure grammaticale d'une phrase, formules logiques représentant son sens, les deux structures étant intimement liées. Certains aspects de cette connexion sont étudiés depuis

l'Antiquité, d'autres sont tout récents.

À la mémoire Vladimir Voevodsky (4 juin 1966—30 septembre 2017).
Je connaissais peu la personne
— un exposé suivi du diner de gala du Logic Colloquium à Evora en 2013 —
mais un peu son travail remarquable sur la théorie homotopique des types.
J'étais justement en train de mettre en forme le présent article
dans lequel je mentionne son nom quand j'ai appris sa disparition prématurée.

Raisonnement et preuves, phrases et formules

Le raisonnement mathématique, dont la forme aboutie est la démonstration mathématique — plus souvent appelée preuve de nos jours — a servi de canon à Aristote pour poser les bases de la logique. Les règles de déduction définissant la logique étaient supposées apporter aux autres sciences, voire à toutes les sciences, la rigueur des mathématiques, qui avaient extrait des principes de raisonnement « évidents ». Ces principes étaient formulés comme des règles de déduction portant sur des énoncés en langage naturel, mais ces derniers n'étaient pas libres, ils avaient une forme figée (aujourd'hui nous parlerions d'un langage « contrôlé », comme dans *Natural Logic* dont il est question ci-après) :

- **A** : Proposition universelle affirmative, *Tout S est P*, par exemple *Tout carré est positif* ;
- **I** : Proposition particulière affirmative *Certains S sont P*, par exemple *Certains entiers sont premiers* ;
- **E** : Proposition universelle négative, *Aucun S n'est P*, par exemple *Aucun carré n'est négatif* ;
- **O** : Proposition particulière négative, *Tous les S ne sont pas P*, par exemple, *Tous les entiers ne sont pas premiers* [1 ***(#nb1)***].

Tout comme les énoncés A I E O, les syllogismes sont formulés en langage plus ou moins naturel (et un exemple) :

1re prémisse E	Aucun S n'est Q	Aucun multiple de 4 n'est premier.
2de prémisse A	Tout P est S	Tout multiple de 12 est un multiple de 4.
conclusion E	Aucun P n'est Q.	Aucun multiple de 12 n'est premier.

Un exemple de règle de déduction ***(javascript:;)***

La nature des formules logiques (A, E, I, O) figurant dans le syllogisme (E A E) a permis aux logiciens scolastiques qui ont prolongé les études aristotéliciennes (4 ***(#KK86)***) de recenser, de classer et de nommer les syllogismes par un mot latin dont ce soient les

voyelles, ici C**E**L**A**R**E**NT.

En étendant à d'autres disciplines le raisonnement mathématique, une particularité importante des mathématiques est occultée : sur quels faits ou principes construire notre raisonnement extra-mathématique ? Les axiomes mathématiques sont connus des mathématiciens et sont les mêmes pour toute la communauté concernée ; les principes des sciences empiriques sont davantage sujet à discussion ; les postulats des sciences humaines sont plus difficiles à identifier — et ceux qui sous-tendent un raisonnement de la vie courante semblent hors d'atteinte.

Un raisonnement peut être mené en solitaire, mais il prend une autre dimension lorsqu'il est utilisé comme un argument, c'est-à-dire lorsque quelqu'un cherche à prouver le bien-fondé d'une affirmation mais aussi à convaincre autrui. L'échange d'arguments entre personnes ayant des opinions différentes conduit au débat.

Y a-t-il une argumentation en mathématique ? Peut-il y avoir débat mathématique [2 *(#nb2)*] ? Il convient de distinguer d'une part l'enseignement des mathématiques ou l'exposition d'un résultat établi et, d'autre part, la pratique des mathématiques, la résolution de problèmes ou la recherche proprement dite. Tout enseignant sait que l'élève qui aura le mieux compris est celui qui conteste la preuve ou le résultat, et qui en débat avec l'enseignant. Mais s'agit il vraiment d'un débat ? L'élève se doute bien que, par exemple, la preuve du théorème des valeurs intermédiaires, enseignée depuis des décennies ou des siècles est correcte ; lors d'un séminaire, la plupart des auditeurs sont *a priori* convaincus de la justesse du résultat. Ce genre de débat est donc un faux débat, un peu comme une question rhétorique est une question dont l'auteur connaît la réponse. L'issue d'un tel (faux) débat est (sauf exception) connue d'avance, et le débat vise simplement à faire le détail de la preuve, à expliquer davantage, et *in fine* convaincre celui qui ne demandait qu'à être convaincu. De plus, dans ce cadre, seuls les arguments mathématiques peuvent être utilisés : dire que Bolzano était davantage un philosophe qu'un mathématicien (attaque *ad hominem*), ou que Brouwer remettait en cause ce résultat (argument d'autorité) ne va pas changer le cours de ce faux débat. La situation est bien différente lorsque deux mathématiciens cherchent à prouver ou a réfuter un énoncé mathématique devant un tableau. Si l'un et l'autre ont un point de vue opposé sur la véracité de l'énoncé, il y a un réel débat, et celui qui convaincra l'autre à l'issue d'une discussion amicale, mais parfois tendue, aura sans doute raison et établi son résultat. Lors de cette discussion les arguments non mathématiques ont leur place. Un argument d'autorité, comme « *J'ai croisé Voevodsky au Logic Colloquium, je lui ai parlé de notre conjecture, il pense qu'elle est fausse.* », va sans doute influer sur cours du débat — sans toutefois en changer l'issue ultime. On notera qu'il y a des situations intermédiaires entre l'enseignement et la discussion mathématique, comme par exemple une conférence sur un résultat que l'orateur souhaite vérifier en le confrontant à l'opinion des ses pairs.

Les mathématiques, l'exposition de résultats établis, tout comme la discussion qui conduira (peut-être) à une nouvelle preuve, se font en français, en anglais, en chinois...

alors que les règles de déduction, garantes d'un raisonnement correct portent sur des formules logiques. Même si quelques formules sont écrites au tableau lors de la discussion, les articulations logiques de la preuve en train de se faire, ou de l'exposé mathématique, qu'elles soient écrites ou orales, le sont dans la langue courante. Une exception importante est Vladimir Voevodsky *(http://fr.wikipedia.org/wiki/Vladimir_Voevodsky)* dont certains articles (comme par exemple *A preliminary univalent formalization of the p-adic numbers* par Álvaro Pelayo, Vladimir Voevodsky, Michael A. Warren, prépublication *(https://arxiv.org/pdf/1302.1207.pdf)*) sur la théorie homotopique des types (voir par exemple cet article *(A-la-croisee-des-fondements-des-mathematiques-de-l-informatique-et-de-la.html)* d'Antoine Chambert-Loir) se terminent par les preuves formelles de l'article écrites avec l'assistant de preuve Coq (sur cet assistant de preuve, voir par exemple cet article *(Coq-et-caracteres.html)* de Jérôme Germoni).

Ainsi, la langue est plus agréable et efficace pour véhiculer la pensée mathématique que de longues formules logiques truffées de parenthèses et de symboles à la *Principia Mathematica* *(http://fr.wikipedia.org/wiki/Principia_Mathematica)*.

Cette souplesse va hélas de pair avec une ambiguïté endémique, problématique pour l'enseignement, comme signalé ici *(Un-langage-necessaire-pour-raisonner-et-prouver-en-mathematiques.html)* par Denise Grenier et là *(Sur-l-impact-de-la-structure-grammaticale-des-langues-dans-l-enseignement-et-l.html)* par Viviane Durand-Guerrier. Dans la discussion mathématique ou la lecture d'un texte mathématique, l'interlocuteur mathématicien comprend généralement de la bonne manière, et sinon, il s'en rend compte, il lui est aisé de se faire préciser tel ou tel passage par quelqu'un, et il osera le lui demander. En revanche l'élève ayant compris un énoncé autrement que ce qui était attendu ne s'en rendra pas forcément compte et, s'il perçoit un problème, osera-t-il en parler ?

L'ambiguïté peut se voir comme le fait qu'une phrase soit associée à plusieurs formules logiques, alors que nous aimerions qu'il n'y en ait qu'une seule (l'unique formule perçue par celui qui comprend bien). En mathématiques l'ambiguïté lexicale n'est guère problématique : certes des mots comme *dérivée*, *régulier* ou *normal* peuvent avoir des sens différents suivant le contexte mathématique (fonction dérivée, catégorie dérivée ; vecteur normal, réel normal, sous-groupe normal ; point régulier, polyèdre régulier) mais peu d'erreurs de compréhension sont imputables à ce type d'ambiguïté. En revanche, les ambiguïtés de portée des quantificateurs, de portée de la négation, l'ambiguïté de la référence d'un pronom sont de fâcheuses causes d'incompréhension, surtout pour celui qui a du mal à suivre comme par exemple un·e élève en difficulté.

1. Un bus ramènera chacun des invités.
2. Un entier divise chacun des nombres de la liste.
3. Un point est incident à chaque segment.
4. Un garde se tient devant chaque porte du musée.

Le présent article va donc détailler le passage d'une phrase ou de quelques phrases à une ou des formules logiques, qui permettent de raisonner correctement, et en particulier à l'expression des opérations logiques dans la langue. Nous essaierons de traiter en parallèle le raisonnement courant et le raisonnement mathématique en montrant les spécificités de ce dernier.

Phrases & formules

Syntaxe et grammaire

Un aspect récent de cette connexion est l'étude mathématique de la structure grammaticale des phrases, les arbres syntaxiques, connue en informatique sous le nom de théorie des langages formels *(http://fr.wikipedia.org/wiki/langages_formels)*. Cette théorie, venue de la linguistique, sert aussi en biologie pour la génomique et en mathématiques dans la théorie des groupes (notamment profinis) : un bel exemple d'interaction entre mathématiques et linguistique. Même si nous souhaitons ne pas trop parler de cet aspect évoqué dans cette série d'articles, notamment dans celui *(Calculer-la-langue-un-enjeu-interdisciplinaire.html)* de Maxime Amblard, il faut tout de même en parler en peu, car la structure syntaxique d'une phrase est intimement liée à son sens. Un exemple trivial d'importance de la syntaxe est que « *Pierre aime Marie.* » et « *Marie aime Pierre* » ne sont pas équivalents, même si les personnes sentimentales le souhaiteraient.

Nous supposerons que la structure grammaticale est connue sans trop nous préoccuper de savoir comment elle a été calculée. Il y a grosso modo deux familles de descriptions syntaxiques et toutes deux décrivent les phrases par des arbres :

1. La grammaire générative *(https://fr.wikipedia.org/wiki/Grammaire_générative_et_transformationnelle)* (aussi appelée transformationnelle) utilise des arbres syntagmatiques dont les feuilles portent les mots qui suivent l'ordre gauche droite des feuilles, les nœuds internes donnant la nature grammaticale du sous arbre qu'il engendre (groupe nominal, groupe verbal, groupe prépositionnel). Bien qu'issues de considérations linguistiques, ces grammaires ont donné lieu à une théorie riche en informatique, grammaires et langage formels *(https://fr.wikipedia.org/wiki/Grammaire_formelle)* (théorie initiée par Chomsky et Schützenberger et développée à leur suite notamment par Maurice Nivat, disparu tout récemment), tandis que d'un point de vue linguistique, la relation très étroite entre structure grammaticale et ordre des mots oblige à des complications formelles : transformation d'arbres par déplacements de sous-arbres, éléments vides, etc. On notera que ce genre de grammaire a été inventé pour l'anglais, langue sans cas et donc dans laquelle l'ordre des mots est assez strict.

2. La grammaire de dépendance *(http://fr.wikipedia.org/wiki/grammaire_de_d%C3%A9pendance)* utilise des arbres de dépendances dont les nœuds sont des mots qui sont reliés entre eux par des arcs étiquetés par des relations syntaxiques, l'ordre des mots dans la phrase étant une donnée indépendante de la structure de

l'arbre. La relation entre ordre des mots et arbre syntaxique étant souple, elle nuit aux qualités mathématiques de ce modèle, dont l'étude mathématique n'a pas eu le même succès que celui de la grammaire générative, par manque de travaux ou par manque de propriétés formelles. On notera que cette famille de grammaires a été développée pour le russe, langue avec cas, et donc dans laquelle l'ordre des mots est relativement libre.

(cliquer pour agrandir)

Analyse syntaxique

Analyse syntaxique de « si un nombre premier divise le carré d'un nombre alors il divise ce nombre » avec une grammaire catégorielle qui étend le calcul de Lambek, acquise sur corpus — plate-forme Grail, Richard Moot *(https://github.com/RichardMoot)*.

Le calcul de Lambek *(javascript:;)*

Les deux sens de « sens »

Parlons maintenant de sémantique, c'est-à-dire du sens des énoncés en langage naturel. Il y a deux aspects du sens, comme l'illustre l'exemple que voici. Si vous vous posez la question suivante : *Geach était-il l'élève de Wittgenstein ?* votre moteur de recherche préféré appliqué à *Geach*, *Wittgenstein* et *élève* fournit l'article de Wikipedia sur Peter Geach *(http://fr.wikipedia.org/wiki/Peter_Geach)* qui nous dit ceci :

> En 1941, il épousa la philosophe Elizabeth Anscombe, grâce à laquelle il entra en contact avec Ludwig Wittgenstein. Bien qu'il n'ait jamais suivi l'enseignement académique de ce dernier, cependant il en éprouve fortement l'influence.

Est-il possible de répondre automatiquement à la question *Geach était-il l'élève de Wittgenstein ?* à partir du paragraphe trouvé sur Wikipedia ? Hélas non, car il manque des informations sur la structure grammaticale, et sur la structure logique qui y est intimement liée : comme chacun peut le constater sur cet exemple, il manque notamment les (co)références des pronoms (7 sur 47 mots), la négation et sa portée (« jamais ») : *Bien qu'il n'ait jamais suivi l'enseignement académique de ce dernier*. Cette dernière proposition permet de conclure que Geach n'a pas été l'élève de Wittgenstein car « Bien que P » présuppose P – cela n'aurait pas été possible avec d'autres expressions comme « Supposons que P », qui ne présuppose pas P.

Sens logique et compositionnalité

Avant de revenir sur le sens des mots et les relations entre les différents sens des différents mots, penchons-nous sur les aspects structurels, syntaxiques et sémantiques du sens. La structure syntaxique évoquée ci-dessus constitue l'ossature de la structure logique de la phrase, elle permet notamment de voir ce que la phrase analysée affirme, nie, suppose... Les propositions sont organisées entre elles suivant des relations logiques : *ou*, *et*, *non*, *si... (alors)....* Les quantificateurs de la langue sont plus variés que ceux des mathématiques, à savoir *pour tout* (∀) et *il existe* (∃) : *20 %, au moins 5, peu, beaucoup, la majorité, la plupart...*

Le sens logique d'une phrase, ou plus généralement celui d'une expression complexe, résulte de la composition des sens des expressions qui la compose, et donc *in fine* du sens des unités les plus simples, à savoir les mots. Ainsi le sens de « *les malades alités attendent un bon médecin* » s'obtient-il à partir de la structure grammaticale de la phrase et du sens des mots qui la composent.

Le sens des mots a, en plus des relations lexicales dont nous reparlerons ci-dessous, un aspect logique connu depuis l'Antiquité. Un nom commun, un verbe intransitif ou un groupe verbal est un prédicat unaire (une propriété) qui peut être vrai ou faux suivant l'argument auquel il est appliqué : $\mathbf{malade}(x)$, $\mathbf{médecin}(x)$, $\mathbf{attendre_un_bon_médecin}(x)$. Un verbe transitif (comme *attendre*) correspond à un prédicat binaire ($\mathbf{attendre}(x,y)$), vrai ou faux suivant les valeurs respectives de ses arguments (x et y). Un adjectif comme *alité* est aussi une propriété : être un *malade alité*, c'est être un *malade* et être *alité*. En revanche être un *bon médecin* ce n'est pas être un *médecin* et être *bon* : bon est une propriété de propriété, qui vient modifier la propriété d'être médecin.

Les quantificateurs comme « *chaque* », « *tous les* », « *les* », « *tout* » (« *pour tout* » en mathématiques) ou « *un* », « *des* », « *quelques* », « *certains* » (« *il existe* » en mathématiques) sont plus complexes : à partir d'un nom commun (une propriété N) et du reste de la phrase (une autre propriété Q), un quantificateur fabrique une proposition qui est vraie ou fausse, « *les [N : malades alités] [Q : attendent le Dr House]* », ou « *[Q : Pierre attend] un [N : bon médecin]* ». Lorsque les deux quantificateurs sont en compétition, il y a ambiguïté : « *les malades alités attendent un bon médecin* » peut être compris de deux manières (*tous les malades alités [par exemple dans un hôpital donné] attendent tous le même médecin, lequel est un bon médecin*, ou *chaque malade alité [par exemple d'une ville donnée] verra un bon médecin, par forcément le même*). Percevoir la différence entre ces deux sens, et le cas échéant savoir reconnaître lequel convient en situation est une difficulté majeure dans l'enseignement des mathématiques. L'ambiguïté est encore plus forte et la compréhension plus difficile lorsque les quantificateurs interagissent avec la négation : « *Tous les malades n'apprécient pas les bons médecins* ».

Sémantique de Montague *(javascript:;)*

Structure logique, pronoms et référents

La compositionnalité est parfois mise en défaut par les pronoms. Dans des phrases, mathématiques ou non, appelées *donkey sentences* cela pose problème :

5.

a. Si un fermier possède un âne alors il le bat [3 *(#nb3)*].

b. Si un nombre premier divise le carré d'un nombre alors il divise ce nombre [4 *(#nb4)*].

Ces phrases sont de la forme « *si* P *alors* Q » mais si le sens était strictement compositionnel la formule logique associée devrait être $[P] \implies [Q]$ où $[P]$ et $[Q]$ désignent respectivement les formes logiques de P et Q. Essayons :

6.

a. $[\exists p\, \exists a\, \mathrm{ane}(a)\ \&\ \mathrm{paysan}(p)\ \&\ \mathrm{possede}(p,a)] \implies \mathrm{bat}(p,a)$.

b. $(\exists p\, \exists a\, \mathrm{premier}(p)\ \&\ p \mid a^2) \implies p \mid a$.

Les variables p et a sont libres dans la deuxième partie de l'implication, on peut aussi bien dire les deux absurdités suivantes :

7.

a. $[\exists q\, \exists i\, \mathrm{ane}(i)\ \&\ \mathrm{paysan}(q)\ \&\ \mathrm{possede}(q,i)] \implies \mathrm{bat}(p,a)$.

b. $(\exists q\, \exists i\, \mathrm{premier}(q)\ \&\ q \mid i^2) \implies p \mid a$.

Il faut donc sortir les quantificateurs existentiels de l'antécédent de l'implication et ceux-ci deviennent des quantificateurs universels dans cette transformation. Ces quantificateurs universels prennent dans leur portée les variables/pronoms de la conséquence de l'implication. Les formules correctes associées à ces deux phrases sont donc :

8.

a. $\forall p\, \forall a\, [[\mathrm{ane}(a)\ \&\ \mathrm{paysan}(p)\ \&\ \mathrm{possede}(p,a)] \implies \mathrm{bat}(p,a)]$.

b. $\forall p\, \forall a\, \left[\mathrm{premier}(p)\ \&\ p \mid a^2\right] \implies p \mid a$.

On notera que la phrase sur les nombres serait moins difficile à comprendre si des variables nommaient et représentaient les nombres, comme cela est courant dans un texte mathématique :

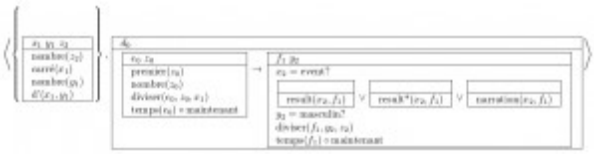
9. Si un nombre premier p divise q^2, le carré d'un nombre q, alors p divise q.

Pourtant cette phrase figure sans variables dans un manuel scolaire. Pourquoi la phrase avec l'âne est-elle moins ambiguë que celle avec les nombres alors qu'elles ont la même structure syntaxique ? Tout simplement parce que notre connaissance du monde, des paysans et des ânes nous dit qui bat qui — mais ce genre d'analyse du

sens est bien difficile à automatiser.

Cette digression sur un phénomène non compositionnel ne doit pas nous faire oublier que le sens est bien, en général, compositionnel.

(cliquer pour agrandir)



Analyse sémantique

Analyse sémantique (DRS, structure de représentation discursive) (Richard Moot) de « si un nombre premier divise le carré d'un nombre alors il divise ce nombre ». Il s'agit grosso modo d'une formule de la logique du premier ordre (on notera l'implication $\longrightarrow$), avec en tête des boîtes des référents de discours, et les variables figurant dans le conséquent de l'implication peuvent accéder aux référents de l'antécédent de l'implication. Cette analyse sémantique a été réalisée en lambda-DRT à partir de l'analyse syntaxique précédente avec la plate-forme Grail, Richard Moot *(https://github.com/RichardMoot)*.

La théorie des représentations discursives *(javascript:;)*

Intégration du sens lexical

Le sens lexical n'est pas pris en compte dans la formule logique construite. De manière plus intuitive cela signifie que les mots sont des prédicats ou des prédicats de prédicats sans aucun lien entre eux. Pourtant il y a des liens évidents : *dormir* ne s'applique qu'à des (mots désignant des) animaux, on ne dérive que des (mots désignant des) fonctions (ou des catégories, mais c'est nettement plus compliqué). Le fait que les arguments appartiennent à la bonne sorte d'entités, soient du bon type, aide grandement à la reconstruction du sens d'une phrase.

Cependant il est possible, dans certains contextes, d'outrepasser ces règles de cohérence ? Ne dit-on pas qu'un ordinateur « *est en veille, dort, se réveille* » ? Ne dit-on pas qu'une *fonction est comprise entre* 1 *et* 2 pour signifier que ses *valeurs* le sont ? — si 1 et 2 désignent les fonctions constantes, il peut s'agir de l'ordre partiel entre fonctions. Il est également possible de dire que « *le sergent a aboyé après la nouvelle recrue* », avec un sens figuré d'*aboyer*.

On notera que les différents sens d'un même mot influent réellement sur le sens logique d'une phrase. Un livre est tout à la fois un contenu informationnel et un objet matériel sur lequel sont imprimées ces informations. Certaines situations font

référence à un aspect, d'autres à un autre. Lire, écrire, mais aussi imprimer, relier, et tout ce qu'un objet physique peut subir : ranger perdre, brûler, et même manger (« *La chèvre a dévoré le livre de Villani.* »). Une telle phrase prend un tout autre sens lorsqu'on dit : « *Mon amie Claire a dévoré le livre de Villani.* » Dans ce dernier cas, le prédicat « *dévorer* » adapte son sens à « *livre* » et à « *Claire* ».

(cliquer pour agrandir)

Réseau lexical Rezo
(http://www.jeuxdemots.org/jdm-accueil.php)
(extrait)

Un extrait du réseau lexical Rezo de JeuxDeMots ***(http://www.jeuxdemots.org/jdm-accueil.php)*** (Mathieu Lafourcade) construit en faisant jouer les internautes à des Jeux de Mots. Ici les mots les actions (sous forme de verbes) que l'on peut faire avec un livre — les 7 les plus probables, les arêtes de ce graphe sont pondérées. Rezo contient 162.535.567 relations de 70 sortes entre 2.102.984 expressions.

Supposons qu'une étagère contienne trois exemplaires d'*Algèbre géométrique* d'Artin, et deux exemplaires de *Topologie* de Choquet. Si quelqu'un dit « *J'ai monté ces livres au grenier, je les connaissais par cœur.* », combien maîtrise-t-il de livres ? — Deux. Combien en a-t-il porté au grenier ? — Cinq. L'aspect de l'objet intervient donc sur son identification et par suite sur son comptage et sa quantification.

Les glissements de sens des mots dépendent de la nature ontologique de ce qu'ils représentent, mais aussi des mots et de la langue. Par exemple « *classe* » et « *promotion* » désignent tous deux un groupe d'élèves ; même si les « *promotions* » sont en bijection avec les salles il est impossible de dire « *la promotion de 2015 a été repeinte durant les vacances de Noël* », tandis qu'il est possible de dire que « *la classe des CE1 a été repeinte durant les vacances de Noël* ». Les glissements de sens

dépendent donc des mots utilisés, même si leur nature ontologique est identique. Elles dépendent donc également de la langue : « *Je suis venu à pieds, car mon vélo est crevé.* » passe bien, mais en anglais c'est épouvantable : « *I walked here because my bike is punctured.* »

Les relations entre sens des mots permettent non seulement de trouver quel sens convient au contexte linguistique et extralinguistique mais aussi de reconstituer un élément sémantique manquant. Par exemple, dans « *J'ai fini mon livre.* », il n'y a pas moyen de retrouver l'action associée à livre que j'ai fini d'exercer sur ce livre à partir des mots de la phrase. En revanche, un réseau lexical permet de savoir que les actions que je peux faire sur un livre sont *lire, écrire* et d'autres bien moins probables qui font référence au sens purement matériel de *livre* : *perdre, oublier, brûler, etc.* — voire *manger* !

Lexique génératif (montagovien) *(javascript:;)*

Applications

Traitement automatique des langues

Le traitement automatique des langues ou linguistique informatique s'est bien sur penché sur l'expression du raisonnement en langage naturel, ainsi que sur son analyse automatique — mais aussi sur la génération automatique de preuves mathématiques en langage naturel à partir de preuves formelles.

L'intérêt de la génération est évident à toute personne ayant vu une preuve formelle ou un lambda-terme de quelques pages : un mathématicien ou un expert de Coq souffrira en essayant de la comprendre, et un individu « normal » n'essaiera même pas. Donc, si l'ordinateur trouve ou aide à trouver une preuve, par exemple la vérification d'un logiciel embarqué ou d'un protocole de communication — ou encore la preuve formelle d'un théorème comme mentionné supra —, la personne ayant sollicité cette preuve formelle sera formellement convaincue mais restera sceptique. En revanche un individu presque « normal » pourra s'en faire une idée si la preuve est divisée en lemmes, etc., et rédigée en langage naturel à la manière des livres de problèmes résolus. Aarne Ranta a écrit des générateurs de textes mathématiques à partir de preuves formelles, notamment pour la géométrie élémentaire *(http://www.cse.chalmers.se/~hallgren/Alfa/Tutorial/GFplugin.html)* et plus récemment pour la théorie homotopique des types *(http://old-darcs.grammaticalframework.org/~aarne/gf-hott/)* ainsi que Muhammad Humayoun et Christophe Raffalli *(http://www.lama.univ-savoie.fr/~humayoun/phd/mathnat.html)*.

Un autre aspect est l'analyse automatique de raisonnements, de la relation de conséquence logique entre énoncés en langage naturel. Cela permet de répondre à des questions : tel article du *Monde* entraîne-t-il ou non que « *le chômage des jeunes a baissé le mois dernier* » ? ou, comme plus haut « *Geach a-t-il été l'élève de*

Wittgenstein » ? Une recherche sur Internet trouve les pages web pertinentes, peut identifier les expressions voisines dans le texte, mais elle ne répondra pas à la question posée, à moins qu'un être humain ou un programme spécifique analyse plus finement les informations trouvées. Cette tâche appelée inférence textuelle (*text entailment, natural language inference*), permet de répondre directement, et elle suscite actuellement un vif intérêt. Deux méthodes bien différentes ont été développées : d'une part une analyse linguistique et logique profonde des énoncés en question suivie d'un appel à un assistant de preuve comme dans le travail de Chatzikyriakidis et Luo (2 ***(#Chatzikyriakidis2014)***) d'autre part un apprentissage sur des couples de textes dont le premier entraîne le second (voir par exemple (1 ***(#cabrio2017phd)***)). Bien évidemment l'analyse profonde est limitée par la décidabilité ou la complexité algorithmique et par l'ambiguïté des énoncés — et bien sûr par l'absence de représentation complète des connaissances du domaine. La *formulation* de la question avec des méthodes symboliques est simple : existe-t-il un sens , une formule ϕ de la prémisse et un sens de la conclusion, une formule ψ ; telles que $T, \phi \vdash \psi$? Toute la difficulté est dans T qui doit inclure les faits et les règles du domaine, ainsi que les équivalences et inférences entre mots/prédicats : *X a vendu Z à Y* $\sim$ *Y a acheté Z à X*, *dévorer un livre* $\sim$ *lire un livre* (mais pas pour une chèvre !), un *chien* peut faire ce que tout *animal* peut faire... À l'opposé de ces méthodes, les techniques d'apprentissage peinent à prendre en compte la structure d'un argument : elle se limitent souvent aux arguments « pour » ou « contre » un énoncé donné et sont mises en difficulté par la négation (qui porte sur une partie du texte difficile à déterminer) au point de laisser de coté tout argument contenant une négation. L'analyse ainsi simplifiée des arguments rejoint alors une thématique à la mode notamment pour des raisons commerciales « sentiment/opinion analysis » qui étudient les avis des internautes sur les produits de sites marchands.

Une manière d'éviter l'analyse logique du langage naturel consiste à se placer dans un fragment du langage naturel et dans un fragment logique où il est possible de raisonner directement sur les phrases — comme le faisait Aristote, avec les énoncés A E I O et les figures de syllogisme ! Cette approche permet en fait d'aller beaucoup plus loin et dans la structure des phrases, et dans la logique utilisée. C'est ce que fait la *natural logic* qui définit les schémas d'inférence valides en langage naturel avec des phrases à plusieurs quantificateurs, des nombres entiers (au plus n, au moins n, exactement n), etc. (6 ***(#Moss09)***) (3 ***(#Karttunen2015)***). Du point de vue de la logique, on peut se limiter à un fragment décidable de la logique du premier ordre, par exemple aux logiques de description.

Cognition, didactique

Les sciences cognitives expérimentales ont montré que l'activité de raisonnement n'est pas un module mais plutôt la combinaison de différents modules et zones de l'activité cérébrale : langage, vision directe ou via des représentations schématiques, prise de décision, etc. (cf. van Benthem, A brief history of natural logic ***(http://projects.illc.uva.nl/lgc/translation/papers/Kolkata.pdf)***).

Les enseignants observent que beaucoup d'élèves éprouvent des difficultés à comprendre mais aussi à formuler clairement les énoncés mathématiques. A fortiori, ces élèves peinent à exprimer les raisonnements mathématiques, souvent riches en quantificateurs et en connecteurs logiques.

C'est pourquoi, en plus des linguistes et des logiciens, les cogniticiens et les didacticiens des mathématiques ont eux aussi beaucoup à apporter aux réflexions sur le lien entre logique (mathématique) et langage (naturel). Certains participent à notre projet e-fran AREN Argumentation et Numérique *(http://www.lirmm.fr/aren)*. Ce projet, porté par l'université de Montpellier, réunit l'académie de Montpellier et trois start-up (IntactileDESIGN, Mezoa, WeAreLearning) autour d'une plate-forme de débats écrits en ligne à partir de textes. Pour le moment des enseignants de philosophie, de sciences de la vie et de la terre, de physique-chimie et d'histoire-géographie participent et les textes à débattre sont des controverses socio-scientifiques ; il est prévu que des enseignants de mathématiques fassent débattre sur des preuves ou sur des algorithmes. L'élève sélectionne un passage du texte ou d'un argument précédent, et remplit une bulle, qui contient trois champs : une reformulation du passage sélectionné, *d'accord*, *pas d'accord* ou *pas compris*, et justifie son choix par un argument. Notre projet a pour objectif de développer des outils d'analyse automatique des débats afin d'aider l'enseignant à restituer le débat, en montrant la justesse ou non de la reformulation, comment les arguments se répondent, et se répartissent en thèmes, la nature des arguments... Il ne s'agit pas tant d'enseigner des contenus que d'apprendre aux élèves à débattre. Pour revenir sur le début de cet article, nous allons cette année faire débattre des élèves sur des preuves mathématiques en classe de mathématiques. Espérons que cette plate-forme et les fonctionnalités d'analyse automatique issues de nos recherches sur logique et langage permettront aux élèves de mieux comprendre les mathématiques ainsi que l'art de débattre — et par effet de bord de mieux maîtriser le français.

Conclusion

Le raisonnement, notamment mathématique, se fait en langage naturel, et une des sources de difficultés pour exprimer ou pour suivre un raisonnement est l'ambiguïté du langage naturel, dont un aspect important est l'absence de correspondances bi-univoque entre formule logique et phrase, notamment en raison de la portée de la négation, des quantificateurs, du sens des connecteurs, etc. Un raisonnement correct est fait de règles de déductions correctement appliquées, mais ces règles portent sur des formules logiques, et non sur des énoncés de la langue commune, et il faut bien faire avec car le style de *Principia Mathematica* n'est souhaitable ni pour pratiquer ni pour enseigner les mathématiques. Cet article a donc insisté sur le passage en partie automatisable d'une phrase à une ou plusieurs formules logiques, utile à l'analyse automatique de raisonnements et donc à leur vérification automatique. Dans l'autre sens, il s'agit de traduire les preuves formelles produites par un assistant de preuve en textes lisibles par un être humain — les rendre naturels n'est guère facile.

Tout au long de la rédaction de cet article une question m'a tracassé : quelles sont les différences et les ressemblances entre le raisonnement mathématique et le raisonnement ordinaire ? J'en ai mentionné quelques-unes, mais je n'ai assurément pas épuisé le débat. Je trouve paradoxal que la logique ait été extraite du raisonnement mathématique par Aristote, alors que le raisonnement mathématique est relativement rare dans la vie et la conversation courantes. Une explication raisonnable serait justement que la moindre ambiguïté des énoncés mathématiques, la meilleure connaissance des axiomes et du domaine, rendent plus facile l'accès aux règles de raisonnement, qui dans la vie courante sont masquées par des principes et des connaissances vagues.

Bibliographie

(1) ***()*** ↑ ***(#cabrio2017phd)*** Elena Cabrio. *Component-based Textual Entailment : a Modular and Linguistically-motivated Framework for Semantic Inferences*. thèse de docorat, université de Trento (Italie), 2001.

(2) ***()*** ↑ ***(#Chatzikyriakidis2014)*** Stergios Chatzikyriakidis and Zhaohui Luo. Natural language inference in Coq. *Journal of Logic, Language and Information*, 23(4), pages 441-480, décembre 2014.

(3) ***()*** ↑ ***(#Karttunen2015)*** Lauri Karttunen. From natural logic to natural reasoning. *In* Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing : 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part I*, pages 295-309, Cham, 2015. Springer International Publishing.

(4) ***()*** ↑ ***(#KK86)*** William Kneale and Martha Kneale. *The development of logic*. Oxford University Press, 3e édition, 1986.

(5) ***()*** ↑ ***(#MootRetore2012lcg)*** Richard Moot and Christian Retoré. *The logic of categorial grammars : a deductive account of natural language syntax and semantics*, volume 6850 des *Lecture Notes in Computer Science*. Springer, 2012.

(6) ***()*** ↑ ***(#Moss09)*** Lawrence S. Moss. Natural logic and semantics. *In* Maria Aloni, Harald Bastiaanse, Tikitu de Jager, and Katrin Schulz, editors, *Logic, Language and Meaning - 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, Revised Selected Papers*, volume 6042 des *Lecture Notes in Computer Science*, pages 84-93. Springer, 2009.

(7) ***()*** ↑ ***(#Pustejovsky1995)*** James Pustejovsky. *The generative lexicon*. M.I.T. Press, 1995.

(8) ***()*** ↑ ***(#Retore2014types2013)*** Christian Retoré. *The Montagovian Generative Lexicon* ΛTy_n *: a Type Theoretical Framework for Natural Language Semantics*

(http://dx.doi.org/10.4230/LIPIcs.TYPES.2013.202). *In* Ralph Matthes and Aleksy Schubert, editors, *19th International Conference on Types for Proofs and Programs (TYPES 2013)*, volume 26 des *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 202-229, Dagstuhl, Germany, 2014. Schloss Dagstuhl-Leibniz—Zentrum für Informatik.

(9) *()* ↑ *(#EK97)* Jan van Eijck and Hans Kamp. Representing discourse in context. *In* Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, chapitre 3, pages 179-237. North-Holland Elsevier, Amsterdam, 1996.

Post-scriptum :

Ce texte appartient au dossier thématique « Mathématiques et langages » *(+-Mathematiques-et-langages-+.html)*.

Article édité par Jérôme Germoni *(_Germoni-Jerome_.html)*

NOTES

[1 *(#nh1)*] L'usage reformule cette phrase que certains locuteurs comprennent comme *Aucun entier n'est premier* en *Certains entiers ne sont pas premiers*, mais cette dernière formulation porte un point de vue légèrement différent : ce n'est plus l'ensemble de tous les entiers qui mis en lumière, mais plutôt les entiers qui ne sont pas premiers.

[2 *(#nh2)*] Il y a pu y avoir des controverses mathématiques par le passé *(http://culturemath.ens.fr/content/etudes-de-controverses-en-histoire-des-mathématique)*, il y a aussi des débats entre les mathématiciens intuitionnistes et les autres sur la validité des preuves d'existence fondées sur le raisonnement par l'absurde, ou entre théoriciens de la mesure sur la pertinence de l'axiome du choix non dénombrable, etc.

[3 *(#nh3)*] Cet exemple non politiquement correct provient de Geach, lui-même inspiré par sa lecture des scolastiques.

[4 *(#nh4)*] Merci à Viviane Durand-Guerrier de cet exemple.