

ENJEUX SÉMIOTIQUES DANS LA CONCEPTION D'UNE AIDE À LA RÉSOLUTION DE PROBLÈME DE PREUVE

VENANT* Fabienne – RICHARD** Philippe R.
GAGNON** Michel – HENRÍQUEZ RIVAS** Carolina

Résumé – Nous montrons comment les questions didactiques et informatiques qui émergent dans la conception d'un système de tuteur intelligent pour l'apprentissage de démonstration interagissent et se complètent. Le développement du système QED-Tutrix a nécessité une collaboration entre didacticiens et informaticiens, que nous présentons à travers la mise en place du processus d'accompagnement et de validation du travail de démonstration de l'étudiant, conformément au contrat didactique de la classe.

Mots-clefs : Tuteur intelligent, géométrie, démonstration, travail mathématique, genèse sémiotique

Abstract – We show how the didactic and computing questions that emerge in the design of an intelligent tutor system for demonstration learning interact and complement each other. The development of the QED-Tutrix system has indeed revealed the need for collaboration between didacticians and computer scientists, of which we give an overview through the implementation of the process of accompaniment and validation of the student's demonstration work, in accordance with the didactic contract of the class.

Keywords: Intelligent tutoring system, geometry, proof, mathematical work, semiotic genesis

Notre réflexion s'inscrit dans le questionnement sur la conception des technologies éducatives proposé dans le texte de cadrage du groupe 7. Nous proposons ici de montrer comment les questions didactiques et informatiques qui émergent dans la conception d'un système tuteur intelligent pour l'apprentissage de la démonstration s'inter-influencent et se complètent. L'élaboration du système QED-Tutrix a, en effet, fait apparaître la nécessité d'une collaboration entre didacticiens et informaticiens dont nous donnons un aperçu à travers la question de l'accompagnement et de la validation du travail de démonstration de l'élève par le logiciel, dans le respect du contrat didactique de la classe. QED-Tutrix est conçu de façon à laisser à l'apprenant une liberté totale dans l'exploration du problème à résoudre, et dans le choix de la preuve, tout en lui proposant un accompagnement adapté à ses choix, et en respectant le contrat didactique habituel de la classe (Leduc et al., 2016). Pour cela, le système doit être capable de suivre l'activité mathématique de l'élève, en s'appuyant sur ses propres résolutions des problèmes choisis. Cette phase de l'accompagnement pose aux informaticiens des défis liés à la représentation et à l'organisation des différents itinéraires de résolution pour un problème donné. Nous montrons ici comment une réflexion didactique peut, et doit, accompagner la résolution de ces défis, qui repose sur un processus d'extraction d'informations et d'appréhension des objets mathématiques à partir de l'énoncé de problème, que nous appelons genèse sémiotique en référence au modèle théorique des espaces de travail mathématiques (Kuzniak et Richard, 2014). Cette réflexion peut déboucher sur la question de l'élaboration de conjectures, de validation et d'acceptabilité d'une démonstration en contexte scolaire.

I. LA DEMONSTRATION : UNE ACTIVITE GEOMETRIQUE

La géométrie, plus particulièrement dans un contexte de démonstration, est souvent abordée comme une science qui constitue une « partie des mathématiques ayant pour objet l'étude de l'espace et des figures pouvant l'occuper » au sein de laquelle « les principes sont simples et

* Université du Québec Montréal – Canada – venant.fabienne@uqam.ca

** Université de Montréal – Canada – philippe.r.richard@umontreal.ca

** École Polytechnique de Montréal – Canada – michel.gagnon@polymtl.ca

** Universidad de La Frontera, Temuco – Chile – email carolina.henriquez@ufrontera.cl

absolument vrais sans aucune restriction » (TLFI). Selon cette définition, on pourrait croire que l'apprentissage de la géométrie s'effectue par adhésion aux concepts ou aux processus issus du modèle théorique existant, sans trop d'égard à la réalité de l'espace et des formes. Cependant l'apprentissage passe par une pratique qui s'ancre dans la réalité, par le fait même d'exercer une activité concrète. Dans cette perspective, résoudre un problème de preuve est une activité finalisée qui permet à l'élève d'effectuer son travail de mathématicien, en développant son sens géométrique. L'activité de démonstration favorise en effet la compréhension de la nécessité épistémique des propriétés (Coutat et *al.*, 2016), l'implémentation des concepts géométriques et le développement des compétences cognitives de l'élève (Richard et *al.*, 2016). Cependant, ainsi que le souligne Brousseau (2011), pour qu'une notion mathématique soit le fruit d'une activité mathématique, « il faut aussi que des alternatives plausibles lui soient opposables et que son choix soit le résultat d'une anticipation possible ». C'est pourquoi nous voulons que, au sein du système tutoriel, l'élève soit confronté à des problèmes complexes. Nous considérons qu'un problème est complexe quand il répond aux différentes exigences définies par Richard et *al.* (2011) : existence de différents processus de résolution (exigence heuristique) qui mobilisent un réseau de concepts et de processus mathématiques (exigence cognitive), reposent sur une approche argumentative et un raisonnement à plusieurs niveaux ou des routines non calculatoires (exigence discursive) et permettent de développer des compétences qui vont au-delà de la simple reproduction (exigence de compétence). Nous centrons notre exposé sur l'exigence heuristique, soit l'existence de différents processus de résolution, et la façon de la mettre en œuvre au sein d'un système tuteur. Pour pouvoir guider l'élève dans la résolution d'un problème complexe, QED-Tutrix doit anticiper l'ensemble de ces processus puis reconnaître celui dans lequel l'élève s'engage. Nous présentons ci-dessous les choix de modélisation informatique qui se sont imposés dès le début de la conception du système.

II. MODELISATION DES PROCESSUS DE RÉOLUTION : GRAPHE HPDIC

Lorsqu'un élève effectue son travail de mathématicien à l'école, le modèle des espaces de travail mathématique (Kuzniak et Richard, 2014) reconnaît spécifiquement l'activation de trois genèses, entre un plan épistémologique et un plan cognitif (**Figure 1**). Ces genèses permettent d'apprécier des questions comme celles de la création du sens, de la validation de propriétés ou de l'usage d'outils techniques en termes de coordination des genèses. Nous intéressons ici aux genèses sémiotiques et discursives sous-jacentes à l'activité de démonstration.

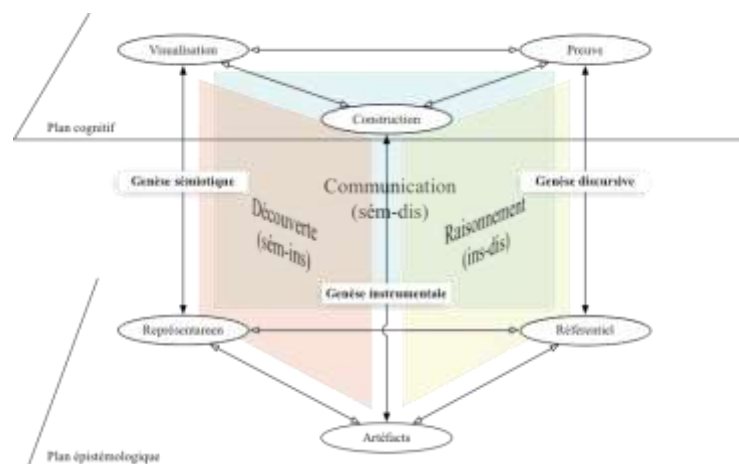


Figure 1 – L'Espace de Travail Mathématique et ses genèses.

La genèse sémiotique est le processus par lequel les signifiants, qu'ils soient textuels ou figuraux, acquièrent leur statut d'objets mathématiques opérationnels. La genèse discursive est le processus par lequel les propriétés et les résultats organisés dans le référentiel théorique sont actionnés afin d'être disponibles pour le raisonnement mathématique et les validations discursives. Au sein du système tutoriel, ces genèses sont prises en charge par deux processus d'extraction et de traitement de l'information : extraction des hypothèses et des conclusions à partir de l'énoncé du problème (genèse sémiotique), et organisation de ces dernières en démonstrations acceptables sous forme d'un graphe d'inférence et d'un texte argumenté (genèse discursive). Une inférence est une opération logique consistant à conclure la vérité d'une proposition à partir d'autres propositions prises comme hypothèses, et d'une propriété ou d'une définition prise comme justification.

Considérons le problème suivant : ABCD est un parallélogramme de centre O ; On appelle M le milieu de [AB] et N le milieu de [DC] ; Démontrer que (OM) est parallèle à (BC).

La figure 2 présente, sous forme de graphe, une inférence à partir de deux hypothèses présentes dans l'énoncé du problème : *ABCD est un parallélogramme*, *O est le centre du parallélogramme*. Or, « le centre de symétrie d'un parallélogramme est le point d'intersection des diagonales », donc O est l'intersection des diagonales [AC] et [BD].

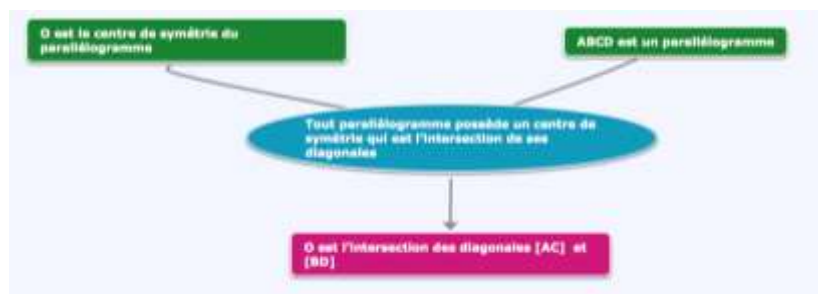


Figure 2 – Représentation d'une inférence sous forme de graphe.

Une preuve complète peut être représentée par un graphe rendant compte de l'enchaînement des inférences qui la constitue. Deux inférences s'enchaînent quand la conclusion de la première devient une hypothèse pour la suivante. Elle prend alors le statut de résultat intermédiaire. Nous appelons un tel graphe un graphe HPDIC car il contient des Hypothèses, des Conclusions, des résultats Intermédiaires et des Conclusions. (Leduc 2016, Tessier-Baillargeon 2016). Les propriétés et définitions acceptables sont consignées dans un fichier qui constitue le référentiel théorique du système. La constitution de ce fichier informatique est le fruit du travail des didacticiens de l'équipe et repose sur des observations en classe et de manuels scolaires.

III. MODELISER LA GENESE SEMIOTIQUE

La construction du graphe HPDIC constitue la pierre angulaire de l'architecture de QED-Tutrix. Le défi est de le générer, à partir d'un énoncé de problème. Il faut pour cela extraire le résultat à démontrer, mais aussi l'ensemble des hypothèses, et résultats intermédiaires qui vont intervenir dans les démonstrations possibles, et de les structurer en un enchaînement d'inférences dont les justifications sont toutes issues du référentiel théorique. La figure 3

présente un graphe HPDIC¹ représentant six preuves possibles pour le problème mentionné dans la partie II.

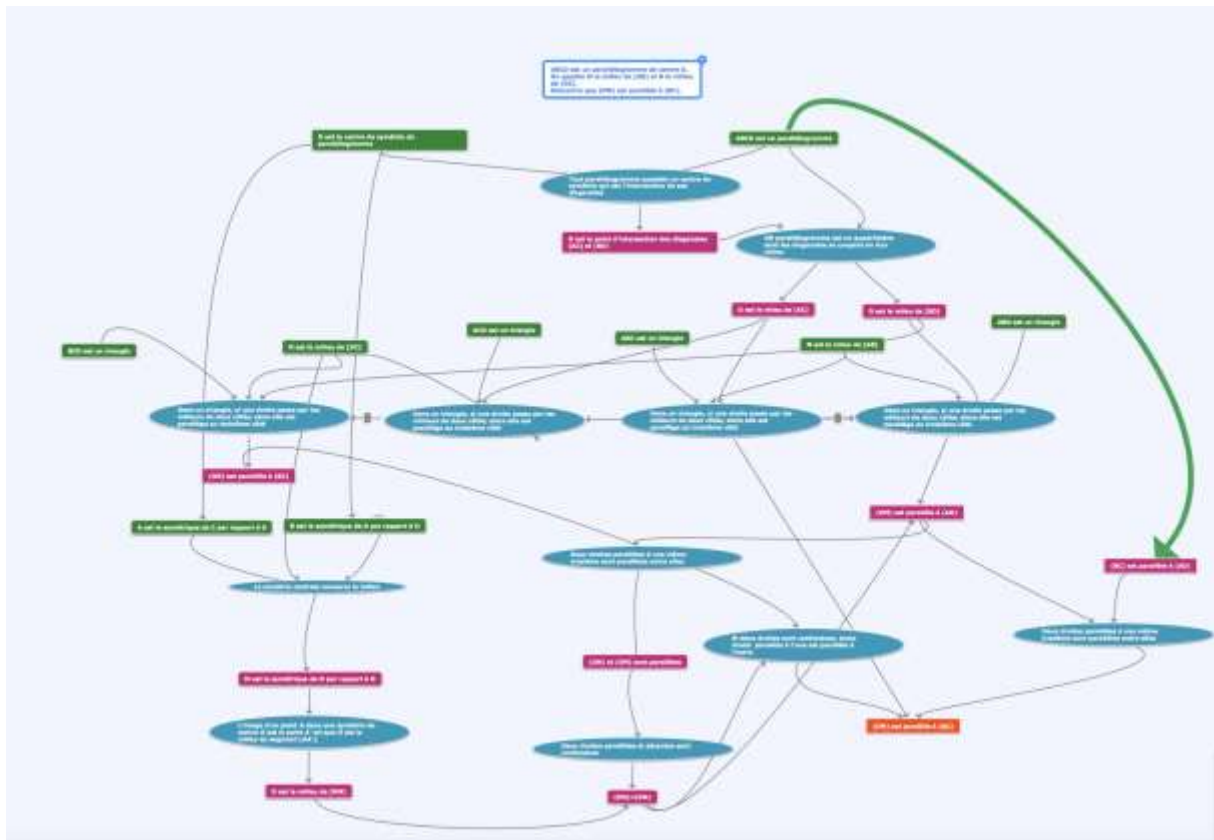


Figure 3 – Exemple de graphe HPDIC

On trouve dans ce graphe plusieurs types d'information :

- Les hypothèses explicites, comme *ABCD est un parallélogramme*. Ce sont les informations qui sont directement accessibles dans le texte, moyennant une analyse linguistique.
- Les hypothèses implicites. Ce sont des hypothèses qui ne nécessitent pas un pas d'inférence, mais dérivent directement des propriétés d'un objet présent dans une hypothèse lexicale. Par exemple *(AB) est parallèle à (CD)* est sous-entendu par *ABCD est un parallélogramme*.
- Les hypothèses dérivatives. Ce sont les hypothèses qui posent des objets intermédiaires que l'on doit considérer ou construire pour avancer dans un pas de démonstration. Ces objets sont potentiellement présents dans une hypothèse lexicale, mais ne sont pas explicitement décrits. Par exemple *ABC est un triangle* est porté par *ABCD est un parallélogramme*, qui sous-entend implicitement que les points ABCD ne sont pas alignés.
- Les hypothèses contextuelles. Ce sont les hypothèses qui dérivent implicitement du contrat didactique. Par exemple, *ABC est un triangle* est aussi une hypothèse contextuelle. Dans cette hypothèse, on suppose en effet qu'on ne considère pas le cas limite où les points A, B, C et D sont alignés. De même, dans le graphe de la figure 2, l'hypothèse *[AD] est parallèle à [BC]* est une hypothèse contextuelle. En toute rigueur,

¹ Une version plus lisible de ce graphe est disponible à l'adresse : <https://www.mindomo.com/mindmap/067937c9f30f47269cd3b9e3b7779d30>

la propriété *être parallèle* ne s'applique pas à aux segments, mais aux droites qui les supportent. Cependant, certains enseignants acceptent la proposition *[AD] est perpendiculaire à [BC]* car elle découle de l'expression « les côtés opposés sont parallèles » utilisée dans la définition du parallélogramme plutôt que « les droites supports des côtés opposés sont parallèles ».

- Une conclusion explicite : le résultat à démontrer est ici clairement exprimé dans le texte. D'autres cas de figure sont envisageables, ainsi que le résume la figure 3 ci-dessous.

La génération du graphe HPDIC à partir de l'énoncé du problème débute par l'extraction des informations textuelles explicites à partir d'une analyse syntaxique. Il peut aussi s'agir d'informations graphiques codées directement sur une figure incluse dans l'énoncé. Dans notre exemple, il s'agit :

- d'une conclusion explicite indiquant clairement le résultat à démontrer : *(OM) est parallèle à (BC)*,
- de trois hypothèses explicites: *ABCD est un parallélogramme*, *M est le milieu de [AB]* et *O est le centre du parallélogramme*.

Cette première tâche, bien que simple en apparence, n'est pas du tout triviale. Sur le plan informatique, elle soulève quelques questions relatives au traitement automatique de la langue. Le logiciel doit par exemple calculer le sens du verbe *appeler* qui indique ici une hypothèse, mais qui pourrait tout aussi bien indiquer une conclusion ou un résultat intermédiaire dans d'autres énoncés comme celui-ci : « Soit ABC est un triangle. La perpendiculaire à (AB) passant par C coupe (AB) en K, la perpendiculaire à (AC) passant par B coupe (AC) en J. Comment appelle-t-on le point d'intersection de (KC) et (BJ) ? Justifie ta réponse. ». Il faut également reconstruire l'information *O est le centre du parallélogramme* à partir de la construction syntaxique *de centre O* rattachée à *parallélogramme*, ou repérer le résultat à démontrer en tant que complément d'objet du verbe *démontrer*.

Une fois les informations lexicales extraites, le système doit prendre en charge l'extraction des hypothèses implicites. Une première étape est le traitement de l'implicite définitoire, c'est-à-dire le fait que tout objet géométrique en sous-entend d'autres qui constituent ses caractéristiques. Dans notre exemple, le parallélogramme sous-entend quatre sommets, quatre côtés consécutifs, quatre angles, deux diagonales (qui sont en fait les caractéristiques du quadrilatère) ainsi que deux paires de côtés opposés parallèles. À cet implicite définitoire, s'ajoute souvent un implicite dénominatif, qui permet d'instancier ces caractéristiques générales dans des objets géométriques particuliers. Dans notre exemple, l'expression *ABCD est un parallélogramme* permet d'instancier les caractéristiques du parallélogramme : les quatre sommets sont A, B, C et D, les quatre angles sont DAB, ABC, BCD et CDA, les quatre côtés consécutifs sont [AB], [BC], [CD] et [DA], les diagonales sont [AC] et [BD], les côtés opposés et parallèles sont d'une part [AB] et [CD], et d'autre part [BC] et [AD]. Dans le même ordre d'idée, l'hypothèse *BCD est un triangle* est porteuse implicitement des informations suivantes : BCD a trois sommets, trois côtés et trois angles, qui sont respectivement A, B et C, [BC], [CD] et [BD], et ABC, BCA et CAB.

Dans la figure 2, la flèche épaisse à droite, représentant la déduction de la proposition *(AD) et (BC) sont parallèles* à partir de la seule proposition *ABCD est un parallélogramme* constitue donc un raccourci inférentiel résultant de ces deux implicites. En réalité, le logiciel devra passer au travers des étapes suivantes :

1. Extraire du texte l'expression *ABCD est un parallélogramme* en tant qu'hypothèse lexicale.

2. Extraire du référentiel théorique la définition de l'objet parallélogramme : « un parallélogramme est un quadrilatère dont les deux côtés opposés sont parallèles ».
3. Extraire du référentiel théorique les caractéristiques du quadrilatère.
4. Reconnaître la dénomination ABCD et réaliser qu'une instantiation de parallélogramme doit être faite.
5. Instancier les caractéristiques d'un quadrilatère en nommant les côtés opposés, côtés consécutifs et diagonales.
6. Instancier la deuxième partie de la définition du parallélogramme en gérant l'implicite contextuel et générer les deux propositions possibles *[BC] et [AD] sont parallèles* ou *(BC) et (AD) sont parallèles*.

Ces considérations informatiques éclairent les considérations didactiques sur la lecture d'énoncé. L'élève qui résout le problème est lui aussi en prise avec l'extraction des objets mathématiques, et de leurs caractéristiques, à partir des informations contenues dans l'énoncé. La question de l'importance du vocabulaire et des pratiques langagières dans l'appréhension des objets mathématiques prend ici tout son sens (Venant et *al.*, 2015, Gobert, 2013). La traduction informatique des propriétés et des hypothèses exige une grande précision dans la définition des objets et l'utilisation du vocabulaire, alors que le langage courant peut parfois accepter d'assimiler certains concepts comme côté et segment, ou même droite et segment, dans l'expression des propriétés de parallélisme par exemple. L'extraction des informations textuelles, et la gestion des implicites, se concrétisent souvent dans la réalisation d'une figure. Ainsi, qu'une figure accompagne ou pas l'énoncé, l'élève va en tracer une pour lui-même. Le rôle de la figure peut éventuellement aller plus loin. L'élève peut en effet construire tout ou partie de son raisonnement de façon instrumentale (Richard et *al.*, à venir). La modélisation informatique que nous avons réalisée permet de mettre au jour des enjeux didactiques présents dès les premières phases de la résolution de problème, incluant l'importance de l'articulation des registres sémiotiques (Duval, 1991). De plus, la question des raccourcis inférentiels soulève celle de la validité d'une preuve, en lien avec le contrat didactique : Quels sont les implicites ? Quel est le niveau de rigueur attendu ou que peut-on considérer comme trivial ? Certaines hypothèses dans une inférence peuvent-elles être considérées comme facultatives ? Quels cas limites traite-t-on ?

Le logiciel, quant à lui, appréhende ces questions en surgénérant des hypothèses, anticipant au maximum celles qui sont implicites. Une fois extraites, toutes les hypothèses et conclusions sont mises en instance discursive, mais seules certaines hypothèses seront retenues dans le graphe complet, c'est-à-dire seulement celles qui produisent un effet dans l'ensemble des démonstrations possibles. L'obtention du graphe complet, c'est-à-dire l'ajout de tous les chemins acceptables au sein du graphe, relève de la genèse discursive. Le graphe HPDIC doit contenir toutes les inférences, ou pas de démonstration, qu'un didacticien ou un enseignant jugerait acceptables en fonction du problème à résoudre, du contexte scolaire et de l'apprentissage visé. Cette exigence d'intégration dans un contexte d'enseignement ou d'apprentissage n'est pas banale. Dans la théorie des situations didactiques en mathématiques (Brousseau, 1998), et plus particulièrement au regard de la notion de contrat didactique, on considère que l'élève et l'enseignant ont des responsabilités réciproques par rapport aux connaissances en jeu. On peut certes parler de règles dans un partage (asymétrique) de responsabilités, certaines étant bien connues de tous, mais la plupart sont implicites, pouvant résulter bien plus d'une habitude de classe que d'une contrainte nécessaire. Ainsi, sur le plan du discours, et plus précisément dans une logique déductive, l'enseignant peut tolérer plusieurs types de raccourcis inférentiels, parce que cela faciliterait la lecture d'une preuve ou la rendrait pédagogiquement plus efficace. De plus, l'élève qui s'initie progressivement au

discours déductif risque de trouver ses propres raccourcis inférentiels, dans un équilibre qui évolue au fur et à mesure que lui et ses compagnons s'habituent au discours déductif en mathématique. La question de la lisibilité d'une preuve ou celle de l'acceptabilité d'un raisonnement sont donc étroitement liées au contrat didactique, c'est-à-dire à la géométrie en tant qu'activité (Font, Richard et Gagnon 2018). Les défis didactiques que soulèvent les genèses sémiotiques et discursives informatiques sont résumés dans la figure 5.

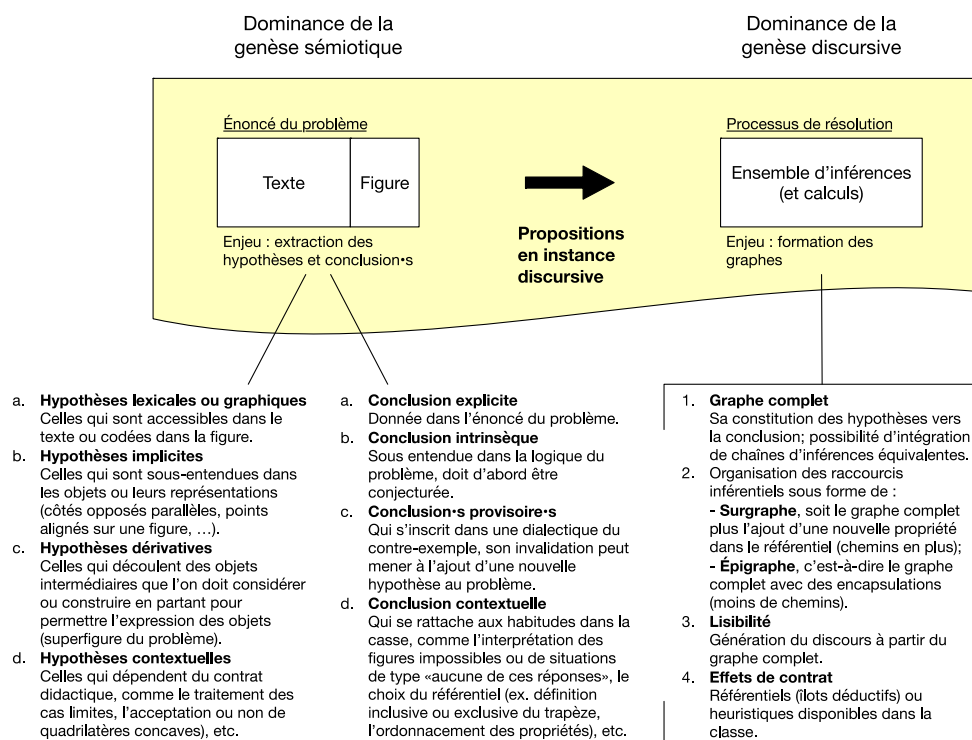


Figure 4 – Genèses sémiotiques et discursives dans QED-Tutrix

IV. CONCLUSION

Notre réflexion se situe au carrefour de l'informatique et de la didactique. La conception de QDEX prend en compte l'utilisateur très tôt dans le processus pour pouvoir le guider au mieux. Dans cette optique, les genèses sémiotique et discursive informatiques doivent reprendre les caractéristiques de celles de l'utilisateur et être à la fois complémentaires et très imbriquées, afin de mener à des preuves acceptables et lisibles, mais aussi de permettre la caractérisation d'un éventuel blocage de l'élève en termes de résultats à démontrer et de propriétés à utiliser. C'est une étape indispensable pour pouvoir proposer à l'élève un problème connexe qui lui permettra de dépasser son blocage. L'idée de répondre à un blocage de l'élève en lui proposant la résolution opportune de problèmes constitue une solution effective à l'une des difficultés majeures de l'enseignement : éviter de donner des réponses en même temps que les questions lorsque l'élève est en difficulté. En ce sens, notre projet soulage théoriquement un paradoxe de Brousseau (1998), que l'on appelle «paradoxe de la dévolution» : tout ce que fait l'enseignant pour faire produire, par les élèves, les comportements qu'il attend tend à diminuer l'incertitude de l'élève et par là, à priver ce dernier des conditions nécessaires à la compréhension et à l'apprentissage de la notion visée; si l'enseignant dit ou signifie ce qu'il veut de la part de l'élève, il ne peut plus l'obtenir que comme exécution d'un ordre et non par l'exercice de ses connaissances et de son jugement.

La notion de dévolution, en tant que levier didactique pour l'enseignant et condition indispensable pour le développement de l'autonomie de l'élève, gagne en force et reprend ici l'idée qu'un problème connexe appartient à l'espace de travail du problème racine et que l'enseignant cherche à rendre cet espace à l'élève en lui laissant la responsabilité de la résolution. Le développement de l'autonomie dans l'apprentissage est un enjeu social majeur.

REFERENCES

- Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble : La Pensée Sauvage.
- Brousseau, G. (2011). « La théorie des situations didactiques en mathématiques », *Éducation et didactique*, vol. 5, no. 1.
- Coutat, S., Laborde, C. et Richard, P. R. (2016). L'apprentissage instrumenté de propriétés en géométrie : propédeutique à l'acquisition d'une compétence de démonstration, *Educational Studies in Mathematics*.
- Duval, R. (1995). *Sémiosis et pensée humaine*. Berne : Peter Lang.
- Font, L., Richard, P.R. & Gagnon, M. (2018). Improving QED-Tutrix by Automating the Generation of Proofs. In *Theorem proving components for Educational software in the Electronic Proceedings in Theoretical Computer Science* (EPTCS Post-Proceedings).
- Gobert, S. (2013). Construire des significations dans et par le langage. Bronner A. et alii. (Coord.). *Questions vives en didactique des mathématiques : problèmes de la profession d'enseignant, rôle du langage.*, La Pensée Sauvage.
- Kuzniak, A., et Richard, P.R. (2014). Spaces for Mathematical Work. Viewpoints and perspectives. *Revista latinoamericana de investigación en matemática educativa*, 17(4).
- Leduc, N. (2016) *Développement d'un tutoriel intelligent pour aider à élaborer des preuves en géométrie*. Thèse de doctorat. École Polytechnique de Montréal.
- Richard, P. R. Fortuny J.M., Gagnon, M., Leduc, N. Puertas, E. et Tessier-Baillargeon, M. (2011). Didactic and theoretical-based perspectives in the experimental development of an intelligent tutorial system for the learning of geometry. *ZDM Mathematics Education* 43: 425.
- Richard, P. R., Oller Marcén, A. M. et Meavilla Seguí, V. (2016). The concept of proof in the light of mathematical work. *ZDM Mathematics Education* 48:843–859.
- Richard, P. R., Gagnon, M. et Fortuny, J. M. (2018). Connectedness of problems and impasse resolution in the solving process in geometry: a major educational challenge. Dans P. Herbst, U. H. Cheah, K. Jones et P. R. Richard (Éds), *International perspectives on the teaching and learning of geometry in secondary schools*. Springer.
- Richard, P.R., Venant, F. et Gagnon, M. (à paraître). Issues and challenges about instrumental proof. In Gila Hanna, David Reid et Michael de Villiers (Eds.) *Proof Technology in Mathematics Research and Teaching*, Springer book series Mathematics education in the digital era.
- Tessier Baillargeon, M. (2016). *GeoGebraTUTOR : Développement d'un système tutoriel autonome pour l'accompagnement d'élèves en situation de résolution de problèmes de démonstration en géométrie plane et genèse d'un espace de travail géométrique idoine*. Thèse de doctorat. Université de Montréal.
- TLFI. Définition du mot *géométrie*. Trésor de la langue français informatisé : <http://stella.atilf.fr/Dendien/scripts/tlfiv5/advanced.exe?8;s=1604356215>; Page consultée le 3 décembre 2017.
- Venant, F., Tremblay, O et Labrecque A.-A. (2015). Le lexique au carrefour des mathématiques et du français: pistes pour travailler vocabulaire courant et mathématique. *Bulletin de l'Association de Mathématique du Québec*.