TASSADIT AMGHAR
BERNARD LEVRAT
Leria, Angers, France
{amghar, levrat}@univ-angers.fr

PIAGET (Plateforme Informatique d'Aide à la Génération d'Enoncés Textuels): Présentation générale

Introduction

La paraphrase et le phénomène de reformulation ont fait l'objet de nombreux travaux qui ont mis en évidence l'influence toujours sensible d'un changement de forme d'une expression langagière, sur le contenu sémantique que celle-ci véhicule. Parmi ces travaux, citons (Mel'cuk, 1992; Fuchs, 1987, 1994) en linguistique, (Amghar & Levrat, 1995; Nogier, 1991; Boyer & Lapalme, 1985; McKeown, 1983) en modélisation linguistique, ou (Mathieu & Sabatier, 1985; Salton & Buckley, 1990; Baeza-Yates & Ribeiro-Neto, 1999) en recherche documentaire.

Nous décrivons ici quelques aspects de la plateforme informatique PIAGET (Plateforme Informatique d'Aide à la Génération d'Enoncés Textuels) en cours de développement qui vise à permettre à des psychologues, à des didacticiens et à des enseignants, d'accéder à un ensemble d'outils d'aide à la reformulation et à la génération d'énoncés d'exercices arithmétiques simples.

Il s'agit de caractériser la corrélation entre des modifications plus ou moins profondes susceptibles d'être faites sur des énoncés d'exercices et la variation de difficulté de résolution qui résulte de ces modifications. A cette fin, la plateforme intègre un ensemble d'outils informatiques permettant la modification et la génération d'énoncés. Notre corpus expérimental est constitué d'énoncés d'exercices arithmétiques simples dédiés à de jeunes enfants âgés entre 7 et 9 ans et scolarisés dans les cours élémentaires 1 et 2 du cycle l'école primaire en France. Il se compose d'une soixantaine d'énoncés d'exercices répartis en 24 types selon les valeurs de différents critères qui sont définis par les psychologues et les didacticiens impliqués dans le projet (R. Brissiaud, n2002), (JF Richard & E. Sander, 2000), (E. Sander, 2001).

En l'état actuel, les outils auxquels la plateforme donne accès sont principalement:

- des outils généraux de manipulation, de structuration, d'importation et de création de lexiques et de thésaurus
- des outils de remplacement de mots d'un énoncé par des substituts lexicaux
- un classifieur qui permet de situer un exercice dans une typologie prédéfinie
- un outil de définition et de stockage de *familles d'énoncés* (structure mathématique abstraite)

Notons que la majeure partie des outils (étiquetage, traitements morphologiques, définition de vocabulaire, création et gestion de thésaurus,…) sont génériques et dépassent le cadre strict de notre application spécifique (la reformulation et la génération d'énoncés d'exercices). En revanche, les outils de classification d'énoncés et le gestionnaire de familles d'exercices sont eux dépendant de notre domaine d'application.

Dans une première partie, caractérisons la problématique du processus de reformulation en déclinant les différentes formes qu'il est susceptible de prendre selon la couche de la langue principalement impliquée par la reformulation. Puis nous nous focalisons sur les aspects de la reformulation fondés sur le lexique. Nous décrivons ensuite le corpus expérimental sur lequel nous travaillons et donnons une brève description du classifieur d'énoncés que nous avons réalisé. Enfin nous esquissons en conclusion les principaux points en cours de développement de notre plateforme[1].

Différents types de reformulation d'énoncés

Les reformulations d'énoncés auxquelles nous nous intéressons, c'est à dire celles susceptibles d'être contrôlées pour moduler la difficulté de résolution des exercices correspondants, peuvent porter sur des éléments de natures différentes selon la couche de langue sur laquelle porte principalement la reformulation:

Au niveau lexical:

La modification du champ sémantique lexical, par exemple, permet de prendre en compte l'influence de l'habillage d'un énoncé, sachant que la familiarité plus ou moins grande des élèves avec un domaine aura des répercussion sur la compréhension de la mise en situation de l'exercice dans ce domaine. On pourra, par exemple, modifier le nature des variables à calculer et passer d'un problème d'effectif, à un problème de prix ou à un problème d'âge considérés comme isomorphes au premier. On peut aussi envisager des transformations d'énoncés, qui en restant dans le cadre d'énoncés consistant à calculer des valeurs de variables d'un type donné, les exercices de calcul de prix, par exemple, consistent modifier la situation qui permet d'introduire ces calculs. On passera par exemple du calcul du prix d'articles pour l'école, au calcul du le prix de vêtements ou de denrées.

Un autre aspect de ces modifications lexicales consiste à prendre en compte le niveau familiarité de l'élève avec le vocabulaire utilisé dans l'énoncé (le vocabulaire peut relever d'un registre de langue plus ou moins recherché, par exemple, et l'on proposera ainsi de passer du mot «coûteux » au mot « cher », ou de « merisier » à « cerisier », pour étudier les répercussions de ce type de modification sur la compréhension de l'exercice), ou d'agir sur le niveau d'abstraction du vocabulaire de l'énoncé pour moduler le degré de correspondance de la situation qui y est décrite avec une entité ou une situation réelle éventuellement rencontrées par l'élève. Ce type de modifications, qui s'appuiera en particulier sur la relation d'hyperonymie utilisera une structuration du

[1] Une grande partie des manipulations morphologiques et de recherche de motifs réguliers réalisés par la plateforme utilisent le logiciel Intex. (M. Silberztein, 1993).

vocabulaire en thésaurus par des relations lexicales. On passera par exemple de « pommier » à l'un de ses hyperonymes «arbre ».

Au niveau discursif

Les transformations dont il s'agit ici modifient la structure de l'énoncé, et parmi celles qui sont envisagées, l'ordre de présentation des différentes phrases qui le constituent est un facteur susceptible de jouer un rôle dans la difficulté de résolution ainsi par exemple: *Jean à 3 ans de moins que Paul. Paul à 5 ans. Quel est l'âge de Paul ?* semble plus difficile à résoudre que *Paul a 5 ans. Jean a 3 ans de moins que Paul. Quel est l'age de Paul ?*
C'est à ce niveau aussi qu'intervient la possibilité d'insérer ou d'éliminer des questions intermédiaires dans les cas où la résolution de l'exercice fait appel à des résultats partiels, ou bien encore d'introduire des informations facultatives pour guider l'élève: *Jean a des billes dans sa poche. Il en a gagné 3 et en a 10 à présent. Combien en avait-il avant la partie ?* semble plus difficile que *Jean a des billes dans sa poche mais ne les a pas comptées. Il en a gagné 3 et en a 10 à présent. Combien en avait-il avant la partie ?*

Au niveau syntaxique

La structure grammaticale des phrases intervient elle aussi dans la difficulté de résolution. La segmentation de l'énoncé en phrases plus ou moins courtes (usage des propositions coordonnée, ajout de circonstancielles, par exemple) ou plus ou moins complexes du point de vue syntaxique avec les répercussion sur la compréhension (rôle du passif dans la modification du thème, introduction d'ambiguïtés, structure actantielle du verbe complétée ou simplifiée). *Paul a gagné 3 billes* vs *3 billes ont été gagnées par Paul* (la thématisation réalisée par le passif met en avant la quantité de bille, au détriment d'une complexité syntaxique plus grande)

C'est pour permettre de réaliser de telles reformulations que nous avons conçu la plateforme PIAGET dont nous donnons une brève description de l'état actuel de développement dans ce qui suit.

La plate forme PIAGET

Il s'agit d'une interface donnant accès à un ensemble d'outils permettant d'effecteur ou de faciliter les différents types de reformulation évoqués précédemment sur des énoncés d'exercices. A terme, PIAGET devrait aussi permettre la génération d'énoncés d'exercices répondant à des critères caractéristiques de leur difficulté de résolution[2] par les élèves.
Dans les principales fonctionnalités des outils offerts actuellement par PIAGET citons:

- La définition de lexiques répondant à des critères fixés par l'utilisateur: Il s'agit ici de permettre d'utiliser et donc de construire des lexiques répondant à des particularités fixées.
- Le remplacement de mots dans les énoncés en ayant éventuellement recours, pour le choix des substituts lexicaux, à la navigation au travers des relations d'un thésaurus dans un vocabulaire choisi.
- L'importation d'un thésaurus préexistant ou bien la structuration d'un vocabulaire par des relations lexicales, que ces dernières soient standard ou que la plateforme ait permis à l'utilisateur de les définir
- La définition et le stockage au format XML de la structure mathématique abstraite des exercices.
- Un outil de classification permettant l'affectation d'un exercice à une famille d'exercices correspondant à une classe d'une typologie préalablement définie en se fondant sur l'énoncé de l'exercice.

Nous esquissons ci-dessous quelques une des fonctionnalités actuellement remplies par PIAGET.

Utilisation des fonctionnalités principales de la plate-forme.

Le remplacement

Un dictionnaire, un thésaurus et un énoncé ayant été chargés, la fenêtre principale se présente de la façon suivante:
La partie gauche est occupée par l'énoncé; partie droite fait apparaître le vocabulaire[3]

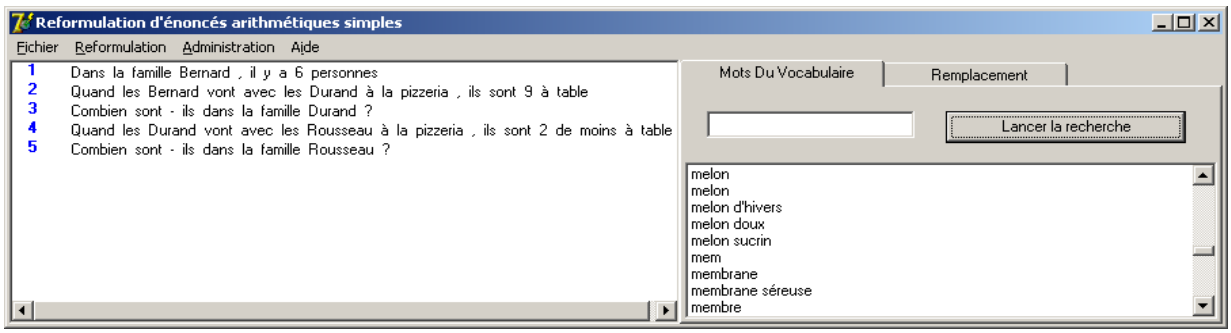


Figure: Fenêtre principale de PIAGET
L'onglet de remplacement à la droite de l'écran permet d'effectuer des substitution lexicales. Il donne accès la fenêtre suivante dans laquelle on se propose de replacer le mot *personne* par un mot qui lui est lié par une relation lexicale[4]

[2] Ces critères font l'objets d'expérimentation par les psychologues et les didacticiens avec lesquels nous collaborons.
[3] Par défaut du choix d'un vocabulaire, c'est le vocabulaire du thésaurus choisi qui est utilisé comme vocabulaire courant.
[4] Notons que l'on pourrait aussi choisir un mot d'un vocabulaire ou un mot quelconque. Nous n'illustrerons pas ces choix en raison de la place maximum autorisée ici pour les articles.

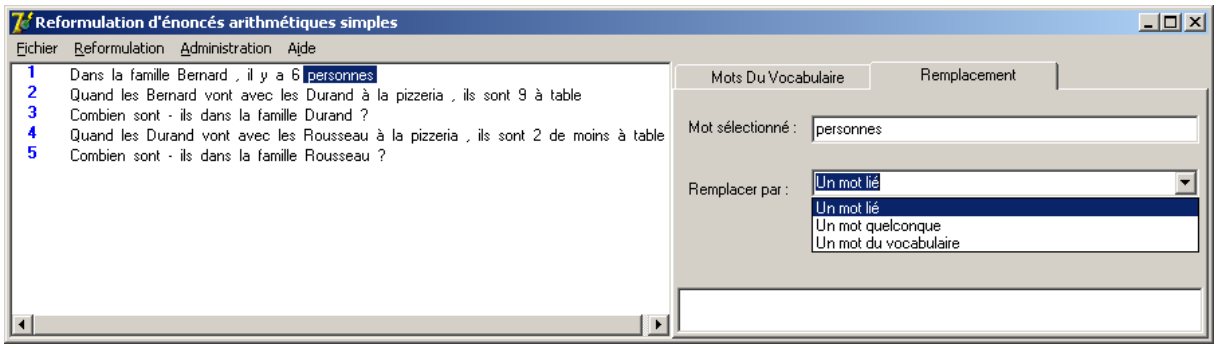
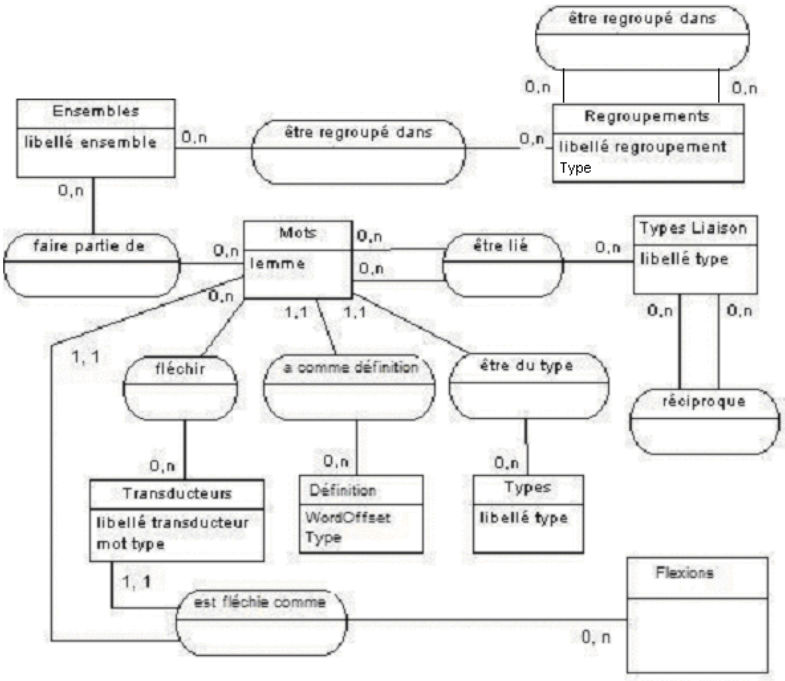


Figure: choix du remplacement d'un mot par un mot lié par une relation lexicale

Le choix de la relation lexicale à utiliser est effectué par l'accès à une autre fenêtre donnant accès à un menu permettant de caractériser la relation lexicale (qui peut, par exemple être simple, ou bien encore être le résultat de la composition, de plusieurs relations lexicales).
En l'occurrence si la relation lexicale choisie est l'hyperonymie, PIAGET proposera *être*, *homme*, *individu*, *mortel* issus du vocabulaire courant comme substituts possibles à *personnes*. On peut alors choisir l'un d'entre eux en cliquant sur celui-ci et réaliser la substitution.

La gestion du vocabulaire

Elle s'appuie principalement sur l'utilisation d'une base de données correspondant au modèle conceptuel suivant:



.

Figure 1: Modèle conceptuel des données pour les vocabulaires

Grossièrement, ce modèle conceptuel comporte --- des informations relatives aux catégories morphosyntaxiques des mots, en fait le nom des transducteurs qui décrivent leurs comportements morphologiques, -- des informations permettant de créer des ensembles de mots dans un premier temps puis de réaliser des opérations entre ces ensembles pour créer des vocabulaires par union ou intersection -- et des informations relatives à la structuration du vocabulaire par des relations lexicales (nommées ici *liaisons*).

Enfin de façon plus spécifique d'un système d'aide à la reformulation d'énoncés d'exercices, finalité principale de notre plateforme, PIAGET donne accès à un outil de représentation des structures abstraites ou *familles* d'exercices et à un classifieur d'exercices dans une typologie prédéfinie par les psychologues impliqués dans notre projet.

Les familles d'exercices

Les familles d'exercices de problème sont des modèles abstraits décrivant les structures mathématiques des exercices. La représentation des familles s'appuie sur:

- des *propriétés* = attributs quantifiables ou variables
- des *collections* ou *entités* = éléments ayant une valeur pour une ou plusieurs propriétés de la famille

– des *formules* = opérations qui expriment les relations entre les valeurs calculées en utilisant des opérations simples dont les opérandes sont des propriétés de la famille: +,-,x,/.

Remarquons ici que du point de vue méthodologique, et même si cela n'a au bout du compte que peu d'incidence sur le résultat en termes de représentation d'une famille, nous avons choisi pour la définition des éléments caractérisant les familles, de définir les collections ou objets de l'exercice comme étant des entités sur lesquelles portent des propriétés et non l'inverse comme l'intuition pourrait le suggérer[5]. En effet, s'agissant d'énoncés d'exercices, les propriétés, en ce qu'elles correspondent à des variables du problème, sont de première importance pour l'exercice; les entités ou collections, quant à elles n'apparaissent qu'en tant qu'elles sont des supports pour les propriété[6]. Ainsi, par exemple pour le fragment d'énoncé *3 crayons à 2* Euros, la famille susceptible de l'engendrer, comportera, 3 propriétés (*valeur*, *nombre* et *prix unitaire*) -- et une collection (un ensemble de *3 crayons*) sur laquelle peuvent porter les trois propriétés.

Lors de la définition d'une famille associée à des exercices dont une partie reprend l'exemple précédent,

– *valeur* sera décrite comme une propriété de type *calculée*, qui peut faire l'objet d'une question, et sera résultat d'un calcul, que l'on spécifiera ici, comme le produit de la propriété *nombre* par la propriété *prix unitaire*.

–*prix unitaire* et *nombre* seront décrites comme des propriétés de type *donnée* dont la valeur sera explicitement donnée dans l'énoncé.

Notons enfin que le stockage des représentations des familles d'exercices est réalisée en XML.

Le classifieur d'énoncés d'exercices

Le corpus sur lequel nous travaillons est composé d'une soixantaine d'énoncés d'exercices arithmétiques, dits « additifs » pour la résolution desquels les seules opérations nécessaires sont des additions et des soustractions. Pratiquement les exercices font intervenir deux entités correspondant à des « Tout », et trois à des partie « parties »[7]. De façon plus schématique, les exercices ont une structure mathématique que l'on peut décrire par -- Tout1 = Partie1 ∪ Partie2; Tout2 = Partie 2 ∪ Partie3 --

Une typologie de énoncés possédant des structures mathématiques similaires, si ce n'est identiques, a été établie.

Pour ce faire, quatre critères de classement ont été définis et permettent de distinguer 24 classes d'énoncés d'exercices:

1 - le type de variable calculée: des effectifs, des prix ou des âges.

2 - le mode de calcul de valeur d'un attribut inconnu de la seconde équation: par complément de la valeur d'une partie à un tout (par exemple: *Quand la famille Robert qui a 6 personnes et la famille Dumas vont au marché ils sont 11 personnes*) ou par comparaison avec une valeur (par exemple: *Dans la famille Robert ils sont 2 de plus que dans la famille Dumas qui a 4 personnes*.)

3 - la valeur calculée dans question finale de l'exercice est-t-elle relative à un attribut portant sur un *tout* ou bien sur une *partie*[8] ?

4 – L'existence dans l'énoncé d' une question évoquant la nécessité d'un calcul intermédiaire.

Nous avons réalisé un classifieur permettant d'affecter un énoncé à l'une des 24 classes d'exercices définies au moyen des critères précédents. Pour ce faire, nous avons défini un ensemble de graphes Intex (M. Siberztein, 1993) décrivant des transducteurs qui recherchent dans l'énoncés des modèles correspondant aux différentes valeurs des critères de classification des énoncés et récupèrent les motifs qui filtrent avec ceux ci pour proposer une structure de l'énoncé. En l'état actuel, si le classifieur permet l'affectation correcte des énoncés de notre corpus il reste cependant très spécifique de ceux et nos travaillons à la généralisation des critères de classification exprimés par les graphes.

Conclusion

Nous avons présenté PIAGET, Plateforme Informatique d'Aide à la Génération d'Enoncés Textuels, plateforme en cours de développement qui vise à proposer des outils permettant de reformuler et à terme d'engendrer des exercices arithmétiques additifs de type « partie-tout », s'adressant à des enfants du cours élémentaire 1 et 2. Cette plateforme est conçue principalement en direction d'un public de psychologues et de didacticiens afin de leur permettre de produire des énoncés d'exercices. PIAGET, vise à leur permettre de bien contrôler les différents paramètres des énoncés susceptible d'avoir une influence sur la difficulté de résolution des exercices qu'ils décrivent.

Dans son état actuel, PIAGET permet de nombreuses manipulations lexicales telles que différentes formes de reformulation lexicales, la définition et la structuration de vocabulaires, la définition et le stockage de « familles » d'exercices, et la classification d'exercices dans une typologie définie par les psychologues et les didacticiens avec lesquels nous travaillons.

La poursuite du développement de PIAGET, passe maintenant

---

[5] On définira par exemple l'attribut prix puis c'est le fait qu'il s'applique à une entité qui en fait une entité du problème.

[6] Une des justification d'un tel choix, correspond à privilégier les propriétés parce qu'elle correspondent aux variables de l'exercice alors que les entités auxquelles elles sont rattachées dépendent de l'habillage de l'exercice: les propriétés arithmétiques (par exemple la distributivité de l'attribut sur les élément d'un ensemble, ou la possibilité de cumuler ses valeurs pour en calculer la valeur sur une ensemble) sont des caractéristique de la propriété et non des objets sur lesquels elles porte.

[7] et éventuellement une « différence », cette dernière servant alors à définir la valeur d'un attribut pour une des entités par comparaison à la valeur de cet l'attribut pour une des autres entités

[8] Un tout signifie que la réponse de l'exercice doit être une addition alors qu'une partie doit être une soustraction.

- par la conception de types de reformulations plus élaborées qui reposent, par exemple, sur une analyse syntaxique des énoncés,
- la généralisation des mécanismes de classement des énoncés,
- l'intégration d'un module d'aide à la résolution en direction de l'élève comportant une mise en évidence dans l'énoncé, des éléments abstraits caractérisant l'exercice (*collections*, *propriétés*, type de propriété…).

Enfin nous visons à terme la mise au point d'un mécanisme de génération d'énoncés d'exercices permettant de moduler la difficulté des exercices en nous appuyant sur les résultats expérimentaux obtenus par nos collègues psychologues et didacticiens.

Bibliographie

1. Amghar, T. (1996) *Une contribution à la modélisation informatique du phénomène de paraphrase: le jugement de paraphrase et la métonymie*. Thèse de doctorat. Université de Nantes.
2. Amghar, T., & Levrat, B. (1995) *Using synonymical rephrasing as a way to cope with polysemy of grammatical categories: an application to the characterization of meaning of some French modal auxiliary*, DIALOG'95, International seminar on computational linguistics and its applications, PP 30-38, Kazan, C.E.I.
3. Baeza-Yates, & Ribeiro-Neto, B. (1999): *Modern Information Retrieval*, 513 pages, Addison-Wesley, 1999.
4. Brissiaud R. (2002) Psychologie et didactique: choisir des problèmes qui favorisent la conceptualisation des opérations arithmétiques. In J. Bideaud & H. Lehalle (Eds): *Le développement des activités numériques chez l'enfant,* 265-291. Paris: Hermes.
5. Porcheron J.-L. (1998) *Production d'inférences dans la résolution de problèmes additifs*. Thèse de doctorat. Université de Paris 8.
6. Fayol,M. & Abdi, H. (1986). Impact des formulations sur la résolution de problèmes additifs chez l'enfant de 6 à 10 ans. *European Journal of Psychology of Education, 1,1,* 41-58.
7. Fuchs, C. (1994): *Paraphrase et énonciation*, Ophrys, 1994.
8. Fuchs, C. (1987): L'ambiguïté et la paraphrase: *opérations linguistiques, processus cognitifs, traitements automatisés*, 330 pages, Presses de l'université de Caen.
9. Hegarty,M., Mayer, R.E. & Monk, C.A. (1992).Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology, 87,1*,18-32.
10. Richard, J.-F., & Sander, E. (2000). Activités d'interprétation et de recherche de solution dans la résolution de problèmes. Dans J-N. Foulin & C. Ponce (Eds.), *Lire, écrire, compter, apprendre: Les apports de la psychologie des apprentissages, pp. 91-102*. Editions du CRDP de Bordeaux.
11. Sander, E. (2001). Solving arithmetic operations: a semantic approach. *In proceedings of the 23rd Annual Conference of the Cognitive Science Society. Edinburgh,* pp. 915-920.
12. Sander, E., & Richard, J.-F. (1997). Analogical transfer as guided by an abstraction process: the case of learning by doing text editing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 1459-1483.
13. McKeown,C.(1983): *Paraphrasing using Given and New information*, AJCL, Vol. 9, n° 1, pp. 1-10.
14. Mathieu, Y. & Sabatier, P.,1986): Interfacile: Linguistic coverage and query reformulation, Colling.
15. Melc'uk,I., (1992): *Paraphrase et lexique: La théorie sens-texte et le dictionnaire explicatif et combinatoire*, in Dictionnaire Explicatif et Combinatoire du français contemporain, pp. 9-58, Les presses de l'université de Montréal.
16. Reusser, K. (1989). From text to situation to equation: Cognitive simulation of understanding and solving mathematical word problems. *European Journal in an International Context, 2, 2,* 477-498.
17. Riley,M.S., Greeno,J.G. &Heller,J.I. (1983). Development of problem solving ability in arithmetic. In H.P. Ginsberg (Ed.).
18. Silberztein, 1993. *Dictionnaires électroniques et analyse automatique de textes: le systeme INTEX*. 240 p., Masson Ed.: Paris.