

# Analyse de forums dans la formation à distance

**Nadine Lucas\***, **Mohamed Sidir** et **Emmanuel Giguet\***

*Laboratoire Paragraphe (Paris8), université de Picardie Jules Verne Amiens*

*\*GREYC, CNRS UMR 6072, Université de Caen, 14032 Caen Cedex*

**Nadine.Lucas@info.unicaen.fr**

**Sidir@u-picardie.fr**

**Emmanuel.Giguet@info.unicaen.fr**

## **Résumé :**

Deux forums d'étudiants en situation d'enseignement à distance sont étudiés dans le cadre de la rédaction collective de devoirs par trinômes. Les discussions accompagnant la réalisation de la tâche ont été analysées manuellement, du point de vue pédagogique pour évaluer le déroulement de la collaboration et du point de vue linguistique pour évaluer le discours collectif. Les forums ont également été analysés automatiquement par le logiciel ThemAgora. Ces trois éclairages convergent et se complètent dans la mise en lumière de phases de développement des forums collaboratifs. L'analyse automatique souligne les moments conflictuels des forums avant l'établissement d'un consensus.

**MOTS-CLÉS :** forums, documents collaboratifs, pédagogie, EAD, analyse de discours, mise en forme matérielle, TAL robuste, thématique, communication.

## **Abstract :**

E-learning students' fora were studied both manually and by computer. Physically distant students were asked to collaborate to write a program in small groups. Tutors evaluated the records of their discussions to assess the collaborative process and its phases. Discussions were also studied as a collective discourse directed by a goal. Computer parsing based on a discourse model was achieved. Results back some of the human findings but also contribute to show how collaborative discussion evolves through disagreement before reaching consensus.

**KEY-WORDS :** discussion groups, e-learning, collaborative writing, computer discourse parsing, text structure, robust parsing.

# 1. Introduction

La diffusion croissante des environnements virtuels partagés, de type synchrone et asynchrone, a renouvelé les problématiques de la communication dans les communautés. Le recours à la collaboration à distance revêt un grand intérêt tant dans le monde professionnel que dans le cadre de l'éducation et de la formation. Ceci pose une série de questions de recherche insistantes du point de vue des changements pouvant survenir dans la communication et leurs impacts sur les processus de collaboration et d'acquisition des connaissances dans des situations éducatives formelles ou informelles. L'analyse automatique des forums est menée dans le cadre d'une équipe de recherche en technologie de l'éducation, Calico<sup>1</sup>. Elle a pour objectif de faciliter le suivi d'un nombre important de formations sur une plate-forme d'enseignement entièrement à distance ou dans un dispositif hybride (enseignement en présence et à distance). Au stade actuel, c'est une aide attendue pour faciliter la réflexion *a posteriori* des tuteurs sur une formation en ligne.

Si les nouvelles formes de communication électronique commencent à faire l'objet de recherche [Anis 1998, Véronis & Guimier 2004, Karlgren 2006], les problèmes de collaboration sont principalement abordés dans une visée pédagogique [Lewkowicz & Marcoccia 2004, Bruillard & Baron 2006]. Très peu de travaux portent sur l'analyse automatique de forums de discussion, qui posent de nombreux problèmes aux analyseurs classiques du fait de leur fantaisie tant orthographique que de ponctuation [Torzec 2004]. Les forums sont ainsi traités essentiellement de manière statistique [Farrel 2001, Newman 2002, Bigi & Smaïli 2002]. La recherche d'information dans les forums à fin d'indexation ou de résumé suit [Vert 2001, Klaas 2005]. D'autres analyses à partir des échanges de contributions mettent en valeur les réseaux sociaux [Reffay 2005]. À notre connaissance, il n'existe pas de traitement automatique portant sur le discours des forums et a fortiori intégrant l'aspect collectif du discours à plusieurs voix.

Nous avons abordé cet aspect et mené des expériences d'analyse automatique sur des forums d'apprenants. Nous nous intéressons ici à l'analyse qualitative de deux mini-forums orientés par une tâche d'écriture d'un cahier des charges informatique. Ces forums ont la particularité de concerner trois participants seulement et d'être limités dans le temps (un mois). Le nombre de contributions échangées par trinôme dépasse cependant largement la centaine (157 et 129 respectivement), ce sont donc des forums très actifs. Un outil d'analyse automatisée, ThemAgora, a été mis en œuvre. Les résultats de l'analyse automatique sont confrontés à l'analyse de discours manuelle et à l'analyse de la collaboration menée par les formateurs.

Nous présentons le contexte des forums et le cadre de l'expérimentation dans un dispositif de Formation Ouverte et à Distance (FOAD) de niveau universitaire, puis les présupposés de l'analyse de discours appliquée aux forums et la mise en œuvre du logiciel ThemAgora. Les résultats sont discutés et comparés avec ceux de l'analyse manuelle. En conclusion, nous envisageons différentes pistes pour améliorer le suivi des activités collaboratives à travers les forums à visée pédagogique.

---

<sup>1</sup> Communautés d'apprentissage en ligne, instrumentation, collaboration. Cette ERTé réunit cinq laboratoires de recherche et six IUFM.

## **2. De la communication aux processus collaboratifs**

### ***2.1. Le forum de discussion : un média, une forme écrite et une pratique de communication***

L'écriture collective en réseau est le résultat d'une tension entre un espace privé d'écriture individuel et un espace public de diffusion et de communication. Le premier espace est le système d'écriture à travers un écran d'ordinateur, basé essentiellement sur un logiciel de traitement du texte et le deuxième espace est présenté ici par le forum de discussion ; un lieu permettant une écriture commune, construite au travers d'une collaboration qui aurait pour propriété de profiter des possibilités techniques du réseau, tout en fonctionnant dans une logique de groupe.

Ainsi, le forum de discussion se présente comme un outil simple d'utilisation et favorisant la collaboration. Il renvoie à des situations de communication interpersonnelles n'obligeant pas le scripteur à manipuler des logiciels d'écriture complexe et de mise en formes spécifique. Le forum de discussion met cependant en scène une situation de communication très particulière qui repose à la fois sur une communication écrite asynchrone sous forme de messages qui tend à rapprocher ce type de communication des situations classiques de production écrite comme le courriel et d'une forme de conversation interpersonnelle qu'on retrouve dans le chat. Il présente une forme d'écriture informatisée qui se conjugue intimement à des processus de communication se trouvant inscrits sur l'écran. Ainsi, les écrits dans les forums de discussion présentent à la fois des marques d'oralité communicationnelle (style informel) et des marques d'écrit soutenu (bibliographie). Marcoccia [1998] explique la relation entre les écrits médiatisés par ordinateur et la communication écrite par un cadrage cognitif permettant « de faire du face à face avec l'écrit ». Pourtant les marques d'une partie du matériau sémiotique disponible dans la conversation en face à face disparaissent naturellement avec la conversation médiatisée par ordinateur pour laisser la place à un certain nombre de procédés permettant de représenter le non-verbal et le para-verbal (l'intonalité et la mimogestualisation) ayant des fonctions comparables aux données en face à face : les smileys ou émoticons, la ponctuation expressive ...

Une autre caractéristique du forum de discussion réside dans son asynchronicité qui relève des dimensions de coterporalité et de simultanité. Utiliser dans un cadre éducatif cette asynchronicité du forum couplé à une permanence des messages permet de parler à la fois d'extériorisation et de partage de la cognition qui semblent favoriser l'apprentissage [Sharples et Pemberton 1990, Legros et Crinon 2002, Mangenot 2002]. Mais la propriété particulière du forum de discussion réside dans son caractère public. En effet, l'écrit asynchrone public fait du forum l'équivalent d'un texte en perpétuelle évolution et enrichissement où les caractéristiques stylistiques de la langue écrite semblent être conservées en majorité, sans doute à cause de ce caractère permanent, public des textes produits [Peraya, 2005], ce qui modifié profondément le cadre communicationnel comme le signale Marcoccia [1998, p. 17] :

*« dans un forum de discussion, il est impossible de sélectionner un destinataire. Toute intervention est « publique », lisible par tous les participants au forum, même si elle se présente comme la réaction à une intervention initiative particulière. L'aparté est impossible : le polylogue est la forme habituelle du forum et le multi-adressage en est la norme ».*

Reprenant à son compte les positions de Marcocchia, Mangenot [Mangenot 2002] considère que le polylogue constitue une des caractéristiques majeures de ce type de dispositif de communication. En revanche, Peraya faisant référence à [Ducrot 1980] et à ses analyses des interactions verbales dans une classe virtuelle, nuance cette affirmation dans la mesure où une approche polyphonique de la communication permet de faire la distinction entre les allocutaires qui sont ceux à qui les paroles sont dites et les destinataires auxquels s'adressent réellement les actes de langages.

Cette description montre l'intérêt des forums de discussion comme outil de communication et de collaboration. Il nous semble important d'observer ces dispositifs d'écriture et de lecture dans un contexte de communication éducative particulière que nous décrivons dans la section suivante.

## **2.2. Contexte de l'étude**

Le dispositif de formation ouverte et à distance (FOAD) exploité à l'université de Picardie Jules Verne a été mis en place depuis 1995. Il est composé de contenus en ligne et de services pédagogiques dont le tutorat constitue la majeure partie. Ce dispositif compte dix formations diplômantes pour plus de 1000 apprenants dispersés géographiquement en France et dans quelques pays francophones.

Le service pédagogique est basé sur un tutorat individuel où le tuteur est considéré comme un accompagnateur, suffisamment disponible pour répondre aux questions de l'apprenant lorsque celui-ci rencontre des difficultés dans sa formation. Or, comme le remarquent [Bruillard, D'Halluin et Weidenfeld 2003], dans un dispositif de formation à distance, la mise en ligne des ressources pédagogiques, assortie d'un simple échange de questions-réponses entre apprenant et tuteur n'est qu'une forme très réductrice du processus d'apprentissage. C'est la raison pour laquelle des expérimentations ont été mises en place pour mettre à l'épreuve des modèles pédagogiques notamment l'Apprentissage Collectif Assisté par Ordinateur (CSCL) [Sidir 2004, 2005, 2006].

La démarche pédagogique consiste à faire travailler les apprenants par groupes de trois sur des projets planifiés sur des périodes de un à six mois, basés exclusivement sur le travail collaboratif à distance. Pour soutenir leurs travaux, nous avons mis en place un environnement technique attaché à la plate-forme INES [Sidir, 2003] permettant aux membres du même groupe de communiquer, d'organiser et de partager des idées, des contributions et des informations. Cet environnement se présente sous la forme d'un ou plusieurs forums de discussion, intégrant à la fois un service d'édition et de publication de documents et un système de vote pour une prise de décision « démocratique » et collective.

Le corpus de notre étude est extrait de ce type de forums inséré dans un scénario d'écriture collective d'un document numérique sous forme d'un cahier de

charges. La suite de cet article s'intéresse aux formes caractéristiques du discours médiatisé s'inscrivant dans une perspective d'analyse pourtant sur les processus collaboratifs. On considère alors l'ensemble des interactions dans le forum comme un texte, un corpus unique prêt à être analysé. La section suivante décrit la méthodologie relative à cette analyse.

### **3. Analyse du discours collectif**

#### ***3.1. Analyse linguistique du discours collectif***

Les forums sont envisagés dans notre approche non comme une somme de contributions individuelles, mais comme un discours collectif [Peraia 1999, 2005]. Les cadres théoriques que nous avons utilisés sont basés sur les propriétés du discours écrit, qu'il soit individuel ou collectif. Les forums sont ainsi analysés de la même manière que des articles ou ouvrages qui peuvent avoir plusieurs auteurs.

Pour mettre en valeur la construction d'un développement collectif à partir d'un problème soumis par les tuteurs, nous nous sommes basés sur la théorie de l'exposition : le modèle de référence est celui de Yamada (1873-1958), un linguiste japonais qui définit des opérations de mise en discours inspirées par la philosophie allemande. Il présente des opérations cognitives, qui peuvent être vues comme une série de contraintes qui limitent *de facto* un exposé qui serait sinon un développement infini [Yamada 1936]. On va donc délimiter des passages de texte, comme ayant une cohésion interne, mais il faudra aussi hiérarchiser ces passages, puisqu'il est possible qu'un exposé contienne des développements subordonnés ou incisés, qui interrompent momentanément la progression du discours. L'intérêt principal d'une telle théorie est qu'elle fait appel aussi à la perception (de la mise en forme du texte et du style). Elle a une portée générale puisque le discours est un phénomène universel et s'applique aisément à des textes en langues occidentales.

Pour mettre en valeur les aspects dialectiques, la construction d'un récit collectif d'expérience, nous nous sommes appuyés sur la théorie de l'énonciation et sur la poétique de Jakobson [Jakobson 1971, 1973]. Ce qui nous intéresse est davantage le rapport des apprenants à l'objet traité que les relations des apprenants entre eux. Dans cette théorie, le problème (l'énoncé) est considéré en tant que protagoniste du discours, quoiqu'il ne s'agisse pas d'une personne mais d'un objet de transaction. Cependant il est moteur de la discussion. Autrement dit, le problème posé est à résoudre et « actif » pendant une certaine période, en particulier, il peut diviser les apprenants. À la fin des forums, on observe le changement de statut du problème, en tant que protagoniste du discours, autrement dit le problème posé devient résolu donc « passif ». Ainsi les divergences et le consensus sont envisagés comme des moments signalant l'émergence de difficultés et de solutions.

### 3.2. *Algorithme et implémentation*

Les forums sont formatés par une DTD partagée minimale, permettant l'échange de forums, XMLForum<sup>2</sup>. Elle décrit le forum comme document, constitué de contributions, elles-mêmes segmentées en paragraphes. La ponctuation n'est pas très stable, aussi nous ne donnons pas une définition normative mais contextuelle de ces segments. Le logiciel ThemAgora traite des forums sous ce format-pivot. Il est intégré à la plate-forme « Wims » [Giguet 2005] et il fonctionne en ligne.

ThemAgora est un logiciel d'étude dérivé à l'origine du logiciel UniThem, basé sur la théorie de l'exposition de Yamada et utilisé pour les articles de presse [Lucas & Giguet 2005]. Il a été adapté à la segmentation particulière des forums et à leur ponctuation. L'analyseur est descendant du discours jusqu'aux composants des contributions, le grain le plus fin équivalant à des phrases.

L'algorithme fonctionne sur le principe de la division, en respectant le modèle cognitif théorique qui suppose une structuration ordonnée ayant un rapport de forme « raisonné » : le thème se trouve avant le rhème et il est proportionnellement plus court que le rhème ; le rhème peut faire l'objet de développements subordonnés, et il peut à ce titre faire l'objet d'une nouvelle division. Cependant, le nombre de sous-développements doit également rester cohérent avec l'ensemble. Le programme établit un segment thématique au niveau global et son développement ou rhème (niveau 1). Le rhème est divisé en sous-thèmes et sous-rhèmes de niveau 2, eux-mêmes divisés le cas échéant en thème et rhème de niveau 3. L'affichage des sorties sous forme graphique indentée et coloriée utilise le module d'UniThem.

La frontière entre les segments thématiques et rhématiques est concrètement déterminée par une fonction d'identification de marques contrastives (en présence/absence). Ces marques sont de nature typo-dispositionnelle (ponctuations, mises en forme matérielle, ...) et linguistique (connecteurs de discours).

ThemAgora a été conçu pour traiter des forums. Pour tenir compte de l'absence d'orthographe, les contraintes morphologiques sur les connecteurs ont été relâchées et les règles exploitant les ponctuations particulières, comme les « émoticons » ou les ponctuations multiples, fréquentes et répétées dans les forums, ont été renforcées. Cette adaptation est possible car l'algorithme n'est pas dépendant d'un lexique fourni en entrée. Les seuils de fréquence sont gérés automatiquement. Les règles reflètent les attendus des modèles théoriques utilisés, mettant l'accent sur des propriétés du discours interprétables à la fois dans le cadre de l'exposition et dans le cadre de la communication à travers un problème.

Dans la conduite de l'expérience, le document analysé est variable, pour tenir compte des attendus de l'analyse de discours et permettre un aller-retour entre interprétation et modèle d'analyse [Lahlou 1995]. Ainsi, les forums ont été analysés séparément par tâche ou ensemble pour un trinôme, et avec ou sans l'énoncé de départ.

---

<sup>2</sup> Les spécifications adoptées dans le cadre de l'ERTé Calico seront prochainement publiées.

## 4. Résultats et discussion

### 4.1. Analyse automatique

Les résultats de l'analyse corroborent en grande partie les analyses manuelles. Lorsque l'énoncé du problème est intégré au document, il est en fonction thématique, mais lorsque les forums sont analysés indépendamment comme tels ou par étapes, la structure de niveau 1 remonte au niveau global. Faute de place, étant donné la taille des forums, nous nous concentrons sur l'analyse des sous-ensembles établis par étapes et en l'absence de l'énoncé de départ. Les résultats sont ici présentés pour les étapes 1 correspondant à l'écriture collaborative d'un cahier des charges pour les deux trinômes. Nous envisageons ici uniquement les niveaux 0 (global) 1 et 2. Les unités de niveau 3 sont plus fines et leur discussion est omise.

#### Résultats de l'analyse du trinôme 1

Les résultats de l'analyse automatique sur l'étape 1 du trinôme 1 travaillant sur l'« école » (85 contributions pour l'étape 1 sur 157 au total) montrent que le segment thématique détecté est très long. Il correspond à la phase d'exploration ou d'initialisation de la discussion collective, à l'appropriation du problème posé.

L'analyse automatique distingue 4 unités de niveau 2 ou sous-thèmes à l'intérieur de ce rhème, correspondant à des moments dans le déroulement chronologique du forum. Les deux premières unités de niveau 2 correspondent à un moment de discussion ou à l'élaboration d'une solution. Les unités 1 et 2 Ce moment se termine par un accord. La troisième unité correspond au remaniement de cette proposition, elle marque un désaccord, c'est le moment de « dramatisation ». Dans cette partie, les participants se disputent. La quatrième et dernière unité marque le consensus final. La comparaison est peu développée et n'apparaît pas en tant qu'unité autonomisée par l'analyse automatique. L'analyseur souligne une répétition de déclaration *Je valide a priori et par avance vos modifications du cahier des charges*.

#### Résultats de l'analyse du trinôme 2

Pour le trinôme travaillant sur la gestion documentaire de bibliothèque (81 contributions pour l'étape 1 sur 129 au total), le schéma est similaire. Le segment thématique détecté en l'absence de l'énoncé de départ est également très long (29 contributions sur 81). Le développement ou rhème commence lorsque la discussion reprend des questions déjà abordées, ce qui est conforme à la sémantique du modèle linguistique de l'exposition et souligne l'intérêt du modèle stylistique sous-jacent. On y trouve des reprises (*Dans ton document...*) interprétables à la lecture, mais rappelons que le programme ne calcule pas d'anaphores. On remarque aussi que le rhème correspond bien au changement de statut cognitif du problème, qui n'est plus « sujet » mais « objet » pour les apprenants.

Le rhème (52 contributions) est subdivisé en 5 unités très inégales en longueur recouvrant, à quelques contributions près par rapport à l'analyse manuelle, des moments principaux : la discussion (G1 et G2), la dramatisation (G3), la comparaison (G4) et la clôture (G5). On note en particulier que le premier thème de niveau 2 (le début du rhème) coïncide en fait avec les premières confrontations sur

la solution, par exemple *je suis absolument contre ... t-ai-je convaincu ??*. Le dernier moment correspond au contraire à l'accord collectif sur la dernière version du cahier des charges.

<i>G0 Thème niveau global</i>	
bonjour ! quand est-ce qu'on se voit pour en parler ? le plus tôt possible !!!! [...] je n'ai qu'un type de compte. il n'y aura qu'un compte ADMIN (id=1 ou 0) par exmple, le reste sera des membres.	
G1	Dans ton document, il semble que les emprunteurs soit également des utilisateurs indentifié du système ? [...] Si la bibliothécaire doit modifier les informations de ses membres (une adresse par exemple) , faut pas qu'elle court appeler l'administrateur., idem pour ajouter un auteur... t-ai-je convaincu ?? [#] Si c'est le cas cela v nous compliquer l'affaire pour la gestion des disponibilités des ouvrages ... [...] et si oui combien.
G2	-> Je ne pense pas que ca soit si simple que cà: si on veut gerer des dates d'emprunts et de retour pour chaque exemplaires d'une même ouvrages, [...] Impressionnant :) Très joli travail [#] j'ai modifie le CDC et j'ai mis mes remarques dedans en rouge et jaune. tu peux dire si t'es d'accord ?
G3	Ben apparemment déjà on est pas d'accord sur ce qu'on doit faire ... [...] qu'un login pour accéder à la partie admin [#] c'est un petit projet. je ne pense pas qu'il faut faire une chose comme ça. [...] à nous de le réaliser au mieux... comme des pros ^^
G4	Je pense que nous pourrions combiner la localisation des ouvrages dans le centre (qui pourrais paraître superflu) et le numéro unique apposé sur les ouvrages (pas très parlant pour le public), [...] roman pour un roman qui parlerai de la vie quotidienne des paysans au 14ème siècle par exemple... [#] pas plus que le placement dans le centre ... [...] les bibliothécaires serai libre de mettre ce qu'il veulent suivant la classification qu'il veulent ... (c juste qu'on leur donne la possibilité ...)
G5	Voici la nouvelle version de mon cahier des charges integrant quelques unes des suggestions exprimés par Xing. [...] Oui. [#] maintenant c'est ok pour moi [...] mais pas 2002 ^^

**Figure 1. Résultat compacté de l'analyse automatique étape 1 trinôme 2**

Dans l'extrait de la figure 1 compressé faute de place, les segments sont évidés. Le symbole [#] indique des coupures de plusieurs contributions entre la fin d'un thème et la fin du rhème correspondant, le symbole [...] indique des coupures de une ou plusieurs contributions à l'intérieur d'une unité.

Ces résultats ou plutôt leur intérêt sémantique peuvent paraître étonnants, ils sont cependant corroborés par d'autres approches stylistiques [Karlgrén 2000, 2006].



## 4.2. Discussion

L'analyse automatique de forums par le logiciel d'étude ThemAgora a permis de dégager utilement des épisodes structurés dans des forums clos, malgré les difficultés posées par ce type d'écrit. Notre étude est centrée sur des forums guidés par la tâche et limités dans le temps. Le résultat d'analyse des forums de trinômes est satisfaisant et conforme aux résultats précédents obtenus sur une étude de cas, avec davantage de participants (une trentaine). Les analyses automatiques sur cinq forums de ce type ont été favorablement évaluées par les formateurs dans le cadre de Calico, en confrontant leurs analyses manuelles et les résultats de ThemAgora. Cela montre l'adéquation du modèle à ce type de forum au cours desquels le discours progresse. Un plus grand nombre de cas devrait être analysé, dans les mois à venir, grâce à la mise à disposition du logiciel sur la plate-forme Calico.

En revanche, pour des forums libres, ressemblant à une liste de questions/réponses, sans progression dans le temps, l'analyse avec cette grille est moins pertinente, ce qui semble normal puisque le présupposé de l'analyse n'est pas respecté [Sidir et Lucas, en préparation]. La discussion à bâtons rompus ne correspond pas au modèle théorique de l'exposition. Nous avons donc entrepris d'implémenter une version différente du logiciel pour ces forums ouverts.

## 5. Conclusion

L'analyse automatique de forums est tout à fait possible, sans passer par une normalisation de l'écrit, lorsque l'algorithme est fondé sur un modèle théorique d'analyse de discours. Elle est satisfaisante lorsque le modèle est adéquat au type de forum analysé. L'analyse automatique par le logiciel d'étude ThemAgora est basée sur les principes de l'exposition, elle segmente les forums guidés par la tâche en unités inégales en longueur mais très informatives sur le déroulement de la discussion collaborative. Elle met en valeur les moments de dissension et de consensus. Lue dans la grille de Jakobson, l'analyse est également informative sur le rapport des participants à l'objet de la discussion, elle montre les étapes d'appropriation puis de résolution du problème posé.

Nous ne prétendons pas avoir résolu tous les problèmes d'analyse automatique de forums. Outre l'implémentation d'une version pour forums libres, beaucoup de travail reste à faire pour adapter cet outil déjà utile aux besoins des formateurs. Ceux-ci se réfèrent souvent à des grilles d'analyse empiriques issues de leur pratique professionnelle ou à des grilles inspirées par des travaux de didacticiens ou encore par des modèles variés d'analyse de discours. Nous avons entrepris l'étude des grilles d'interprétation familières pour les utilisateurs, dans le cadre de Calico. Une autre direction de recherche plus générale est le calcul dynamique du grain d'analyse pour traiter des forums, en fonction de leur taille et des objectifs des utilisateurs.

Les perspectives de cette recherche sont à court terme de fournir des outils objectifs et partageables permettant d'asseoir et mutualiser les pratiques pédagogiques en ligne. À long terme, le but pratique serait de suivre l'évolution d'un

groupe en contexte d'apprentissage pour gérer les formations en cours et non plus seulement d'étudier les traces de la discussion *a posteriori*.

## 6. Références bibliographiques

- [Anis 1998] *Texte et ordinateur, l'écriture réinventée* J. Anis. Bruxelles, De Boeck.
- [Bigi & Smaili 2002] "Identification thématique hiérarchique : application aux forums de discussion". *TALN 2002*, Nancy. pp. 115-124.
- [Bruillard & Baron 2005] "Study and design of new modalities of learning and work for pre-service teachers assisted by ICT: collaborative learning and case studies" *WCCE 2005*, IFIP, juillet 2005.
- [Bruillard & Baron 2006] *Technologies de communication et formation d'enseignants : vers de nouvelles modalités de professionnalisation* G. L. Baron et E. Bruillard (coord.) Paris, INRP.
- [Bruillard, D'Halluin et Weidenfeld 2003] "Comment appliquer l'apprentissage collaboratif assisté par ordinateur à la formation à distance ? L'exemple du campus numérique Ape-Lac" Actes *Campus numériques et Universités numériques en région Montpellier*, France.
- [Ducrot 1980] *Les mots du discours*. O. Ducrot. Paris, Minuit.
- [Farrel 2001] "Text Summarization and Question Answering: Summarization of discussion groups" R. Farrel, P. G. Fairweather, K. Snyder *Proceedings of the tenth international conference on Information and knowledge management* October 2001 ACM.
- [Giguet 2005] "Modélisation de l'activité expérimentale du chercheur en traitement des langues sur corpus multilingues" E. Giguet *Journée de l'ATALA Articuler les traitements sur corpus*, 12 février 2005.
- [Jakobson 1971] *Word and Language* R. Jakobson. The Hague, Paris, Mouton.
- [Jakobson 1973] *Questions de poétique* R. Jakobson. Paris, Seuil.
- [Karlgrén 2000] *Stylistic Experiments for Information Retrieval*, J. Karlgrén, PhD thesis, Stockholm, Université de Stockholm, 2000.
- [Karlgrén 2006] "Workshop on New Text" April 3, 2006 Trento, Italy J. Karlgrén (ed) <http://www.sics.se/jussi/newtext>
- [Klaas 2005] *Toward indicative discussion fora summarization* M. Klaas UBC CS Technical Report TR-2005-042005
- [Lahlou 1995] "Vers une théorie de l'interprétation en analyse des données textuelles." S. Lahlou *JADT 1995, 3rd International Conference on Statistical Analysis of Textual Data*. S. Bolasco, L. Lebart and A. Salem Eds.. CISU, Roma, 1995. 221-28. Vol. I.
- [Legros & Crinon 2002] *Psychologie des apprentissages et multimédia*. D. Legros et J. Crinon (eds) Paris, Colin université.
- [Lewkowicz & Marcoccia 2004] "The Participative Framework as a design model for newsgroups" M. Lewkowicz et M. Marcoccia. *PartRoOM. COOP 2004* pp. 243-256
- [Lucas 2005] "Les procédés d'exposition et de développement collectif dans un forum pédagogique : le cas Maxime" N. Lucas Symposium *Symfonic*, 20-22 janvier 2005 Amiens. <http://www.dep.u-picardie.fr/sidir/articles/index.php>
- [Lucas & Giguet 2005] "UniTHEM, un exemple de traitement linguistique à couverture multilingue" N. Lucas et E. Giguet *Cide 8 Conférence internationale sur le document électronique* Beyrouth 25-28 mai 2005. K. Zreik (ed). Paris, Europia, pp. 115-132.
- [Mangenot, 2002] "Forums et formation à distance : une étude de cas", in *Education permanente* 152, pp. 109-119.

- [Marcoccia 1998] "La normalisation des comportements communicatifs sur Internet : étude sociopragmatique de la netiquette" M. Marcoccia In *Communication, société et Internet*, N. Guéguen & L. Toblin (éds.) pp. 15-22. Paris, L'Harmattan.
- [Newman 2002] Exploring discussion lists: steps and directions. P. S. Newman. *Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries*, 2002, pp. 126–134.
- [Peraya 1999] "Vers les campus virtuels. Principes et fondements techno-sémio-pragmatiques des dispositifs de formation virtuels" D. Peraya In *Le Dispositif. Entre Usage et concept*. G. Jacquinet et L. Montoyer (eds) Paris, CNRS Editions Hermès, pp. 153-168.
- [Peraya 2005] "Axes de recherches sur les analyses de communication dans les forums" Daniel Peraya *Symfonic*, 20-22 janvier 2005 Amiens. <http://www.dep.u-picardie.fr/sidir/articles/index.php>
- [Reffay, 2005] "Réseaux sociaux et analyse de traces des forums d'une communauté d'apprentissage – Calculer la cohésion. Symposium *Symfonic*, 20-22 janvier 2005 Amiens. <http://www.dep.u-picardie.fr/sidir/articles/index.php>
- [Sharples & Pemberton 1990] "Starting from the Writer : Guidelines for the Design of User-centred Document Processors" In *Computer Assisted Language Learning Vol. 2 / 1990*, Oxford, Intellect. pp.37-57.
- [Sidir, 2003] "e-formation : quel choix technique ?" In *l'enseignement à distance, théories et pratiques*, Acte 8 (ed.) Paris, France. pp. 47-58
- [Sidir, 2004] "Modes de collaborations au sein de groupes d'apprentissage dans une formation à distance universitaire" M. Sidir *TICE 2004*, Compiègne.
- [Sidir, 2005] "La médiation par les TIC de la communication éducative" M. Sidir *H2PTM'05*, Hermès-Lavoisier, pp. 395-408.
- [Sidir, 2006] "Interaction communicationnelle à travers les forums de discussion dans un dispositif de e-formation" M. Sidir In G-L., Baron et E., Bruillard (coord.). *Technologies de communication et formation d'enseignants*, INRP, pp. 235-249.
- [Torzec 2004] "Contribution à l'étude des messages électroniques francophones Quelques résultats et leurs conséquences pour le TALN". N. Torzec *Journée de l'ATALA* du 4 juin 2004. (<http> voir Veronis)
- [Vert 2001] Text Categorization Using Adaptive Context Trees J-P. Vert *Proceedings Computational Linguistics and Intelligent Text Processing CICLing 2001*, Mexico-City, Mexico, February 18-24 2001, A. Gelbukh (Ed.) Berlin / Heidelberg Springer.
- [Veronis & Guimier de Neef 2004] "Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)" J. Veronis et E. Guimier de Neef (eds) *Journée de l'ATALA* du 4 juin 2004 <http://www.up.univ-mrs.fr/veronis/jenfce/index.html>
- [Yamada 1936] *Nihon bunpôgaku gairon [Somme sur la grammaire japonaise]* Y. Yamada, Tôkyô, Hôbunkan, 1936 (ré-imp. 1989).