



HAL
open science

Modeling of a pragmalinguistics analysis of a discussion forum

Goritsa Ninova, Hassan Atifi

► **To cite this version:**

Goritsa Ninova, Hassan Atifi. Modeling of a pragmalinguistics analysis of a discussion forum. 2011 International Conference on Collaboration Technologies and Systems (CTS), May 2011, Philadelphia, United States. 10.1109/CTS.2011.5928693 . hal-01849093

HAL Id: hal-01849093

<https://hal.science/hal-01849093>

Submitted on 16 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modeling of a pragmalinguistics analysis of a discussion forum

Goritsa Ninova, Hassan Atifi

ICD/Tech-CICO, Université de Technologie de Troyes

Abstract-In this paper we present, a modeling of an expertise in pragmatics. We follow knowledge engineering techniques and observe the expert when he analyses a social discussion forum. Then a number of models are defined. These models emphasize the process followed by the expert and a number of criteria used in his analysis. Results can be used as guides that help to understand and annotate discussion forums. We aim at modeling other pragmatics analysis in order to complete the base of guides; criteria, process, etc. of discussion analysis

Keywords- Traceability; Knowledge Engineering; Human-knowledge Interaction; Interactions analysis; Pragmatics; linguistics.

I. INTRODUCTION

This study is a part of a pluridisciplinary research project driven in the Tech-CICO laboratory¹. Our work meets the ever-increasing need for development of a transdisciplinary methodology of digital corpora² management and analysis. In this sense, our main objective is to facilitate sharing and analysis of corpora between specialists in communicative interaction. We are not aiming to create one (too) more hyper interaction model (theory) but to provide an operative framework to an empirical analysis of authentic data by respecting the specificity of each chosen discipline. Our starting point is the description of real practices of constitution and analysis of corpora in interaction disciplines in order to find the appropriate method of human scientists' knowledge capitalization through corpora and corpora interpretation charring.

A. Context of the study

The specific practices of research in human sciences we have observed are mainly characterized as qualitative [Miles and Huberman, 1994]. (1) We have to deal with descriptive and/or comprehensive practices of analysis in which the process of corpora structures elaborating is a part of the interpretative work of the scholar who made pair evaluation [Have, 1986]. (2) Usually researchers don't make use of big technical equipment and a very

large amount of data. On the contrary, detailed handmade analysis often is observed (3) Detailed analysis is made through tools heavily influenced by the culture of the field. (4) Contrary to the practices on corpus linguistics [McEnery and Wilson, 2001], which aims a very strong generalization of results, analysis performed in the Interactional Linguistics field, for example, are local, situated, and dependent on analysts expertise [Mondada, 2006]. Several questions arise about corpus sharing and collaborative analysis: Are these possible? What are the best conditions? And finally, how can we support sharing and circulation of corpora and comments on them between several researchers in these conditions?

For the need of the study we developed a methodology, in order to describe and understand sociological, psychological and linguistic interpretations of verbal interactions and the conditions of theirs exchanges. This methodology combines ethnographic enquiry and knowledge engineering. The stages of our research include: (1) An initial comparative analysis of theoretical pattern of verbal communicative interaction, the chosen criteria being: applied models, aim, granularity of analysis, data approach, etc; (2) A study of real practices of interaction data gathering and analysis which are done in three disciplines: psychology, sociology and pragmatics. Our main approach is internal investigation [Atifi et al, 2006]. These interviews also revealed the tension between the obvious need for cooperation among different researchers around the understanding of the communicative interactions and the difficulties of setting data sharing. (3) We compare the results of the previous investigation with the observations of concrete cases of collaboration between scientists interested in workplace communication: for example we have investigated the complementary work of two researchers, a pragmatic linguist and a sociologist of organizations, both dealing with documents creation, use and circulation in a telecom company. (4) In order to complete the results of this investigation, we used a modeling of analysis practices of communicative interaction [Sidoummou, 2007]. This modelling is based on the approaches related to the knowledge management field.

¹ Tech-CICO is a laboratory who associate humanities and computer scientists, in particular concerning professional organization oriented intervention research. For more information see <http://tech-cico.utt.fr/>

In this paper we focus on some results of stage 4. We chose to present our modeling of an analysis whose object is a socially orientated discussion forum³. We report the choice of models, the protocol of the study and the modeling we are obtaining using CommonKADS methodology. The objective of this modeling is to highlight certain techniques followed in pragmalinguistic analysis of discussion forums. We observed a pragmalinguistics researcher in his work and we modeled the steps to set up script types of interactions in a discussion forum⁴. The models defined will be used to provide elements in order to define guides for tags definition and other type of metadata facilitating comprehension of mediated interactions.

B. Pragmatics of interactions

Pragmatics of interactions, also called interactions analysis, is a new multidisciplinary approach, a synthesis of various disciplines whose object of study is primarily language in situations [Kerbrat-Orecchioni 1998]:

- New communication: transdisciplinary Anglo-Saxon approach that developed in the psychosocial sciences: ethnography of communication, ethnomethodology, Palo Alto School, etc.
- Pragmatics: born with the work of Austin and Searle, which made a radical break with Chomskyan linguistics [Searle, 1979]. [Austin, 1970].

The beginning of the pragmatics of interactions is recent. It is the result of a double movement which affected psychosocial sciences during the early 1960s, and linguistics from the 1970s:

- Progressive breakdown of boundaries between traditional disciplines that deal with language in context (anthropology, sociology, psychology). These American researchers have shown the weakness of autonomous disciplines to properly handle the language in context. “Social scientists asking relevant questions have usually not had the training and insight to deal adequately with the linguistic face of the problem. Linguistics, the discipline central to the study of speech, has been occupied almost wholly with developing analysis of the structure of language as a

referential code, neglecting social meaning, diversity, and use.” [Hymes, 2003].

- Evolution of traditional linguistics towards a new oral, pragmatic and communicative linguistics. This change relates primarily to two dimensions of the language: its communicative vocation of exchange and dialogue and its behavioral and pragmatic nature. Pragmatics of interactions values a sociolinguistic perspective of the object of studies by asking the question: Who speaks? With whom? About what? When? Where? How? And Why? This approach develops an ethnographic methodology based on the exchanges observation, description and interpretation of the exchanges. One of the branches of pragmatics of interactions is conversation analysis. This is about producing a thorough study of what happens in verbal interaction with focus on the sequential organization of turns of speech.

C. Our positioning

We are interested in the way the expert applies these theories on his corpora analysis according to research subject. It is important to understand that the applied theories are chosen by a well noun expert of computer-mediated communication (CMC). We want to know how this expert makes choice of theories according to the corpus specificity, the research topic and his experiential knowledge. We think that the knowledge engineering modeling methods may give us some answers to this question.

The originality of our approach presented in this article is in application of a modeling from knowledge engineering to a scientific work characterized by a qualitative approach. To do this, we needed to apply a top-down method offering a library of templates (models) for modeling analytic tasks. From these models we can interpret the data collected from experts. CommonKADS is a standard for Knowledge Based Systems (KBS) in Europe. We have chosen to ease it offers in the task of knowledge elicitation and reasoning used.

³ The research we observed is part of a pluridisciplinary project of analysis of discussion forum. The common topic of studies is online social support.

⁴ In this paper, we use the term “discussion forum” to refer to “message board”.

II. ANALYSIS FRAMEWORK

The pragmalinguistics researcher “Michel Marcoccia” analyzed extracts from the forum Doctissimo⁵ were messages requesting social support. This analysis was carried out in collaboration with “Nadia Gauducheau”, a cognitive psychology researcher. Michel Marcoccia's aim is to provide corroboration for analysis through psychosocial theories regarding social support request messages. He has as main aims on the one hand, to validate the psychosocial categories on the exchanges of social support and on the other hand, to define standard scripts for each category [Gauducheau, Marcoccia, 2007].

The extracts of exchanges corresponded mainly to a request (the initial post of the thread) and discussions concerning this request (the displayed list of the other posts). As an example of these exchanges, we can quote: exchanges on the addiction to drugs or exchanges about allergies to milk, etc. Titles of these exchanges correspond of threads' titles.

Discussions on addiction to drugs:	
Orchidée	kikou,
Noire	Voilà je me lance enfin à ouvrir mon premier post. Ca fait des semaines que j'erre sur le forum, poste des messages par-ci, par-là, mais je n'arrive pas à me poser. Je ne connais encore que très peu de personnes, donc je vais faire une p'tite présentation : J'ai 33 ans et je suis Après avoir consommé pas mal de produits, j'ai fini par trouvé mon bonheur (et mon malheur aussi) dans l'héro... à laquelle je suis dépendante maintenant depuis 8 ans. [...] Je suis aussi dépressive et hyper-angoissée. Quand j'arrête, je prends conscience de la réalité et ca m'est insupportable. Je ne demande rien, je me sens seule avec mes problèmes. Vous pouvez venir ici pour me parler, me raconter votre histoire ou juste pour dire ce que vous avez sur le coeur... Amicalement,
Profil sup primé	Bienvenue Orchidée. Je ne suis pas très loquace aujourd'hui, mais je connais bien le problème de l'héro. Je suis encore loin de m'en être "déarrassée", bref ... Si j'ai un peu de courage plus tard, lolll.

⁵ <http://forum.doctissimo.fr/>

This forum is part of an extremely popular French website dedicated to health information, Doctissimo. This forum is considered by researcher as a forum where the charring information activity is central: the displayed aim of Doctissimo, the exchange of medical information is announced upon entry into the site.

	Encore bienvenue et @+ Marie
DianaC	Bonjour, désolée, je ne connais pas ton problème, mais si c'est un appui ou une écoute que tu cherches, je t'ai lue. Je te souhaite bon courage, pour moi c'est l'alcool et le shit (en conso excessive et "famélique"). Voilà. Diana Bon courage!
Discussions on milk allergies:	
cartoon181	Salut à tous, Je m'interroge sur une éventuelle intolérance aux produits laitiers (quel composant exactement, je ne sais pas) chez moi. Je m'explique : J'ai longtemps remis en cause mon alimentation dans ma légère acné qui persistante malgré mes 22 ans. Je ne mange pourtant pas très gras, mais je me tappe souvent des éruptions cutanées sur le front, les tempes, qui sortent de je ne sais ou, sans raison precise [..]
go	C'est difficile à affirmer que les produits laitiers provoquent de l'acné systématiquement, mais ... cela peut en induire chez certaines personnes plus facilement [...]
joely	Bonjour, je suis a peu près persuadé que le lait en particulier de vache, est mauvais pour la santé. Une idée reçue, répandue par la puissance des lobbys pousse à croire qu'il apporte du calcium [...]

Fig. 1. Extracts of the studied exchanges ⁶.

Psychosocial categories used in this analysis are: requests for emotional support, request for sharing experiences, request for evaluation, and request for informational support and for advice. The categories are derived from psychological and sociological theories about the forms of social support [Gauducheau, Marcoccia, 2007].

III. OBSERVATION OF THE ANALYSIS

We used knowledge engineering techniques to analyze and model the expertise of forum analysis. Knowledge Engineering (KE) provides essentially a process of analysis and modeling of problem solving [Charlet, 2003], [Aussenac et al, 1996]. Methodological guides are given in this discipline. These guides make possible to collect an expertise and to model it in order to highlight the “why”, “what” and “how” of a behavior.

⁶ For a translation, see Appendix 1.

In our case, we followed knowledge engineering method of collecting and modeling. We observed the analysis activity performed by the pragmalinguistics expert following a well-established scenario based on ergonomics' activity observing techniques. The first step is to clearly identify the objective of the task to be observed, followed by a series of observations, and an analysis of the observations in order to define a set of behavioral patterns, the final stage to be performed being a validation of those observations with the expert in pushing him to explain the reasons of his behavior..

We alternated observations with validation and explanation interviews in order to progressively understand the reasons of the expert's behavior at each stage of analysis

IV. MODELING OF THE ANALYSIS

The reasoning followed by the researcher expert in pragmalinguistics, is mainly abduction oriented. In fact, he analyses a sample of messages on social support requests, he makes a number of assumptions in the form of scripts of analysis and finally he validates these assumptions' application on other scripts (Figure 2).

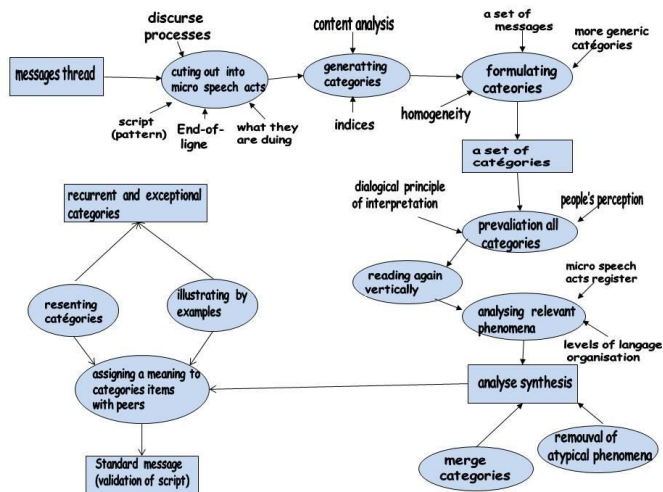


Fig. 2. Analyses of interactions.

If we look more closely at researcher's scripts identifying analysis, we find out that he proceeds by a series of assessments based on pragmalinguistics theories [Kerbrat-Orecchioni, 1998], [Hymes, 2003]. These theories serve him as the criteria in his analysis. The generic CommonKADS assessment model [Breuker

et al., 1984] defined in knowledge engineering enabled us to identify that the expert establishes classes of decisions (which are scripts relative to each type of request) on the basis of standard models. Once the interactions are cut out into threads (request, replies and replies to replies), he compares the thread with the standard models by using the criteria of analysis (Figure 3.). We note that this analysis is done in a manual way.

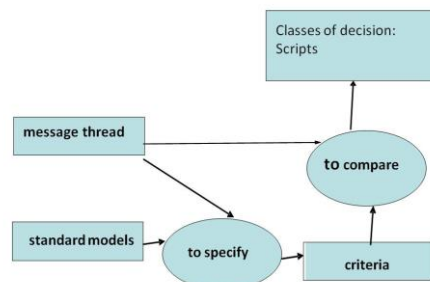


Fig. 3. Assessment of interactions threads

While analyzing the messages, he seeks at the same time to identify those nonclassable in the identified categories or which are exceptions scripts. That allows him to state new hypotheses on social support message boards.

D. The used pragmalinguistics criteria

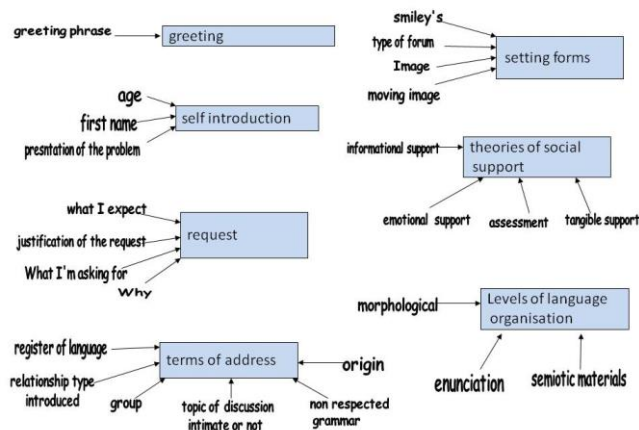


Fig. 4. Assessment of interactions threads

In our modeling we studied the detailed analysis of the messages. After identifying the outline of his reasoning and the criteria he uses, we observed how the

researcher proceeds in his analysis for cutting out interactions and establishing scripts. We present in illustrations the traces of the drafts of his analysis in order to show the steps that he follows

E. Exchanges analysis performed by the expert

The interactions are cut out in the thread of exchanges corresponding to each request. Each thread includes a request and the replies to this request. The thread is then quickly analyzed in order to identify its type: Request for Emotional Support, Request for Information, Request for Advice or Request for Sharing Experiences, etc. The request is then analyzed in order to identify its type and various indices such as: recipients of the request,

according to the contents of reactions. For example, some types of reactions were identified as: encouragement and compliments, criticisms and disagreements, advice and contributions to the information, evaluation of the situation and request additional information, evaluation of the expertise and sharing close experience etc.

F. Interactions scripts identification performed by the expert

	Thread 1 Initiative message Request for Emotional Support	Thread 2 Initiative message Request for Sharing Experiences	Thread 5 Initiative message Request for Tangible support	Thread 6 Initiative message Request for Information support and advice	Thread 8 Initiative message Request for Information support and Sharing Experiences
Recipient of the request	X	X	X	X	X
Greeting conversation opening	X	X	X		X
Address term	X	X			
Description of the activity in the forum	X	X			
Identity (age, origin, etc.)		X			
Presentation of the problem		X		X	X
Failure in problem solving		X			
Psychological state	X	X		X	
Health status					
Formulation of the query	X	X	X	X	X
Description of expected benefits	X				
Terms of the exchange and the support					
Vows (countergift)					
Thanks by anticipation			X		X
Closing of the exchange		X	X		X
Signature	X				
Sayings, proverbs, quotes					
Formating smiley's, images		X			

Fig.5. Cross grid of initiative messages

terms of address, presentation of the transmitter, clarification of what is required, greetings, signature, etc. [Gauducheu, Marcoccia, 2007].

The reactions to the message are then analyzed to identify its canvass: terms of address, greetings and signatures, setting forms (Smiley, etc...) sayings and proverbs. Then, other indices are identified by the expert

Once the exchanges trade are analyzed, the expert draws up a cross grid (Fig. 5) in order to identify scripts by request type. This grid allows to first identify recurring indices in each type of request, and then validate the scripts by analyzing some others threads of exchange in order to validate its assumptions.

The researcher notes for example, by analyzing the grid crossed above that the requests of the emotional support type are characterized by a detailed self-presentation and a description of the anticipated benefits while in the other types of support requests, it is rather the problem which is more detailed in detriment of a personal presentation and benefit. In the same way, the researcher notes that in the experience sharing, a personal presentation and a presentation of the problem are well cared for.

V. STATEMENT OF THE MODELING

The aim of the modeling of this expertise is to highlight, on the one hand, the type of reasoning by the expert and on the other, to observe indices and criteria used in its reasoning. These indices can help thereafter the analysis and the understanding of other interactions of the same type, the mutual aid.

We have found that the followed reasoning is a continuous assessment at several levels:

- Cutting out the interactions into sequences of messages.
- Identification of the queries types: In this case, types of psychosocial criteria are used: query for assessment, query for information support and for advice, etc.
- Definition of scripts: pragma-linguistics indices related to the structure of the messages are used. These indices follow well defined scripts: Greeting, introducing oneself, recipient of the message, presenting the problem, explanation of the request or response, signature and fence.
- Classification of the interactions: pragmalinguistics indices related to the contents of the messages are exploited. We can mainly quote: encouragement and compliments, criticisms and disagreements, advice and contributions of information, assessment of the situation and complementary information request, evaluation of the expertise and similar experiences sharing.

VI. CONCLUSION

The researchers in pragmalinguistics study the interactions in order to understand the mechanisms which govern these interactions. They have developed

techniques and criteria to guide the analysis of these interactions.

We modeled an analysis of a discussion forum performed by a pragmalinguistics researcher in order to highlight the selection and the use of these criteria. Our goal is to provide guides for analyzing and understanding the interactions in a discussion forum, for example: to help define tags and other metadata to facilitate understanding of the interactions. Answering to three questions - what, how and why, this modeling proves to be very useful for the practices understanding mentioned above, as well as the initiation of reflexive posture from researchers.

We studied the analysis of a precise type of messages which is the social support. The theories mobilized in this analysis also concern social psychology. We showed the specific scripts dedicated to this type of interaction. We modeled another expertise of pragmalinguistics analysis of the interactions concerning the evaluation of an activity of a company. Certain criteria like the personal presentation and the greetings are the same ones, but the contents analysis requested other criteria in direct relationship with the speech acts.

Compared to ethnographic investigation the contribution of the knowledge engineering methodology have played its role - to reveal procedural knowledge, in opposition to prescribed knowledge in manuals that appears in some discourse during interviews. This methodology is also very useful to the aim of supporting knowledge sharing, especially about representing reference knowledge, implemented in each individual analysis.

This is only the preliminary stage of our study and we plan to complete and extend to other types of interaction in order to determine the criteria employed and the corresponding scripts for each type of interaction. We plan to consolidate all of these guides in a support tool for the charring analysis and understanding of the interactions in discussion forums. The way we try to explore is to memorize traces of each analyze and follow evolution of these traces replicated by another researcher and applied to another corpus.

VII. APPENDIX 1: English translation of exchanges

Discussions on addiction to drugs:	
Orchidée	Hi all,
Noire	I decided to take the plunge and start my first post. I have been lurking on this board for weeks, I have posted replies, without setting my own thread though.
	I only know a few persons, so I thought I'd introduce myself:

	<p>I am 33 and I am a... After having consumed loads of different stuff, I finally found happiness – and misery too – with heroin, to which I have been addicted for 8 years now. [...]</p> <p>I am depressive and hyper anxious as well. As soon as I stop taking it, I see reality and can't cope with it.</p> <p>I am not asking for anything, I feel I am alone with my problems. You can come here to just have a chat, share your own story or get off your chest whatever you need to...</p> <p>Best,</p>
Profile deleted	<p>Welcome Orchidée,</p> <p>I am not in a talkative mood today, but I do know about heroin problems.</p> <p>I am myself far from being rid of it, so...</p> <p>See you later, if I get the courage, lolll.</p> <p>Welcome to the forum, see you,</p> <p>Marie</p>
DianaC	<p>Hello, I am sorry, I am not familiar with such problems, but if it is support or a sympathetic ear you are looking for, I am here to read.</p> <p>I wish you strength, my own problem is alcohol and weed (excessive compulsive consumption).</p> <p>That is for now.</p> <p>Diana</p> <p>Take care!</p>
Discussions on milk allergies:	
cartoon181	<p>Hi all,</p> <p>I wonder whether I have an intolerance to dairy products (no idea which ingredients though).</p> <p>What I mean is that I thought that my diet had to do with the acne I still have in spite I am 22 years of age. Although I don't eat much greasy food, I often have rashes on the forehead, temples, they just come out, with no apparent reason [...]</p>
go	<p>It is difficult to say whether dairy products are responsible for acne, but that could be easily the case with certain persons [...]</p>
joely	<p>Hello, I am almost convinced that milk, in particular cow milk, is bad for your health. A common misconception, probably diffused by certain influent circles, is that it is rich in calcium [...]</p>

REFERENCES

Atifi, H., Lejeune, C., Ninova G., Zacklad, *Méthodologie transdisciplinaire de gestion du corpus pour les disciplines de l'interaction : recherche de principes directeurs* in F. Rastier et M. Ballabriga (dir.), *Corpus en lettres et sciences sociales : des documents numériques à l'interprétation,*

actes du colloque international CALS, Albi juillet 2006 pp. 185-190. Texto! vol. XI no. 2. 2006.

Aussenac-Gilles N., Laublet P., Reynaud C., *Acquisition et Ingénierie des connaissances, Tendances Actuelles*, Cepadues Editions, 1996. (Knowledge Acquisition and Engineering, Current Trends)

Austin J., *How to do things with Words*, (Traduction par Gilles lane), William James Lectures delivered at Harvard University in 1955, Ed. Urmson, Oxford, 1962, Éditions du Seuil, Paris, 1970.

Breuker J., Van de VeldeW., *Common-KADS Library for expertise modelling Reusable problem solving components*, Frontiers in Artificial Intelligence and Applications, J. Breuker and W. Van de Velde (EDS), Amsterdam: IOS.Press 1994.

Charlet J., *L'ingénierie des connaissances, Développements, résultats et perspectives pour la gestion des connaissances médicales*, Mémoire d'Habilitation à Diriger des Recherches, université Pierre et Marie Curie, 2003. (Knowledge engineering, developments, results and prospects for management of medical knowledge)

Gauducheau, N., Marcoccia, M. , « Le soutien social dans les forums de discussion : réalisation interactionnelles et contrats de communication », *Colloque Psychologie Sociale et Communication*, Dijon, 22-23 juin 2007.

Have, P. T., *Methodological issues in conversation analysis*, in Bulletin de Méthodologie Sociologique, Nr. 27 (June): 23-51. 1990

Hymes D., *Models of Interaction of Language and Social Life, Sociolinguistics, The Essential Readings*, Ch. Pratt Paulston et G. Richard Tucker (Eds.), Blackwell, 2003,

Kerbrat-Orecchioni C., *Les interactions verbales*, tome 1 (1990), tome 2 (1992), tome 3 (1998), Paris : A. Colin, 1998.

Miles, M. B, Huberman,A., *Qualitative Data Analysis: An Expanded Sourcebook*, SAGE Publications Inc; Édition : 2nd Revised edition, 9 février 1994.

McEnery T. and Wilson A., *Corpus Linguistics*, Edinburgh University Press. 1996. Édition : 2 (29 mars 2001)

Mondada, L., *L'analyse de corpus en linguistique interactionnelle : de l'étude de cas singuliers à l'étude de collections*, in A. Condamines (dir.) *Sémantique et corpus*, Hermès, Londres. 2005.

Sidoummou K., Matta, N., Atifi, H., Ninova, G. *La modélisation d'un savoir faire lors d'une interaction conversationnelle médiatisée*. Rapport de Master-recherche RACCOR, Université de technologie de Troyes, 2007

Searle, J.R. , *Expression and Meaning : Structures and Theory of Speech Acts*, London: Cambridge U.P. 1979