HAL
open science

# Mediated electronic discourse and computational linguistic analysis: improving learning through choice of effective communication methods.

Rachel Panckhurst

## To cite this version:

Rachel Panckhurst. Mediated electronic discourse and computational linguistic analysis: improving learning through choice of effective communication methods.: ascilite conference, Sydney, Australia, 3-6 December, 2006. ascilite conference: "Who's Learning? Whose Technology?", Dec 2006, Sydney, Australia. pp.633-637. hal-00291738

**HAL Id: hal-00291738**

**https://hal.science/hal-00291738**

Submitted on 29 Jun 2008

**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.

**[Proceedings, ascilite conference, http://www.ascilite.org.au/conferences/sydney06/proceeding/html_abstracts/16.html, p. 633-637. , Sydney, Australia, 3-6 December, 2006.]**

# Mediated electronic discourse and computational linguistic analysis: improving learning through choice of effective communication methods.

The author conducted research on the ways in which electronic mail, forums and chats are used within an online distance, open and virtual learning environment (WebCT) at a French University, for both on and off-campus students. This article briefly describes research on how computational linguistic analyses help us understand language evolution in the context of higher/further education and research. Results may lead teachers and tutors to choose more effective communication methods, thereby improving overall learning.

Keywords: mediated electronic discourse, computer-mediated communication (CMC), computational linguistics, virtual-learning environments (VLE)

## 1. Introduction

We have conducted research on computer-mediated communication (henceforth CMC) within a French University (Université Paul-Valéry, Montpellier 3) since 1996. Discourse appearing in email messages, forums (i.e., asynchronous discussion groups within a closed VLE) and chat sessions seems to be shaped in a particular way, precisely because one uses a computer. The computer becomes a tool, a sort of mediator, indirectly modifying the discourse within a CMC environment. A new discourse 'genre' which we call *mediated electronic discourse* (henceforth MED) is created. Others refer to Netspeak, Weblish, Cyberspeak (Crystal, 2001), oral-written hybrid forms (Anis, 1999), computer-mediated communication (Herring, 1996), electronic communication (Anis, de Fornel, Fraenkel, 2004), new forms of written communication (Guimier de Neef & Véronis, 2004, 2006). In this paper, after specifying several main features related to MED, we describe how computational linguistic tools help to perceive language evolution in the context of further/higher education and research. Although our research is conducted entirely on the French language, we believe that computational linguistic techniques are readily applicable to other languages (Herring, 1996). Using results from these analyses may serve as a guide for deciding which communication methods to use in particular pedagogical contexts, thereby influencing overall learning.

## 2. Mediated electronic discourse

In several articles (*cf.* Panckhurst 1999; Panckhurst & Bouguerra 2003), we posit that MED is similar to oral forms for some aspects and similar to written forms for other aspects, and that in other cases, features may appear to be MED-specific. In this brief paper, our interest lies with the types of language evolution appearing within a CMC environment, whether related to written or oral forms or neither.
Some of the main features of MED are listed below (*cf.* Panckhurst, 2006a for more detail):

- *smileys* to introduce non-verbal semiological aspects; specific typography: words in uppercase, lengthening or repetition of letters, (which, in certain cases may simulate intonation, and therefore indicate some paraverbal information), marks such as '>'or '|' (indicating a repetition of discourse between sender and recipient);
- spelling, grammatical mistakes and absence (or reduction) of punctuation (*cf.* Panckhurst, 1998; Véronis & Guimier de Neef, in Sabah, 2006);
- neology or neography (*cf.* Véronis & Guimier de Neef, in Sabah, 2006), for instance, SMS abbreviations or words borrowed from foreign languages.

More linguistic features include:

- predominant usage of the present tense (often over 60-70%) as opposed to imperfect/past, future, conditional, imperative;

- high usage of first person deictic pronouns (as compared to second and third person pronouns);
- lower percentage of verbs (under 20%) compared to other written forms (over 20-25%), and among verbs used, frequent usage of modals (between 20 and 30% of overall verb usage);
- increased usage of ellipsis (for instance: *Vous remerciant/Thanking you*; *Impossible de trouver le document à enregistrer sur disquette/Impossible to find the document to save on disk*)

Other more extra-linguistic aspects which are typical of online communication include:

- relational: conciseness, rapidity, anguish/worry (if a long silence is observed before responding to messages), aggressiveness, impulsiveness, an (illusionary) impression of proximity, protective barriers (no direct face-to-face contact), etc.
- communication context: reduction or absence of introductions and closures, non-observance of conversational rules (turn-taking, floor-taking, adjacency pairing, etc).

## 3. Case studies

### 3.1. Situation & context

Since 2001, our initial research (on MED & email) has been broadened in order to take into account forums and chat sessions, within an online distance, open and virtual learning environment (WebCT) at our University, for both on and off-campus students. Over a ten-year period, we have gathered an important amount of data (corpora totalling almost 500,000 words) and used it specifically to study email messages, forums and chat sessions (*cf.* Panckhurst 1999, 2001, 2003, 2005, 2006), between students and teachers on the one hand and between student peer groups on the other hand. Students are of course informed of this and all messages are rendered anonymous before analysis. Both on-campus and distance-education students' messages are analysed with a computational linguistics tool for French morpho-syntactic analysis, Cordial (by Synapse: http://www.synapse-fr.com). In the present research, morpho-syntactic analysis is essentially used for determining syntactical categories of words (verbs, nouns, adjectives, adverbs, etc.) and for reducing ambiguity at the sentence level; this extends beyond a solely lexical/statistical/concordancing text-analysis approach in which word frequencies are indicated (see the Xerox website for an online demonstration of morpho-syntactic tools in various languages: http://www.xrce.xerox.com/competencies/content-analysis/toolhome.en.html). In the present paper, we briefly compare data from 3 corpora used in recent years: 1999 (solely email corpus), 2005 and 2006 (forums and chats).

### 3.2. Presentation of the 2006 corpus

The 2006 corpus includes forum and chat messages related to three courses: two undergraduate courses, one for off-campus students (L3E57-chat), and one for on-campus students (L3E63-forum, L3E63-chat); one off-campus Masters' course (M2-chat).

**Table 1: Statistics for the 2006 corpus: dates, participants, messages**

| | L3E63-forum<br>undergrad.<br>on-campus | L3E63-chat<br>undergrad.<br>on-campus | L3E57-chat<br>undergrad.<br>off-campus | M2-chat<br>postgrad.<br>off-campus |
|---|---|---|---|---|
| Date & participants | Jan-Feb 2006<br>90 participants, in 3 groups | Jan-Feb 2006<br>90 participants, in 3 groups | Nov-Feb 2005-2006<br>4 participants, 2 sessions | March-April 2006<br>6 to 8 participants, 3 sessions |
| No. of messages | 186 total (13 to 67 messages per group) | 716 messages | 432 messages | 1,219 messages |
| Volume (no. of words) | 14,934 | 3,836 | 3,220 | 7,573 |
| Volume (average no. of words per message) | 80.3 | 5.4 | 7.5 | 6.2 |

In Table 1 above, the average number of words per message is much lower in chat sessions (5.4 to 7.5) as compared to forums (80.3). This accords with recent corpora (2005: 47.9 to 97 words per forum message, 10.7 for chats). Variation within forum messages can depend on the nature of the pedagogical work and students' habits associated with particular communication methods; the 2005 corpus varied from 47.9

words (undergraduate students) to 97 words per message (Masters' students having used forums for a long time). However, there is always an important difference between the averages for forum messages and chat messages because of the way in which the quasi-synchronous nature of chat sessions may simulate oral communication and the fact that one types quickly in a reduced typing space.

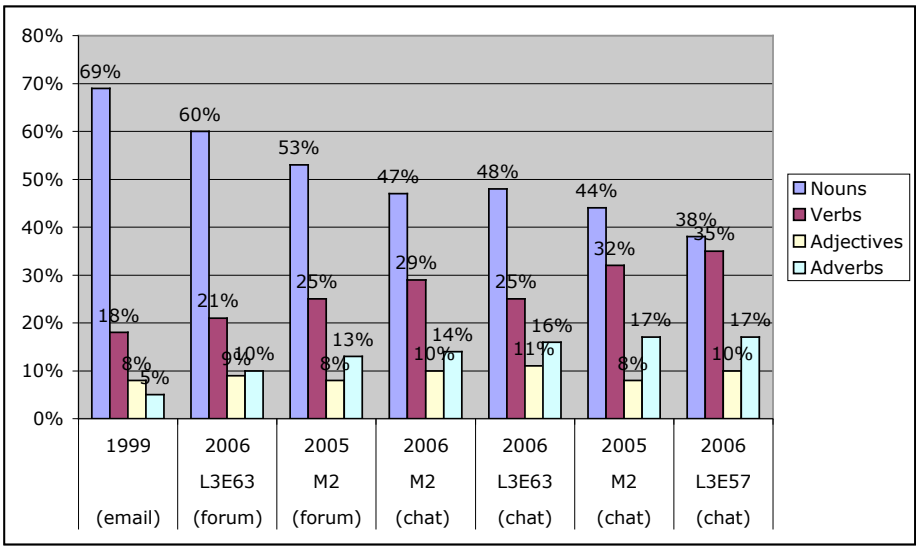
### 3.2. Results and language evolution

In ten years, we have noticed a certain number of changes in language within MED, concerning linguistic issues: question and negative forms, tenses and types of verbs, deictic pronouns, syntactic categories. In this short paper we briefly describe just two issues: question/negative forms and syntactic categories.

Question and negative forms have remained traditional in French online communication (for emails and forums), i.e., either inverted question forms such as *Le partiel aura-t-il lieu ? Will the exam be held?* or those using a particle such as *est-ce que — Est-ce que vous pouvez me faire un résumé du cours?/Can you summarise the lecture for me?* appear massively and *ne* appears fairly systematically with *pas*. This is contrary to French oral communication, where intonation is often used (*Tu viens?)* and *ne* is usually eliminated (*Je sais pas*). However, in recent analysis of chat sessions, interrogatives using solely a question mark have increased remarkably. In our 2006 corpus, 41.4% of question forms used question marks, compared to only 6.1% of those appearing in forum messages. However, the chat messages usually coincide with abbreviated SMS type usage, such as *C fini ?* (*C* abbreviates *c'est — Is it finished*?), compared to more formal usage in emails and forums: *Je suis très préoccupée par la grève de ce jour, le partiel est-il toujours maintenu?/I'm very concerned about today's strike, is the exam still going to be held?* Concerning negative forms without the *ne* particle, our 2006 corpus shows only 3 to 9.1% for forums. However, in one particular instance, where students were put into peer groups without teacher intervention, this usage increased dramatically to 60.4%. Again, this automatically coincides with SMS usage: *Le truc ke g pas compris c kil fo faire un résumé/Le truc que j'ai pas compris c'est qu'il faut faire un résumé/The thing I haven't understood is that the summary is compulsory*. In French SMS, *g* replaces *j'ai* through phonetic usage, therefore writing *je n'ai* is much longer, hence the abbreviated form and elimination of *ne*.

Syntactic categories used in different communication methods have evolved remarkably since the 1999 corpus (see Table 2, below):

**Table 2: Syntactic categories (occurrences) used in email, forums, chats (comparison 1999-2006)**




Up until 2005, our corpora showed that syntactic categories used in MED were very similar to other written forms (i.e., high usage of nouns, low number of verbs). In Panckhurst (2006b), we indicated an important change in the 2005 corpus: reduction of nouns, adjective stability, increase of verb and modal adverbs usage, but new corpora were needed to confirm this tendency. The 2006 corpus provides this

confirmation and demonstrates that linguistic usage has evolved quite dramatically; the results of the analysis for syntactic categories indicate that mediated electronic discourse may be either closer to speaking or to writing, and varies according to the communication method adopted. In Table 2, chat can be seen as the more "oral" method, whereas forums and email are closer to the "written" method.

## 4. Conclusion

Computational linguistic analysis is important in order to help us perceive certain changes. In our study using automatic analysis with Cordial, syntactic features indicate that chat sessions, which contain an overall higher percentage of verbs and adverbs and a lower percentage of nouns, may be more appropriate for oral, social communication, whereas forums and emails, which contain an overall higher percentage of nouns and a lower percentage of verbs and adverbs, may be more readily used for exchanging information (*cf.* Crystal 2001, p.168). Many of our colleagues choose chat sessions (rather than forums) with distance-education students because they see them as an important tool for creating a virtual "community" instead of simply a "group" of students, as well as for maintaining links and reducing student dropout. However chat sessions are not the right tool to use in all pedagogical situations (as forums may be more appropriate in specific contexts) and teachers may not necessarily perceive this. The choice and use of communication methods needs to be thought through carefully. Choosing the right communication tool for a particular pedagogical context is important for coherent learning; linguistic analysis which demonstrates features related to various communication methods (i.e. syntactic indications highlighting oral/written, informal/formal usage, etc.) can help in making the most effective choice, but broader comparative linguistic, extra-linguistic (*cf.* Panckhurst & Bouguerra, 2003) and cross-disciplinary research is necessary in order to understand more about current language and communication situations. MED has changed over the past ten years, according to communication methods chosen and as a result of overall language evolution. In 1996, we could not have imagined that SMS-type abbreviated messages would invade communication spaces in higher/further education. Within several years, University lecturers may well receive SMS-type messages not only in chat sessions but also in forums and email; up until now, this has never been the case in our experience. More importantly, as Véronis & Guimier de Neef (2006) note, most of the linguistic phenomena in SMS messages appear simultaneously: syntactic modifications, spelling, abbreviations, phonetic incorporation, etc. They indicate examples for French, e.g. *1dpdte* (*indépendante*), including a combination of problems: numbers, vowel elimination, morphemes. The architecture of current software, which is mainly based on sequential processing will thus need to be totally redesigned in order to analyse new forms of written communication efficiently.

## 5. References


Anis, J. (1999), *Internet, communication et langue française*, Paris, Hermès.

Anis J., de Fornel M., Franckel B. (organisers) « La communication électronique : Approches linguistiques et anthropologiques », International colloquium, EHESS, Paris, 5-6 February 2004.

Crystal, D. (2001), *Language and the Internet,* Cambridge, Cambridge University Press.

Herring, S. C. (ed), (1996), *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives,* Amsterdam/Philadelphia: John Benjamins.

Panckhurst R. (1998), « Marques typiques et ratages en communication médiée par ordinateur », Proceedings, CIDE 98, INPT, Rabat, 15-17/04/98, Paris, Europia Productions, 31-43.

Panckhurst, R. (1999), « Analyse linguistique assistée par ordinateur du courriel », in *Internet, communication et langue française*, Anis, J. (ed.), Hermès, 55-70.

Panckhurst R. (2001), "Distance, open and virtual lifelong learning: shaping the transition within a French University" Proceedings, 20th World conference on open learning and distance education, Düsseldorf, Germany, 1-5 April.

Panckhurst R. (2003), "Computer-mediated communication and linguistic issues in French University online courses", Online Educa, Berlin, 3-5 december, Proceedings, 454-457, ISBN 3- 9808909-2-9.

Panckhurst R. (2006a), « Discours électronique médié: quelle évolution depuis une décennie? », Proceedings, CMT, Bordeaux, 18-20/5.

Panckhurst R. (2006b), « Le discours électronique médié: bilan et perspectives », in A. Piolat (ed.). *Lire, écrire, communiquer et apprendre avec Internet.* Marseille: Éditions Solal.

Panckhurst R., Bouguerra T. (2003), “Communicational and methodological/linguistic strategies using electronic mail in a French University”, Proceedings, 8th International Symposium on Social Communication, Santiago de Cuba, 20-24 january, 548-554.

Piolat A. (2006, ed.). *Lire, écrire, communiquer et apprendre avec Internet*. Marseille : Éditions Solal.

Véronis, J., & Guimier de Neef, E. (2006), « Le traitement des nouvelles formes de communication écrite », in Sabah, G. (ed.), *Compréhension automatique des langues et interaction*, Paris : Hermès Science.