

A Tentative Typology of Net Mediated Communication

*Dr Adriano Allora*¹

Università degli Studi di Torino
all.adr@e-allora.net

Introduction

This work is part of a wider project aimed at collecting and publishing a considerable amount of texts written for the Internet— especially NewsGroups — in Italian, German, Spanish, French, and English: about 600,000,000 words per language were collected (some tagged Italian NewsGroups and some raw Spanish NewsGroups are now available at [4]).

Such a wide ranging project required a variety of preliminary studies on vocabulary, grammar, and textual varieties of Italian. One of the several case-studies under way originated an abstract model for the description of the textual features peculiar to Computer Mediated Communication (CMC). The present analysis will show the main characteristics of the model which quantifies the parameters of space, time, and accessibility of selected texts and defines indexes of attention for competition, interactionality, and connexity.

At the end, the values obtained from analysis are compared with text-message, forum and NewsGroup data.

1 Terminology

1.1 CMC, CMR and CMM

The suggested parameters are based on the idea that a text is a mass of verbal matter independent from meaning or stylistic choices: such a presupposition implies the solution of few preliminary problems which are not to be overlooked. The first merely terminological question requires a distinction within the broad field of Computer Mediated Communication (CMC).

This label generally describes the restricted area of Net Mediated Communication (NMC), i.e. e-mails, Internet Relay Chats (IRC), NewsGroups (NG), multi-user dungeons (MUD), blogs, mailing lists, and forums): Herring [7], Jones [9], Reid [14], Rheingold [15], Turkle [17], Pedemonte [11], Paccagnella [10], and the authors of the articles included in the JCMC [6] use the acronym CMC with the same meaning used here.

However, CMC does not include static texts such as web-sites and other types

¹The results displayed in the present article would not have been achieved without the indispensable contributions of Carla Marengo, Manuel Barbera, and Marco Tomatis. This paper starts from the General Theory of Medial Variation (Allora [2]).

of NMC whose terminals do not belong to the category of the computer, even though they are conceptually and physically similar, mobile phones, video-game consoles, and TV-decoders are complex programmable computers, i.e. they have a processor, a memory, and output/input devices.

A typology of CMC is carried out on the basis of the presupposed type of network employed during the execution of the communicative act. The network may be medial or social: communication would be impossible without some kind of network. The medial network includes telephones and computers (see Berruto [3] and Pistolesi [12] for a further analysis of the recent convergence of linguistic studies on text messages and IRC.), whereas the social network links producers and receivers of digital messages outside the computer.

In the medial network sender and receiver do meet and the fruition/production of the message resides in the encounter itself². In the social network a support is exchanged (e.g. a floppy disk or a file) and its content is received in a later moment. It is assumed that documents developed for an independent fruition belong to the social network, i.e. .doc and .pdf documents, CD-ROMs, off-line video games, DVDs, and applications in general. As long as documents remain within the relation sender-receiver – as it happens with a file written with a word processor – or between the user of a document and the document itself – as it happens with video games – the network is external to the communicative act implied by the document. This communicative act is to be seen *per se*, i.e. detached from the natural context constituted by the network. Communicative acts which presuppose a medial network are classified as NMC, whereas communicative acts which presuppose a social network only are classified as Machine Mediated Communication (MMC).

1.2 Visible and Invisible Verbality

The material nature of the text presupposes another remarkable distinction. The verballity of CMC is expressed by means which are different from those of natural/historical languages.

An HTML page, as well as any formatted text (e.g., when a .pdf file is read in its \LaTeX form), is more than meets the eye: the surfer of a site does not only read the information included in the page but also perceives – more or less consciously – the invisible text through which the natural/historical language text is edited. If the HTML code is badly written (e.g. non-standard syntax or non-precise vocabulary), the understanding of the entire text is jeopardized. Besides visible verballity (e.g. natural/historical language), invisible verballity also exists. Invisible verballity may be separated into intelligible (e.g. HTML) and non-intelligible (e.g. binary codes or encrypted texts). It depends on whether the human reader – after a specified

²This definition has nothing to do with the synchronicity of the exchange or the proximity of participants: the author of web-sites “meets” the receiver in the moment in which he/she is experiencing the site/text on the Internet.

training – is able to understand the meaning or not.

Besides, any type of on-line verballity is presented in a binary form. It is a different state of the same matter which, at first sight, may be confusing and unclear (e.g. binary code), structured according to a non-immediate logical form (e.g. programming languages which are characterized by invisible but intelligent verballity), or perfectly orderly and understandable (e.g. text and audio executed by computers). This distinction between different types of verballity is enriched by the distinction between NMC and MMC: Net Mediated Communication (NMC) is generally rich in intelligent verballity, whereas Machine Mediated Communication (MMC) is rich in binary verballity.

2 Parameters

2.1 Basic Notions

The model presented here is useful in defining some features of IRC, the visible verballity of chat-lines, NewsGroups, e-mails, mailing lists, forums, blogs, and text messages visible verballity, so the analysis is focused on expression constrictions. Although they are peculiar to invisible verballity, positive values also exist (e.g. HTML frames enlarging the page format).

However, the model does not start from the existing dichotomy speech vs. writing, which is considered illusory and ineffective. A text is regarded as oral when it mechanically propagates, as well as sound waves and sometimes electromagnetic waves do. A text is regarded as written when it resides in a support. The notion that both speech and writing may be static and dynamic will not affect the above definition.

For the sake of simplicity, time and space aspects are dealt with in terms of constrictions to the manifestation of the message. Some sub-parameters, whose value is easy to determine, were elaborated in order to assign specific values.

Time sub-parameters are **noise**, **permanence** after the end of the interaction, and **actuality** of the message.

Space sub-parameters are limits to the **dimension**, **structure**, and **form** of the message.

As to accessibility, only potentially uttered messages were defined as private, forwardable, semi-private, and public. The following paragraphs will illustrate how sub-parameter values were quantified: numerical values were specified between parentheses next to the possible sub-parameter values.

2.2 Definition of Time Sub-parameters

Noise may be private (0), both private and public (-1), and public (-2). There is an apparent overlapping between the parameter of noise and the parameter of publicity: the difference is that publicity is measured according to

the number of recipients, whereas noise is constituted by the simultaneous presence of few or many senders who create noise and make the request for an answer more urgent. A modifier (+1) must be added to these scores for those media which provide a noise reduction, such as a list of message titles.

Permanence of the message may have three possible values: the message remains at the end of the interaction (0); the message may be saved after the interaction (-1), the message is likely to disappear during the interaction (-2).

Actuality of the message is the possibility of a communicative act to include an entire or partial communicative act made by another sender. It is assumed that the highest degree of independence from time takes place in the case of messages which may contain the entire message of another sender. Nothing is left to human memory and not even a considerable delay would imply loss of information. This sub-parameter has three possible values: a message can contain messages from another sender (0); a message cannot contain messages from another sender (-1); a message may be updated (-2). This last case denotes the maximum dependency on time: if a message is not seen at the right time, it might never be seen again in the same form.

	noise	perm.	actual.	total
text-msg	0	0	-1	-1
chat-line	-1	-1	-1	-3
NG	+1	-1	-2/0	-2/0
ml	-2+1	0	0	-1
forum	-2+1	-2	0	-3
MUD	-1	-2	-1	-4
e-mail	0	0	0	0
blog	-2+1	-2	0	-4

Table 1: Time Constrictions

2.3 Definition of Space Sub-parameters

No further explanation is necessary for **dimension** constriction: either there is a limit (-1) or not (0).

In addition, even though there is a positive value (+1), this is not a constriction but a device to enlarge available space. It is worth noticing that other documents or files may be attached, inserted, or queued according to the limits of the space structure.

The lowest value given to dimension depends on intrinsic and ubiquitous limits which are due to the media used in producing and receiving messages: space constrictions always exist even though, in some cases, they are so wide-ranging that they do not bias the analysis.

Constrictions to the **structure** are to be set according to the possibility of dividing an utterance in smaller units (0) or not (-1): subdivisions are caused by the presence of characters which signal end of a line, paragraph, tabulation, and space. For instance, the tabulation feature is absent from mobile-phone keyboards and is not usually used in chat-lines or web-interfaced forums.

Form constrictions are characterized by the presence of a metalanguage, i.e. visible/invisible verballity. In addition to natural/historical language words, the medium may require (-2), accept (+1), or ignore (0) the use of encoded words useful to obtain particular effects. The mandatory requirements of a metalanguage are stricter than other options since they set out norms on the usage and efficacy of a natural/historical language, i.e. the easiest mean the users have to communicate.

	dim.	struc.	form	total
text-msg	-1	-2	0	-3
chat-line	0	-2	+1	-1
NG	0	0	0	0
ml	+1	0	0	1
forum	0	0	0	0
MUD	+1	-2	-2	-3
e-mail	0	0	0	0
blog	0	0	-1	-1

Table 2: Space Constrictions

2.4 Definition of Accessibility Sub-parameters

The absence (zero degree) of accessibility is represented by binary codes or encrypted texts. The following degrees are: private messages, forwardable or quotable messages, semi-private messages (with the explicit selection of one receiver only in a public context), and public messages.

Obviously, every private message is quotable and in some media this is a well-established practice (e.g. the phenomenon of quoting) which generates specific verbal-interaction behaviors. Other cases require the implementation of technical procedures, which are not easily accessible, or the resorting to devices which are

external to the computer (e.g. the human memory). Consequently, every public message may also be semi-private. Once again, this is a well-established practice specific to certain media (e.g. chat-lines). In other cases there may simply be an improper use of the medium. The publicity value of a channel equals the sum of all the values concerning the channel itself. For instance, e-mails, which can be private (1) and quotable (2), have a value which corresponds to that of semi-private media (3).

Such values (considering the four degrees of publicity and excluding degree zero which follows separate rules) may be modified:

- on the one hand, if messages are encrypted by the RSA method (See Singh [16]), they automatically become private (1); on the other hand, if the encrypting method is different (e.g. Caesar shift, Vignere square or others), the value is reduced by one unit;
- if the channel requires resorting to the jargon (or techspeak, see [13]) or if expert users of a channel resort to the jargon according to custom, the value is reduced by one unit, i.e. the message is accessible to anyone but not everyone can understand it.

	priv.	cit.	semi-p.	pub.	jargon	total
text-msg	1	-	-	4	-1	4
chat-line	1	2	3	4	-1	9
NG	-	2	3	4	-1	8
ml	-	2	3	4	-1	8
forum	-	-	3	4	-1	6
MUD	1	-	3	4	-1	6
e-mail	1	2	-	-	=	3
blog	-	-	-	4	=	4

Table 3: Accessibility

3 Textual Indexes

3.1 Competition for Attention

3.1.1 Definition of Competition for Attention

A fundamental concept for the description of NMC is competition for attention, which characterizes some communicative situations even to a great extent (e.g. IRC, see [18]). This kind of competition derives from the proportional ratio between accessibility and time constrictions. The two variables interact as follows:

Time the greater the time constrictions imposed by the channel, the more likely it is that competition exists. If interactants have all the time they need to communicate, the lack-of-attention parameter disappears;

Accessibility the greater the number of participants, the more likely it is that a situation of competition for attention takes place. If only few people are to be paid attention to, it is easier to pay and draw attention since little is required.

Competition for attention, which in some cases may be viewed as interactional craze, is to some extent a reformulation of time constrictions, i.e. a representation of the way in which time is perceived by interlocutors. It may be modified so that its value increases in proportion to the degree of publicity which characterizes the channel. By defining the competition as π (polemos), accessibility as δ (doxa) and time as χ (chronos), the formula is as follows:

$$\pi = -\chi + \delta$$

3.1.2 Verifying Competition Index

As a first attempt to verifying the proposed formula on competition, it was applied to the channels considered in the present paper.

text-msg	1+4	5	V
chat-line	3+9	12	I
NG	2/0+8	10/8	II/IV
ml	1+8	9	III
forum	3+6	9	III
MUD	4+6	10	II
e-mail	0+3	3	VI
blog	4+4	8	IV

Table 4: Competition for Attention

It is legitimate to view the value 9, shared by forums and mailing lists, as a phase or threshold value. Starting from this value, the competition of a channel is meaningful. Above this value media features are able or likely to display a degree of competition which effects the verbal production that takes place.

3.2 Interactionality

3.2.1 Definition of Interactionality

Interactionality, i.e. the ability of the channel to yield maximum general interaction among its users, is – at least to a certain extent – a function of the ratio of all three parameters, whose reciprocal ratios still have to be defined³.

Since the maximum interactionality of a channel equals the maximum number of messages produced thanks to that channel, accessibility is defined as a fundamental value. It must be kept in mind that interactionality is greater for channels allowing public messages.

However, the publicity parameter is already included in the competition-for-attention parameter, which undoubtedly turns out to be useful in defining interactionality and definitively substitutes the time parameter. As to the space parameter, a greater freedom in dealing with textual space ends up in a reduction of the number of messages during communicative practice. It is then logical to evaluate as negative a freedom of interactionality value of this kind.

The incidence of this parameter is in inverse ratio to its value.⁴

This will be added to competition: $\frac{1}{space}$.

The representation of this index may appear as follows:

$$\pi + \frac{1}{\tau}$$

τ being space (topo) and π the competition for attention.

3.2.2 Verifying Interactionality Index

It is worth trying to compute interactionality indexes and then verify them by confronting the indexes with the previously obtained types of NMC.

Even in this case there is a set of channels, i.e. the media channels, which involve – more or less permanently – groups of users which are rather close-knit and homogeneous. This may be predicted when considering the following: 1) centrality in calculating the degree of interactionality and publicity; 2) media which may convey public messages, i.e. high publicity score, are very likely to create communities of users.

³There are other parameters which cannot be computed a priori but need to be considered at the moment in which concrete linguistic realizations take place.

⁴The reason for this inverse ratio is not easy to motivate. There are two phases of constriction distribution: social constrictions (optional) and medial constrictions (non-avoidable). In general, medial constrictions are stronger, but, as to space, in a certain number of cases medial constrictions are less effective than social ones. Therefore, the greater the amount of space constrictions present, the least likely they are to be relevant.

text-msg	$5 - \frac{1}{3}$	4,6	V
chat-lines	$12 - 1$	11	I
NG	$10/8 + 0$	10/8	I
ml	$9 + 1$	10	I
forum	$9 - 0$	9	III
MUD	$10 - \frac{1}{3}$	9,6	II
e-mail	$3 + 0$	3	VII
blog	$8 - 1$	7	III

Table 5: Interactionality

3.3 Connexity

3.3.1 Definition of Connexity

Connexity is constituted by the cohesion and coherence of each single message related to the co-text. Since this parameter, differently from interactionality (which is the number of potential messages produced by the channel), belongs to a certain extent to the structuring of the message, the starting point for this analysis is space. The incidence of accessibility on the computing of connexity is ambiguous (e.g. the size of the co-text may generate dispersion or yield reference), so the use of competence-for-attention indexes must be excluded. In fact, this may reluctantly lead to the introduction of an ambiguous parameter. Instead, time constrictions are effective modifiers: the maximum availability of time and freedom in dealing with textual space represent the theoretical case of maximum connexity.

The time value must then be subtracted to the space value:

$$\tau + \chi$$

3.3.2 Verifying Connexity Index

Even in this case, distribution represents a trend and such a trend may also be used to select a significant threshold, which is signaled by forums, chat-lines, and text messages.

4 Comparing Indexes with Data

The most complex indexes, i.e. interactionality and connexity, were compared with three corpora derived from three different types of NMC: forums, NewsGroups, and text messages.

The corpus of text messages is a small monitor corpus of 4,392 words distributed along 251 messages with a textual-search tagging (available on [1]).

text-msg	-3	-1	-4	IV
chat-lines	-1	-3	-4	IV
NG	0	-2/0	-2	II
ml	1	-1	0	I
forum	0	-3	-3	III
MUD	-3	-4	-7	VI
e-mail	0	0	0	I
blog	-1	-4	-5	V

Table 6: Connexity

The forum corpus is made of a sample of 702 messages (39,925 words) taken from a much-attended web-site which mainly deals with computer science but also includes non-technical forums (see [5]). Linguistic research was carried out using this sample, whereas raw figures were extracted from the general corpus.

The corpus of NewsGroups is a fragment taken from [4] and is made of 11,435,385 words. All data are updated to May 10th 2005.

4.1 Interactionality

The computation of interactionality indexes provided the following results: NewsGroups 10/8, forums 9, text messages 4.6. The values for forums and NewsGroups dealing with specialized registers (e.g. computer science or cuisine) may be reduced by one unit. The following were selected as interactionality signals, feasible for a large-scale search and manually verifiable: presence and quantity of explicit answers/replies; amount of first-person singular and first-person plural personal pronouns; amount of first-person singular and first-person plural possessive pronouns and adjectives, i.e. highly deictic persons which imply a high degree of interchange.

Out of 12,971 conversations, 100,184 were posts (6.7 answers/replies per message).

Out of 371,090 forum discussions, 3,382,903 are messages (8.2 answers/replies per message).

Out of 251 text messages, 56 are answers/replies (3.4 answers/replies per message).

As to NewsGroups, out of 11,453,385 words, 197,973 are pronouns and deictic adjectives (1,7%).

As to the forum sample-corpus, out of 399,925 words, 2,312 are pronouns and deictic adjectives (5,7%).

In the monitor corpus of text messages, out of 4,392 words, 55 are pronouns and deictic adjectives (1,25%).

In the NewsGroup corpus there are 445 explicit interactional signals (0.0003%).

Forums only include 3 explicit signals (0.0007%).
No signal was found in the text-message corpus.

Supposing that the two types of data have the same importance, it is worth trying to attribute a specific weight to the data: if every minimum value is considered as a single unit, the following values are obtained. As to the quantity of answers/replies: NewsGroups 1.9, forums 2.4, and text messages 1. As to the presence of deictic persons: NewsGroups 3, forums 1, and text messages 2. As to explicit interactionality signals: NewsGroups 4.2, forums 1, and text messages 0. The sum of the above values is: NewsGroups 9.1, forums 4.4, and text messages 3.2. Results do not perfectly fit the initial hypotheses of the model, i.e. NewsGroups and forums should roughly have the same interactionality index, but do comply with the general guidelines of the model itself.

4.2 Connexity

The computation of the connexity index provided the following values: NewsGroups -2, forums -3, and text messages -4.

Explicit meta-textual relatives were searched (e.g. reference to texts produced by the author or others, textual deixis, explicit reformulations such as: *mi spiego meglio, ribadisco, come dicevo, come ho/hai detto/scritto, questo intendo*); quoting.

Within NewsGroups, there were 5,962 forms of reformulation, explanation, meta-textual reference (0,05%).

As to the forum corpus, the same analysis provided 37 forms (0,009%).

As to text messages, only one dubious occurrence was found (0,39%).

As to NewsGroups, out of 100,184 posts there are 134,593 blocks of quotations (134%).

As to the forum sample-corpus, out of 702 messages there are 472 quotations (67%).

No quotation is present in the corpus of text messages.

As to the distribution of weight, the following results are obtained. As to meta-textual indicators: NewsGroups 5.5, forums 1, and text messages 0. As to the amount of quoted lines: NewsGroups 2, forums 1, and text messages 0. As to the total: NewsGroups 7.5, forums 2, and text messages 0.

These results are in line with the initial hypotheses set by the model.

References

- [1] A. Allora (2004) Sms Monitor Corpus Available on-line from http://www.e-allora.net/SMS/ms_index.php (accessed May 10th, 2005).

- [2] A. Allora (2005) Per una tipologia della comunicazione mediata dalla rete. variazione diamesica generale. *Bollettino dell'Atlante Linguistico Italiano* 29, 1–30.
- [3] G. Berruto (2003) *Italiano parlato e comunicazione mediata dal computer* Hannover In Hoelker / Maaß [8], 109–124.
- [4] corpora.unito Group (2004) L'italiano nella varietà dei testi. Available on-line from <http://www.corpora.unito.it> (accessed May 10th, 2005).
- [5] HTML.it editors (2001) Html.it, il portal italiano sul web publishing. Available on-line from <http://www.html.it> (accessed May 10th, 2005).
- [6] S. Herring. Journal of computer mediated communication. Available on-line from <http://jcmc.indiana.edu> (accessed May 10th, 2005).
- [7] S. C. Herring, editor (1996) *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives*. (Amsterdam/Philadelphia: John Benjamins).
- [8] K. Hölker and C. Maaß, editors (2003) *Aspetti dell'italiano parlato. Tra lingua nazionale e varietà regionali* Hannover, Atti del convegno (12- 13 may).
- [9] S. G. Jones, editor (1995) *Cybersociety, Computer-Mediated Communication and Community* (Thousand Oaks: Sage).
- [10] L. Paccagnella (2000) *La comunicazione al computer* (Bologna: Il Mulino).
- [11] E. Pedemonte (1998) *Personal media. Storia e futuro di un'utopia* (Torino: Bollati Boringhieri).
- [12] E. Pistolesi (2004) *Parlar spedito. L'italiano di chat, e-mail e SMS* (Padova: Esedra editrice).
- [13] E. S. Raymond et al. The jargon file, vers. 4.4.7. Available on-line from <http://www.catb.org/esr/jargon> (accessed May 10th, 2005).
- [14] E. Reid (1995) *Virtual Worlds: Culture and Imagination*. Sage. In Jones [9], 164–183.
- [15] H. Rheingold (1993) *The virtual community. Homesteading on the electronic frontier* (Reading: Addison/Wesley).
- [16] S. Singh (1999) *The Code Book. The Secret of History of Codes and Code-Breaking*. Trad. it. (2001), Rizzoli, Milano.
- [17] S. Turkle (1995) *Life on the Screen. Identity in the Age of the Internet* (New York: Simon & Shuster).

- [18] C. C. Werry (1996) *Linguistical and interactional features of Internet Relay Chat*. John Benjamins. In Herring [7], 47–80.