

LEXICAL PHRASES IN TELECOMMUNICATION BLOGS: A GENRE ANALYSIS

Irina Argüelles Álvarez

César Benayas Arroyo

EUIT de Telecomunicación

Universidad Politécnica de Madrid

irina@euitt.upm.es

The study presented in this paper aims to provide a description of telecommunication blogs as a genre. Lexical phrases are analysed in order to reach conclusions regarding the nature of the language in these texts and the extent to which the results obtained are comparable with other written or conversational discourse types. Although the departing hypothesis is that the articles in blogs are basically transactional and their main objective is to transfer information, the conclusions point to interaction as a distinctive characteristic of this type of discourse and also to a more careful organization in the comments to the blogs entries than originally expected.

Key words: *corpus, genre, lexical phrases, blogs, Telecommunications.*

Este trabajo presenta una descripción de los blogs de Telecomunicación como género. A través del análisis de frases léxicas se llega a conclusiones sobre algunas de las características que definen la lengua que se usa en este tipo de textos y se comparan los resultados obtenidos con otros previos sobre el discurso escrito y el conversacional. Aunque la hipótesis de partida es que los artículos de los blogs son básicamente transaccionales, donde el principal objetivo es transmitir información, los resultados llevan a conclusiones sobre la importancia de la interacción en este tipo de discurso y también apuntan a una mayor organización de la esperable en las entradas de los comentarios al artículo principal del blog.

Palabras clave: *corpus, género, frases léxicas, blogs, Telecomunicación.*

1. INTRODUCTION

Conversation and written language have long been characterized as different kinds of discourse (Brown and Yule, 1983). As a general rule, writing is basically described as a transactional means of communication because its main purpose is to communicate factual information while conversation, in contrast, is more interactional since it is basically used for maintaining social relationships. According to Nattinger and DeCarrico (1992), this distinction is made explicit in the use of lexical phrases which tend to be more idiomatic and variable in spoken language. Spoken discourse is in general less structured than written discourse, and this can be easily exemplified by repetitions (*I would rather have... rather have*), false starts (*I don't want..., eh, I would rather have...*) or incomplete sentences. On the other hand, written discourse makes more use of devices such as specific phrases to give structure and direction to its content. There are, however, exceptions for some types of more particular conversational discourse with clear transactional aims such as business meetings, academic lectures or conferences where fewer interactional marks are found compared to average conversation. The opposite can be said about some types of written discourse like informal letters, which usually imply some kind of dialog with the audience. In the extreme, web applications such as e-mail, blog or chat rooms are still mostly text-based while their main function is frequently to establish and maintain interaction among individuals.

In Argüelles et al. (2009), blogs were studied in comparison with news reports through stance adverbs and adverbial phrases. There, the main aim was to obtain data regarding their use in both genres, the initial hypothesis being, based on observation, that a similar use of these devices would lead to preliminary conclusions about the extent to which the two types of discourse had common features. The analysis of data led to some conclusions that differed from the initial intuition and established more similarities among blogs with academic discourse than among those with news reports. In order

to confirm these findings and departing from Nattinger and DeCarrico's (1992) definition and classification of lexical phrases, this paper explores the use of these sorts of devices in blog articles and their comments as a means to help the characterization of what is here understood as an electronic text subgenre.

Although there have been more recent studies of lexical phrases (Li and Schmitt, 2009; Moon, 1997; Sinclair and Mauranen, 2006) Nattinger and DeCarrico's (1992) has clearly been an important reference for former studies. Recent papers such as the one by Ling et al. (2009) have also based their research design on Nattinger and DeCarrico's classification of lexical phrases from a functional perspective, to compare their use in spoken discourse by native and non-native speakers of English, but there seem to be no previous studies regarding the use of these devices in blogs. In this particular case, the functions analyzed by these authors in both spoken and written discourse and the examples illustrating their use in different languages provide an adequate framework for the study presented here and for further eventual analysis of the results in comparison with blogs in other languages.

2. BLOGS AS A GENRE

There are a number of reasons why our departing point is that blogs are in fact a genre. The idea that discourse communities define common conventions that imply a social behavior (Herzberg, 1986: 1 in Swales 1990: 21) is adopted here to justify that there is in fact a specific discourse community around blogs: based on Swales (1990: 24), the medium used, the predominance of functional needs and the users' tendency to create special-interest groups support this vision. Thus, it can be useful and interesting to demonstrate that the language used in blogs can in fact be characterized by sufficient features to understand discourse used in the group as a distinct genre.

The format of blogs is quite regular. Part of a website, "[...] they are usually maintained by an individual, with regular entries of commentary, descriptions of events, or other material such as graphics or video" (Wikipedia). Original articles in blogs are initially considered

to be basically transactional where the main objective is to transfer information. Because the writers do not have a direct relationship with the audience, they have to provide the necessary information in the text so that the content is totally understandable in itself. Written transactional discourse also implies distance and detachment from the audience, because the communication is asynchronous and unidirectional although there must also be some kind of formal indication of dialog on the part of the author because the information is targeted to an audience, and it is still a social event. In addition, it can usually be observed that language is integrated, since the writer has time to plan and structure the text.

But interaction plays an important role in blogs too. Most blogs allow visitors to leave comments and even to exchange messages, and this is actually a distinctive characteristic of this type of discourse. Nevertheless, blogs are still primarily text-based, so there must be a number of features closer to interactional, conversational discourse which characterize the comments which follow the initial article.

In what follows, the extent to which the study of lexical phrases in blogs helps the characterization of these particular types of text as a genre will be explored. In Section 3, some of the main functions of lexical phrases as described by Nattinger and DeCarrico (1992) are reviewed. Then, the methodology and the results of the study are presented and discussed in order to reach some conclusions about the nature of blogs and their possible classification as a distinct genre.

3. LEXICAL PHRASES

Nattinger and DeCarrico (1992: 59-66) analyze, among others, social interaction and discourse devices as functions of lexical phrases.

Social interactions relate to the establishment and maintenance of a conversation where lexical phrases that function as clarifiers (*Excuse me? What did you mean by X?*), topic shifters (*by the way, that reminds me of X*), topic nominators (*do you remember X? have you heard about X?*), etc. are found. Social interactions can also convey a conversational purpose and they can be used for questioning (*do you X?*

is there Y?), expressing politeness (*thank you (very much), if you don't mind*) or compliance (*of course, I'd be happy to*).

Discourse devices are lexical phrases used as links or connectors to build the structure in discourse. Some examples are fluency devices such as *so to speak, as a matter of fact, I think that X, you know* or exemplifiers such as *for example | for instance, in other words, this is*; and evaluators like *I guess, I'm (not) absolutely sure | positive| certain (but) X, there's no doubt that X*.

Nattinger and DeCarrico (1992) also highlight the important function of some lexical phrases for organizing texts. Macro-organizers are lexical phrases that give direction and structure to the text by exemplifying, shifting and summarizing topics, evaluating, qualifying, etc. They are basically transactional and should be distinguished from social interactional markers:

- Global macro-organizers coordinate the flow of the main topics in the discourse by introducing (*let me start with X, what do you think of X*), shifting a topic (*that reminds me of X, by the way*) and summarizing the ideas presented in the discourse (*in a nutshell, to make a long story short*).
- Local macro-organizers have a similar function but for topics subordinated within the global framework in the discourse: exemplifiers (*to give an example, something like that*), evaluators (*I'm absolutely sure that X, there's no doubt that X*), qualifiers (*at least in X, it depends on X*), relators (*however, it has to do with X*) or asiders (*I got off the track, where was I?*)
- Micro-organizers are lexical phrases that signal relations between adjacent sentences. Some examples of these micro-organizers are fluency devices such as *you know*, the conjunctions *and* or *but*; or other signals found in interactional discourse such as *well, yeah, sure*, used mainly to give the speaker time to think and elaborate the following part in the discourse. Micro-organizers are frequently shorter in length compared to macro-organizers since they are more typically used in spoken discourse.

With regard to spoken / written discourse, most lexical phrases that are used in written discourse will also frequently appear in spoken discourse: logical connectors (*however, nonetheless*), exemplifiers (*for example, in other words*), temporal connectors (*a final / last word*) or summarizers (*in summary, in conclusion*). However, lexical phrases commonly used in spoken discourse are very seldom found in written discourse: logical connectors (*but how about X?*), exemplifiers (*something like that, a little (bit) like*), temporal connectors (*last but not least, the last thing is*) or summarizers (*to make a long story short, in a nutshell*). Another typical characteristic of lexical items found in spoken discourse is the use of contractions such as *I'm, you'll, and it's*, that would be inappropriate in written texts.

4. THE STUDY

In Argüelles et al. (2009), the analysis of stance through adverbs and adverbial phrases departed from a corpus consisting of 25 blogs of telecommunications (300,000 words in the blog entries and about 50,000 words in the comments). The results obtained after a preliminary analysis of lexical phrases with the same corpus were not relevant enough and therefore new blogs of the same topic were added to the original corpus with the aim of increasing the chance of obtaining more coincidences of lexical phrases and, hence, more reliable results. The number of blogs included in the corpora was consequently increased to 52, multiplying the number of words compared to the initial corpus almost by three, and reaching 700,000 words for the blog articles and 200,000 for the comments.

The lists of lexical phrases used as a basis for the study was adapted from the original published in Nattinger and DeCarrico (1992) and focuses here on social interaction and discourse devices, and how they are used as global and local organizers to produce more conversational or rhetorical styles. The tagged blogs and comments were analyzed with the software AntConc, using the options “concordance” and “collocation”. In addition, in those cases where there was a variable part within the lexical phrase as in *Not only X, but also Y*, the option “regex” in AntConc was used. The results have been presented in number of occurrences per 10,000 words to help the comprehension of graphics shown in the following sections.

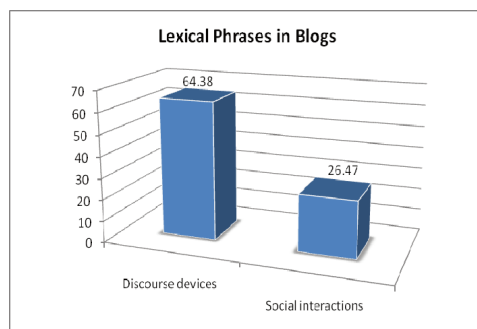
Departing from the medium used by this discourse community as an alternative communication channel, the initial hypotheses were the following:

- For blog articles, high usage of discourse devices is probable and few social interactions are expected. Although blogs may be more interactive than other examples of written text, blog articles are initially understood to exhibit most features of written discourse with a transactional purpose where the main aim for the writer is to give a structure to the text.
- On the contrary, and regarding the use of lexical phrases in the comments following the blogs, discourse devices are not expected to be frequent. Since comments are initially thought to be of an interactional nature similar to that of spoken discourse, lexical phrases for social interactions could appear more frequently, in particular, those used for questioning in order to clarify the content of an entry, or those with a responding function to expand previous comments.

5. RESULTS

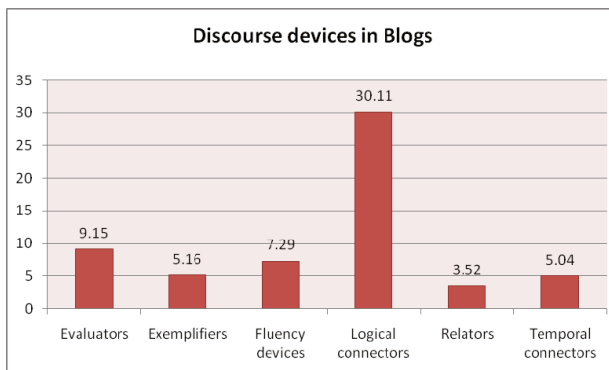
5.1. Discourse devices and social interactions in blog articles

Graph 1 shows how discourse devices (64.38) are predominantly used in English blogs. They account for more than twice the number of social interactions (26.47).



Graph 1. Main functions of Lexical Phrases in blogs of telecommunications.

According to the results obtained, the most frequent discourse devices, as shown in Graphic 2, are logical connectors with 30.11 occurrences. In particular, there is a high usage of some causal connectors such as *because of* or *due to*. Evaluators also appear quite frequently (9.15) with lexical phrases such as *I guess/My guess is* or *at least*.



Graph 2. Most frequent subfunctions of Discourse Devices in blogs.

- Logical connectors = 30.11

Example 1: *Because of* X (Blog 15)

“Carriers will adopt voice services *because of* the increased capacity and reduced cost of delivering voice over 4G networks”

- Evaluators = 9.15

Example 2: *My guess is* (Blog 42)

“*My guess is* that convergence applications will become more important over time - they are an exciting development but haven’t quite achieved critical importance for the industry.”

- Fluency Devices = 7.29

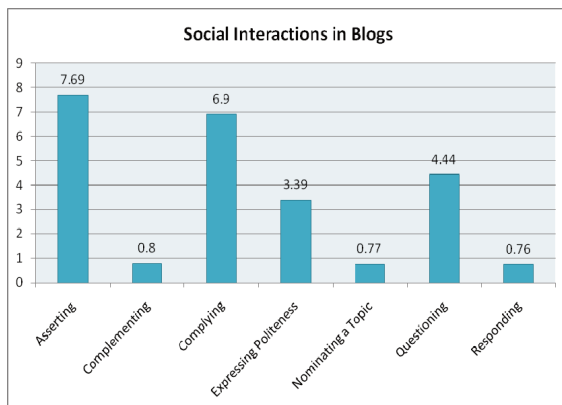
Example 3: *And so on* (Blog 16)

“There are so many fine prints related to how many lines you are using, which location are you calling from, which location are you calling to, the features you are using, *and so on*”

When it comes to other subfunctions such as spatial connectors, summarizers and qualifiers, they are not shown in the graph due to a

very low occurrence compared to the rest of the categories.

As for social interactions, asserting (7.69) and complying (6.90) are the two main categories with the most prominent results. Examples such as *I think/I believe that* or *sure (thing)*, respectively, are the most commonly used lexical phrases in this category.



Graph 3. Most frequent subfunctions of Social Interactions in blogs.

Some examples for each of the subfunctions shown in Graph 3 are:

- Asserting = 7.69

Example 4: *It is a fact that* X (Blog 50)

“Low technical barriers to entry: *It is a fact that* web programming is easier to get into than more typical native application programming (that typically requires C, C++ or Java skills)”

- Complying = 6.90

Example 5: *Of course* (Blog 3)

“The premise, *of course*, was that three areas stood out as being likely to generate decent revenues for anyone involved in the mobile content value chain”

- Questioning = 4.44

“Most of the cases here are rhetorical questions as shown in example (9). As this could also be considered as an evaluative element, these lexical phrases are here understood to have a double function.”

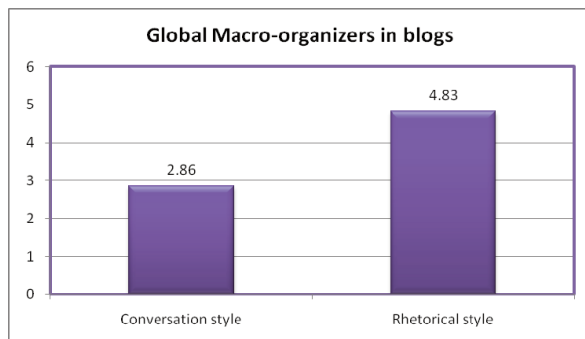
Example 6: *How do you X?* (Blog 20)

“How do you sort it out and come to a simple purchasing decision? The best way is to compare your T1 and SIP telephone options and get complementary consulting help for your particular circumstances”

The rest of the subfunctions occur less than 0.5 and some of them—responding to summons, checking comprehension and shifting turns—result in no occurrences at all.

5.2. Global and local macro-organizers in blog articles

Rhetorical style is the most frequent in blog articles (with 4.83 occurrences per 10,000 words), almost duplicating the results for the conversation style (2.86). See Graph 4 below.



Graph 4. Frequency of global macro-organizers in blogs of telecommunications.

Regarding the function of the lexical phrases for the global macro-organizers, the results show that there is much higher usage of summarizers than topic markers or shifters.

As an example, we can find the following results in the rhetorical style:

- Topic Markers = 0.30

Example 7: *You may recall* (Blog 35)

“You may recall that at last year’s EComm, we unveiled Fonolo for the first time, and won “Best New Product”. (Video of that presentation is here.)”

• Topic Shifters = 0.67

Example 8: *The second X is Y* (Blog 1)

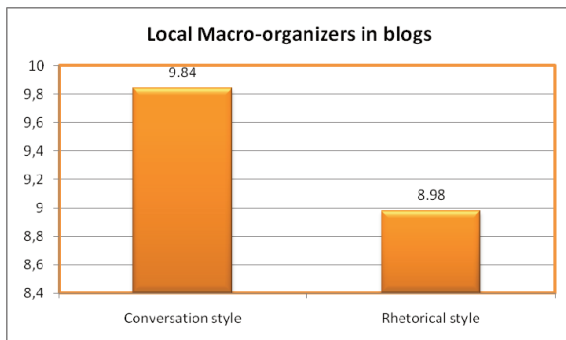
“*The second* big disadvantage for 8x8 *is* that they don’t own their own network and don’t manage the quality of the voice packets on the network.”

• Summarizers = 3.85

Example 9: *In other words* (Blog 19)

“The underlying factor in both the above examples is that the players are venturing into MVNO space to support their products thus having a strong value proposition. *In other words*, data was required to provide a better experience to their customers and hence Amazon and Dell ventured into the MVNO space.”

In the case of local macro-organizers, conversational style offers more frequency of lexical phrases (9.84), closely followed by rhetorical style (with 8.98).



Graph 5. Frequency of local macro-organizers in blogs of telecommunications.

When it comes to the function of the lexical phrases for local macro-organizers, the most commonly used are the exemplifiers, followed by relators and evaluators. See some detailed results for conversational style below:

• Exemplifiers = 3.42

Example 10: *One way is X* (Blog 34)

“Those who follow me know that I continue to stay active in the smart grid space, and *one way is* by participating at Connectivity Week, which takes place next week in Santa Clara.”

• Relators = 2.68

Example 11: *It has to do with X* (Blog 42)

“For TSTV, *it has to do with* the fact that the subscriber has less flexibility than DVR, since the operator decides what content is available to them.”

• Evaluators = 1.99

Example 12: *It is worth mentioning* (Blog 3)

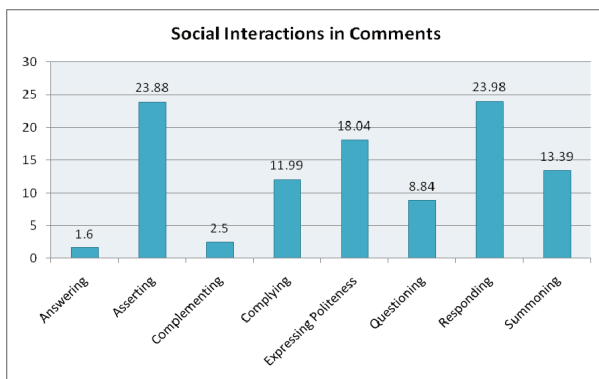
“In the case of the latter, however, *it is worth mentioning* that the Comstar-UTS group, a leading provider of integrated telecommunication solutions in Moscow and other cities, is controlled by Sistema, which is also the parent company of mobile market leader MTS.”

5.3 Discourse devices and social interactions in blog comments

Concerning blog comments, social interactions showed the highest results (109.01), closely followed by discourse devices (82.74).

Graph 6. Main functions of Lexical Phrases in comments to blogs of telecommunications

In what follows, data regarding the distribution of the main subfunctions in the comments to blogs are detailed. Graph 7 shows how responding (23.98), asserting (23.88) and expressing politeness (18.04) are the three main subfunctions mostly used within social interactions.



Graph 7. Most frequent subfunctions of Social Interactions in blog comments

Other subfunctions with significant results are summoning (13.39), complying (11.99), questioning (8.84), complementing (2.50) and answering (1.60).

Examples can be found in the list below:

· Responding = 23.98

Example 13: *I absolutely agree* (Comments 49)

"I absolutely agree with your analysis that apps are the gravy to the steak."

· Asserting = 23.88

Example 14: *I think that* (Comments 19)

"I think that you have well outlined the actual potential winning business models which can boost the mobile payment ecosystem"

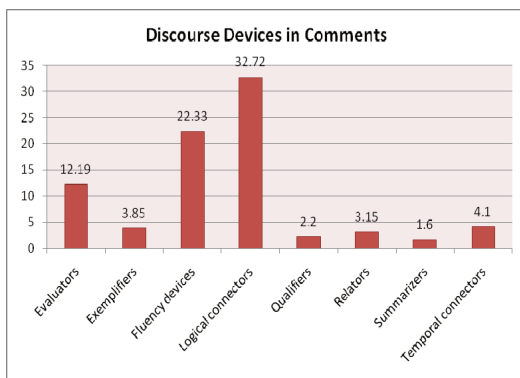
· Expressing politeness = 18.04

Example 15: *Thank you very much* (Blog 19)

"Comments and articles from quality professionals like you, dear Mohit, are vital to my business...and anybody else's here. *Thank you very much*"

When it comes to subfunctions with lower occurrences, the following can be mentioned: responding to summons, nominating a topic, clarifying, checking comprehension, shifting topic, shifting turn, closing, parting, requesting and offering and expressing sympathy, all of them with results lower than 1.0.

Regarding discourse devices, logical connectors are the lexical phrases with the most matches (32.72), followed by fluency devices (22.33), evaluators (12.19) and temporal connectors (4.10) to highlight some.



Graph 8. Most frequent subfunctions of Discourse Devices in blog comments

Examples for each of these main discourse device types can be found below:

- Logical Connectors = 32.72

Example 16: *Rather than* (Comments 19)

“TV Subscribers today have HDTV with over 200 channels. They sit in their easychairs and say, Nothing’s on! I believe that OTT video will be a reasonable means of entertaining ourselves through tough times, *rather than* watching the news and thinking of our fathers’ stories about the Depression.”

- Fluency Devices = 22.33

Example 17: *It seems to me that* (Comments 50)

“Volume is ** critical ** *and it seems to me that* its like saying the emperor has no clothes ie the volume is an obvious aspect but yet not talked of .”

- Evaluators = 12.19

Example 18: *At least* (Comments 12)

“Without knocking Cisco or bolstering Nortel there is quite a price difference between the two, or *at least* there was the last time I priced out some equipment”

Spatial connectors are on this occasion the only ones with very low results (>1.0) and they are not included in the graph above.

5.4 Macro-Organizers in blog comments

Results for global and local macro-organizers in comments are scant and similar in conversational (1.99) and rhetorical style (2.22). Regarding the function of lexical phrases for global macro-organizers, the results show again that there is a higher usage of summarizers than topic markers and shifters. As an example, we can find the following results under rhetorical style:

- Topic Markers = 0.45

Example 19: *First of all* (Blog 12)

“Hi Michael,

First of all thank you for the script. It will really help. Secondly let me recap and correct me if I am wrong: [...]”

- Topic Shifters = 0.50

Example 20: *Back to* (Blog 49)

“Getting *back to* RIM, which is why we are here, I found the article to be very well thought out. RIM is already being bashed about its two OS strategy. It seems as though they will drop their current OS in favor of the Playbook sometime in the future, for all devices.”

- Summarizers = 1.05

Example 21: *In short* (Blog 49)

“*In short*, RIM should just be themselves and keep on evolving from the base they have established over the last decade.”

With regard to local macro-organizers, conversational style (7.84) and rhetorical style (8.19) are again used similarly. Looking at the function of the lexical phrases for local macro-organizers, the ones that stand out are exemplifiers, followed by relators and evaluators. These are the detailed results for conversational style, for reference:

- Exemplifiers = 2.99

Example 22: *Something like that* (Comments 47)

“Earlier I guess the cities or *something like that* required that antennas are camouflaged to disappear in the environment here in Finland.”

· Relators = 1.80

Example 23: *How about?* (Comments 49)

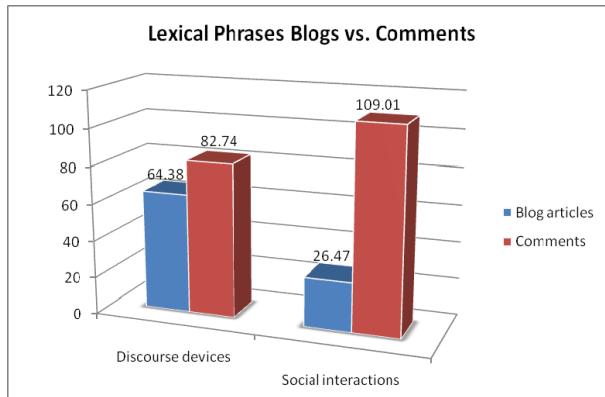
“So they must 1) rebuild their entire software suite 2) create a killer app such as HyperCard for a mobile device or combine it with lotus notes type corporate software 3) create suite just for meetings 3a) easy conference calls 3b) *how about* auto voice recognition and meeting notes software 3c) task identification and status completion built in — via Voice command!”

· Evaluators = 1.55

Example 24: *It is very important that X* (Comments 39)

“[...] we are a translation agency and also provided language training *it is very important that* we can jump from one language to another... and no matter how great 3Cx is, that feature is not there yet.”

5.5 Comparison between lexical phrases in Blog articles and Comments.



Graph 9. Comparison of lexical phrases in blogs vs. comments

Graph 9 shows a very high use of social interactions in comments (109.01) when compared to their use in blogs (26.47), which is the most noteworthy difference. Discourse devices surprisingly appear more frequently in the comments than in the blogs.

6. DISCUSSION

The final part of this study will explore the results obtained for the general hypothesis stated in Section 4 and recalled here, that a higher usage of discourse devices was expected in blog articles in which few social interactions were also probable. Blog articles were initially understood to exhibit most features of written discourse with a transactional purpose where a main aim would be to give structure to the text. On the contrary, discourse devices were not expected to be frequent in comments, as comments were initially thought to have an interactional nature similar to that of spoken discourse.

From the results, it can be observed that the initial hypothesis that the frequency order for the functions would be Discourse Devices > Social Interactions is not confirmed. Social interactions present a very high number of occurrences, definitely higher than expected. Authors in blogs expect comments to their texts in the short term. The comments to the blogs allow quick interaction between the writer and the reader in an asynchronous way but still with a good deal of interaction. For the comments, the order expected was Social Interactions > Discourse Devices, an initial hypothesis again not confirmed. Comments to the entries in the blogs are in fact more carefully structured than was originally thought.

When it comes to the macro-organizers, the following are the conclusions that can be drawn. First, that for both comments and blogs, high figures for local macro-organizers are found. This could mean that the presentation of ideas in the body of the blogs is less structured than in other genres, as could be the case with academic articles, which present more fixed structures and conventions in the use of these macro-organizers. Second, and in the case of comments, local macro-organizers were expected to be highly used, as here the need to present a complex general structure is preliminarily not as great. Third, exemplifiers and evaluators present the highest results, as was expected.

Comparing the results of blogs and their comments, social interactions were expected to be much higher in comments than in blogs because of the hypothesized similarities with spoken discourse

that is more interactional in nature. However, the results for discourse devices were not expected to be higher in comments than in blogs.

7. CONCLUSIONS

This study offers new data to add to those results that were obtained in Argüelles et al. (2009), which led to the conclusion that the use of evaluative adverbs in blogs was in fact different from their use in news or academic articles. That quantitative difference which was discussed from a more qualitative perspective, led to establish preliminary conclusions that blogs had specific characteristics that differentiated them from other genres which, a priori, could have been considered to share many characteristics.

The aim of this second study was to confirm from the quantitative point of view that other different discourse devices also led to similar conclusions. From the study of lexical phrases in the original corpus of extended blogs, we derive the following conclusions. First, blog articles are more interactional and less structured than originally expected, which leads one to think that they are different from news or research articles, both from the point of view of the global structure and the writer's intention. Second, comments are not as similar to spoken discourse as was initially expected, and these also have an important transactional component as seen in the use of language to provide structure for the presentation of ideas.

WORKS CITED

- Argüelles Álvarez, I., S. Molina Plaza, X. Xu. 2009 "La actitud del hablante a través de adverbios y expresiones adverbiales en los blogs de telecomunicación". *Revista Electrónica de Lingüística Aplicada* 8: 174-198.
- Brown, G. and G. Yule. 1983 *Discourse Analysis*. Cambridge: Cambridge University Press.
- Li, J. and N. Schmitt. 2009 "The acquisition of lexical phrases in academic writing: A longitudinal case study". *Journal of Second Language Writing* 18 (2): 85-102.
- Ling, J. and W. S. Hong. 2009 "Lexical phrases found in spoken

- discourse: A contrastive study of native and non-native speakers of English". *Education Technology and Computing Science* 1: 598-601.
- Moon, R. 1997 "Vocabulary connections: multiword items in English". *Vocabulary: Description, Acquisition and Pedagogy*. Eds. N. Schmitt and M. McCarthy. Cambridge: Cambridge University Press. 40-63.
- Nattinger, J.R. and J.S. DeCarrico. 1992 *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Sinclair, J. McH. and Mauranen. 2006 *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam: John Benjamins.
- Swales, J. M. 1990 *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.