Journal of Advances in Education Research, Vol. 2, No. 2, May 2017
https://dx.doi.org/10.22606/jaer.2017.22007

Cybercorpora and Textual Production

Maria Cecilia Mollica[1*] , Hadinei Ribeiro Batista[2]

[1,2]Department of Linguistics, Institute of Language, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
Email: ceciliamollica@terra.com.br

Abstract. In this paper, we discuss new possibilities for e-learning and studies on language from a virtual tool. We emphasize the use of computational tools for the creation and treatment of data in order to elucidate and put into thesis advantages and challenges from such methodology, especially in relation to textual production. This study is an experiment with the virtual tool Sabere[1], designed specifically for the purpose of the research in thesis. Sabere was thought for pedagogical purpose, supporting e-learning and setting up cybercorporaii with fine-grained control of social variables. This tool, besides providing social services in education, allows the construction of corpora for different purposes and areas of knowledge. In this study, we will focus on research conducted in the field of language (especially those regarding social variables), particularly on studies on textual production and revision in different genres and text types. As an innovative goal, we seek to point out new guidelines in sociolinguistics, cognitive linguistics and policies on education, carrying out analysis on language and mind. Beyond this, we look forward to discussing new public policy on e-learning and academic research.

Keywords: Corpora, corpus linguistics, sociolinguistics, cognitive linguistics, e-learning, language variation and change, technology, education

1 Introduction

In a previous article, Batista & Mollica (2014a) discussed remaining gaps in education on the use of virtual learning environment as a way to expand the network of interaction between teachers and students (classroom without barriers). The traditional system in force - with one teacher to several students - is not enough to meet the specific demands of the students, who have particular needs in developing learning. There is a challenge in having students as protagonists of their knowledge with the help of virtual environments. And there still remains the challenge of tooling to compile cybercorpora from virtual mechanisms of communication. Such gaps boosted the creation of the public learning virtual environment - Sabere, described in detail below, for the purpose of conducting an experiment able to show the advantages and challenges from such methodology. In addition to providing students with a wide range of teachers from different educational levels and areas of knowledge, the platform aims to compile (mega)cybercorpora correlated with a wide control of social indicators that can serve to academic research in general.

In this paper, we focus our attention on research within Sociolinguistics, Cognitive Linguistics and new methodology for e-learning, as social factors monitored enable research with wide set of social variable.

2 Brief Overview on Corpus Linguistics and Sociolinguistics

On the definition of Corpus Linguistics (CL), McEnery & Hardie (2012: 1) state:

We could reasonably define corpus linguistics as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions.

[1] Platform built for the experiment in thesis.

Friginal & Hardy (2014) describe the CL as a methodological approach that has as main objectives to compile, based on some principles, lots of data from empirical natural language that can be analysed automatically by technological tools, combining both quantitative techniques as qualitative analysis. Although the corpora usage practice for linguistic research is old, from the twentieth century, the popularity has become outstanding from the 60's with the compilation of SEU (Survey of English Usage), followed by Brown (written texts of American English) and Lob (written British English texts). Later, in the 80s and 90s, with the advent of individual computers, the CL has become even more popular with the establishment of the Brown Family and corpora references as BNC (British National Corpus).

Since then, many corpora, written and spoken (these less representative due to the complexity involved in its compilation) were created to suit many different purposes of research. Annotations for computer data reading have become sophisticated, including the use of automatic labellers, enabling research on different linguistic levels, even phonetic and prosodic. According to Baker (2010), CL, from the 80, received focus primarily to research on lexicology and pedagogy. Currently, CL subsidizes various areas of linguistics by providing corpora for lexical semantic research (Stubbs, 2001), applied linguistics (Hunston, 2002), stylistic (Semino & Short, 2004), translation (Olohan, 2004), discourse analysis (Baker, 2006), cognitive linguistics (Gries, 2006) and metonym and metaphor (Stefanowitsch & Gries, 2006) and Sociolinguistics (Baker, 2010).

Turning our attention on building corpora, it is observed that more care is given to written texts. As we have mentioned, this is due to the complexity in the compilation and annotation of speech data. The BNC, for instance, contains a percentage of speech data considerably lower than the amount of written texts. Another important challenge is to build data from virtual environment regarding a more effective control of social variables.

There is still a gap on the construction of cybercorpora, when it comes to virtual environment. King (2009) compiled a corpus[2] from chat rooms of gay men to compare different uses of language by Americans and Australians. To explore the data, he had to send e-mails to users requesting consent. Taking up the issue concerning the age group, we highlight the challenge in building megacybercorpora in order to keep up with the continuous use of the language to meet the main premise on language variation and change, which is to investigate the use of the language in real time. It is noteworthy that several studies have been developed from web data (virtual interactions), which point out many advantages. One example is Deakin and Wakefield's (2014) analysis, which explored the Skype for purposes of demonstrating that web-based interviews can be so accurate as to research conducted by face-to-face ones. Another is Computational Sociolinguistics (Nguyen, D. et al, 2015), new field of linguistic studies in order to help shape and understand the human relationship. Much research has been done in this area from data from social networks like Facebook, Twitter and many other social media for the purpose of investigating the impact of social context on aspects such as feeling, emotion, authority, gender, location, etc. Anyway, remains the abyss as to the identity treatment, real-time monitoring of language use, production and textual revision and availability of data specific to the demand in the educational context.

3 Cognitive Linguistics (CL)

Silva (1997) explains that the CL addresses the language as a way of knowledge and connection to the human experience of the world. The issues with which this theory deals are varied, although the focus is on the relation between language and thought. The CL was institutionalized in the 90s, although their remains have appeared in the 70s, strongly influenced by psycholinguistic studies. The most notorious work in CL are American authors Langacker (1987, 1990, 1991), Lakoff (1987), Lakoff & Johnson (1980), Lakoff & Turner (1989) and Talmy (1978, 1983, 1985, 1988, b) and Europeans, whose exponent is Fauconnier & Sweetser (1996). CL has interdisciplinary character and cross different areas: categorization and prototypes, metaphors and conceptual metonymy, imagistic schemes and their transformations, cognitive and cultural models, grammar as conceptual organization system. Cultural models include the frame notion, whose definition is quite diffuse and changes depending on the

[2] Corpus – a collection of written or spoken material in machine-readable form, assembled for the purpose of linguistic research. (Oxford dictionary of English)

investigation. In this research, we adopted the definition of Fillmore (1975, 1977, 1978), who considers frame as a set of lexical and syntactical means to refer to a scene or scenario and each lexical and syntactic option reflects a certain perspective on a situation that scene. In this analysis, we assume frame as a lexical and syntactic ways used to reflect the prospect that the speakers have addressed the issue as a way of representing their experiences and knowledge on the topic. Resuming, the focus of research is to capture how this representation or perspective is 'photographed' by language, more precisely, the lexical and syntactic ways.

4 Experiment with Virtual Environment to Build Cybercorpora

Sabere is a public virtual learning tool, but with a purpose quite different from other virtual environments of e-learning. First, it is public. It was thought, besides serving as a research experiment in the construction of cybercorpora, to serve as a pedagogical tool, a free space, for any education agent (teacher, student or anyone interested in learning) has a free, spontaneous, informal, independent and flexible interaction. Constituted of interaction rooms for each area of formal education, Sabere aims to compile megacybercorpora for the purpose of supporting research in different academic fields, especially those that require a more refined knowledge of the social profile of the subjects.

When building Sabere, it was decided to meet the social identity of users through a form required to get logged in, which filters a wide range of social variables. Such variables were drawn from previous work (Batista, 2014; Batista & Mollica, 2014a,b) in which it was argued that the social identity of an individual takes from identity features (Brandão, 1986; Damata, 1998; Hall, 1999 translation: Silva, T. & Louro, G.; Santos, 2008; Marcia, 1980; Oliveira, 1976). In this first version, designed for a controlled experiment, the Sabere just groups people in a social profile. There is no intention in identifying the users. They may use nicknames or numbers. For the experiment in thesis, the users are only put together in social variables like sex, genre and many others.

Sabere consists of virtual rooms of interaction, one for each area of knowledge of basic education. The platform allows the user to create new rooms, which can be public or private, to form study groups, presentation of papers or to meet more specific demands. The screen below gives an overview of the interactive environment.

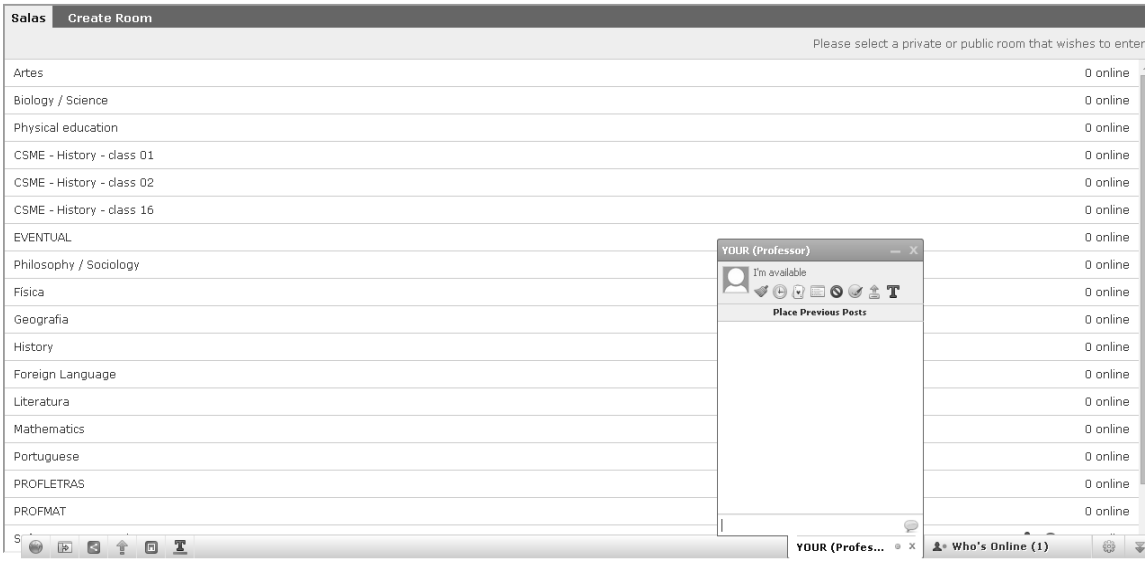

Figure 1. General and particular interaction

Once logged in, users can change the profile, make abuse allegations, change password or even delete their registration. The Sabere has in the submission screen a consent form, authorizing the use of data for research. The spaces created on the tool represent an advance in the collection and use of corpora,

since the consent of users, especially in online network, it is not easy to achieve as search King (2009) mentioned earlier.

This first version of Sabere has made possible some investigations. The tool currently has over 150 members, including several elementary school students. The platform has already made available for analysis over 15000 words.

The latest study, published in (MOLLICA, BATISTA and GUIMARAES, 2015), deals with involvement strategies in virtual interactions. Batista examined, through a qualitative analysis, a group of turns used by early adopters in their first contacts with other participants. The analysis considered factors such as sex and gender. It should be noted here that, unlike the traditional variationist sociolinguistics, Sabere allocates the variables differently: sex as biological information (male x female) and gender as sexual orientation (homosexual, heterosexual, etc.). Batista found that heterosexual men tend to type more concise turns when contrasted with hetero women's. Women usually use discourse markers like 'oi' (hi), 'olá' (hello), 'bom dia' (good morning), among others, before typing the main turn. Men, in contrast, appear more concise, avoiding such discourse markers around the turns.

New researches from the corpus are still forthcoming and seek to explore a more complex set of social factors. By being in the initial phase, Sabere does not yet have a large number of occurrences that enable quantitative analysis, subject to programs like Goldvarb or tools designed to work in Corpus Linguistics.

New projects are being implemented on the platform. The most current, focused on working with Portuguese, is for the production and textual revision. In Sabere, teacher can invite students to produce texts of different genres and types. The call can be made to all students of a room or the choice can be made at random. See the screen below:

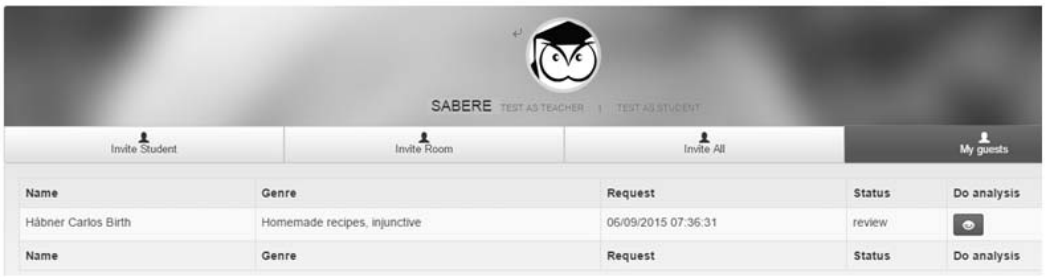

Figure 2. Invitation screen/writing

Once invited, a link appears in the student environment, containing the genre and textual type that must be produced as well as instructions for the writing task. The teacher can type instructions and also attach text or images as support for the activity.

All the processing of the production is monitored in detail by both teacher and student. Teachers are informed about the status of each student, whether the task is or not yet done, and the student, in turn, is informed about the status of the review.

For the writing, the student uses an editor without spelling or grammar checker. This alternative was chosen only to prevent the system overload, once nothing prevents the student use editors such as 'word' for a first review of the text and then paste it in Sabere's editor.

For textual revision, the teacher has available a set of descriptors (took from the ENEM[3] matrix) to justify any intervention. Such frame was broken into subcategories for better guiding the review work. The table below shows the most descriptors available on the platform:

[3] ENEM matrix:
http://download.inep.gov.br/educacao_basica/enem/guia_participante/2013/guia_participante_redacao_enem_2013.pdf

Table 1. Descriptors

Descriptors	Components
C1. Formal writing modality	Oral mark and informal writing Lexical precision Nominal agreement Verbal agreement punctuation Nominal and verbal inflexions Pronoun placement ((un)stressed) Spelling of words (including graphical accent and use of upper and lowercase letters) Hyphenation in word wrap Verbal and nominal regency
C2. Structural limit of argumentative text	Compliance with textual type Presentation of thesis statement Topic sentence Argumentative strategies (the items below should only appear if the teacher chooses this alternative) 1. Examples 2. Statistics data 3. Research 4. Verifiable facts 5. Quotes or testimonials from experts in the subject 6. Historical allusions 7. Comparisons with diverse facts, situations, times or places. 8. Other (if the teacher choose another, must be open a box to specify it, and then write the problem) Coherence between introduction and conclusion Information from different areas of knowledge Predictable reflection / Originality Theme total escape Theme partial escape
C3. Selection and organization of information (coherence)	Thematic progression (logical order) Semantic relationship between parts of the text Relationship text content and the real world
C4. Textual organization (cohesion)	Structuring of paragraphs (articulation between a paragraph and others) Structuring of periods 1. Fragmented phrases 2. Juxtaposed sequence of ideas 3. Phrase with only subordinate clause 4. Wrong connector uses (preposition, conjunction, relative pronoun, some adverbs and adverbial phrases) 5. Use of relative pronoun without preposition, when required Referencing strategy (repeat or inadequate replacement words)
C5 – Proposal for intervention	No proposal Proposal detailing Respect for human rights

Therefore, teachers must only select the part they judge there is deviation and a selection box is shown. Then, the teacher chooses, from the C1-C5 options, the set containing the rationale for the deviation. Then choose one of the alternatives presented. The system also allows the teacher just to make comments or spouse comments and descriptors. For each part selected by the teacher, the editor

tags it with a different colour. Students, in turn, may locate the parts commented by the reviewer, proceeding with adjustments.

The tool has a researcher space where they may generate various reports. For any survey data, it can be made a cross with any variable controlled by the system. Mollica, Batista and Guimaraes (2015), for example, for the study on the introduction of turns from virtual interactions, generated a report to verify the most frequent patterns per sex / gender. The system allows searches on general or specific expressions and terms.

The same goes can be done working with textual productions. You can map both the teacher and / or student as interaction rooms. For example, one of the system's objectives is to investigate the main demands of students to propose improvements in instructional materials. One can then generate reports to show the 'keywords' or 'concordances' for each room. Suppose, for the Portuguese Language room, a 'keyword' is 'verbal agreement'. This shows students are hard on this issue and new proposals for teaching and learning need to be thought. All other analysis can be made to any entry in the system, including any issue related to the textual productions of the students.

As for the text production, the option to map the review by descriptors due to the need to generate complex reports on the major problems presented by students. Thus, a report can be generated to verify the descriptors 'rank' of diagnosed deviations. If only the option comments become available to reviewers, manual analysis (from the point of view of the CL, human analysis) and time consuming should be requested.

5 New Studies from Textual Production Environment

A new study was conducted from texts production environment of Sabere, all in Portuguese. The collected texts[4] were produced by students from basic education whose proposal asked those students to produce a text / comment on the subject of bullying. The aim was to investigate how students structure their texts and the use of verbs and adjectives distribution. Following the theories mentioned, this analysis aims to verify the way the language is used by the users to 'take picture' of their thought/cognitive processing of the theme according to the social profile they are tagged in. It was found that male students employed more frequently the verb "call", revealing an experience more direct and aggressive with bullying. As for the girls, they used more stative verbs such as "to be", to contextualize the act suffered and to demarcate a more indirect relationship to the insult. As for adjectives distribution, their use also varied. Boys showed themselves more creative and free in the use of these words and girls emphasized adjectives that relate to own physical image. Beyond such study, we are carrying out many others on the relationship of the structure and use of the language, the subject, the text type and the speaker's mind from the variables that the system records. The images below show the analysis described above. It was carried out by the Corpus Linguistics tools: Antconc and Graphcoll.

As we may see, in male's production, the verb 'chamar'(to call) was the most frequent and in female's ones the stative verb 'estar'(to be). Note that the stative verb in female's texts was followed by prepositions 'na/no'(in/at) and 'com'(with). It shows that females tend to contextualize the bullying or offence they suffered. The data reveal that males normally experience the bullying in a different way in relation to females, proving that the social profile and the experiences have important impact to the use of the language. The sentences below were taken from the sample:

Males:

a. um menino chamou di macaco (a boy dubbed me monkey)

b. Me chamaram de PAC-MAN (they dubbed me PAC-MAN)

Females:

c. eu estava na escola (I was at the school)

d. estava na hora do recreio (It was break for recreation)

[4]We emphasize that the data from Sabere are directly authorized by the users (aged over 18) for academic research through the consent term present in the platform registration environment. In order to use the data that compose the present research, a collection of terms of consent was collected separately, signed by the parents of the students because they have not yet completed the age of majority.

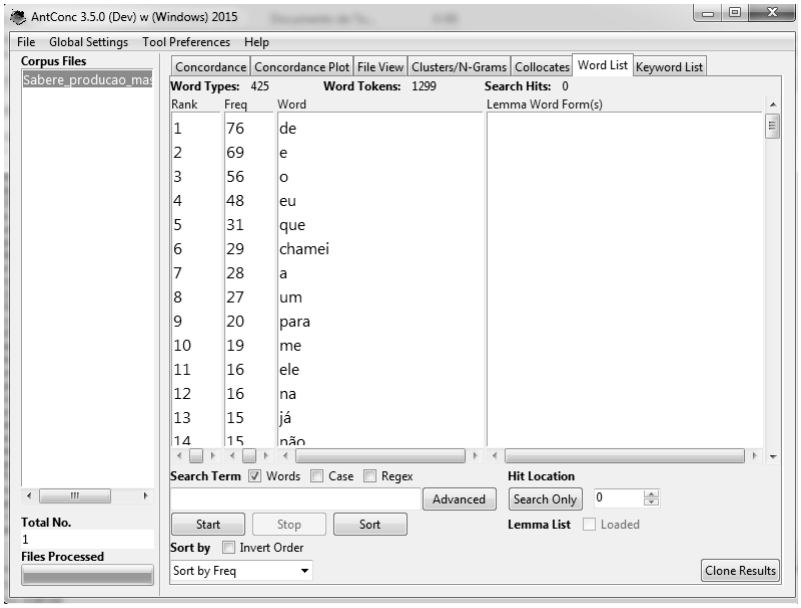

Figure 3. More frequent word in male's production

Figure 4. More frequent word in female's production

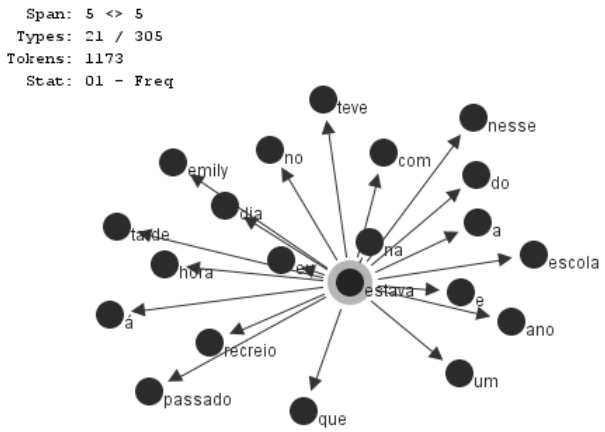

Figure 5. Relationship between the verb 'estava'(to be) and prepositions

As for the use of adjectives, we find that male students are more productive and use freely in the texts exactly terms they have heard.

Female students were much more timid and selected the lexical items that characterize the physical image, which reveals a greater concern of the women with the appearance.

In addition to showing the difference in the use of language in relation to sex, it is observed that minorities continue to be the target of negative judgments, even in a Western country with modified habits and customs, with a diverse range of family cell configurations. The picture presented to us is of a population with successive failures in national disputes, unprepared linguistically to reach high levels of literacy and, consequently, to overcome social barriers.

Table 2. List of adjectives used by males

Types: 425 **Tokens: 1299**		
Adjectives	**Frequency**	**Total**
esqueleto (skeleton)	5	1
baleia (whale), gordo (fat), viado (fagot), biladen (bin laden), ceará (who was born in Ceará), doida (mad), mendingo (beggar), pilhoenta (lousy)	4	8
baiano (who was born in Bahia), calopicita (kind of bird that has a crest), cearense (who was born in Ceará)	3	3
esponja (sponge), estreçado (stressed), ipopotamo (hippo), macunba (witchcraft), padeiro (baker), queloide (keloid).	2	6
balofo (fat), cheinha (fat), gordinho (fat), gordinha (fat), magrelo (thin), macaco (monkey), negro (black), orelhodo (big eared), piroquento (nickname for penis), tromba preto (black trumpt), pretão (black) , doida (mad), pac man, leite azedo (sour milk), carvão queimada (burned coal), Luciano hulk (celebrity in brazil), girafa (giraffe)	1	19
Total	**15**	**37**

Table 3. List of adjectives used by females

Types: 303 **Tokens: 1178**		
Adjectives	**Frequency**	**Total**
esqueleto (skeleton)	6	1
baixinha (small), gorda (fat), pereba (who is not a good player)	3	3
banquela (white)	2	1
feiosa (ugly), gorda (fat), gordinha (fat), macaca(monkey), magrela (thin), monica (comic book character), dentuça (toothy)	1	7
Total	**12**	**12**

6 Digital Literacy and Spelling Questions

Many other investigations can be done from the screen experiment. For this analysis, we present some surveys made for difficulties with orthographic processes. We have identified a number of emerging processes that require careful classroom work. The data reveal residues that should have been overcome in earlier stages of learning. We will move on to new investigations to diagnose the extent to which such difficulties correlate with social factors.

Vocalic raising: mininu (menino/boy), eli (ele/he), elis (eles/they), minino (menino/boy), outru (outro/other), muntos (muitos/many), alemau (alemão/german), caxa (caixa/box), pediu (pediu/asked)

Lateralization in verbal termination: achol (achou/found), procurol (procurou/looked for), resolvel (resolveu/solved)

Use of -s: coizas (coisas/things).

Intervocalic process: corerao (correram/run), pessoas (pessoas/people) – (rotacism)

Use of J/G ⇔ x/ch: deija (deixa/leave), suxa (suja/dirty), gegou (chegou/arrived) (fricative voicing)

Hipossegmentation: diaeli (dia ele/ one day he), coum (com um/ with one)

Hipersegmentation: cal sada (calçada/sidewalk)

Desnasalization: encotrou (encontrou/found), etrar (entrar/enter), enbora (embora/thoug), pergutou (perguntou/asked), tabei (também/also)

Consonantal meeting: poblema, ploblema (problema/problem)

Ditongation: veis

Voicing: sovreo, sovri (sofreu, sofri) (suffer)

Morophological awareness: cearence (cearense/who is from Ceará state)

Although we are referring to a generation that dominates technology, we observe that it is a group without digital literacy. All texts were produced in text editors with full text formatting capabilities. However, students generally did not resort to such resources as we can see from the passage below:

> foi na escola as crianças me chamava de monica,tendusa e muitas coisas diferentes.e eu também cha bresensiem as crianças falando que eu era feia,macaca...,ai eu tomei uma aditute e falei com meus familiares...
>
> It was in the school the children called me monica, toothy and many different things. And I also witnessed the children saying that I was ugly, monkey ..., then I took an atitude and I talked with my relatives ...

7 Challenges for Building Tools and Mapping Social Factors

The alternative of research now discussed presents a set of questions and challenges. Many other questions can be studied and answered from such methodology:

1. What are the most regular patterns in language in restricted chat rooms for teachers and students? And in textual production?
2. How these standards are distributed by social variables?
3. How is the process of teaching-learning of essays on the web? What are the main student's difficulties and challenges? How is the intervention of the teacher?
4. As the analysis above may impact the production of more efficient instructional materials?
5. What are the advantages of the methodology employed in this research regarding research involving "manual" collections?
6. What are the advantages and disadvantages of using technology in education for collecting data in the construction of knowledge, in teacher training and in achieving better results in national tests?
7. How can such research contribute to discussions on web security? Tools like that allow diagnosing and mapping cases of deviant behavior, enabling identify the most vulnerable social groups in the use of virtual platforms?
8. Can platforms like Sabere modify the traditional system of education, assisting students in a wide way once it provides to students a larger group of teachers?

8 Conclusions

The exploitation of virtual environments for studies in different academic spheres has been emphasizing since the popularity of virtual networks. The discussion undertaken here sought to point out new guidelines on research in sociolinguistics and cognitive linguistics. By methodology which is based on building computational tools and networks in the education, further discussions, beyond the linguistic field, show up timely and challenging. The social monitoring of users does not only map uses, variations and changes in the use of spontaneous language but also makes possible a more careful look at how different social groups are presented and constituted in this kind of environment, which facilitated and gave voice to groups previously considered marginalized in making turns. Beyond this, such experiment showed promising since it is efficient in determining learning difficulties and in producing more effective instructional materials. New research and updates are in progress from Sabere data. We cannot ultimately fail to point out the lack of financial resources for the implementation of new proposals.

References

1. Baker, P. (2006). Using Corpora in Discourse Analysis. London: Continuum.
2. Baker, P. (2010), "Sociolinguistics and corpus linguistics," Edinburgh : Edinburgh University Press.
3. Batista, H. R. (2014), "Identidade Social e Sociolinguística," Revista Querubim. , v.10, p.89 – 94.
4. Batista, H. R.; Mollica, M. C., Guimaraes, L. S.(2015), "Cybercorpora e Inovação: com práticas de ensinagem," Curitiba: CRV, v.1 p. 218.
5. (2015), "Perfil social e estratégias de envolvimento em interações virtuais," In: cybercorpora e inovação com práticas de ensinagem.1 ed. Curitiba: CRV, 2015, v.1, p. 11-26.
6. Batista, Hadinei R. & Mollica, Maria Cecília(2014a), "Public Virtual Rooms of Learning: an emerging technology resource," Creative Education. vol.5, n. 8, May 2014.
7. (2014b), "Sociolinguística, corpora e educação," Revista Letras., v.90, p.221 - 231, 2014.
8. Brandão, Carlos Rodrigues(1986), "Identidade e etnia: construção da pessoa e resistência cultural," São Paulo: Brasiliense.
9. Damata, Roberto(1998), "O que faz do Brasil, Brasil?," Rio de Janeiro: Rocco.
10. Deakin & Wakefield(2014), "Skype interviewing: reflections of two PhD researchers," Qualitative research, 2014.
11. Friginal, E. & Hardy, J.(2014), "Corpus-based sociolinguistics: a guide for students," London : Routledge.
12. Gries, S.(2006), "Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntatic and Lexis," Berlin: Mouton de Gruyter.
13. Hall, Stuart(1999), "A identidade cultural na pós-modernidade," Rio de Janeiro: DP&A. Translation:Silva, T. T. and Louro, Guacira Lopes.
14. Hunston, S.(2002), "Corpora in Applied Linguistics," Cambridge: Cambridge University Press.
15. King, B.(2009), "Building and analyzing corpora of computer-mediated communication," In P. Baker (ed.) Contemporary Corpus Linguistics. London: Continuum, pp. 301-20.
16. Marcia, J.E.(1980), "Identity in adolescence," In J. Adelson (Ed). Handbook of adolescent psychology. New York: Wiley.
17. McEnery, T. & Hardie, A.(2012), "Corpus Linguistics: Method, theory and practice," Cambridge: Cambridge University Press.
18. Nguyen, Dong et al. (2015), "Computational Sociolinguistics: A survey," Avaiable at: http://arxiv.org/abs/1508.07544.
19. Oliveira, Roberto Cardoso(1976), "Identidade, etnia e estrutura social," São Paulo: Biblioteca Pioneira de Ciências Sociais.
20. Olohan, M. (2004),"Introducing Corpora in Translation Studies," London: Roudtledge.
21. Renouf. [s.d.], "Corpus development 25 years on: from super-corpus to cyber- corpus," Rasunho, [ ]. Disponível em: http://rdues.bcu.ac.uk/publ/Corp_dev_25.pdf). Acessado em nov/, 2014.
22. Santos, E.N.(2008), "Adolescence, homossexuality, gender: Social-historical psychology as a new path," Revista de Psicologia da UNESP, 7: 1-11.
23. Semino, E. & Short, M.(2004), "Corpus Stilistics," London: Routledge.

24.Stefanowitsch, A. & Gries, S. (eds) (2006), “Corpus-Based Approaches to Metaphor and metonymy,” Berlin: Mouton de Gruyter.
25.Stubbs, M.(2001),”Words and Phrases: Corpus studies of lexical semantics,” Oxford: Wiley-Blackwell.

