ALFA
Revista de Linguística

ALFA
Revista de Linguística

The articles published in Alfa: Revista de Linguística are indexed by:

PRESENTATION

Alfa opens 2018 with a bulky issue, due to the number of articles and to the thematic and theoretical-methodological diversity of these studies. Such diversity is representative of the ever-increasing movement of questioning Language, which has increasingly placed Linguistics in dialogue with other areas of knowledge. The originality of the questions underlying the studies now published, and the richness of the answers they bring lie in those dialogues.

The two articles that open this first issue of 2018 analyze how strategies of argumentation are mobilized to construct certain meanings and discursive effects. In the first one, Freitas scrutinizes a document produced in the context of the legal controversy surrounding the approval of abortion of anencephalic fetuses, to show how the articulation of strategies such as explanation, naming and representation was paradoxically used to, convey support to a justified feminist demand, without departing from the androcentric paradigm that characterizes Law. In the study of Alves Junior and Tomazi, based on the approaches of the New Rhetoric, Sophistic and Semiological Theory of Discourse, we find the defense of an integration between the three rhetorical proofs of persuasion - *logos, pathos* and *ethos*, built from the analysis of a journalistic chronicle.

The issue also includes three studies that share the focus on language teaching and learning, on different objects and within different approaches. Nielsen Niño brings us a picture of the learning process of Spanish by foreign students in Colombia, highlighting the phenomenon of translanguaging, as a contact effect and fundamental strategy in learning L2. Still within the L2 teaching and learning field, Fontana and Leffa elaborate an in-depth analysis of the structure, pedagogical basis and functioning of MOOCs - "Massive and Open Online Courses". In investigating two courses offered in this modality, exemplars of an expanding tendency, the authors show how this tool, certainly useful, still needs to be improved, in order to offer effective spaces of interaction in the teaching process.

The third study in the teaching field deals with the process of acquisition of writing, more specifically the acquisition of orthographic competence. In a qualitative analysis of the interactions between two learners during the production of invented histories, Calil identifies and interprets the process of early recognition of orthographic problems and the search for solutions on the part of the students. The work emphasizes the importance of thinking orthography, not as an autonomous domain, but rather integrated with the other components of writing (and, therefore, of the language).

Following is the article by Souza and Di Felippo, which proposes to provide subsidies for automatic multidocument summarization. The study stands in the field of Automatic Processing of Natural Languages. This trend of linguistic research is no longer so new, but it meets the increasing demands of information processing and management, one of the defining characteristics of our time. Such demands are always bringing new challenges, one of which is the characterization of attributes and relevant phenomena in the texts to be processed and the translation of this information to the computational resources that must detect them. In the study in question, the authors explore mainly one of these phenomena - the temporal complementarity between events.

Three other articles investigate structural aspects related to the morphology and phonology of Portuguese from the point of view of processing and acquisition. Villalva and Pinto discuss the concept of morphological complexity, by analysing the reading processing of derived words, testing the performance of the compositionality parameter in the cost of this processing.

The study by Goulart and Matzenauer analyzes the acquisition of Brazilian Portuguese irregular verbs by native speaking children, from the perspective of Lexical Phonology and Morphology. They note that the complex morphophonological relationships present in these inflections may explain their late acquisition and a tendency to verbal regularization in these contexts.

The last article of this issue, by Bohn and Santos, in turn, deals with the acquisition of pretonic vowels in Brazilian Portuguese. As in the study by Goulart and Matzenauer, this is a late process (compared to the acquisition of tonic vowels), which would be explained by the instability of the pretonic subsystem. The authors analyze data from São Paulo children based on the Contrastive Trait Hierarchy.

As I mentioned earlier, we are facing a set of studies that are very expressive of the recent paths that Linguistics has undergone in Brazil and in other research centers.

Good reading!

Rosane de Andrade Berlinck

CONTENTS

ORIGINAL ARTICLES

THE STF DECISION ON ABORTION OF ANENCEPHALIC FETUS: A FEMINIST DISCOURSE ANALYSIS

Lúcia Gonçalves de FREITAS*

- ABSTRACT: This article presents an analysis of the Federal Supreme Court appellate decision in the Action for Breach of a Fundamental Precept number 54 (ADPF 54), which sought to resolve the controversy about the possibility of voluntary interruption of pregnancy in the case of fetuses with anencephaly. Through discourse analysis with a feminist approach, the paper focuses on the paradoxical way in which the Court met a feminist demand and women's rights movements: (1) concealing the bonds with activists of these groups and with the discourses they defend about the autonomy of women in relation to their bodies; 2) maintaining the traditionalist and androcentric standard of law. Thus, the analysis describes three language devices that stand out in the articulation of this paradox: explanation, naming and representation. The paper points the persistence of the traditional and androcentric paradigms in the Law field that Justice values and with which it operates. This study aims to contribute to the debate on the decriminalization of abortion in Brazil, as well as to discuss the relationship between language / gender / law.

- KEYWORDS: Discourse. Gender. Feminism Law. Abortion. Anencephaly.

Introduction

In this text, I propose to discuss the relationship between language, gender and law with a discursive analysis of a decision of the highest court of the Brazilian Justice on abortion, a recurrent theme from the claim of feminist groups and social movements that fight for the implementation and widening of women's rights. It is the Action for Breach of a Fundamental Precept No. 54 (ADPF 54), which sought to resolve the controversy about the possibility of voluntary termination of pregnancy in the case of fetuses with anencephaly[1].

---

* State University of Goiás (UEG), Post-graduation Program in Education, Language and Technology, Anápolis - Goiás - Brazil. luciadefreitas@hotmail.com

[1] "Anencephaly is a neural tube closure disorder that is diagnosable in the first weeks of pregnancy. For various reasons, the neural tube of the fetus does not close, leaving the brain exposed. The amniotic fluid gradually dissolves the encephalic mass, preventing the development of the cerebral hemispheres" (DINIZ; VELEZ, 2008, p.648).

In its conclusion, in 2012, the decision was acclaimed as a significant progress in the materialization of the reproductive rights of women within the Brazilian constitutional jurisdiction. Nevertheless, in a previous study (FREITAS; LOIS, 2015), when analyzing the vote of the ADPF 54 rapporteur, Justice Marco Aurélio de Mello, I pointed out that the STF approved the legalization of voluntary interruption in case of anencephaly, in response to a claim of feminist theories, excluding from the debate the discourses of this field and even concealing the effort of its representatives.

This happens because the perspective of the decriminalization of abortion, a celebrated feminist claim, which is tainted throughout the text of the decision, is not addressed in the ADPF 54. On the contrary, there was a textual effort very committed to evidence that the decision restricted its scope, exclusively, to cases of anencephalic pregnancy, and that any consideration regarding the right of women to freely interrupt their pregnancies was excluded.

Recently, in face of the considerable increase in cases of microcephaly in infants whose mothers were affected by the Zika virus epidemic, the debate on the topic was re-ignited. Anthropologist Débora Diniz, a professor at the University of Brasilia, a feminist activist and researcher at ANIS[2], an entity that co-authored ADPF 54's petition and which has fought, throughout the entire process, to grant pregnant women with anencephalic fetuses the right to decide on the anticipation of childbirth, said in an interview to the BBC Brazil that the organization is preparing an action to ask the Supreme Court to authorize the abortion of fetuses with microcephaly associated with the Zika virus.

The discursive analysis I propose in this article aims to contribute to the debate on the decriminalization of abortion in Brazil, as well as to discuss the language / gender / law relationship, a topic to which I have dedicated myself in research since 2010 (FREITAS; PINHEIRO, 2010, 2013; FREITAS, 2011a, 2011b, 2013, 2014). For this, I adopt a theoretical-methodological clipping that I call Feminist Discourse Analysis, following feminist epistemologies (HARAWAY, 1995; PAREDES, 2010) and modes of discourse analysis of feminist linguists, such as Michele Lazar (2005, 2007), Viviane Heberle, Ana Cristina Ostermann and Débora Figueiredo (HEBERLE; OSTERMANN; FIGUEIREDO, 2006). Throughout the text, I clarify this partnership in more detail.

ADPF 54, abortion of anencephalics and disputing camps

The document I bring for analysis is a textual legal genre called Action for Breach of a Fundamental Precept (ADPF) and is used to avoid or repair injury to a fundamental precept, which in the legal sphere is directly linked to the supreme values of the State and Society. It is a type of action, filed exclusively in the Federal Supreme Court, which

---

2   Anis - Institute of Bioethics and Human Rights is a feminist, nongovernmental and nonprofit organization of federal public interest, founded in 1999 in Brasília (Information available on the official Anis page: <http://anis.org.br/sobre>).

may be proposed by the following agents: I - the President of the Republic; II - the Bureau of the Federal Senate; III - the Bureau of the Chamber of Deputies; IV - the Bureau of the Legislative Assembly or the Bureau of the Legislative Chamber of the Federal District; V - the Governor of State or the Governor of the Federal District; VI - the Attorney General of the Republic; VII - the Federal Council of the Brazilian Bar Association; VIII – a political party with representation in the National Congress; IX – a Union confederation or class entity of national scope.

The Action for Breach of a Fundamental Precept 54 (ADPF 54) was filed in 2004 by the National Confederation of Health Workers (CNTS) before the Federal Supreme Court (STF), with advice from ANIS (Institute of Bioethics, Human Rights and Gender) to question the constitutionality of the interpretation of articles 124, 126 and 128, sections I and II, from the Criminal Code, which deal with the crime of abortion, in face of the possibility that pregnant women with anencephalic fetuses voluntarily interrupt pregnancy.

Brazil is the fourth country in the world in number of deliveries of fetuses with anencephaly (DINIZ; VELEZ, 2008). There is no treatment or cure and, in more than half of the cases, the fetuses do not resist the pregnancy and the few who come to term have a short period of survival. Prior to the trial, pregnant women with anencephalic fetuses who wished to react to the trouble of such pregnancy had to turn to the judiciary individually[3]. There was no standardization of jurisprudence and, in most cases, the decision only occurred after birth.

Until the final judgment of ADPF 54 in 2012, the uncertainty over the nature of the procedure had been dragging on for more than a decade. Also in 2004, at the time of the trial, Justice Marco Aurélio de Mello granted an injunction authorizing the anticipation of birth of anencephalic fetuses. This injunction, however, was annulled in a Supreme Court plenary session, four months after it came into force, resuming the obligation of women to remain pregnant despite the diagnosis of fetal impotence.

In reconstructing the history of ADPF 54, Camargo (2011) clarifies that, shortly after the revocation of the injunction, the National Confederation of Bishops of Brazil (CNBB) petitioned the Court to participate as *amicus curiae*[4]. However, at the time the request was denied with the justification that Law 9882, which regulates the ADPF, does not anticipate such a hypothesis. However, other requests followed, most of them from religious entities contrary to the ADPF 54 proposal. In 2005, the pressure of this

---

[3]   It is important to emphasize that Brazil has one of the most restrictive legislation in terms of abortion. Until the ADPF 54 trial, our Penal Code (CP) only permitted abortion: I - If there is no other way to save the life of the pregnant woman and; II - If the pregnancy results from rape and the abortion is preceded by consent of the pregnant woman or, when incapable, of her legal representative.

[4]   The Latin term *amicus curiae* means "court friend", that is, representatives of parts of society, as well as authorities, technicians and scientists, called to speak at public hearings, based on their knowledge and experience, collaborating with the Court in the decision-making process. According to Carvalho (2011, p.92), the law provides as attribution of the Supreme Court president and the rapporteur to convene a public hearing "to hear the testimony of people with experience and authority in a particular subject, whenever they understand it necessary to clarify fact questions or circumstances, with general repercussions and relevant public interest, discussed within the Court.".

mobilization led to a polarization of opposing forces and in favor of the litigation, time when the Justice Nelson Jobim suggested holding a public hearing. Notwithstanding the importance of the theme, it was only in July 2008 that hearings were finally determined[5] (CARVALHO, 2011).

In this dispute, three main segments stood out, representing positions, in certain points, quite antagonistic: the doctor, the religious and the feminist[6]. The position of the Churches in general, and of the Catholic Church specifically, is that abortion is a sin before God, and it hurts the right to life, which is considered since the fertilization (ALDANA, 2008). In contrast, abortion is seen by feminist movements as a matter of individual right to women's free choice over their bodies, including maternity (SCAVONE, 2008). Among medical professionals, the issue of abortion is controversial, however, there was a consensus in the CNTS regarding abortion of anencephalics for the sake of protecting the professionals who intervene in these cases from the criminal consequences they may incur when practicing the procedure.

This intricate scenario required a very complex argumentative effort on the part of the STF Justices, who had to address, discursively, fully ideologically opposed social segments and deal with disparate situations, such as: on the one hand, to give pregnant women of anencephalic fetus the right to legally abort without significantly changing abortion legislation in force in the country; on the other hand, to guarantee the secular position of the Brazilian State, without disregarding the importance of religious segments and their beliefs about life and death.

As Miranda Netto and Camargo (2010) point out, the Justices of the Constitutional Court must justify their every decision-making act in the constant search for popular adhesion, in order to guarantee their role as representative in face of their represented (the people), that is why they invest a lot of argumentative effort in their votes. In the case of ADPF 54, this effort is expressed both in the length of the decision-making text, which occupies 433 pages, and in the time demanded for the judgment, almost a decade.

My analysis falls precisely on the discursive effort, in this emblematic case of women's rights, considering this intricate field of disputes. I have as a guide the

---

[5] The participants in these hearings were from twenty-two institutions, represented by twenty-nine people. Carvalho (2011) reports that of the twenty-two institutions, fourteen were in favor (63.6%); seven, against (31.8%); and one (4.5%) presented arguments in both directions, in this case, the Legislative. From the twenty-nine representatives, fifteen were men (51.7%) and fourteen were women (48.3). In those moments, there was a very emotional polarization of views for and against the decriminalization of abortion.

[6] The first one can be considered the arguer himself, included in the National Confederation of the Workers in Health - CNTS and represented by it. Several entities of this segment and even named personalities are listed in the judgment report, as *amicus curiae*. They are: Brazilian Federation of Gynecology and Obstetrics; Brazilian Society of Clinical Genetics; Brazilian Society of Fetal Medicine; Federal Council of Medicine; Federal Deputy José Aristodemo Pinotti former Dean of Unicamp, founder of Campinas Maternal and Child Research Center - CEMICAMP and specialist in pediatrics, gynecology, surgery and obstetrics. In the group of religious entities are: National Conference of Bishops of Brazil; National Association Pro-life and Pro-family; Association of Family Development and Universal Church, although the last one has stood in favor of the lawsuit. In the feminist bloc or aligned with it are the ANIS, the Institute of Biotechnology, Human Rights and Gender that, in the initial petition, is considered co-author of the action; the National Feminist Health Network; Social Rights and Representative Rights; the School of People (human rights). Also aligned with this segment are Catholics for the Right to Decide, an entity composed by groups of theologians within the Catholic Church, contrary to the position of the institution in the case of ADPF 54.

questions of Castilho (2008) and Pimentel (2009): when deciding, do the Justices recognize the gender perspective and contemplate the discourses that seek to evidence the subsistence of patriarchy, the relations of domination between the sexes and the material inequality between men and women? With this direction, I develop a discursive analysis articulated by the issues of gender / sexuality (HEBERLE; OSTERMANN; FIGUEIREDO, 2006) and Feminist Theory (HARAWAY, 1995; PAREDES, 2010), which I call here Feminist Discourse Analysis (LAZAR, 2005, 2007; BAXTER, 2003, 2008), on which I dedicate the next topic.

Feminist Discourse Analysis

The feminist discourse analysis that I propose to develop on the text of the ADPF 54 is an assumption based on a growing field of discursive studies focusing on the gender and sexuality theme that assume a declared feminist perspective and that claim the inclusion of the term "Feminist" to Discourse Analysis works.

Although in Applied Linguistics there has been, since the last three decades, a large bibliography of studies based on gender / sexuality and discourse (HOLMES; MEYERHOFF, 2003; EHRLICH; MEYERHOFF; HOLMES, 2014) and many authors of these papers join Feminist theories, the term Feminist itself only begins to appear in partnership with the Discourse Analysis in the 2000s. One of the first examples of this junction is the article by Ann Weatherall and Anna Priestley, published in 2001 in Feminism & Psychology, whose title was: "A Feminist Discourse Analysis of Sex 'Work'" (WEATHERALL; PRIESTLEY, 2001).

Defining a linguistic approach that can be labeled as "Feminist Discourse Analysis" is a complex task, as Mary Bucholtz (2003) observed. Both because the discursive studies articulated by the "gender" and "sexuality" categories are not necessarily feminist, and that one single form of feminism, to which these studies affiliate, cannot be measured.

The "feminist" label, in the singular, actually covers a plurality of theories with their own specificities, such as cultural feminism, liberal feminism, postmodern feminism, radical feminism, etc. (SOUSA, 2015). Nevertheless, even in plural, feminism, in its different strands, converges to the common interest of understanding and overcoming social inequalities related to gender and sexuality (BUCHOLTZ, 2014).

As feminism itself is not unified, the growing field of discursive studies with a feminist perspective is not unified either. Examples of this attempt are proposals like the Feminist Talk Analysis (KITZINGER, 2000), Feminist Style (MILLS, 1995), Feminist Pragmatics (CHRISTIE, 2000), Critical Feminist Discourse Analysis (LAZAR, 2005, 2007) and Post-Structuralist Feminist Discourse Analysis (BAXTER, 2003, 2008). These theories are somewhat unified in their general political objectives, but are divided in the theoretical-methodological forms that follow to reach them.

It is observed that such proposals develop from theoretical theories already considered canonical within language studies, such as Critical Discourse Analysis (WOODAK; MEYER, 2001), Conversation Analysis (SACKS, 1992), the Pragmatics (MEY, 2001), the Style (BRADFORD, 1997) and so on. Recurring discursive studies on gender and sexuality of feminist bias were being produced under these headings without, however, being made visible. Gradually, a stated stance was taken to include the term "Feminist" in these labels as a political strategy to establish and strengthen feminist representation within the mainstream of Applied Linguistics.

Basically, these studies undertake a reappropriation of the theoretical-methodological points of view and the analytical tools of those canonical theories, with specifically feminist ends: the challenge of social inequalities related to gender and sexuality, recognizing the intersections with categories such as race, class, generation, ethnicity, etc.; the deconstruction of codes that naturalize and perpetuate these inequalities and the effort to overcome the sexist systems. Reappropriation and recreation are recurrent and recommendable dynamics to feminist practice, as Audre Lorde (2007) notes. In her famous essay on the risks of appropriation of "master's tools", the author warns on the power exercised by the dynamics of patriarchy, of which science is also invested, and proposes the reappropriation of knowledge and the use of the creative force, in feminist research and activism.

The feminist discourse analysis that I develop in this article is also a reappropriation of theory assumptions already in force in language studies with a feminist bias. Basically, I seek support in the work of Michele Lazar, for whom the goal of a critical feminist discourse analysis is:

> To show the complex, subtle, and sometimes not so subtle forms in which the often assumed gender assumptions and hegemonic power relations are produced discursively, sustained, negotiated, and challenged in different contexts and communities. (LAZAR, 2005, p.145).

The approximation of this author's proposal is due to my experience with Critical Discourse Analysis, in which she lends the theoretical-methodological tools to her feminist approach and of which I have been using in my own works (FREITAS, 2013, 2014). I do not, however, adopt the same name as the author, because I assess, in a first moment, that the term "criticism" is redundant in feminist studies, considered in advance as critical approaches (GUBA; LINCOLN, 1994), thus, I prefer a more succinct form: Feminist Discourse Analysis.

Secondly, I justify this choice also by the understanding that feminist research operates from the outset within a program politically invested in the struggle for recognition and to that extent it is imperative to claim own labels that represent us and highlight our location (HARAWAY, 1995) on the pathways of knowledge. For some time now Brazilian linguists such as Ana Cristina Ostermann, Débora Figueiredo, Viviane Herberle (HEBERLE; OSTERMANN; FIGUEIREDO, 2006), Carmen Rosa

Caldas-Coulthard (1996), Suzana Funck (2007), just to name some, have published works under language / gender / sexuality axes with feminist perspectives, without, however, assuming them nominally.

In this article, therefore, I follow in the footsteps of Audre Lorde, towards a reappropriation and recreation of knowledge in feminist studies, and of Julieta Paredes (2010), who preaches the need for the autonomy of Latina feminists in the face of western feminist epistemologies. Thus, I adopt this label, both for the search of self-identification and for its potential to shelter, under the same naming, different forms of discursive analysis united in the feminist effort to challenge sexist knowledge systems.

The discursive analysis that I develop on the text of the ADPF 54 turns precisely on the knowledge system that bases this well reasoned decision on the scope of Law. On this system, authors affiliated to Feminist Theories of Law (BARTLETT, 1991) denounce that it incorporates ontological notions that underpin modern institutions, in which man is the universal referent and woman is the special and the derivative. As a consequence of this relationship, women's rights are subordinated to this subject, and always in reference to the same places: of sexuality, of conjugality and of procreation. Such dynamics mean that legal reforms are merely palliative, delude reality, but do not transform it properly (SOUSA, 2015).

ADPF 54, in deciding on the possibility of abortion, a right that has been claimed for decades by women's and feminist movements, is therefore a special object for an analysis that seeks to unveil, in language, the ideologies and the acting ways that structure the decisions in such an emblematic case of women's rights. This is what I propose next, by the combination of Discourse Analysis and Feminist Theory.

A feminist analysis of the Supreme Court Justices discourse in the ADPF 54

The starting point for my analysis of the ADPF 54 appellate decision (BRASIL, 2012) is the paradoxical perception that, although the Court has met a feminist demand and women's rights movements, it has done in a way to conceal the bonds with these groups and the discourses they defend about the autonomy of women in relation to their bodies and without breaking with the traditionalist and androcentric standard of law. On the contrary, the appellate decision reveals the resistance that exists in this field in facing these paradigms that Justice values and with which it operates. Thus, my analysis focus on the description and discussion of the language devices that articulated this paradox, of which I highlight three specifically: explanation, naming and representation, as I develop below.

Approving anencephalic abortion in a Christian country: justifications and explanations

One of the problems of any communication is the risk of compromising the social image of the participants, and therefore the need for strategies that soften this

impairment, that is, an elaborate "face-work", as Goffman named (1967) the effort we make to be well evaluated. This concept has been more widely explored in studies that define their academic objects as face-to-face interaction, thus, excluding asynchronous and written communication.

Nevertheless, appellate decisions are textual genres of highly interactive nature, as they record the argumentative exchanges that the Court members, in their votes, repeatedly address among themselves and to more participants. Therefore, although it is a written piece, the appellate decision is constructed based on those interactions which, at times, are undertaken face-to-face, as was the case in ADPF 54, where public hearings were part of the decision-making process and in which, as Camargo (2011, p.14) observes, "the orality and the presence of the public prevail".

Thus, throughout the text, we perceive the Justices concern with the evaluation of their interlocutors who, ultimately, comprehend the Brazilian "Nation" itself, as it is textually assumed by Justice Celso de Mello:

> Recalling the late Justice LUIZ GALLOTTI and considering the high significance of the decision to be taken by this Supreme Court, in this Action for Breach of a Fundamental Precept, on the intended right, in favor of pregnant women to the therapeutic anticipation of childbirth, in the exceptional situations of fetal anencephaly, I bear in mind his grave warning that, in such emblematic cases as this, the Federal Supreme Court, in rendering its judgment, may itself be "judged by the Nation.". (Vote of Justice Celso de Mello, 317).

Considering this judgment, various discursive resources are mobilized by the Justices to protect themselves from possible negative evaluations. Among these resources, I highlight the "explanation" or "accountability" (PASSUELLO; OSTERMANN, 2007; OSTERMANN; ANDRADE; FREZZA, 2016). These terms are translation attempts for the English-language correlate *accountability*, coined by Garfinkel (1967 apud OSTERMANN; ANDRADE; FREZZA, 2016) to refer to the notion of normative responsibility that, once broken, makes room for apologies, explanations and rendering of accounts.

On the first page of the appellate decision, just below the heading, we read the following sequence: STATE - SECULARISM. Brazil is a secular republic, appearing absolutely neutral regarding religions. Considerations.

This textual arrangement obeys the norms of thematic content indexing by keywords that are proper to the schematic conventions of the jurisprudential summary (GUIMARÃES, 2004). The sequence of words emphasizes the State and Secularism relation. Immediately afterwards, the two sentences that also observe the norms required for the summaries, privileging conciseness and clarity (GUIMARÃES, 2004), highlight Brazil's neutral position as a secular republic in relation to religions. Since the separation of state and religion is an essential condition of modern democracies, which require

secularism as a logical consequence of the application of its principles, affirming the secularism of the state would be a redundancy. However, the sequence serves as an early explanation of the authorities signing the decision, on dealing with a lawsuit that goes against fundamental religious principles such as life and death. The emphasis on religious neutrality of the Court beforehand, at the very beginning of the text, has the functionality of an account.

In the words of Ostermann, Andrade and Frezza (2016), accounts are attempts to justify yourself, to explain yourself or to undertake some other action that demonstrates the orientation of participants to a possible problem, whether moral or rational, even, of practical order, in face of what was said. Still according to the authors, these explanations may be spontaneous or required. In the first case, those who provide them try to anticipate the possible moral implications that certain discursive follow-up may generate.

Such a perspective is latent throughout the decision, the text of which is composed of recurrent explanations of how the position of secularity of the State is not incompatible with respect for religious beliefs, as the following excerpts show:

> In the secular State, marked by the separation of State and religion, all religions deserve equal consideration and profound respect; however, there is no official religion, that is transformed into the only state conception, to abolish the dynamics of an open, free, diverse and plural society. It is the duty of the State to guarantee the conditions of equal religious and moral freedom, in a challenging context in which, if, on the one hand, the contemporary State seeks to penetrate the domains of the State (e.g. religious seats in the Legislative). Two strategies are highlighted here: a) to reinforce the principle of state secularism, with emphasis on the Declaration of Elimination of All Forms of Discrimination based on Religious Intolerance; and (b) to strengthen progressive readings and interpretations in the religious field, so as to respect human rights. (Vote of the Justice Joaquim Barbosa, p.229-230, citing: *Direitos Humanos* (Coord.). Curitiba: Juruá editora, 2007, p.24-25).

In his vote, the then Minister Joaquim Barbosa justifies his position favorable to the decriminalization of abortion of the anencephalic fetus, emphasizing the need to guarantee the laity of the State and, for this, makes a long explanation, in which he appropriates the words of jurist, about how laity does not constitute disrespect or disregard for religions. The same emphasis on respect is also given by Minister Celso de Mello in the explanation he lays about secularism to justify his vow, in line with that of his colleague:

> Indeed, pluralism is one of the essential characteristics of contemporary societies. Within a single state, there are people who embrace different

religions - or do not adopt any-; who profess different ideologies; who have disparate or even antagonistic philosophical moral conceptions. And today, it is understood that the State must respect these life choices and orientations, not being allowed to use its repressive apparatus, not even its symbolic power, to coerce the citizen to adjust his conduct to hegemonic conceptions in society, nor to stigmatize the "outsiders". (Vote of the Justice Celso de Mello, p.336, citing: Daniel Sarmento "Legalização do Aborto e Constituição". In: Nos Limites da Vida: Aborto, Clonagem Humana e Eutanásia sob a Perspectiva dos Direitos Humanos, p.03/51, 26-27, 2007, Lumen Juris).

Explanations and justifications such as those in these excerpts are identified in critical discourse studies in the category of *implicit*, as Fairclough (2003, p.42) states: "what is said in a text is always in relation to what is not said". Thus, the laity of the Brazilian state is affirmed in relation to the enormous symbolic force that religions, especially those of Christian bias, still hold on our society and its institutions. Indeed, as feminists as Vuola (2001) argue, the triumph of secularization over religion, a promise of modernity, did not materialize as a global phenomenon. On the contrary, it is evidence of Western Europe that, in the rest of the world, and there it emphasizes Christian Latin America, represents one more exception to be explained than a pacified rule.

The state is secular, but not atheistic: the pervasiveness of Christian culture

Throughout the decision secular / religious form an antagonistic pair indicative of the duality that crosses ADPF 54 in deciding in favor of women's right to abortion in a country heavily dominated by the symbolic power of religion. Feminist theory denounces the eminently androcentric, hierarchical and sustaining character of a patriarchal structure (ROSADO, 2001), which in turn also shapes the field of law (SOUSA, 2015).

Without articulating a rhetoric to highlight this subsistence, the ministers of the Supreme Court eventually accommodate the decision in the same way without confronting them. On the contrary, in the appellate decision there are many language arrangements that in face-to-face interaction studies would be associated with engagement marks that collaborate in the management of positive impressions (GOFFMAN, 1967). One of the tactics I identified in this management, besides the explanations about the compatibility between lay state and religiosity, was an approximation with the religious discourse itself. The best example is the opening of the vote of the rapporteur of the ADPF54:

JUSTICE MARCO AURÉLIO (RAPPORTEUR) - Father Antônio Vieira told us: "As time has not, nor can it have any consistency, and all things

from the beginning were born together with time, so neither it, nor they can stop for a moment, but with a perpetual moto, and an unsurpassable resolve to pass, and always go by "- Sermon on the First Sunday of Advent. (Vote of Justice Marco Aurélio, p.32).

It is noted in this section that Minister Marco Aurélio de Mello, inserts a passage in a quotation to the First Sunday Advent Sermon by Father Antônio Vieira, in the first lines of his text, seeking an approximation with the religious discourse, guided by a Christian and Catholic bias. The rapporteur's vote has a guiding role in the decision, since it is from this that the other members of the Court stand against or in favor. Thus, in supporting a lawsuit that runs counter to religious dogma, especially the Catholic Church, which undertakes a real moral crusade against abortion in any situation, it seeks protection in the words of a religious.

The judge could have directly triggered a view of the liberal tradition of the state, in which self-ownership is the indispensable basis for access to citizenship, and to assume that freedom of choice to carry on or interrupt unwanted gestation is an autonomy of about half the population. But instead, he chose to accommodate his rhetoric in less direct ways and clothed in a perspective that is not incompatible with Christian bias. Thus, he appropriated the words of the religious priest, who emphasize the need to accept changes that are invariably demanded according to each historical time, referring, therefore, to the object of the ADPF54, that only in the current stage of advancement of medical studies on anencephaly, led to a change of perspective of the judiciary over abortion of anencephalic fetus.

At the end of his vote, after a long exposition on the premises that sustained his decision, the minister concludes, emphasizing the separation between state and religion, but with a caveat:

> It is concluded that, despite the preamble, devoid of normative force - and could not be different, especially with regard to divine protection, which could never be judicially demanded -, Brazil is a secular state tolerant, due to Sections 19, item I, and 5, item VI, of the Constitution of the Republic. Gods and Caesars have separate spaces. The state is not religious, nor is it an atheist. The State is simply neutral. (Vote of the Minister Marco Aurélio, p.39).

The manner in which the minister closes his discussion, in which secularity and religion are central themes, corroborates the critique of feminists such as Vuola (2001), who I have already commented on and who is supported in Montero's work on the distinctive characteristics of the State national in Brazil, determined by a particular conformation with the religious field. In affirming that the secular state is not an atheist, the minister protects himself from the stigmas that fall on this figure in our society, the opposite to the religious man, who is distrusted for his lack of faith, perceived "as

a refusal to establish relations of reciprocity and alliance with the supernatural sphere and, ultimately, with human congeners" (MONTERO, 2011, p.3).

For the author, understanding what defines the peculiarities of the national state in Brazil demands to consider the historical hegemony of the still pervasive Christian culture in our society. To this extent, the ADFP judgment54 is illustrative, because throughout the text this bias is latent and in some moments emerges verbatim, revealing personal engagements with this field, as can be seen in the following excerpt:

> This morning, I woke up and thanked God for being able to contribute to humanity through a decision that could avert sorrows, anguish, pains, afflictions and, at the same time, I asked God for reason and passion to accompany me in the exercise of this highest apostolate that a human being can dedicate himself in this world of God: the magistracy. (Vote of the Justice Luiz Fux, p.154-155).

This clipping seems to provide support for what the Rapporteur has pointed out in a previous clipping about the fact that the state being secular does not imply it is contrarily being an atheist. This is what can be inferred from what was stated in this passage of the speech of Minister Luiz Fux, who, in the wake of Judge Marco Aurélio de Mello, at the beginning of his explanation accompanying his colleague's vote, assumes an explicit religious articulation.

From what I have shown in these analyzes, I understand that the State's secular argument, so recurrent in the ADPF54, as opposed to attesting a genuine and pacified incorporation of that value by the Court, reveals more ambivalence about such incorporation. The volume of explanations provided to justify such a premise is proportional to the danger of breaking the normative responsibility that the argument imposes on the STF when deciding on a subject such as abortion, so dear to the Christian tradition of our society.

But beyond the Brazilian society itself, I still consider another scope of dialogue to which the decision is addressed, although not in an assumed way. According to Diniz and Velez (2008, p.649), despite the laity of the Brazilian State, "it causes little political controversy the existence of religious congressmen or confessional political base, whose legislative agenda is to promote and defend the specific accommodated interests of their moral communities of origin and not an idea of reasonable moral pluralism.". Although the text does not allow more evidence on this interlocutor, he has a brief mention in that clipping that I set pages back, of the vote of the Minister Joaquim Barbosa, on the "religious benches in the legislative".

In any case, what stands out most about the textual arrangements that have manipulated the secular / religious pair, is that the Court has triggered them in order to dribble the tensions that decide on the decriminalization of the abortion of anencephalics imposed on its members in the context of a certain cultural order. Tensions that the

ministers and ministers of the FTS accommodated by a discursive strategy that, contrary to openly defying the hegemonic discourse, it aligned itself.

Deciding about abortion without talking about it: management of naming

This strategy of accommodation is what stands out in the judgment, and here, to continue my analysis, I highlight another discursive device that articulated it, management with appointments to deal with other dualities: life / death, fetus / woman. This is the pair: abortion and the therapeutic anticipation of childbirth. Demonstrating that, paradoxically, to approve the abortion was necessary before talking about it. I begin to demonstrate this paradox with a excerpt in the vote of Minister Celso de Mello, the only one to outline an argument that could have opened an interpretive gap for the extension of the right to interrupt pregnancy beyond cases of anencephaly:

> **The Federal Supreme Court,** Mr. President, **at the stage** of this trial, **is recognizing** that a woman, based on reasons directly based on her reproductive rights **and protected** by the undeniable effectiveness of the constitutional principles of the dignity of the human person, freedom, of personal self-determination and of intimacy, **has the insurmountable right** to opt for the therapeutic anticipation of childbirth, **in cases of** proven fetal malformation due to anencephaly, or, therefore, legitimized for reasons deriving from its private autonomy, **the right to express** its will the physiological process of gestation. (Vote of the Minister Celso de Mello, p.315, author's highlighting).

It is observed that the minister is based on the constitutional rights of the dignity of the human person, self-determination and intimacy, following a liberal tone, in which ownership of oneself is the basis for citizenship. These same principles are highlighted by other colleagues of the judge. However, he is the only one who comes to coordinate an extension of the right claimed by pregnant women of anencephalics to all women. Nonetheless, in her argument, she confines herself to women's freedom to "manifest" individual "will" for the continuation of any pregnancy, but fails to develop a more articulated defense to freedom from abortion, a term he even avoided.

As I introduced it further, not talking about abortion to approve it was one of the discursive strategies used in the judgment. In substitution, the Court adopted the nominal structure anticipating the therapeutic delivery, creating a set of meanings established in the initial petition, proposed by the current STF minister, Luis Roberto Barroso, at the time the lawyer of the National Confederation of Health Workers (CNTS). By a very instrumental rhetorical feature, Luis Roberto Barroso, in the previous note of the petition, states: "Therapeutic anticipation of delivery of anencephalic fetus is not abortion."

In language studies, naming is to consider the relation between name and thing symbolically. Subjects name from their position in a discursive formation, thus a name functions not simply as a label, but it produces meaning historically and ideologically (FAIRCLOUGH, 2003) and social effects. In this way, namings have discursive and ideological functionality. The lawyer creates a distinction between abortion and [7]*therapeutic anticipation of childbirth*, arguing that, in the context of the ADPF 54 decision, the first is discarded.

The term *therapeutic anticipation of childbirth* is a nominal structure composed of lexical elements specific to the field of medicine. After Foucault's (2005) studies, the medical discourse had its neutrality questioned, even though the formal rules of power and of knowledge enunciation exercise in this field continue to act legitimizing the most varied social practices. This is how, in ADPF 54 an articulated naming from this order of discourse was appropriately functional to a discursive alignment along the lines of legal argumentation, which seeks what Ferraz Júnior (2013) calls subsumption.

The author explains that the subsumption refers to the submission of the case to the proper rules of law for its application. Thus, by disqualifying the practice of interrupting a pregnancy as an abortion, this practice is removed from the illegality field and from all its associations to the heinous world, such as homicide, relocating it into the hygienic field of medicine, where procedure acquires therapeutic status.

In this sense, according to Pires (2013), in ADPF 54, the opportunity to address the collision between the interests of the unborn child and the reproductive autonomy of women has been lost. The decision, on the other hand, eliminates any further discussion on this topic, strongly excluding the topic of voluntary abortion, as the following excerpts show:

> I emphasize the allusion made by the arguer herself to the fact that the proclamation of abstract unconstitutionality of the criminal types is not postulated, which would remove them from the legal system. It is only intended that the mentioned statements are interpreted according to the Constitution. In this way, it is entirely unwarranted to convey that the Supreme Court will examine in this case the decriminalization of abortion, especially since, as will be seen, there is a distinction between abortion and therapeutic anticipation of childbirth... (Excerpt from the vote of Rapporteur Marco Aurélio, p.33).

> [...] (c) since there is nothing that can be done for the fetus, its withdrawal is the only therapeutic indication for the pregnant woman; (d) the withdrawal of the fetus by a qualified physician constitutes therapeutic anticipation of childbirth, and not abortion according to the Penal Code,

---

[7]   The term was proposed by Débora Dinis who represented ANIS.

a crime whose characteristic is the death of a fetus that is viable for extrauterine life caused by an abortive procedure (Excerpt from the vote of Justice Rosa Veber, p.90).

Therefore, it is apparent that, in the final analysis, the present ADPF takes care of the freedom of choice of the woman in disposing of her own body in the specific case in which she carries in her womb a fetus whose independent extrauterine life is absolutely impracticable. Thus, it is important to emphasize, the wide possibility of interrupting pregnancy is not discussed. The question here refers exclusively to the interruption of a pregnancy that is doomed to failure, since its outcome, even if all possible efforts are made, there will invariably be the death of the fetus. (Excerpt from the vote of Justice Joaquim Barbosa, p.147).

I also want to point out that this Federal Supreme Court, this afternoon, is not deciding or allowing abortion. This is a question posed to society. What we are dealing with here is essentially whether the interpretation that is possible to be given to the provisions of the Penal Code are compatible or not with the interpretation that has been given in the sense that it is also considered a crime to interrupt the pregnancy of an anencephalic fetus. (Excerpt from the vote of Justice Carmen Lúcia, p.172).

In these excerpts, we see again a series of explanations and justifications based on the normative responsibility of the Court in debating on a taboo subject in our society. As Warat (1985) proposes, a judicial decision is a persuasive piece that employs all kinds of argumentative resources, which tend to have an importance, not logically derived, but which obtains its acceptance by psychological and emotional association. In this sense, the persuasive nature of legal discourse inevitably determines the presence of [8]fallacies in its content, as was the case of the strategy articulated by naming and denial exchange, which favored an interpretative opening to the reception of the request seen in the ADPF. The success of this tactic demonstrates, on the other hand, the degree of marginalization in the legal environment and in Brazilian society itself of the feminist argument that abortion constitutes a woman's moral right of autonomy over her own body and over her own conscience. On the other hand, the discursive articulation of the ADPF 54 mirrors the Brazilian feminist articulation in its historical struggle for the legalization of abortion, of strong negotiating nature that, according to Bila Sorj (2002), is rooted in the Brazilian "political culture" itself to avoid conflicts and seek conciliatory solutions.

---

[8] The term fallacy is being used here in line with the pragma-dialectic perspective (EEMEREN; GROOTHENDOR, 1992), which broadly understands that the fallacy is a speech act that violates one or more rules of critical discussion, which, when used as an argument, promises to be decisive in the problem in question, while in reality it is not.

Maternal suffering, health and female autonomy: management of representations

The management of namings shaped a feminist cause to legal hegemonic discourses. However, in order to guarantee the traditional and patriarchal bias of this field, it was still necessary to manage certain female representations, emphasized not in the autonomy and freedom of women, but in maternal suffering and its harmful consequences to the health of pregnant women. The term representation in language studies with focus on gender refers to the discursive construction from the ways of being woman / man (LAZAR, 2005). These studies also highlight the fact that representations are shaped from particular perspectives of specific communities of practice in the interest of maintaining certain power relations. Along these lines, the female representation profile emphasized by the Court supports such relations within the framework that the Law values.

According to Pires (2013, p.581), in ADPF 54, "the mental health of women has acquired a greater scope of protection, on the grounds that the science of anencephaly on the part of the pregnant woman generates a state of psychic disturbance to a high degree, that their interests must prevail over the fetus's right to life as a constitutional value. "The following excerpts illustrate this evidence:

> [...] although in the context, there are other people involved, nobody's suffering is greater than that of the pregnant woman, because the anencephalic fetus is an event in her body. The pregnant woman, in this case, will not even become a mother, because there will not be - or there is not even - a child. By forcing a woman to keep a dying fetus or a technically dead one, the state and society meddle in her right to bodily integrity and to make decisions about her own body. In the case of healthy fetuses, it is still possible to discuss whether the woman is obliged to have the child, since the fetus will be a person and therefore presumed to have the right to be preserved. But the anencephalic fetus will never be a person, it will not have a human life, it is not even a subject of potential rights... (Doctors Telma Birchal and Lincoln Frias discourse in public hearings transcribed in the ADPF Report 54, p.65).

> Anyone (who does not even need to read legal literature), anyone who has had the opportunity to read "Manuelzão e Miguilim", from Guimarães Rosa, will know that perhaps the greatest example of human dignity that God has given was exactly that of Mom - and this considering I have a super dad! The dignity of the mother goes beyond herself, beyond her body. When Guimarães Rosa puts the woman carrying a dead son in her arms, who had a piece of cloth tied at his feet, hurt a few days before, she seeks to bathe the small body of her dead son and almost bumps into the basin; She then takes care that, even dead, he does not have any bumps

because it would be a suffering imposed on that little body. Whoever has read so much will know that when a choice is made of interrupting what could be the life of a moment or life for another month, it is not an easy choice, it is always a tragic choice; It is the choice that is made to continue and not to stop; it is the choice of the possible in an extremely difficult situation. Therefore, I think it must be known that all options like this, even this interruption, are of pain. The choice is which is the smallest pain; it is not to not hurt, because the pain of living has already happened, the pain of dying as well. She only makes the possible choice in this sense. (Anticipation of Carmen Lúcia's vote, p.174).

These excerpts summarize the main argument in ADPF 54 in favor of abortion / abortion of anencephalic pregnancy: the enormous suffering of women in this situation. In this direction, a discourse of solidarity is generated for these women. The representation of maternal suffering in the appellate decision generally precedes arguments in favor of female autonomy and their reproductive rights, as can be seen in the excerpt from the discourse of the doctors, in which they resort to this device and then defend the bodily integrity of the women and their rights to make decisions about their own bodies. The woman / fetus pair is very recurrent and is directly associated with the life / death pair in the emphasis given to the inevitable death of the fetus, its vital unfeasibility in direct relation to maternal suffering.

Although the suffering of women and the detriment of their health are almost indisputable arguments, there was an effort to carry emotional weight and dramaticity to the condition of the anencephalic pregnant, engendered by a representation of femininity based on the historical configuration of the emotional woman, with a certain appeal to the pathologization of the body, to the centrality of motherhood, which may even include self-denial, suffering, etc. In both clippings, a number of explanations and justifications also helps to compose this representation, "the mother in this case not even come to be a mother, for there will be - not even there - a son," "the choice is which lower pain; not to hurt, because the pain of living has already happened, the pain of dying too. "

This engendering, articulated to meet claims of women's reproductive rights, does so in order to frame the discourse in the hegemonic patterns of the patriarchal structure, in which the power of medicine has always had a regulatory role of women's autonomy for the control of their body. Control that the legal discourse also regulates, and that doctors Telma Birchal and Lincoln Cold do not make a point of answering when they ponder: "in the case of healthy fetuses, it is still possible to discuss if the woman is obliged to have the child, therefore it will be a person and therefore is presumed to have the right to be preserved." This is, in my view, the great duality expressed in the decision: to deal with women's rights without touching on the regulatory ways that maintain the status quo of our society and its gender asymmetries.

Final considerations

In developing this analysis, I have always been aware that ADPF 54 has been acclaimed in some media as a major achievement of women's rights, providing an opening for the decriminalization of abortion in Brazil. However, as I endeavored to demonstrate, such a conquest was engendered in a paradoxical way, in which to deal with a demand for feminist struggles, and which had the direct commitment of representatives of these groups, it was necessary to silence about such struggle, and to disregard the fruitful scientific production on the theme from the feminist academic milieu.

Simone de Beauvoir was mentioned in the decision by the rapporteur, Justice Marco Aurélio de Mello, and by Justice Carmen Lúcia, a fact that at first sight gives the impression that there was a discursive alignment with the author's ideas. Nevertheless, as I tried to explain, rather than approaching the discourse defended by this icon of feminism, the Court triggered a textualization supported by settings that the author herself attacked.

In her referential work, the Second Sex, the same from which the mentions in the decision were withdrawn, Simone de Beauvoir highlights how abortion is treated in bourgeois society in a hypocritical way and as a repugnant thing. And she exemplifies: "That a writer describing the joys and sufferings of a parturient is perfect; if he mentions an abortive he will be soon accused of wallowing in filth and describing mankind in an abject way" (BEAUVOIR, 1967, p.248). Paradoxically, the strategies triggered in the decision to (not) speak of abortion, leave latent the implicit idea that it is "a repugnant crime that is indecent to allude" (Ibid.), as the feminist criticized.

As noted by Scavone (2008), the debates and feminist political actions in favor of the liberalization of abortion in our country were marked by numerous political negotiations and, above all, by advances and retreats. Thus, I understand that the discursive strategies that define the decision reflect the negotiating and conciliatory nature of the Brazilian feminist approach. It seems that there was, as Miguel (2012, p.671) pointed out, "a 'realistic' accommodation to conditions of debate in the political field."

Thus, I understand that, on the one hand, the feminist movement was able to account the Supreme Court's decision as a progress, although limited, in the struggle for the expansion of women's rights. On the other hand, the strategy of accommodating this decision to the conditions molded by a structure that, as denounced by the feminist theorists of the Law area (BARTLETT, 1991; CASTILHO, 2008; PIMENTEL, 2009), traditionally gave normative strength to the inequalities of gender, ends up not investing in the dismantling of such structure and, to the contrary, to some extent, collaborates with its permanence.

It is necessary, therefore, now at the end, to reflect on the warning made by the feminist Audre Lorde (2007), already cited in this text, on the dangers of using the "Master" tools to dismantle his "house." Tools that, according to her, may even allow us to beat him temporarily at his own game, but will never allow us to bring up genuine changes.

FREITAS, L. A decisão do STF sobre aborto de fetos anencéfalos: uma análise feminista de discurso. **Alfa**, São Paulo, v.62, n.1, p.11-33, 2018.

- *RESUMO: Este artigo traz uma análise do acórdão do Supremo Tribunal Federal na Ação de Descumprimento de Preceito Fundamental n° 54 (ADPF 54), que buscou solucionar a polêmica acerca da possibilidade de interrupção voluntária da gestação em caso de fetos com anencefalia. Por uma abordagem feminista de análise de discurso, o trabalho foca a forma paradoxal pela qual a Corte atendeu uma demanda feminista e de movimentos de luta pelos direitos das mulheres: 1) escamoteando os vínculos com ativistas desses grupos e com os discursos que elas defendem sobre a autonomia das mulheres em relação a seus corpos; 2) mantendo o padrão tradicionalista e androcêntrico próprio do Direito. Assim, a análise descreve três artifícios de linguagem que se sobressaem na articulação desse paradoxo: explicação, nomeação e representação. O trabalho aponta a persistência no campo do Direito de paradigmas tradicionais e androcêntricos que a Justiça valoriza e com os quais opera. Este estudo visa contribuir para o debate sobre a descriminalização do aborto no Brasil, bem como discutir a relação entre linguagem / gênero / direito.*

- *PALAVRAS-CHAVE: Discurso. Gênero. Feminismo. Direito. Aborto. Anencéfalo.*

REFERENCES

ALDANA, M. Vozes católicas no Congresso Nacional: aborto, defesa da vida. **Rev. Estudos Feministas**, v.16, n.2, p.639-646, ago. 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-026X2008000200018&lng=en&nrm=iso>. Acesso em: 30 set. 2014.

BARTLETT, K. T. Feminist legal methods. In: BARLETT, K. T.; KENNEDY, R. **Feminist legal theory**. Colorado: Westview Press, 1991. p.370-403.

BAXTER, J. A. **Positioning gender in discourse:** a feminist methodology. New York: Palgrave Macmillan, 2003.

_____. Feminist post-structuralist discourse analysis: a new theoretical and methodological approach? In: HARRINGTON, K.; LITOSSELITI, L.; SAUNTSON, H.; SUNDERLAND, J. (Eds.). **Gender and Language Research Methodologies**. Basingstoke: Palgrave Macmillan, 2008. p.243-255.

BRADFORD, R. **Stylistics**. London and New York: Routledge, 1997.

BRASIL. Supremo Tribunal Federal. **Acórdão na Ação de Descumprimento de Preceito Fundamental (ADPF) 54**. Relator: MELLO, M. A. de. Publicado no DJ de 12/04/2012, p.433. Disponível em: <http://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=TP&docID=3707334>. Acesso em: 17 jan. 2017.

BEAUVOIR, S. **O segundo sexo:** a experiência vivida. Tradução de Sérgio Milliet. São Paulo: Difusão europeia do livro, 1967.

BUCHOLTZ, M. Theories of discourse as theories of gender: discourse analysis in language and gender studies. In: HOLMES, J.; MEYERHOFF, M. (Eds.). **The handbook of language and gender**. Oxford: Blackwell Publishing, 2003. p.43-69.

_____. The feminist foundations of language, gender, and sexuality research. In: EHRLICH, S.; MEYERHOFF, M.; HOLMES, J. **The Handbook of Language, Gender, and Sexuality.** New York: Wiley-Blackwell, 2014. p.23-47.

CALDAS-COULTHARD, C. "Women who Pay for Sex. And Enjoy It": Transgression versus Morality in Women's Magazines. In: CALDAS-COULTHARD, C.; COULTHARD, M. (Org.). **Texts and Practices**. London: Routledge, 1996. p.248-268.

CAMARGO, M. M. L. As audiências públicas no supremo tribunal federal brasileiro: o exemplo da ADPF 54. **Publica Direito** [*online*], 2011. Disponível em: <http://www.publicadireito.com.br/artigos/?cod=01d8bae291b1e472>. Acesso em: 2 nov. 2016.

CARVALHO, F. M. **Audiências Públicas no Supremo Tribunal Federal:** uma alternativa democrática? 2011. 118 f. Dissertação (Mestrado em Direito) – Universidade Federal do Rio de Janeiro, Centro de Ciências Jurídicas e Econômicas, Faculdade de Direito, Rio de Janeiro, 2011.

CASTILHO, E. W. V. A criminalização do tráfico de mulheres: proteção das mulheres ou reforço da violência de gênero? **Cadernos Pagu** [*online*]., n.31, p.101-123, 2008. Disponível em: <http://dx.doi.org/10.1590/S0104-83332008000200006>. Acesso em: 4 jan. 2017.

CHRISTIE, C. **Gender and language**: towards a feminist pragmatics. Edinburgh, UK: Edinburgh University Press, 2000.

DINIZ, D.; VELEZ, A. C. G. Aborto na Suprema Corte: o caso da anencefalia no Brasil. **Revista Estudos Feministas**, Florianópolis, v.16, n.2, ago. 2008.

EEMEREN, F. H. V.; GROOTENDORST, R. **Argumentation, communication and fallacies.** A pragma-diatectical perspective. Hillsdale: Erlbaum, 1992.

EHRLICH, S.; MEYERHOFF, M.; HOLMES, J. **The Handbook of Language, Gender, and Sexuality**. New York: Wiley-Blackwell, 2014.

FAIRCLOUGH, N. **Analysing discourse**. New York: Routledge, 2003.

FERRAZ JÚNIOR, T. S. **Introdução ao estudo do direito**. Técnica, decisão, dominação. 7. ed. São Paulo: Atlas, 2013.

FOUCAULT, M. **Microfísica do poder**. Rio de Janeiro: Edições Graal, 2005.

FREITAS, L. G. Representações de papéis de gênero na violência conjugal em inquéritos policiais. **Cadernos de Linguagem e Sociedade**, v.12:1, p.128-152, 2011a.

_____. Violência contra a mulher no sistema penal de uma cidade do interior do Brasil. **Discurso & Sociedad**, n.5, p.701-722, 2011b.

_____. Análise crítica do discurso em dois textos penais sobre a lei Maria da Penha. **Alfa**: Revista de Linguística (UNESP. *on-line*), v.57, p.11-36, 2013.

_____. Argumentação e discurso sobre Lei Maria da Penha em acórdãos do STJ. **Bakhtiniana**, São Paulo, n.9 (1), p.71-89, jan./jul. 2014.

FREITAS, L. G.; LOIS, C. C. A Ação de descumprimento de preceito fundamental nº 54 e a necessária inserção do paradigma de gênero no Direito: uma análise a partir da Filosofia da Linguagem. In: OLIVEIRA JUNIOR, J. A. de; VILLATORE, M. A. C. (Org.). **Encontro de Internacionalização do CONPED.** Barcelona: Ediciones Laborum, v.14. p.11-29, 2015.

FREITAS, L. G.; PINHEIRO, V. Atores da violência de gênero: suas narrativas no inquérito policial. In: ZYNGIER, S.; VIANA, V. (Org.). **Avaliações e perspectivas:** estudos empíricos em Letras. Rio de Janeiro: UFRJ, 2010. p.223-243.

_____. **Violência de gênero, linguagem e direito**. Análise crítica de discurso em processos na Lei Maria da Penha. São Paulo: Paco, 2013.

FUNCK, S. B. A (in)visibilidade da mulher na mídia impressa: uma análise discursiva. **Comunicação & Inovação**, v.8, p.15-22, 2007.

GUIMARÃES, J. A. C. **Elaboração de ementas jurisprudenciais**. Elementos teórico-metodológicos. Brasília: Conselho da Justiça Federal, 2004.

GOFFMAN, E. **Interaction ritual**: essays on face-to-face behavior. New York: Anchor Books, 1967.

GUBA, E. G.; LINCOLN, Y. S. Competing paradigms in qualitative research. In: DENZIN, N. K.; LINCOLN, Y. S. (Eds.). **Handbook of qualitative research.** Thousand Oaks: Sage, 1994. p.105-117.

HARAWAY, D. Saberes localizados: a questão da ciência para o feminismo e o privilégio da perspectiva parcial. **Cadernos Pagu,** v.5, p.07-41, 1995.

HEBERLE, V.; OSTERMANN, A. C.; FIGUEIREDO, D. (Org.). **Linguagem e gênero no trabalho, na mídia e em outros contextos**. Florianópolis: UFSC, 2006.

HOLMES, J.; MEYERHOFF, M. **The handbook of language and gender**. Oxford: Blackwell Publishing, 2003.

KITZINGER, C. Doing feminist conversation analysis. **Feminism & Psychology**. n.10:2, p.163-193, 2000.

LAZAR, M. Politicizing gender in discourse: feminist critical discourse analysis as political perspective and praxis. In: LAZAR, M. (Ed.). **Feminist critical discourse**

analysis. Gender, power and ideology in discourse. New York: Palgrave Macmillan, 2005. p.1-28.

_____. Feminist Critical Discourse Analysis: Articulating a Feminist Discourse Praxis. **Critical Discourse Studies**, v.4:2, p.141-164, 2007.

LORDE, A. **Sister outsider***: essays and speeches. Berkeley: Crossing Press, 2007.

MEY, J. L. **Pragmatics**: an introduction. 2. ed. Mass., EUA e Oxford, Reino Unido: Blackwell Publishers, 2001.

MIGUEL, L. F. Aborto e democracia. **Estudos Feministas**, Florianópolis, v.20:3, p.657-672, set./dez. 2012.

MILLS, S. **Feminist stylistics**. London: Routledge, 1995.

MIRANDA NETTO, F. G. de; CAMARGO, M. M. L. Representação argumentativa: fator retórico ou mecanismo de legitimação da atuação do Supremo Tribunal Federal? **Anais do XIX Encontro Nacional do CONPEDI**, Fortaleza – CE, 09-12 de junho de 2010. Disponível em: <http://www.conpedi.org.br/manaus/arquivos/anais/fortaleza/3589.pdf>. Acesso em: 18 fev. 2018.

MONTERO, P. O campo religioso, secularismo e a esfera pública no Brasil. **Boletim CEDES**, PUC-RJ, 2011. Disponível em: <http://www.cis.puc-rio.br/cis/cedes/PDF/out_2011/campo.pdf>. Acesso em: 02 out. 2017.

OSTERMANN, A. C.; ANDRADE, D. N. P.; FREZZA, M. A prosódia como componente de formação e de atribuição de sentido a ações na fala-em-interação: o caso de formulações no tribunal. **DELTA**, São Paulo, n.32:2, p.481-513, ago. 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502016000200481&lng=pt&nrm=iso>. Acesso em: 9 fev. 2017.

PAREDES, J. **Hilando fino**. Desde el feminismo comunitario. La Paz: Creative Commons, 2010.

PASSUELLO, C. B.; OSTERMANN, A. C. Aplicação da análise da conversa etnometodológica em entrevista de seleção: considerações sobre o gerenciamento de impressões. **Estud. Psicol.**, Natal, n.12:3, p.243-251, dec. 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-294X2007000300006&lng=en&nrm=iso>. Acesso em: 1 fev. 2017.

PIMENTEL, S. A superação da cegueira de gênero: mais do que um desafio – um imperativo. **Revista Direitos Humanos**, n.2, p.27-30, jun. 2009.

PIRES, T. I. T. Uma abordagem interpretativa dos fundamentos jurídicos do julgamento da ADPF 54, dignidade humana, liberdade individual e direito à saúde. **Revista de Direitos Fundamentais e Democracia**, Curitiba, n.14:14, p.577-598. jul./dez. 2013.

ROSADO, M. J. O impacto do feminismo sobre o estudo das religiões. **Cadernos Pagu**, Campinas, n.16, p.79-96, 2001. Disponível em: <http://www.scielo.br/scielo. php?script=sci_arttext&pid=S0104-83332001000100005&lng=en&nrm=iso>. Acesso em: 28 fev. 2017.

SACKS, H. **Lectures on Conversation**. Oxford: Blackwell, 1992.

SCAVONE, L. Políticas feministas do aborto. **Revista Estudos Feministas**, v.16:2, p.675-680, ago. 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_ arttext&pid=S0104-026X2008000200023&lng=en&nrm=iso>. Acesso em: 30 set. 2014.

SORJ, B. O feminismo e os dilemas da sociedade brasileira. In: BRUSCHINI, C.; UNBEHAUM, S. **Gênero, democracia e sociedade brasileira**. São Paulo: FCC/ Editora 34, 2002. p.98-107.

SOUSA, R. M. **Introdução às teorias feministas do direito**. Porto: Edições Afrontamento, 2015.

VUOLA, E. **God and the government:** women, religion, and reproduction in Nicaragua. Meeting of the Latin American Studies Association (LASA), Washington DC, setembro de 2001. Disponível em: <http://lasa.international.pitt.edu/Lasa2001/ VuolaElina.pdf>. Acesso em: 17 fev. 2017.

WARAT, L. A. As falácias jurídicas. **Revista Seqüência** (UFSC), Florianópolis, v.06, n.10, p.123-128, 1985. Disponível em: <https://periodicos.ufsc.br/index.php/sequencia/ article/view/16702/15255>. Acesso em: 10 dez. 2017.

WEATHERALL, A.; PRIESTLEY, A. A Feminist Discourse Analysis of Sex 'Work'. **Feminism & Psychology**, v.11:3, p.323-340, 2001.

WOODAK, R.; MEYER, M. (Orgs.). **Methods of Critical Discourse Analysis**. Londres: Sage Publications, 2001.

RHETORICAL-DISCURSIVE PERSPECTIVES FOR THE STUDY OF PATHETICIZATION

Mário Acrisio ALVES JUNIOR[*]
Micheline Mattedi TOMAZI[**]

▪ ABSTRACT: Taking as a starting point some reflections on the pathetical dimension of discourse, this article focuses the discussion of the three rhetorical modes of persuasion – *logos*, *pathos*, and *ethos* – in an attempt to argue in defence of their integration, with contribution from the new rhetoric, sophistics and the semiolinguistic theory of discourse. Refusing the dichotomous perspective of the Cartesian paradigm, our main purpose is to draw attention to the need to reconfirm a relationship of interdependence between *pathos* and *logos*. To illustrate our proposal, after a theoretical exposition on the subject, we analyse excerpts from a journalistic chronicle in the hope of confirming the legitimacy of the subject under consideration and the need for more consistent debates and demonstrations. Once the categories of intergenericity and metaphor have been chosen as discursive structures representative of the *logos*, the analysis allows us to observe their deploymenting in the effects of patemization, considered here as meaning effects.

▪ KEYWORDS: *Pathos*. *Logos*. New Rhetoric. Sophistics. Semiolinguistics.

Introduction

In recent decades, discussions regarding emotion in discourse have been taking shape, especially based on the postulate that certain emotions can be intentional and, therefore, rational. Such a posture calls into question a traditional view in which reason and emotion are understood in a relationship of opposition — an idea derived from the classical Platonic dualism *body/mind* but which was actually established by Descartes in the seventeenth century. It advocated the suppression of passions on the grounds that only by being free of them does the mind remain healthy.

In the wake of Plato and the Cartesian paradigm, from which paradoxes such as *body* and *mind*, and *reason* and *emotion* came to be crystallized, other dichotomies were established. Among the many formulations, we can cite some that are both widely known and widely disseminated, for example, argumentation x rhetoric, objectivity x subjectivity, exact sciences x humanities, and one that particularly interests us:

* Federal University of Espírito Santo (UFES), Post-graduate Program in Linguistics, Goiabeiras campus, Vitória - Espírito Santo - Brazil. Department of Languages and Letters. marioalwes@hotmail.com

** Federal University of Espírito Santo (UFES), Post-graduate Program in Linguistics, Goiabeiras campus, Vitória - Espírito Santo - Brazil. Department of Languages and Letters. michelinetomazi@gmail.com.

logos x *pathos*/*ethos* dualism, which encompasses the three means of persuasion covered by Aristotle in *Rhetoric* (ARISTÓTELES, 2012).

On the basis of this problem and in light of the finding that very little has been produced to support further discussion and revision of the dichotomy, we believe it is necessary to bring the debate to the fore. A review of the literature shows that there has been some progress with regard to the assumption of a less rigid idea, according to which there emerges a more harmonious movement between artistic proofs of the Aristotelian triangle, which we interpret as discursive perspectives in interactions or — in Galinari's words (2014, p.257) — as "dimensions of the same discourse". Further, because our position is to avoid dichotomous formulas, understanding that many of them put up barriers to the building of knowledge and the integration of different disciplines, we will seek to establish our discussion by articulating notions derived from different approaches to language studies that are apparently distinct, but which, in our view, complement one another. Such notions include the following: discourse analysis, in its semiolinguistic aspect; new rhetoric, with its most prominent representatives; and a debate with sophistics, whose contributions are so fundamental as to effectively exceed the limits imposed by Cartesian rationalism.

Although our proposal for theoretical evaluation and analysis had as the main target *pathos* — a means of persuasion concerning the emotions aroused by the orator/speaker in the audience/co-speaker, in this introductory section it is worthwhile to anticipate the inevitability of using the concept of *logos*. Unlike the model that considers *logos* (rational/objective) to be the opposite of *pathos* and *ethos* (less rational/more subjective) from the viewpoint of a fragmenting rationalism, we understand the latter as the development of the former. Given that *logos* is related to discourse itself — or to argumentative construction — we cannot ignore that the construction of a representation of the self (*ethos*) or the linguistic expression of certain emotions (*pathos*) must be the result of discourse (*logos*).

Thus, admitting that *pathos* (as well as *ethos*) always results from *logos* — the latter materialized in texts through different rhetorical-discursive strategies — and understanding *pathos* as a property of "discourses that work on emotional effects for strategic purposes" (CHARAUDEAU; MAINGUENEAU, 2014, p.372), after a theoretical exposition this article investigates processes for managing emotions in a journalistic chronicle. To continue with this theoretical-analytical path, we assume the following:

- in a dialogue between the elements of discourse studies and sophistic doctrine, it is possible to understand *pathos* as an outcome of *logos* and thus to more didactically and intelligibly systematize a framework of emotions (*pathos*) as an effect of discourse (*logos*);
- accordingly, certain stereotypes or social representations are activated in discourse with strategic purposes so that they achieve — and also reflect —

pathetic effects. However, if Charaudeau (2013) understands emotions as being linked to knowledge based onbelief (one of the ways to construct and/or maintain representations), there is no way to dissociate them from the way in which discursive activities condition the appeal to emotions based on social representations in discourse, particularly in the scope of journalistic media;

- because the journalistic chronicle is one of the most emblematic textual media genres with respect to the appeal to all types of subjectivity, including emotions in discourse, it is a privileged discursive space for the examination of the possible pathetic effects reflected and potentially produced in the interlocutor, above all because of the openness enjoyed by the speaker to "take a chance" on various strategies, whether they are those guaranteed by the language itself, or through certain representations to which the chronicler resorts to achieve his or her purposes in the discourse.

The article continues with a brief presentation of the state of the art on the subject before the reappearance of rhetoric in the twentieth century; a proposal for the revision of the subject in focus, with the contributions of the new rhetoric, sophistics, and semiolinguistics; a brief qualitative and interpretative analysis of excerpts from a journalistic chronicle by Lya Luft to illustrate the proposal under discussion; and some final considerations on the articulation between the theoretical trajectory and the analysis.

Patheticization and language: some problems

The issue of the importance of emotions in the study of discourse is nothing new. It was first treated by rhetoricians as a phenomenon concerning the passions, and according to Aristotle (2012, p.13), it is one of the three persuasive modes present in discourse: "some depend on the moral character of the speaker [*ethos*]; others putting the listener in a certain frame of mind[*pathos*]; and others in the discourse itself, in so far as it proves or seems to prove [*logos*]".

Of the three modes of persuasion, *pathos* is linked to the passions, affections, and emotions aroused in listeners by a speaker. Addressing such a subjective aspect — as is the affective disposition of the listener — may give the impression that *pathos* is a phenomenon that escapes the control of the discourse because the emotional effects produced in another person are always dependent on the other's receptivity. However, not even for this reason can we attribute a less essential role to *pathos*. Aristotle himself (2012, p.XXX) states that *ethos*, *pathos*, and *logos* are the "three elements of proof that *together* contribute to enthymematic reasoning"[1]; that is, the rationale assumes shared knowledge between speaker and audience, treating certain assertions and assumptions as non-problematic.

---

[1]   Our emphasis.

Although Aristotle's framework suggests the integration of the three modes of persuasion, in the classical period there were already those who, similar to Plato, elaborated their canonical postulates based on dichotomies such as *body/mind* — which was possibly the most prominent in the "golden age" of rhetoric, having served as a paradigm for so many other oppositions such as *conviction/persuasion* and, along the same lines, *reason/emotion*. It is important to highlight that the latter was sustained for centuries, gaining even more force in the seventeenth century under the aegis of Cartesian thought:

> [...] With Plato we have a sample of the exclusion to which emotion has been relegated throughout history — a thought followed by various philosophers (e.g., St. Augustine) from other times. However, it is with some philosophers, such as Descartes and Kant, that such a split was shown to be even more acute. If, as Descartes postulated, passions are a sign of illness, only if they were jettisoned would the mind be in perfect health. This dichotomous way of thinking about emotion and reason contributed to centuries of delay in debates on the subject. (LIMA, 2007, p.140).

In later centuries, this rational-mechanistic idea was the foundation that resulted in almost the annihilation of the Greek rhetorical tradition. However, the Western world of the Age of Enlightenment — dominated by scientism — could not accept the plausible forms of thought or credible demonstrations predicted by Aristotle's rhetoric. The rational and the absolute were required to arrive at truly valid statements that privileged logical forms of reasoning and notions such as the assertion that the value of truthis established in the space of argumentative theory. This is how the great conceptual schisms occurred because the influence of philosophers such as Descartes would be responsible for the definitive separation between argumentation and rhetoric, which carries

> [...] in its "DNA", the property of reproducing in cascade, bringing to light other dualisms, respectively analogue, such as: "**convince** x **persuade**", "**reason** x **emotion**", "**logic** x **rhetoric**", "**not misleading** x **misleading**", "**valid** x **invalid argument**", "**good rhetoric** x **bad rhetoric**", "**objectivity** x **subjectivity**", "**logos** x **ethos/pathos**". (GALINARI, 2014, p.261, bold highlighting by the author).

In view of the specific purposes of our proposal, it should be observed that such a split ended up affecting the harmony of the Aristotelian triangle formed by the three modes of persuasion due to the domination exerted by the Cartesian philosophical paradigm on epistemological models in various sectors of knowledge, which has persisted to the present day.

Thus, among the studies that articulate emotion and discourse, we must note the reflections of Galinari (2007, 2014), who indicates some barriers to the reintegration of the modes of persuasion, describing the paradigmatic influence as a "Cartesian fetish" (GALINARI, 2014). Another way to describe such influence is as a mania or a culture of dichotomization, whereby the Aristotelian triangle is often the target of a bipartition in which *logos* — shown as objective proof — is opposed to *ethos* and *pathos*. In his analysis of certain authors, Galinari (2014) observed obstacles that impede the effective integration of rhetorical proofs (although such authors were not aware of the hindrance caused by these obstacles):

> [...] Eggs (2008)[2], for example, based on his reading of Aristotle, separates the modes of persuasion into 'two blocks of conviction': on the one hand, we find *logos* (classified as inferential, as reasoning, and as argumentation); while in another block of the dichotomy, *ethos* (habitus, virtue, character) and *pathos* (passion, affection) come together. In turn, Plantin (2005)[3], when mentioning the three paths for validating an opinion, executes his divorce within the loving/rhetorical trio: to one side goes, solitarily, *logos* (objective, propositional proof); and to the other, *ethos* and *pathos* (subjective non-propositional 'proofs') follow entwined. (GALINARI, 2014, p.260).

With respect to Eggs' stance, we note a contradiction because, given that his proposal is based on his reading of Aristotelian rhetoric, which seems to offer no room for a dichotomous view, the division into blocks would be an almost declared paradox. The purpose of Plantin's idea — the *logos* x *pathos*/*ethos* opposition — probably comes from another, more comprehensive view, between *objectivity* and *subjectivity*. Galinari emphasizes that although the reflections of these and other authors offer ample contributions to the analysis of discourse, there is still a preference for strict separations.

Although we can understand that such authors are guided by an epistemological model suited to their purposes, dichotomies, in this case, can be real obstacles, so that studies related to rhetorical proofs go beyond the barriers of theory to become more operational, for example, in the practices of reading and discursive analysis.

It is important to emphasize that our positioning coincides with that of Galinari (2007, 2014), in the sense of accepting the interrelationship of proofs of persuasion. Thus, instead of considering that *logos* is opposed to *pathos* and to *ethos*, we deem it imperative to revise this proposal, thereby abandoning old dichotomies. In the integrative outlook adopted here, both *pathos* and *ethos* would consist of a semantic-discursive development of *logos* (see GALINARI, 2014), or better — of discourse as

---

2   Reference to: EGGS, E. *Ethos* aristotélico, convicção e pragmática moderna [Aristotelian ethos, conviction, and modern pragmatics]. In: AMOSSY, R. (Org.). **Imagens de si no discurso:** a construção do *ethos* [Images of oneself in discourse: the construction of ethos]. São Paulo: Contexto, 2008, p.29-56.

3   Reference to: PLANTIN, C. **L'argumentation:** histoire, théories et perspectives. Paris: PUF, 2005.

the result of an enunciative action, which naturally involves the action of the subject of the enunciation.

In the next section, we will show how the reformulations performed by the new rhetoric and how the sophistic doctrine regarding *logos* are productive sources in supporting the proposal adopted here.

Contributions of the New Rhetoric and Sophistics

As noted in the section above, despite rhetoric having suffered — from the seventeenth century onward — as a discipline or art of argument, it reappeared with considerable energy in the twentieth century. The work *The New Rhetoric: A Treatise on Argumentation* (PERELMAN; OLBRECHTS-TYTECA, 1996), first published in 1958, marks, in particular, the (re)integration of rhetoric and argumentation, which would — among other contributions — offer to language studies the following: the incorporation of all discursive forms of persuasion, such as advertising and other media (e.g., the consideration of non-verbal matter, whether in purely imagistic discourses or in syncretic forms); and the insertion of the non-rational, thus opening up the discourse to the passionate universe (*pathos*). Consequently, the new rhetoric indicates the path for escape from an outdated rationalism, giving way to a logic of the reasonable, the probable, and the credible (see FERREIRA, 2015).

In the wake of this integrative posture, the neo-rhetorical postulates point to the fact that the three components of persuasion — *logos*, *pathos*, and *ethos* — complement more than detract from one another. They are — before being isolated perspectives — fully interactive elements. However, it is not difficult to admit that, in a situation of communication, an orator/speaker is capable of connecting with the affectivity (*pathos*) of the audience/co-speaker through a representation of him- or herself (*ethos*) as the effect of the discourse itself (*logos*).

Thus, the modes of persuasion can productively be compared to integrated perspectives or, in the case of a didactic application, different reading perspectives that act simultaneously. This does not mean that each of them cannot be examined separately; however, this must be done, clearly and solely, for the sake of systematization, so that the three elements are understood within the particularities that characterize each of them.

Although all of the sophistication provided by the new rhetoric had been part of a very important turning point, some legacies of the Cartesian paradigm remain strong. We must explain — once again based on the studies of Galinari (2011, 2014) — that a more solid understanding of the interrelation of the rhetorical proofs requires a deepening of the notion of *logos*. In Galinari's view, the sophistic element of rhetoric involves a theoretical-conceptual framework capable of conceiving *logos* as more than a set of reasonable mental operations, which, as we have seen, would have been a great innovation of the neo-rhetoricians. Based on a study of the sophists — who

were considered to be great thinkers and educators in the fifth century B.C.[4], despite being somewhat marginalized — and contemporary works about the subject, Galinari understands that *logos* "is not restricted to the field of mental operations or the reasonings outlined in the materiality of the discourse, since, thinking such as this, one can interpret that this would exclude the field of the **saying**, of the **enunciation**", of which "the sophists were great scholars, covering several of its dimensions, and always — which is most important — from a rhetorical and communicative perspective" (GALINARI, 2014, p.263, bold highlight added by the author).

Pinto's study (2000), titled "A Doutrina do *Logos* na Sofística" [The Doctrine of Logos in Sophistics], is based on a deep theoretical foundation and investigates the way in which *logos* was created, and how it developed with the fostering of the school of the first sophists. Among many viewpoints, Pinto cites a work by Emmanuele Riverso about the influence of these thinkers:

> With sophistics, a crisis arises with *logos*, as had been elaborated by the preceding philosophies. Its validity for discovering an absolute truth — situated in the structure of the real — was demolished, and with it, the primitive objectivity of the truth was demolished. On the other hand, its persuasive force was perfected; that is, its capacity for modelling interpretations of the things which — by their internal coherence and the ability to elicit the adhesion of others — were constituted as truths created by the human being. (RIVERSO apud PINTO, 2000, p.100).

Pinto (2000) also notes that there was a sophistical concern with language — a concern centred above all else on the use of the words, which is evidence of a less rigorous idea of *logos*, given the view of the dynamicity of the language itself. Thus, the author cites the examples of Protagoras, who "elaborated on grammar issues, dealing with the gender of names and the classification of parts of speech", as well as the "criticism of mismatched expressions and the syntactic composition of speech" (PINTO, 2000, p.173); Prodicus, because his activity "in the field of spoken *logos*-related issues is documented by numerous sources who attribute to him [...] the mastery in the art of the division of names" (PINTO, 2000, p.182); and Antiphon, whose teachings include the assumption that "the concepts we use in discourse [...] do not have a single meaning because the ultimate constituents of nature — discernible through meaning or thought — are singular things" (PINTO, 2000, p.208), in a possible glimpse of consideration of the dependence of contextual factors for the interpretation of discourse.

Based on the activities developed by these and other sophists, who, as we can see, offer an expanded view of *logos*, seizing it — in the case of *logos* in relation to *saying* — by means of different categories of language, Galinari (2014, p.264)

---

[4]   We refer to sophistics as practiced in its first era — ancient sophistics or first sophistics — in the second half of the fifth century B.C., and whose most prominent representatives were figures such as Gorgias, Protagoras, and Hippias, among others. It differs from the second sophistics, which was developed in the second century A.D.

suggests a subdivision of *logos* into different dimensions such as "word *logos*", considering "the whole rhetorical weight of vocabulary formation and of lexical selection"; "syntax *logos*"; "prosodic *logos*"; "reasoning *logos*"; "narration *logos*"; "description *logos*";"enunciation *logos*" (in the first-, second-, and third-person). The author further explains as follows:

> Obviously, all these and other dimensions of *logos*, including — why not! — the reasonings, reinforce or construct opinions, "theses", or worldviews, based on their cognitive and referential structure; however, they also develop — that is the question — in the *ethos* and in the *pathos* during the interaction. In other words, this is equivalent to saying that *ethos* and *pathos* (the images of the selfand the responsive emotions in the audience) only become reality based on discourse; that is, from the use of its structure and its reasonings; in short, from everything that was referred to above as *logos.* (GALINARI, 2014, p.264).

Taking the quotation above as a parameter— which records, for the purposes intended here, the presuppositions with which we agree and from which we shall proceed later to the analysis — we now more clearly reaffirm our initial hypothesis, which is that *pathos* — an aspect of discourse related to emotions — must be defined and examined from its overlap with *logos* and with *ethos*; and we establish a second hypothesis, according to which the orator/speaker has an ample *strategy space* — of a linguistic and rhetorical-discursive nature — to put into practice his project of action and influence over the other, in the most varied exchange situations (oral or written).

Having reached the end of the contributions of the new rhetoric and of sophistics, we turn to the subsequent section, in which we proceed to some considerations of Charaudeau (2007, 2010) about his discursive proposal on the study of emotions.

The notion of patheticization: contributions from Charaudeau

In this section we will show that the study of emotions must be complemented by a theory on discourse. We believe that the contributions presented thus far are essential; however, they lack explanation as to how the analyst should grasp emotions.

Charaudeau (2010) warns that the viewpoint of an analysis of discourse should be distinguished, for example, from the viewpoints of a psychology of emotions, whose perspective reflects upon that which the subjects actually feel, or a sociology of emotions — the social relationships and behaviours of social groups. Without underestimating these realities, Charaudeau emphasizes that emotions must be studied from an interdisciplinary perspective, or rather, with an approach involving analysis of integrative discourse, which is not strictly in the field of the relations between language and discourse but which seeks in disciplines such as philosophy, psychology,

and sociology points of contact that allow a multidimensional and, therefore, broader understanding of the phenomenon of emotions in discourse. Thus,

> [...] despite differing from psychology and sociology, discourse analysis needs them, insofar as their analyses show the intentionality mechanisms of the subject and the mechanisms of social interaction, and the way in which social representations are constituted. Certain notions are more conducive to interdisciplinarity than others, precisely because they are at the centre of different mechanisms. This is the case of "emotion". (CHARAUDEAU, 2010, p.26).

For this reason, Charaudeau places his proposal in a psycho-socio-discursive dimension. Supported, on this account, in these different disciplines, particularly with regard to emotions, the author proposes some reflections toaddress what he calls the "pathetic effects of discourse", focusing, therefore, on two presuppositions, to which we briefly return below.

Emotions are intentional

Discourse always arises according to a want-say, a make-think, a make-make, and/or a make-feel. If, as we saw in the previous section, emotions can be included in discourse along with reason, we must believe, along with Charaudeau, that "it is due to the fact that emotions manifest themselves in a subject 'for the purpose' of something that the subject represents for themselves that emotions can be referred to as intentional" (CHARAUDEAU, 2010, p.28).

In accordance with the semiolinguistic "map" of the author, we cannot fail to mention tha t*pathos* will always be a game, abet, since, for it, to communicate is to venture into the staging (*mise en scène*) of an act of language. Thus, for semiolinguists, the intended effects (intentions) overlap with the effects that are actually produced.

Emotions are at the same time linked to knowledge based onbelief and included in a problematic of representations

In the definition of Charaudeau (2013, p.198), knowledge based on belief is that which has the purpose of proposing a "judgment on the world", according to "values that we attribute to it", through "an evaluation movement". Different from factual knowledge, which imposes itself on the subject, above all by virtue of scientific reason, knowledge based on belief is constructed from choices made by the speaking subject, "according to a logic of the necessary and the credible, in which either reason or emotion

can intervene" — a trait that naturally indicates the relativity of knowledge based on belief, which finds its counterpart in the indisputable character of factual knowledge.

Emotions and beliefs are linked to the extent to which it is not sufficient that the subject has the capacity to perceive the intentions of others or to capture some information or some knowledge. It is necessary that through socially shared values, the subject *evaluates* this knowledge, and, for this, he "mobilizes one or more of the inferential networks proposed by the universes of belief available in the situation in which he finds himself, which is susceptible to triggering in him an emotional state", a state that "puts him in contact with a social sanction that will culminate in various judgments of a psychological or moral order" (CHRARAUDEAU, 2010, p.30).

For this aspect of emotions, Charaudeau (2010, p.28) states: "Any individual can perceive a lion, recognize its morphology, know its habits [...], but as long as he does not assess the danger that it may pose [...], he will not experience any emotion of fear", and it is belief — individual or collective — that allows the subject such an assessment of danger.

The pathetical (affective) relationship of the subject with a fact, phenomenon, situation, or event leads him to a reaction, according to the social norms to which he submits himself, whether they have been previously internalized or arise from his representations. These representations

> [...] can be called 'socio-discursive', when the process of the symbolic configuration of the world is done by means of [...] statements that signify [...] [and] witness [...] how the world is perceived by subjects living in community, the values they attribute to the perceived phenomena, and what the subjects themselves are. These statements circulate in the social community, become the object of sharing, and contribute to a common knowledge and, in particular, a knowledge based on beliefs. (CHRARAUDEAU, 2010, p.31-32).

In other words, if social representations are in the (socio)cognitive plane, we can say that socio-discursive representations are characterized by the way in which collectively shared beliefs and knowledge are constructed and discoursed from credible utterances, based on the common experiences internalized by more complex social groups.

The media, for example, make use of these representations in journalistic articles and publicity pieces as inciting components to connect with the affectivity of their target audience. In a news piece, for example, the headline that announces the death of a baby, the accusation of a priest committing paedophilia, or the acquittal or condemnation of a corrupt politician activates representations capable of arousing the same feelings in the interlocutors — obviously to different degrees, given that we must consider the dependence of the status of each interlocutor and his pathetic dispositions at the time of the announcement.

In any case, we cannot deny that in the illustrated case, socio-discursive representations are always "staged" to reach the greatest possible number of interlocutors in a similar manner.

It should also be clarified that the link between emotions and representations is attested to by the terminological choices of Charaudeau, who — instead of "emotions" — opts for terms derived from the word *pathos* — "pathetic", "patheticalness", and "patheticization". According to the understanding of Mendes, S. and Mendes, P. (2007, p.274), patheticization can be thought of "as an effect of meaning, which, therefore, must be taken in a specific context of exchange, in accordance with the inference capacity of the interlocutor and the social representations that underlie this exchange". With regard to the determinants of the effects of patheticization, the representations would consist of the evaluations of a given proposal about the world, whose judgement of value is assigned in a particular situation of enunciation.

We present above only a few of Charaudeau's considerations with regard to emotions, which seem to be applicable to any discourse. We will not dwell on his considerations because his analysis focuses on a corpus of television media, while our analysis focuses on written journalistic media.

Pathetic effects in the journalistic chronicle

Because the space of this article does not permit a lengthy investigation and examination, our analysis will take into account two rhetorical-discursive strategies that clearly stand out in the chronicles under consideration: *metaphor* and *intergenericity*. We can see the rhetorical activity of these strategies, considering that both engender what we will call here a *representation logos*, with the *potential* to arouse certain emotions, given that *pathos* is always a probability, an expectation.

Some clarifications are needed before we proceed to the proposed analysis. In 2014, the chronicler Lya Luft — who writes biweekly for *Veja* magazine — published four chronicles in which the word *nau* (ship) appears in the title, always in reference to the Brazilian nation. The chronological order of the publications is indicated in the following table:

Date	Title
12/02/2014	*We cannot be a rudderless ship*
05/11/2014	*The ship for all*
19/11/2014	*The faulty ship*
03/12/2014	*Prayer of the drifting ship*

As we explain above, despite it not being possible to proceed with the analysis of the four chronicles, it is necessary to restrict ourselves to the first three titles, whose texts to which they refer are — by the dialogical nature of the language — returned to

in the last chronicle, "Prayer of the drifting ship", which is the focus of our analysis. In other words, to ignore that it is a series of chronicles — the most evident common element of which is the repetition of the word "ship" in the four titles — would be the same as to disregard the heterogeneous property of discourse.

In accordance with our proposal, the *representation logos* will show the pathetic effects of indignation, dissatisfaction, hope, and optimism. This demonstration is elaborated through different linguistically marked discursive strategies[5]. In view of the fact that *pathos* is always a consequence of aspeaker's choices, which, in the discourse, govern the processes of production, comprehension, and persuasion of the co-speaker, we will first consider two strategies that are striking in the chronicle under analysis: *metaphor*, focusing on its persuasive function; and *intergenericity*, based on the ideas of Marcuschi (2008). At the same time, we will discuss how these strategies engender certain representations, reflecting, in the enunciative scene, the effects of a pathetic order.

Thus, the first of the strategies that catches the eye is the use of the metaphor "ship". When mention is made of this term throughout the text, one should not only consider a direct reference to the Brazilian nation, but — depending on the chronicle's temporal production context — one should also bear in mind the circumstances experienced by Brazil and its people at the time. After all, it was a year marked by the outbreak of the *Lava Jato* operation of the Federal Police, which involved a series of investigations into a major money laundering scheme involving Petrobras and renowned contractors as well as Brazilian politicians. In addition to this scandal, which involved billions missing from the public coffers, Brazil was still recovering from a crushing defeat in the 2014 World Cup.

In the titles of the chronicles listed in the table above, the Brazilian nation — this great "ship" — is already qualified as "the **rudderless** ship", "the **faulty** ship", "a **drifting** ship", and throughout the text under analysis, it can be seen that a series of expressions, words, and statements reinforces the negative image suggested by the titles.

In the chronicle "Prayer of the drifting ship", this emphasis is achieved by means of statements such as "indecisive and often rough sea" (1st paragraph), "I float slowly, inclined, on a dead sea, on the verge of being shipwrecked" (5th paragraph), and "debris from a shipwreck" (8th paragraph). In these cases, the representation, which is easily assimilated by the readers of the chronicle — of an inert country in the midst of discoveries of scandals about diversions of millions —which lacks a cohesive, incorruptible and common sense administration — is capable of eliciting, more or less similarly in readers, the emotions of uncertainty, indignation, pessimism, and even a restricted optimism, given that improvement would depend on overcoming the mediocrity and the resignation of so many missing "passengers" from this "uncertain and threatened ship" (6th paragraph).

Thus, it is not difficult to consider some associations between the content of the chronicle and the reality experienced by Brazil, specifically in 2014, which is a moment characterized by a succession of discoveries that justified the economic imbalance of the country, in which, moreover, the rumours about a probable recession were already being echoed, even in the international media. This analogy between the "ship" and concrete reality is in line with an affirmation by Pinto (2000), who, quoting an anonymous author, returns to definitions of the rhetorical and aesthetic functions of metaphor:

> Generally speaking, in the realm of the former, the dominant aims are those of the mode and persuasion, while the latter — mimetically evoking possible universes — arouses in the listeners effects more or less profound that unravel at the margin of any criterion of true and false. (PINTO, 2000, p.277).

Typical of the journalistic chronicle — a genre characterized by the integration of reality and current events with a relaxed prose, and by the freedom given to the chronicler to express her or his point of view through various linguistic-discursive mechanisms — the use of figures such as metaphors enables both rational (in the sense of being reasonable and credible) and aesthetic effects, and contributes — perhaps more than dry, ornament-free language — to the production of pathetic effects. In the case of the chronicle under analysis, comparing Brazil with a "ship" or a "large ship" can produce in the reader's imagination a closer representation of empirical reality because they are objects that are easy to cognitively assimilate, as are the specifiers used in the titles — "drifting ship", "rudderless ship", and "faulty ship" — as well as the correlated terms identified throughout the text — expressions such as "storm clouds" and "fatal rocks" (2nd paragraph); "rotting waters" (3rd paragraph); and "fatal whirlpools" (8th paragraph).

As Charaudeau reminds us, when proposing that emotions are related to knowledge based on belief, evaluating situations experienced by the subject (in this case, a chronicler in complicity with the reader) is essential for certain emotions and representations to be activated. In the passage transcribed below, we can see certain elements related to fear: *"[...] in the fog, and with so many storm clouds threatening, perhaps fatal rocks under the water line, I am afraid of keeling over"* (LUFT, 2014, p.24. **Prayer of the drifting ship**).

The dangers that the country is experiencing and the obscure environment in which it finds itself due to the poor management and corruptibility of some of its rulers, engender *worry* and *fear* — again through easily identifiable images such as the "fog", the "storm", and the "rocks". The belief that the country is in danger, and the representation of a nation that suffers (or that should be concerned) with the uncertainties about the future, is reactivated in this chronicle.

The title — Prayer of the drifting ship — registers exactly the way that the text presents itself: a prayer. In other words, it is a journalistic chronicle in the form of a

prayer, from beginning to end. When addressing such cases of genre hybridization, Marcuschi (2008) states that in some situations — to meet specific communicational goals — "it is common for us to circumvent the canon of a genre by mixing forms and functions" (MARCUSCHI, 2008, p.164). Such a procedure characterizes what the author calls *intergenericity*.

It should be clarified that in the specific case of the chronicle "Prayer of the drifting ship", the intergenericity observed in the mixture of this genre with a prayer is perfectly legitimate because the chronicler's *strategy space* allows this transgression, which, therefore, does not compromise the chronicler's *restriction space*.

Returning to the chronicle in question, we transcribe below excerpts that show a strong dominance of the appeal and cry to God:

> "Lord, give me a competent and highly skilled crew that will alleviate me of these difficulties and afflictions [..])"

> "Lord, give me an experienced, reliable and honest helmsman with a firm hand and clear, coherent, and decisive ideas, who knows what he does and wants to do what it takes to correct course so that this voyage ends well [...]"

> "Lord, give me responsible people who choose their crew according to their merits and preparation — good, courageous, and tireless people who can cleanse the rotten waters in which I am now immersed."

> "Lord, give me clear waters to navigate [...]"

> "Lord, give me people who believe that it is worth changing, that inconveniences, annoyances, even fears that any transformation imposes are essential and beneficial in this hour [...]"

> "Lord, give me the gift of renewal so that I may be a respected and beautiful ship, and not debris of a shipwreck on some forgotten island."

The passage reveals itself to be a personal petition, in which the speaker, by assuming the form of the ship itself in the first person singular, summons divine direction and intervention. In this case, the intergenericity transcends the limits of the form and function of the genres and causes the text to be produced in an interdiscourse whose exchange of distinct spheres — the mediatic and the religious — evokes emotions such as the *sensation of fragility* of mortal man and the *hope* that the divine being will come in response to the cry for help.

The repetition of the "Lord, give me" formula throughout the text may indicate both *hope* — if we consider that God, being omnipotent, will have to attend to the insistent request — and its opposite, *anguish*. This anguish/hope theme (CHARAUDEAU, 2010) seems to dominate the content of the chronicle. The representations oscillate between

positive and negative, depending on whether the focus is on the dangers and doubts or on the trust that, through divine intervention, the country may find itself freedoftorment.

Intergenericity, as well as the use of metaphor, converges in the *representation logos*, which, in turn, puts into play plausible expectations of pathetic effects. The very fact of subscribing to an interdiscursive space — thus allowing the mixture of different spheres and genres — suggests a specific representation: that of the Christian reader because prayer to God can suggest a view of capturing as many readers as possible in a country with a Christian majority, as Brazil is, although (we speculate) a part of the readership is not adept or sympathetic to Christianity.

Although the analysis focuses on the development of *logos* in *pathos*, which is still the purpose of the interrelationship of the rhetorical evidence, we believe it is possible to speculate on the question of *ethos in pathos* proposed by Galinari (2007, 2014). In this analysis, Galinari describes a case in which a columnist — when projecting a certain image of himself — ends up "contributing to the establishment of feelings" (GALINARI, 2014, p.281).

With this in mind, we could reflect on an *ethos* of a sacred nature, capable of establishing hope and a sense of fragility. In the doctoral research of Alves Jr. (2015), in which a corpus of 54 chronicles of Lya Luft was examined, indications of the construction of a Christian *ethos* — were identified, expressed through different linguistic configurations (e.g., nominal expressions, portions of text of various extents, isolated vocabulary).

Thus, it would not be novel for the chronicler to make use of, deliberately or not, an image of herself that is committed to Christianity. However, the overlapping of the persuasive evidence found at this point in our analysis shows, in Galinari's words (2014, p.261), the "ability of *ethos* to act by empathy or identification, ethically and emotionally including the interlocutor". Consequently, in the case of the chronicle "Prayer of the drifting ship", this Christian ethos — projected by the chronic/prayer intergenericity — would be the inciting component of several of the effects of patheticization possibly sought by the speaker, with an emphasis on hope — suggested, above all, by the pleading tone of the prayer.

In conclusion, we hope that this brief analysis fulfils the purpose of illustrating the articulation among rhetorical proofs, especially with regard to the conversion of *logos* into *pathos* but also with evidence of the development of the *representation logos* into a Christian ethos — the latter with the potential to engender certain pathetic effects.

Final considerations

What we presented in this article was an approach — that had already been proposed by the new rhetoric — to review the relationships between rhetorical proofs, particularly with regard to the interrelationship between *pathos* and *logos*. It was a study of the emotions in discourse, which we conceive as a result not necessarily of the desire of an orator/speaker to affectively connect with the target audience, but of a

possibility glimpsed by the discursive act itself, based on acts of enunciation produced in effective situations of exchange.

As with the new rhetoric, sophistics provides important contributions; however, they are not as widespread. This is likely due to the lack of written records attributed to the sophists. Hence, there is a need to seek, in the work of other researchers, some contributions left by neo-rhetoricians of antiquity.

Charaudeau's considerations are equally fundamental for a broader and more integrated study of *pathos* and patheticization. His assumption, according to which emotions are linked to social representations — socio-discursive in cases in which the enunciation acts that are actually produced are analysed — is, in our view, an innovative and essential perspective for understanding the pathetic phenomenon.

The theoretical itinerary covered, followed by the examination of the chronicle, attempted, among other aims, to stimulate greater debate on rhetorical proofs and their very close relationship, which could also serve as an opening to a less fragmented analysis model, considering the different dimensions that are part of discourse.

▪ *RESUMO: Tomando como ponto de partida as reflexões sobre a dimensão patêmica do discurso, este artigo focaliza a discussão sobre as três provas retóricas de persuasão – logos, pathos e ethos –, buscando argumentar em favor de sua integração, a partir de contribuições advindas tanto da Nova Retórica e da Sofística quanto da Teoria Semiolinguística do Discurso. Recusando a perspectiva dicotômica do paradigma cartesiano, nosso objetivo principal é realçar a necessidade de concebermos uma relação de interdependência entre o pathos e o logos. Para ilustrar nossa proposta, procedemos, após exposição teórica sobre o tema, ao exame de excertos de uma crônica jornalística, na expectativa de reforçar a legitimidade do assunto em apreço e de confirmar a necessidade de debates e demonstrações mais consistentes. Elegidas as categorias da intergenericidade e da metáfora como estruturas discursivas representativas do logos, a análise permite observar seu desdobramento em efeitos de patemização, concebidos aqui como efeitos de sentido.*

▪ *PALAVRAS-CHAVE: Pathos. Logos. Nova Retórica. Sofística. Semiolinguística.*

REFERENCES

ALVES JUNIOR, M. A. **O intertexto bíblico como expressão de um *ethos* em crônicas de Lya Luft**. 2015. 183 f. Tese (Doutorado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

ARISTÓTELES. **Retórica**. Tradução de Manuel A. Junior, Paulo F. Alberto e Abel N. Pena. São Paulo: WMF Martins Fontes, 2012.

CHARAUDEAU, P. *Pathos* e discurso político. In: MACHADO, I. L.; MENEZES, W.; MENDES, E. (Orgs.). **As emoções no discurso**. v.1. Rio de Janeiro: Lucerna, 2007. p.240-251.

_____. A patemização na televisão como estratégia de autenticidade. In: MENDES, E.; MACHADO, I. L. (Orgs.). **As emoções no discurso**. v.2. Campinas: Mercado de Letras, 2010. p.23-56.

_____. **Discurso político**. Tradução de Fabiana Komesu e Dilson F. da Cruz. São Paulo: Contexto, 2013.

CHARAUDEAU, P.; MAINGUENEAU, D. **Dicionário de análise do discurso**. Coordenação da tradução: Fabiana Komesu. São Paulo: Contexto, 2014.

FERREIRA, L. A. **Leitura e persuasão**: princípios de análise retórica. São Paulo: Contexto, 2015.

GALINARI, M. M. As emoções no processo argumentativo. In: MACHADO, I. L.; MENEZES, W.; MENDES, E. (Orgs.). **As emoções no discurso**. v.1. Rio de Janeiro: Lucerna, 2007. p.221-239.

_____. A polissemia do *logos* e a argumentação: contribuições sofísticas para a análise do discurso. **EID&A** – Revista Eletrônica de Estudos Integrados em Discurso e Argumentação, Ilhéus, n.1, p.93-103, nov. 2011.

_____. *Logos*, *ethos* e *pathos*: "três lados" da mesma moeda. **Revista Alfa**, São Paulo, v.58(2), p.257-285, 2014.

LIMA, H. Patemização: emoções e linguagem. In: MACHADO, I. L.; MENEZES, W.; MENDES, E. (Orgs.). **As emoções no discurso**. v.1. Rio de Janeiro: Lucerna, 2007. p.140-149.

LUFT, L. Oração da nau à deriva. **Veja**. São Paulo, 03 dez. 2014, p.24.

MARCUSCHI, L. A. **Produção textual, análise de gêneros e compreensão**. São Paulo: Parábola Editorial, 2008.

MENDES, S. S.; MENDES, P. H. A. Uma análise discursiva das emoções em *Laranja Mecânica*: o estranhamento, a humilhação e a indignação. In: MACHADO, I. L.; MENEZES, W.; MENDES, E. (Orgs.). **As emoções no discurso**. v.1. Rio de Janeiro: Lucerna, 2007. p.272-289.

PERELMAN, C.; OLBRECHTS-TYTECA, L. **Tratado de argumentação**: a nova retórica. Tradução de Maria Ermantina de Almeida P. Galvão. São Paulo: Martins Fontes, 1996.

PINTO, M. J. V. **A doutrina do *logos* na sofística**. Lisboa: edições Colibri, 2000.

FOREIGN STUDENTS AND THEIR LEARNING SPANISH PROCESS: FIRST HINT OF TRANSLANGUAGING

Jaddy Brigitte NIELSEN NIÑO[*]

■ ABSTRACT: This article is about the study of translanguaging in the sociolinguistic performance in a group of 22 foreign students who acquire Spanish as a foreign language in a Colombian University. This article explains translanguaging as a strategic resource that refers to the combination of changes so happens, from the translation, changing code-switching and language-switching, for that reason it is defined in the process of learning a language other than the mother tongue, in this case Spanish. Data were collected through direct observation recorded in a field journal. Semi-structured interviews were conducted from the sample of participants and artifacts were used by the students in their learning process of the Spanish lessons. Data were analyzed inductively by certain categories of analysis from the concepts, issues and relationships, in order to give optimal answer to research questions. Concluding that translanguaging is a phenomenon that greatly influences the development of the foreign language.

■ KEYWORDS: Translanguaging. Spanish in Colombia. Strategic resource. Students. Teachers of languages.

Introduction

Learning and teaching a foreign language and the complex multicultural processes that take place in the contexts in which they occur need to be adequately understood through careful study. One of the most deserving phenomena of study in these contexts is the use of the mother tongue in the processes of learning a foreign language by teachers and trainees. It is impossible to think that adult people who already effectively manage their own language do not use it to learn a new language, but we know that there are diametrically opposed ideas about the usefulness and the goodness of doing so. There are foreign language teachers who think that the mother tongue is only a source of interference and an obstacle to the foreigner's learning and banish their use within the classroom; And there are those who believe that, as acquired linguistic knowledge,

* Corporación Universitaria Minuto de Dios (UNIMINUTO). FUNIFELT International Foundation, Academic Department Zipaquirá – Colombia. brigittebrigitte9@gmail.com

the mother tongue can greatly facilitate the learning of the foreigner and must be an essential element in the process.

More in the background of this discussion are the problems related to the cultural content of language learning. Those who defend the unique use of the foreign language in the contexts of their learning usually overlook the fact that the use of the foreign language brings with it the content of the culture of those who speak it. This cultural content can prevail if it is handled in a unique way and restricts the use of the mother tongue, the cultural comparison consequences, affect in the learner its own cultural knowledge and consequently also its identity. The reflection on how to make the learning of a foreign language result in truly additive processes (LAMBERT, 1974), both linguistic and cultural, is at the centre of the pedagogical decisions that are faced for a foreign language teacher day by day.

This study focuses in particular on the strategic processes used by learners and teachers of Spanish as a foreign language in a context in which this new language is necessary to live in community. For this purpose, the researchers investigate what happens with the native languages of the apprentices in three different types of Spanish classes that are taught for foreign students at Universidad Nacional de Colombia, Bogota. It is observed the use that these students make of their mother tongues and the attitudes and pedagogical actions of the teachers in relation to this use; it is also inquired about the effectiveness of practices involving the use of the mother tongue in the classroom. To carry out this study, the concept of translanguaging is mainly used, a concept which has been developed during the last years in various educational contexts where the reference point is English and the population under study has been mostly children between the ages of 3 and 17 years. It is a process that has not yet been investigated in Latin America in depth and in which the teaching-learning of Spanish as a foreign language has not been taken as a research context.

The article is organized as follows. In section two (2), describes the research methodology that was carried out; and, in section three (3) some of the partial results obtained. Finally, section four (4) presents the conclusions derived from this research study.

Extraction of terms from the background

The approach of Translanguaging as a concept in literature is still very scarce, and therefore its definition remains vague. In recent years an attempt has been made to conduct research on this subject in the field of education, and especially studies in basic and secondary education in the United States and Europe. The use of the term part of Wales, in 1980, when it was decided to teach lessons in the classroom using two languages, the Welsh mother tongue and the official English language, in order to achieve learning more effective from English. This pedagogical use that related the two languages was denominated ten years after *trawsieithu* in Welsh and *translingyifying* in

English. Over time the experts saw the need to write it in a more understandableway, for what was called translanguaging. Although there is a translation established in a document in Argentina called translanguage "characterized by the flexible use of more than one language to adjust the interaction to the recipients" (Subsecretaría de Educación, 2016) the term has not been formalized by governmental educational entities in Latin America, omitting for its part the sociocultural aspect of the individuals, therefore in this investigation it is called translanguaging.

In this study, translanguaging is defined as a strategic resource that is strengthened within multilingual classrooms, allowing the development of the student's linguistic abilities in both languages (language of instruction-language of origin) based on a significant pedagogical approach that enhances the multiple practices pedagogical and discursive in which individuals (students and teachers) participate in order to generate lasting learning and a natural contextualization of the multicultural environment in which they find themselves (LEWIS, 2012; LI, 2011; LI; HUA, 2013; WILLIAMS, 1994, 1996).

In this way the translanguaging is not simply a mixture of linguistic forms of diverse languages or dialects, since it really implies a variety of articulations and negotiations of the language and developed in social spaces, besides the linguistic identities are not static or monolithic, on the contrary they are dynamic and complex (NIELSEN, 2017).

For this reason, Williams (2002) coined the term translanguaging as the transition from one language to another within the classroom, in order to reinforce and increase the understanding of two languages in use by the student. Then, Williams (2012) determines that the process of translanguaging should focus on the students and not the teachers, but without excluding that it is a process that must be controlled by the latter. From this approach, six years later the term translanguaging had generated impact among educators and experts in both North America and Europe especially in bilingual education classrooms. Taking as reference a particular case, García (2009a) considers that translanguaging should be used by emerging bilinguals to make decisions about the language they learn and organize their previous knowledge; In other words, translanguaging can become a mechanism for the construction of knowledge with others and mediation between linguistic groups. In subsequent works García (2009b) focuses on the study of different immigrant communities in New York City where he believes that if there is no use of translanguaging it is impossible to generate assertive communication.

However, García (2011) states that he does not agree that translanguaging is designated as a simple pedagogical practice. The author defines the term translanguaging as a pedagogical strategy that should be constructed starting from four specific objectives: the first of them is that the students "translanguajeen" in academic contexts, without any limit; the second is to make use of materials and resources that help the student in the learning process in order to identify the difference between the languages, to develop solid bilingual processes; thirdly, it is about students constructing consolidated learning processes in any subject, so that all teachers can practice translanguaging regardless

of the specialty; and finally translanguaging should be the opportunity to practice the languages learned both at home and at school.

At the same time, García (2013) is totally in agreement with people who use a certain language as a negotiation of knowledge and situations, but at the same time they have to think about the language or the language of instruction. García (2013) defines translanguaging as the decisions of bilingual practices with meaning from the perspective of the speaker, but which in turn involves all students and teachers in the use of those bilingual or multilingual discursive practices in a natural and "meaningful" way. "That manages to generate learning and teaching in the different classrooms.

In this way, translanguaging in education can be defined as a process by which teachers and students engage in discursive practices that include linguistic processes to substantially maintain communication and appropriate knowledge of the concepts developed.

Thus, it can be said that the teaching of a language should depend less on the grammatical structures and yet provide the student the opportunity to have lasting learning without the need to demonstrate the learning of a foreign language mechanically or isolated from the reality in which students live. This means that teachers must take on the role of facilitator and that in turn they should encourage their students to use the language of instruction more frequently. According to García (2014), teachers must be able to identify the use of different linguistic practices that impact on students in the classroom, this also allows understanding when, how, where and why students are making use of said practices and for what purpose.

The translanguaging has also been studied in some European nations, including the United Kingdom, where a study was conducted with Asian students who had to make constant use of English. Li (2011) who was a university professor and who belonged to the Department of Applied Linguistics of the University of London, conducted a case study with three young Asians living in London, United Kingdom, also as an initial characteristic the ages of young people ranged from the 19 and 20 year olds, all of them were in their first year as undergraduate students in a mathematics program at one of the universities in London. To carry out the research, the information was collected through semi-structured interviews, observations and recordings of all the social interactions these students had with their families, classmates and off-campus.

In the analysis of this study, the university professor was able to identify that Asian youth mixed the use of the two languages (English-Chinese) when interacting with people of the same origin or of other nationalities, even several of them affirmed that they chose to study Mathematics because, although they did very well to speak English, they found it difficult to write it. On the other hand, they were creative with the use of language, an example that was given in this research, was the use of Chinese interjections in English, that is, young Asians said all the time Cake seller! That upon hearing it an English-speaking person would not understand the true use until they explained to him that the translation into the Chinese language was 麦 高 德 (Màigāo

dé) which to them sounded similar to English when it says My God! But they preferred to use the literal translation from Chinese to English. However, Asians as a rule when they are in a country where the linguistic code is completely different change their original names according to the context so that others can pronounce it, for example, in this investigation it was evidenced that one of them was called 張龍 (Zhānglóng) and in the English translation is Long John, but his relatives and friends from the foreign country knew him as John.

This study determines that being multilingual generates networks and spaces that help to build social ties with different people and even mixing languages generates a multilingual identity, if it is done naturally because multilingualism does not mean being able to speak different languages separately, but easily unfold among them. In conclusion, the term translanguaging is determined as the use between structures and linguistic systems that go beyond the mere definition, because it was useful to create a social space for multilingual speakers in different dimensions such as personal, historical and attitudinal, as well as includes a range of linguistic performances of multilingual speakers with purposes that transcend the combination of structures, the alternation between systems, the transmission of information, the representation of values, identities and social relations.

In addition, taking the previous study as a reference; two years later Li and Hua (2013), studied translanguaging in the multilingual practice used by transnational individuals; That is, foreign subjects who move to a different nation of their own. In particular, students with different Asian ethnicities, as a result this research focuses on a group of university students, including the three students of the previous study and two more that have been part of a multilingual and transnational academic network with different members of the London community. In addition, the collection of information from this study is carried out through student narratives and ethnographic observations, with the aim of exploring aspects such as the process of socio-cultural identification of students, the interactions between their linguistic ideologies and policies, multilingual practices and what they have learned since they have been part of this type of social and academic networks.

This study recognizes the ability of transnational individuals to mobilize their linguistic resources and the re-construction of different relationships and meanings within a specific social context, in this case London. In turn, it enhances the creative qualities of the mixture of languages, hybridization and miscegenation that can occur in this context (LI; HUA, 2013). For this reason, the term translanguaging is taken as reference, which is defined as the best form, both of the natural dynamics of the use of the language and multilingual practices of various kinds, as well as the ability of the speaker to redefine the information for mobilize their linguistic resources and thus create new social spaces.

Over time, the increased use of the term translanguaging has drawn attention to language researchers, educators and linguists, and some experts have given several pedagogical and methodological interpretations. In fact, several authors advocate the

importance of the translanguaging concept with a view to developing bilingualism not only in specific communities, but also regionally and globally.

In the first instance, Baker et al. (s.f.) rely on Williams (1996) to discuss how to use the term within the educational setting. It defines translanguaging as a pedagogical theory and as a cognitive process that implies the exchange in the use of two languages to achieve important educational results. As a pedagogical theory, translanguaging allows the use of diverse cognitive processing skills in the development of communication skills. Baker et al. (s.f.) mention, for example, cognitive assimilation and accommodation in comprehension processes, and the selection and storage of information in memory, essential for communication. Based on the above, translanguaging is a much more complex process than simply translating from one language to another, which is limited to the parallel recognition of words.

Williams (1994, 1996) had recognized the phenomenon of translanguaging as a pedagogical practice by indicating that the use of two languages generates a bilingual classroom in which students receive information through the first language and then through the second. In fact, the National Assembly of Wales, according to Williams (2012), had the intention of generating bilingual students from early childhood, that is, from kindergartens to primary school, considering that the implementation of translanguaging as a pedagogical practice was really a method appropriate for children to acquire or maintain proficiency in their two languages. In this way true bilingualism was developed, more than the mere teaching of a foreign language.

Baker (2011) establishes four educative advantages of this process conceptualized by Williams: first, he considers the specific purpose of translanguaging as promoting deeper and more complex understandings of a certain topic from the establishment of the Zone of Near Development (ZDP) by Vygotsky (1978); indicates that this purpose is met if the first language is used as a mediator for new learning and as a basis for prior knowledge in a relationship of interdependence between languages that establishes a true support scaffolding for the learning of the new language; It also indicates that the higher the level of proficiency in the second language, the more linguistic transfer will be possible.

Secondly, Baker (2011) states that translanguajear is a pedagogical practice that can help the advancement of the less developed language especially in oral communication and reading and writing in academic context, thus giving way to a true bilingualism. As a third instance, Baker (2011) speaks of translanguaging as a facilitator of links between home and school through cooperation between both contexts, since the educational community can provide materials and support resources in the first or second language; it points to this as an aid to all members of the educational community, including parents, teachers, administrators and students, who can share the knowledge of the curriculum when the latter are school-level students. As a last advantage Baker (2011) indicates that translanguaging can help the integration of speakers who have a higher level of appropriation of the language with those who do not, in order to strengthen the languages through the creation of working groups in the classroom.

Creese and Blackledge (2010) also endorse the idea that translanguaging is a flexible bilingual pedagogy, especially for learning and teaching based on contextual perspectives, that is, one that establishes links between the social, cultural, community and linguistic domains that characterize to each of the students. In this case the authors add that the flexible bilingual pedagogy is of great importance, because it allows responding to the local needs of the context, strengthening the language of origin together with the language of instruction, both outside and inside the classroom.

Stathopoulou (2013a) considers Interlingua mediation as a practice of translanguaging. In his doctoral thesis, he distinguishes mediation based on the standards established in the Common European Framework of Reference for Languages (CEFR), defined as "professional translation and interpretation", of a mediation that must correspond to active participation in intercultural communicative events. The author, referring to the term translanguaging since the affirmation of Guerra (2004), relates it to literacy practices in different languages and dialects learned in different cultural contexts. Translanguaging, then, not only involves managing the linguistic repertoire of two languages (linguistic competence), but also being competent in the communicative management of languages (communicative competence). Stathopoulou (2013b) also indicates that translanguaging can be used with students of different levels of linguistic competence.

This is where two different contexts that serve as background are included, since they make use of translanguaging in the educational system and specific bilingual programs. Mr. Williams (2002) is retaken together with a work team that was part of the National Assembly of Wales at that time, generated a document that focused on the development of bilingualism through a skill called translanguaging. It is argued that this type of skill is used in everyday life and in turn can be part of the entire education system in which students are exposed to effectively change the language code from one language to another.

The skill called translanguaging according to the author offers wide possibilities in the development of two languages effectively, that is why for a person who is an active bilingual, translanguaging is defined in this aspect as a natural form of development and strengthening of two languages Simultaneously, at the same time it generates a deeper understanding of languages and the environment, but Williams (2002) in turn asserts that few teachers are willing to try to use two languages in a single class session through said ability.

In this point of view, teaching-learning languages should not be seen as the sole responsibility of the members that make up the educational entities (language teachers, managers, coordinators, students, among others) each teacher must be a language teacher, according to the statement given by the educational institutions of Wales, languages have a central role in the construction of knowledge, because, when learning to use any language whatsoever, it must focus on some stages, which Lemke (2001) calls how to re-name, describe, record, compare, explain, analyze, design, evaluate and theorize. This can be developed, if the educators are aware that

each teacher, without discriminating the academic level, becomes a language teacher and has to be able to control the teaching process through situations that enrich and develop the languages.

Accordingly, a research group from the University of the City of New York known in North America as The City University of New York-New York State Initiative on Emergent Bilinguals (CELIC; SELTZER, 2011) and that will be named from now on their acronym CUNY-NYSIEB, they have generated a collaborative project with the Research Institute for the Study of Languages in urban society in 2011 that offers a guide for educators who focus especially on New York schools, located in the United States of America. It is important to clarify that the members of this project work focused on the translanguaging through a teaching guide where it is explained how translanguaging can help with the specific aspects in the investigation especially in the learning-teaching of the language.

With all that said, especially by the members of CUNY-NYSIEB they emphasize the Council of School Officials of the North American State (CCSSO), which is an entity that generates the basic academic standards for teaching in elementary and secondary schools, especially. The entity seeks to generate high quality in English categorized as the art of language and ELA literacy through learning objectives. These objectives describe what a student should know and be able to do at the end of each grade of primary and secondary school, through a regulation that ensures that all students achieve high school graduation with the knowledge and skills necessary to succeed both in the university and in the daily life, regardless of where they live or their origin. This regulation is governed for forty-three states of the District of Columbia consisting of four territories and the Department of Defense to the Educational Activity determined in its acronym in English as DoDEA, it is worth clarifying that this regulation has been voluntarily adopted by educational institutions that have Progress was made especially in the Basic Education Core as reported by the Common Core State Standards Initiative (2013).

Therefore, Celic and Seltzer (2011) who are members of the CUNY-NYSIEB project, within the teaching guide have not only included the aforementioned regulations but also include some sections where they explain the process step by step involving the strategies of translanguaging in practice, through examples of how teachers in their teaching practices use some strategies in their lessons and provide an academic context where these strategies can be applied. Undoubtedly, Celic (2012) considers that the translanguaging strategies are necessary because they allow generating a stepped learning that helps the bilingual students, but at the same time the author affirms that it is a way of promoting the bilingualism of the students when they use the language in its entirety as a learning resource. In addition, it can be seen through the development of thinking, social and attitudinal skills.

In fact, Witt (2012), who is one of the members of the CUNY-NYSIEB project, complements the idea through the evaluation rubrics, so that teachers can reflect on the frequency with which the translanguaging strategies are used and identify what their

Strengths and weaknesses through gradual incorporation into schools and sessions taught within classrooms. This rubric has eleven (11) characteristics that teachers must take into account in their teaching practice, but in this project only those that are consistent with the unit of study will be selected. Taking as reference that they are purely qualitative characteristics where it allows teachers and specialists on educational policies to examine teaching practices, curricula, class lessons, among others. In short, the proposed guide goes beyond conceptualization and the educational needs of the context.

Now, it cannot be determined that the act of translanguaging develops only through the change of linguistic code, defined in English as code-switching and explained by Nilep (2006) as the practice of selection or alteration of linguistic elements in order to contextualize the talk within the interaction process. The same happens with another linguistic concept focused on language such as translation, defined by Crystal (1992) as the process or result of converting the expressions of one language to the expressions of another, in order that the meanings correspond to each other. While the change of linguistic code and translation are essentially linguistic concepts, the translanguaging goes further. Kano (2012) through her doctoral thesis conducts a study on translanguaging as a pedagogical process and tool for students who are learning to write in English within the context of New York, the author states that the study proposes a definition of translanguaging as a complex process that includes changes in the linguistic modes and also the change of the codes, in other words, the translanguaging implies the code change (code-switching), the translation, the language change (language-switching) and the combination of the three. This theory is developed because the results of the study showed that the translanguaging act that the students practiced included the previous concepts.

Materials and methods

This study is framed within the qualitative research approach, since it is based on a logic and an inductive process, that is to say that they go from the particular to the general, starting from an exploration and detailed description of the unit of study (the class of Spanish For foreigners) to generate a theoretical perspective on the subject of this research. The purpose of this is, as indicated by Hernández Sampieri, Fernández and Baptista (2010), in "rebuilding" a reality as observed by the participants in some classes framed in the basic level of Spanish for foreigners at Universidad Nacional de Colombia in Bogotá, educational context that has been previously defined.

This study is part of Boyle's (in ÁLVAREZ-GAYOU; SAMPIERI; MENDOZA, 2008) approach to what educational ethnography is. The purpose of educational ethnography is to discover what happens daily in the context chosen for research, based on providing meaningful data in the most descriptive way possible, then interpret them and be able to better understand the phenomenon observed in that ecological

niche, in this case the classrooms. Such data deal with the contexts in which the various interactions, activities, values, ideologies and expectations of all participants (teachers, students and even the researcher) in those determined school scenarios take place (TORRES, 1988).

This research method will allow the description of the elements that are framed within the social processes, which will be analysed in a functional way, thus giving an explanation of how certain parts of the culture and the educational system interrelate within a certain time And in specific situations, to finally be able to describe the learning and teaching of the Spanish that interests.

To achieve these objectives, it is necessary to frame the observation in four dimensions according to Goetz and Lecompte (1988): the first one is the inductive one, in which part of the data collection, through the observations of some classes with a Time period of seven (7) months and are taught by Colombian teachers within the campus of Universidad Nacional de Colombia, and theory is built on translational. Next is the generative dimension, centred on the discovery of constructs and propositions from various sources of evidence, which in this case will be the observational notes, records of the teaching activity, document files and artefacts discovered by the researcher in the classes during the period of observation. The constructive dimension is the third, where a classification and ordering of the data is developed, so that the constructs and categories of analysis can be generated. Finally, a subjective dimension will be taken into account in order to describe the cultural and behavioural patterns of the group investigated, with the final purpose of reconstructing the specific categories that the participants use for the conceptualization of translanguaging in Spanish.

In turn, this study takes as a reference the foreign students who attend at Universidad Nacional de Colombia, located Bogota, and take Spanish courses in person there. This university is a public higher education institution, in which foreign students have access to Spanish classes through the Extension program of the Department of Foreign Languages known as the DLE, the Department of Foreign Relations identified as DRE and also Through the Office of International Relations – ORI, all of them belonging to Bogota.

It is worth noting that Universidad Nacional de Colombia, Bogotá, has an approximate population of between 120 and 180 foreign students who are linked to these three different programs offered by the University, and their Spanish language ranges from basic, intermediate and advanced.

With all of the above, the author have identified that of the 180 foreign students, 80 of them were registered in the courses offered in total, but for the present study, data were used of three (3) teachers and twenty-two (22) foreign students attending Regular form to the courses offered since these courses were not obligatory nor did not generate a quantitative note in their academic reports, in addition all of them responded actively to a direct invitation based on the previous explanation of the

purposes of the study. This unit of study has been selected because they meet the established criteria such as:

- Active participation and constant stay in at least one of the courses selected for this research.
- Acceptance of the signed consent on the part of the participants.
- Foreign students whose mother tongue is different from Spanish.
- Teachers directly linked to the National University of Colombia.
- Teachers who have a work stay within the university for more than three (3) years.

The data were collected by means of direct observation that was recorded in a field diary, it was contemplated what happened within the context studied and the facts were recorded on the same one whose characteristic is the reduction of the interactions between the Researcher and unit of study. This allowed the study to focus attention, not instructively, but rather the possibility of generating analysis, prosody, kinetics and interaction. Goetz and Lecompte (1988, p.154) argue that the kinetic aspects are really useful "when the focus of research is on the use of classroom space or the patterns of teacher-student interaction".

Semi-structured interviews were also carried out with the participants, these interviews were applied in the two groups of participants: the first one comprised five teachers in order to determine the actions or procedures that the teachers carried out to enhance the learning of the Spanish and the following group is for foreign students with the aim of describing the procedure they use to learn Spanish within a Colombian educational context.

Finally, documents (guides, workshops, etc.) used by the students in their process of learning Spanish were collected. These allow us to know sources of information that are very valuable within the research and to identify the antecedents of the context where the study is based on.

Results and discussion

In this study, each of the Departments from Universidad Nacional de Colombia, such as the DLE, DRE and ORI, will be described from this moment on, making use of the previously proposed acronyms (see Table 1). This in turn allows the number of sessions and hours of the observation record to be determined during the 7 months. It should be noted that when observing each of the classes have been recorded in the field diary and voice recordings of each of the sessions in order to complement information that has not been transcribed during the class observation.

Table 1 – Class Observation Record.

Observation Class Record		
Department	**Number of sections**	**Number of hours**
DRE	6	18
DEL	30	57
ORI	14	26
Total	50	101

Source: Classified by Departments from Universidad Nacional de Colombia.

In the same way, interviews were conducted randomly with foreign students, who in some cases were asked in English or Spanish so that they would feel comfortable and calm in answering each of the previously established questions, since the interview has four types of questions: eight (8) general, two (2) examples, two (2) structural and three (3) of contrast. For this reason the interview script was validated, which allowed some modifications in the contrast questions, since all interviewees had never previously had contact with Spanish. In the first eight (8) general questions it was identified that the participating students were enrolled in different programs of the University such as: Mechanical Engineering, Industrial Engineering, Architecture and Human Sciences, in addition several of them were of American origin United States), Asian (China) and European (Belgium, Germany, Netherlands and France) in turn commented that they had a stipulated period of time to remain in Colombia from six (6) months to one (1) year.

On the other hand, the students were questioned as to why they decided to study in Colombia, and among them they found different answers such as: some of them have a sentimental relationship with a Colombian / a, their parents have settled in the country, Universidad Nacional de Colombia has an international exchange program to conduct research stays between the National University and European Universities. Here it is evident that the human being establishes basic social contacts innately. As Harris (2007, p.19) puts it, social contact is framed in "the socially learned way of life found in human societies and encompassing all aspects of social life, including thought and behavior".

In a second moment, I was able to show a total of 437 occasions in the classroom that correspond to the use and promotion of the student's mother tongue, use of the language of instruction that corresponds to Spanish and finally use of English as a third language, everything I will describe it in a discriminatory way in the following sections.

In the use and promotion of mother tongue I found that there are 222 occasions of use of the mother tongue by the student and promotion of the same by the teacher, identifying that on six (6) occasions in the DRE students used German and The Russian turn the teacher towards a comparison of the pronunciation of the letters in German and French. On the other hand the DLE with a total of 215 occasions made use of both teachers and students of English, German, French, Russian, Italian, Flemish, Mandarin and Portuguese.

The previous data confirm the answers given by the students in the interview, since they assured that several of the teachers are very recursive in the moment of teaching, especially the foreign students confirmed that one of the teachers speaks many languages being one of the Main resources in the learning-teaching of Spanish because it allows the teacher to explain easily by emphasizing the languages of origin of foreign students. In addition, the guides that were delivered during the development of the classes serve as a complement to a specific topic and the annotations on the board in several languages in order to explain the complex words of a text and even games related to grammar and vocabulary allowed a connection between the language of instruction and the mother tongue.

Table 2 – Record of use of the mother tongue in the classes.

Department	English			German			French			Russian			Italian			Flemish			Chinese			Portuguese			Korean			TOTAL
	B	T	S	B	T	S	B	T	S	B	T	S	B	T	S	B	T	S	B	T	S	B	T	S	B	T	S	
DRE	x	x	x	1	x	3	1	X	x	x	x	1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	6
DLE	11	24	10	10	21	12	4	10	5	6	8	6	5	12	7	4	3	7	7	10	9	5	13	6	x	x	x	215
ORI	x	x	x	x	x	1	x	X	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	1	2
Total	11	24	10	11	21	15	5	10	5	6	8	7	5	12	7	4	3	7	7	10	9	5	13	6	x	x	1	222

Source: Classified by Departments from Universidad Nacional de Colombia.

Finally, in the ORI according to the data that I was able to register, it was evidenced that there were only 2 occasions in which the students made use of their native languages, in this case German and Korean. These data allow a comparison between classes according to the answers given by foreign students who have been able to enrol alternately in the three or two courses offered by the Departments from Universidad Nacional de Colombia as a free option in the learning process From Spanish to foreigners. For the great majority of foreign students who were in levels I and II of Spanish as a foreign language and who were also interviewed, they considered that it is necessary to have specific class lessons, systematized and that allow them to help to feel good learning Spanish in Colombia, on the contrary, make some recommendations and suggestions to one of the Spanish courses, since in the great majority they have decided to leave the classes given in that course since they felt uncomfortable because the basic level of Spanish in The ones they met and several of the proposed activities do not allow them to participate actively because they omit their native languages.

For a better understanding, a table has been designed (see Table 2) in which I have classified the information minutely, making use of some acronyms for clarity to the reader, described as follows:

- Teacher equivalents to T.
- Student equivalents to S.
- Both equivalents to B.

This last acronym allows show the number of occasions in which there was an interaction between the student (s) and the teacher in the moment of using or promoting the mother tongue of the students as the case may be. In turn, I have inserted in the table there is an x in order to indicate that no record was found according to the classification I have done.

As a second aspect, in the use of the language of instruction, I registered 132 occasions of which in the DRE has a total of 20 occasions in which the interaction in Spanish prevails between the student and the teacher, as for the DLE I identified 96 occasions In which also the interaction between the teacher and the student in Spanish. On the other hand, in the ORI, I recorded 16 occasions of which the use of Spanish as a language of instruction by the teacher and also by the student prevails in isolation (see Table 3).

Table 3 – Record of use of the language of instruction in the classes.

Record of use of the language of instruction in the classes.			
Department	**Spanish**		**Total**
DRE	11	3	6	20
DEL	45	19	32	96
ORI	2	7	7	16
Total	58	29	45	**132**

Source: Classified by agencies of the National University of Colombia.

Although opportunities for language use are not so frequent compared to the use and promotion of the foreign language of the mother tongue, the vast majority of them stated in the interview to the question whether they believe that their level of Spanish has improved since they arrived in Colombia stated that before they came they did not speak anything and in two months approximately that they are in the country and can understand what most Colombians speak, bearing in mind that it is a large percentage of Colombians in the country who do not speak another language fluently Outside their mother tongue because in the country for every ten citizens, one speaks perfectly the English language according to some information given by the Minister of Commerce, Industry and Tourism in Colombia (MinCiT, 2010).

In addition, students state that the Spanish courses offered by Universidad Nacional de Colombia have largely helped them not only because they are the first course they take within the country, but also because they consider that it is possible to learn Without having a solid foundation on Spanish, and some of them have had

a great empathy with teachers because they consider that they make a huge attempt to include everyone in learning because several of them is not limited to correct from a "Is wrong" but explains the reasons and gives opportunities in the midst of correction. Many foreign students even consider it a space for them to ask questions about the correct use of grammar, to solve doubts about a concept, to practice with people who have the same level of knowledge of the language and to know others People of different nationalities.

Finally, in this study I managed to record the use of a third language within the classroom that both teachers and students had in common, that language was English. I found a total of 27 occasions of use, of which in the DRE is recognized 5 occasions of use, in the DLE with 15 and ORI with 7 occasions respectively. Apparently the teachers of each of the units use the Spanish in the process, but as a language bridge there is this third language, there is a great difference in the frequency of use between the languages of origin of the students and the language of instruction.

Table 4 – Record of use of English as a third language in classes.

Record of use of English as a third language in classes				
Department	English			Total
	B	T	S	
DRE	2	1	2	5
DEL	2	9	5	16
ORI	0	1	6	7
Total	4	11	13	28

Source: Classified by agencies of the National University of Colombia.

According to reports provided by the European Union and the British Council (2000) in the world there are three types of English speakers, the first of who speak English as a mother tongue whose population lives in a country where English is spoken as a first language, being the main means of communication. On the other hand, there are those who use English as a second language, Crystal (1997) reported that 377 million speakers of English as a second language have been calculated in countries where they determine that they should not only know their language of origin but Must also be competently communicative an additional language in this case English, evidenced in government plans and educational curricula in their countries.

As a third type there are people who use English as a foreign language, even though in their countries of origin they do not have foreign language policies as compulsory through the use of a language other than the mother tongue. Use of another language, although it has rapidly increased the number of people who are learning English especially in underdeveloped countries in their government policies still prevails the mother tongue as a source of information.

Conclusion

In this study it is evident that the vast majority of foreign students are from the group of people who learned English as a second language because there are a high percentage of students whose nationalities are different from the North American, British and / or Australian. As a result, participants confirmed in the answers given in the interviews that English is a required language in all classrooms at different academic levels (primary-secondary and university) in their countries of origin.

Accordingly, many times the communication between the foreign students attached to the programs without regard to nationality or accentuation was mediated by their mother tongues. Even one of the reasons why foreign students were motivated to learn Spanish in the different spaces offered by the University was because they consider that Spanish is the second language spoken in the world, and this allows them to communicate with Latin Americans not only In a formal way, but also informal when traveling to different places in South America, because today it is not enough to speak a single language or even two because they confirm that many colleagues and friends speak three to four languages.

In identifying the mother tongues of the students, the promotion of these languages by the teacher, the use of the language of instruction and the level of frequency, it is evident that the language is a fundamental tool that allows the internalization and practice of the community Learning.

Acknowledgement

■ *RESUMEN: El presente artículo tiene como propósito estudiar el uso de translanguaging en el que en esta investigación se ha decidido traducir como translanguajeo determinado de acuerdo al desempeño sociolingüístico y el aprendizaje de un grupo de 22 estudiantes extranjeros que adquieren el español como lengua extranjera en una universidad colombiana. En este artículo se define el translanguajeo como un recurso estratégico determinado dentro de los procesos*

de aprendizaje de una lengua distinta a la materna, en este caso el español. Se recogieron los datos por medio de la observación directa, se realizaron entrevistas semiestructuradas a los participantes y se recolectaron algunos documentos que utilizaron los estudiantes en su proceso de aprendizaje del español. Los datos se analizaron de manera inductiva por medio de categorías de análisis determinadas desde los conceptos, temas y relaciones, con el fin de dar respuesta óptima a las preguntas de investigación. Concluyendo que el translanguajeo es un fenómeno que influye grandemente en el desarrollo de la lengua extranjera.

- *PALABRAS CLAVE: Translanguajeo. Español. Colombia. Recurso Estratégico. Estudiantes Extranjeros.*

REFERENCES

ÁLVAREZ-GAYOU, J. L.; SAMPIERI, R.; MENDOZA, P. **Metodología de la Investigación**. México: McGraw-Hill, 2008.

BAKER, C. Translanguaging: origins and development from school to street and beyond. **Educational Research and Evaluation**, p.1-16, 2011.

BAKER, C.; LEWIS, G.; JONES, B. Translanguaging: origins and development from school to street and beyond. **Educational Research and Evaluation**, 1-16, [n.p.], [s.d.].

CELIC, C. Translanguaging: strategies for the Classroom. **NYSIEB Seminar.** New York: CUNY-NYSIEB, 2012. p.1-30.

CELIC, C.; SELTZER, K. **Translanguaging:** a CUNY-NYSIEB Guide for educators. New York: CUNY-NSY, 2011.

COMMON CORE STATE STANDARDS INITIATIVE. **About the Standards**. Washington D.C, United States of America, 2013.

CREESE, A.; BLACKLEDGE, A. Translanguaging in the bilingual classroom: a pedagogy for learning and teaching. **The Modern Language Journal**, p.103-115, 2010.

CRYSTAL, D. **An encyclopedic dictionary of language and languages**. Oxford, UK: Blackwell, 1992.

_____. **English as a Global Language**. Cambridge: Cambridge University Press. 1997.

GARCÍA, O. **Bilingual Education in the 21st Century:** a global perspective. Oxford: Wiley-Blackwell, 2009a.

_____. Education, multilingualism and translanguaging in the 21st century. In: MOHANTY, P. P. K. M., **Multilingual education for social justice:** globalising the local. New Delhi: Orient BlackSwan, 2009b.

_____. Theorizing translanguaging for educators. In: CELIC, C.; SELTZER, K. **Translanguaging:** a CUNY-NYSIEB Guide for Educators. New York: CUNY-NYSIEB, 2011. p.1-6.

_____. Reimagining Bilingualism in Education for the 21st Century. **NALDIC (National Association for Language Development in the Curriculum)**. London: NALDIC, 2013. p.4-11.

_____. TESOL Translanguaged in NYS: Alternative Perspectives. **NYS TESOL Journal**, p.2-10, 2014.

GOETZ, J.; LECOMPTE, D. **Etnografía y Diseño Cualitativos en Investigacion Educativa**. Madrid: Ediciones Morata, 1988.

GUERRA, J. Emerging representations, situated literacies, and the practice of transcultural repositioning. In: BALESTER, V.; KELLS, M. **Latino/a Discourses:** on language, identity, and literacy in education. Portsmouth: Heinemann, 2004. p.7-23.

HARRIS, M. **Teorías de la cultura en la era postmoderna**. Barcelona: CRITICA S. L., 2007.

HERNANDEZ SAMPIERI, R.; FERNÁNDEZ, C.; BAPTISTA, M. **Metodología de la investigación**. México: McGraw-Hill, 2010.

KANO, N. **Translanguaging as a process and a pedagogical tool for japanese students in an English writing course in New York**. 2012. 218 f. Dissertation Thesis (Bilingual Education) – Columbia University, Columbia, 2012.

LAMBERT, W. Culture Factors in Learning. In: MEADE, A. I. **Culture and language as factors in learning and education**. Bellingham: Western Washington State College, 1974.

LEWIS, H. **Body Language:** a Guide for Professionals. India: SAGE Publications, 2012.

LEMKE, L. Foreword. In: WELLINGTON, J.; OSBORNE, J. **Language and literacy in science education**. Buckingham: Open University Press, 2001.

LI, W. Moment analysis and translanguaging space: discursive construction of identities by multilingual Chinese youth in Britain. **Journal of Pragmatics**, p.1222-1235, 2011.

LI, W.; HUA, Z. Translanguaging identities and ideologies: creating Transnational space through flexible multilingual practices amongst Chinese University Students in the UK. **Applied Linguistics**, p.516-535, 2013.

NILEP, C. "Code-switching" in sociocultural linguistics. **Colorado Research in Linguistics**, p.1-22, 2006.

NIELSEN, J.B. **El Translanguajeo como recurso estratégico para el proceso de enculturación en el idioma español colombiano**: Estudio etnográfico con estudiantes extranjeros y profesores de la Universidad Nacional de Colombia.2018. 341 f. Mexico (Unpublished PhD thesis) Doctorado en Educación, Universidad Internacional Iberoamericana, UNINI, 2018.

STATHOPOULOU, M. **Task dependent interlinguistic mediation performance as translanguaging practice:** the use of KPG data for an emporically based study. 2013. 279 f. PhD Thesis – Faculty of English Language and Literature, University of Athens, Athens, 2013a.

_____. Investigating Mediation as Translanguaging Practice in a Testing Context: towards the development of Levelled Mediation Descriptors. **Language Testing in Europe:** time for a New Framework, 2013b. p.209-217.

SUBSECRETARÍA DE EDUCACIÓN. Actividades sugeridas para el abordaje del Cuaderno de trabajo para el aula de Inglés de 4° EP. Retrieved 10 10, 2016, from Buenos Aires Educación. Available in: <http://servicios2.abc.gov.ar/lainstitucion/organismos/lenguasextranjeras/plurilingue/propuesta_4ep/propuesta_4ep_imagenes.pdf>. Access date: 3 nov. 2014.

TORRES, J. La investigación etnográfica y la reconstrucción crítica en educación. In: GOETZ, J.; LECOMPTE, M. **Etnografía y diseño cualitativo en investigación educativa**. España: Ediciones Morata S.A., 1988. p.11-21.

VYGOTSKY, L. **Mind in society:** the development of higher psychological processes. Cambridge, MA: Harvard University Press, 1978.

WILLIAMS, C. **Arfarniad o ddulliau dysgu ac addysgu yng nghyd-destun addysg uwchradd ddwyieithog**. Bangor: University of Wales, 1994.

_____. Secondary education: teaching in the bilingual situation. In: LEWIS, G.; WILLIAMS, C. **The language Policy:** takings stock. Wales: CAI, 1996.

_____. **A language gained:** a study of language immersion at 11-16 years of age. Bangor: School of Education, 2002.

_____. **Extending Bilingualism in the Education System**. Retrieved November 3, 2012, from Education and lifelong learning committee. Available in: <http://www.assemblywales.org/3c91c7af00023d820000595000000000.pdf>. Access date: 3 nov. 2014.

WITT, D. **Strategies for Translanguaging**. New York: CUNY-NY, 2012.

MOOCS FOR LANGUAGE TEACHING: A STUDY ON CALL FROM THE CONNECTIVIST PERSPECTIVE

Marcus Vinicius Liessem FONTANA*
Vilson José LEFFA**

- ABSTRACT: The increasing offer of Massive Online Open Courses (MOOCs) in the World Wide Web generated the need to investigate their dynamics, pedagogy and structure in more depth to judge the possibilities of their effective use in the teaching of foreign languages. In this study, based on investigations in Computer Assisted Language Learning (CALL), we start from the perspective of the Connectivist Learning Theory, established, mainly, from the studies of Siemens and Downes to analyze two foreign language courses in the MOOC modality offered in two different digital platforms: one course in Spanish as a foreign language offered by UNED on its own platform and a Mandarin course for Spanish speakers from the UNIMOOC platform. In both cases, it is evident that the courses did not fulfill what they promised, especially regarding the need to create spaces conducive to interaction. This study is part of a broader research proposal, whose objective is to establish adequate criteria for the creation of MOOCs as tools for the teaching of foreign languages.

- KEYWORDS: Second language teaching. CALL. MOOCs. Connectivism.

Introduction

In mid-2013, Brazil's largest circulation newspapers announced that the University of São Paulo (USP) had begun offering its first MOOC courses, an acronym for the Massive Online Open Course. Some communication vehicles, including Folha de São Paulo[1], reported that they would be the first courses in Latin America, although the National University of Mexico had already signed a contract for this purpose with the Coursera platform since February of that year, with courses on *Scientific Thinking*, *ICT in Education* and *How to Be More Creative*. Even in Brazil, other experiences had already been developed as early as 2012. Unesp had launched the Unesp Open platform,

* Federal University of Santa Maria (UFSM), Santa Maria - Rio Grande do Sul - Brazil. Department of Modern Foreign Languages. marcusfontana2011@gmail.com

** Catholic University of Pelotas (UCPel), Pelotas - Rio Grande do Sul - Brazil. leffav@gmail.com

1 Available in: <http://educacao.estadao.com.br/noticias/geral,usp-lanca-primeiro-curso-virtual-gratuito-de-nivel-superior-da-america-latina,1041146>.

providing content of its courses in digital media, while PUC-SP, in partnership with the Brazilian Association of Distance Education (ABED), launched the MOOC DL course, coordinated by Professor João Mattar. The only restriction of these initiatives is that they did not offer certificates to the participants, an aspect that USP has effectively innovated.

Thus, although it was not the first to offer MOOCs, in fact, USP started two courses certified in 2013: *Basic Physics* and *Probability and Statistics*. The problem, and misconception, as concerns media was to announce them as the first undergraduate courses in this format in Brazil. Although offered by an institution of higher education, the two courses had an extra-mural character.

The novelty celebrated by the mass media, in fact, was already known to many Brazilians with digital technologies and invariably thirsty for innovations, who already participated in courses in the MOOC format offered by universities in the United States and Europe. The MOOCs, as we will see below, have a History that begins at the dawn of the 21st century, with the Connectivist studies of Siemens and Downes, who in 2003 already theorized about them.

In 2008, the first course in this format was offered. In 2011, from a Stanford University initiative, the Coursera platform was born, completely focused on this technology, soon followed by the edX platform, daughter of a partnership between the Massachusetts Institute of Technology (MIT) and Harvard University. Since then, universities around the world have embraced the initiative. The Coursera platform, for example, now has partnerships with universities in South Africa, Argentina, Australia, Brazil, Belgium, Canada, Chile, China, Singapore, Colombia, South Korea, Denmark, Spain, the United States, France, Holland, India Israel, Italy, Japan, Mexico, United Kingdom, Russia, Sweden, Switzerland, Taiwan and Turkey.

As we can see, MOOC platforms have advanced inexorably in a relatively brief period of time, and the trend is to further expand over the next few years. There are, however, issues about the format. An emblematic example is the interview given by Sebastian Thrun, a Stanford University professor and visionary creator of the Udacity MOOC platform, to journalist Max Chafkin in 2013, just a year after the *New York Times* had published a headline claiming 2012 to be the year of the MOOCs (PAPPANO, 2012). Among other things, Thrun stated: "I'd aspired to give people a profound education, to teach them something substantial, but the data was at odds with this idea." (CHAFKIN, 2013).

Another authority on the subject, Professor García Aretio, linked to the UNED (National University of Distance Education of Spain) and Director of the Unesco Chair for Distance Education, published an article in 2015 where he raised around 90 doubts about MOOCs, questioning its mechanisms, pedagogy and efficiency. The researcher himself, however, tried to answer many of these questions in a new article published in 2017 entitled *Los MOOC están muy vivos*, which already gives an idea of the content of the text. Other aspects, however, remain open, including problems of interaction and feedback, which are crucial when it comes to teaching and learning foreign languages in this modality.

This paper aims to discuss some of those issues. This is a preliminary analysis about two language courses offered by different institutions on different platforms, complemented by some data considered relevant in ten other courses that touch on the theme, either because they deal with cultural aspects linked to a foreign language, or because they discuss computer-mediated teaching.

This study is part of a larger project that links researchers from the Federal University of Santa Maria, the Federal University of the Southern Frontier and the Catholic University of Pelotas in search of the best criteria for the development of courses of this type. The final objective of the project is to develop a MOOC course that discusses exactly what a MOOC should be like when geared for language teaching, defined as Language MOOC (MARTÍN-MONJE; BÁRCENA, 2015).

In the next sections, we introduce Connectivism, as a theory to explain MOOCs, describe the main characteristics of some courses analyzed, compare them with the principles established by the theory, and conclude with some suggestions on how to design what should be seen as a good MOOC for language teaching.

Connectivism and MOOC's

In a 2003 article entitled *Learning Ecology, Communities, and Networks: extending the classroom*, George Siemens began to outline the first draft of what would become, in his view, a new learning theory: Connectivism. Later, in 2006, Siemens launched *Knowing Knowledge*, published with a *Creative Commons* license, in which he explains his theory in more depth, adding important reinforcements mainly from the collaboration of Stephen Downes and reaching the world with the launching of the first known MOOC in 2008, in which more than 2,200 people dedicated themselves to discussing the subject in an online course.

In the words of those who created the concept:

> A MOOC is an online course with the option of free and open registration, a publicly shared curriculum, and open-ended outcomes. MOOCs integrate social networking, accessible online resources, and are facilitated by leading practitioners in the field of study. Most significantly, MOOCs build on the engagement of learners who self-organize their participation according to learning goals, prior knowledge and skills, and common interests. (SIEMENS et al., 2010, p.10).

With respect to Connectivist Learning Theory, the first and most important concept of Siemens is that of the Ecology of the Learning. According to the author, ecologies are shared and collaborative means of building knowledge, in which one of the main factors for development is the existence of people connected to each other, such as nodes composing a web, similar to the Web of Life conceptualized by

Capra (2006), a kind of living organism that is self-organized, dynamic, adaptable, and sometimes confused and chaotic, but where the emergence of knowledge is fully possible. From this point of view, Siemens (2010) identifies a series of elements necessary for learning to emerge:

- Learning and knowledge rest on diversity of opinions.
- Learning is a process of connecting specialized nodes or sources of information.
- The promotion and maintenance of connections are fundamental movements to facilitate continuous learning.
- The ability to see connections between ideas, concepts and areas of knowledge is a crucial competence.
- The ability to continually expand knowledge is more important than what we know at any given time.
- Decision making is, by itself, a learning process.
- Learning can reside in non-human devices.

Although Siemens and his colleagues seek to emphasize the fact that Connectivism addresses the need for a theory that looks at learning in digital media, it is important to have it clear that there are many critics who insist that, in terms of learning theories, the ideas that underline connectivism are nothing new (KOP; HILL, 2008). Responding to these researchers, Siemens (2010) seeks to highlight some aspects which, in his view, would be exclusive to his theory

- The contemplation of principles that govern learning in different scopes: conceptual, biological, neural, social, etc.
- Attention to information abundance, speed and impact on understanding, interpretation, construction of knowledge today.
- The importance given to the concept of networking and how the connections that are established with other humans and even with databases interfere in learning.
- The emphasis on digital technologies as a key part of the distribution of cognition and knowledge.

Siemens and his colleagues, thinking about how these principles apply to online education, list some characteristics that they consider to be crucial in courses developed in digital platforms, as the case of MOOCs, in an effective way:

- High levels of student control over modes and places of interaction.
- Weekly synchronous sessions with facilitators and guest speakers.
- A daily newsletter email as a regular point of contact for course participants.

- Use of Really Simple Syndication (RSS, an automatic information updating system) to follow the participants' blogs.
- Emphasis on student autonomy regarding selection of learning resources and level of participation in activities.
- Emphasis on social systems as effective means of self-organization of students.

Although these seem to be required characteristics for a mass online course to be effectively considered a MOOC, researchers such as Bates (2014) suggest that, currently, MOOCs can be divided into two types: xMOOCs and cMOOCs. The xMOOCs would have little or no flexible organization, centered on content and the figure of the teacher. Using xMOOCs, we would work exclusively on a platform specifically designed for the purpose of the course, with video-readings, automatic response activities, peer review, some space to share opinions and discussions, little or no moderation in the discussions. They would be the most common type, present in most popular platforms like Coursera, edX, Miriada X. In fact, the "x" of xMOOC would refer precisely to this type of platform, which would force the courses to a monolithic format.

The cMOOCs, on the other hand, would be more in line with the criteria proposed by Siemens and his partners. The relationship is transparent, since the "c" alludes precisely to Connectivism. Using cMOOCs, we would have more space, there would be more space for learner's autonomy, prioritizing the connection between students and accepting the use of a diversity of tools and means, at the discretion of the students themselves, including social media and tools for collaborative creation. There would be no room for formal evaluations and the teacher would have an auxiliary role, guiding learning without directly interfering.

Student protagonism, in this model, becomes clear especially from the Principle of the Criticality of Creation, according to which "[...] learners create and share their understanding of the course topics through blogs, concept maps, videos, images, and podcasts. Creating a digital artifact helps learners to re-centre the course discussion to a more personal basis." (SIEMENS et al., 2010, p.23).

In the next section, we analyze two language courses, offered as MOOCs, based on the criteria defined from the Connectivist principles. We have no doubt that there are other learning theories that could account for a study on MOOCs. Our aim, however, is to verify whether the theory built under the aegis of MOOCs is robust enough for a profound analysis of this course model and whether, in fact, MOOCs fulfill the promise of educational innovation that has followed them since their creation.

Between Spanish and Mandarin

The present investigation was developed during the second semester of 2016 and depended on the work of ten assistant researchers who became involved as students

in different MOOCs offered in different platforms, following the norms of participant observation for data collection (FLICK, 2009). The initial intention was to evaluate language courses in this format. At the time the group was involved in data collection, and due to the criterion of gratuity, only two language courses were available, Spanish and Mandarin. In order not to take advantage of the time available and be able to collect data from the platforms, the group decided to enlarge inclusion criteria, allowing for the analysis of courses focused on literature, foreign language culture and online teaching pedagogies. All courses were attended by two or three researchers, who made their observations and notes individually. The collected data were then shared and debated in research group meetings in the light of Connectivism, which led to the conclusions shown below.

In this section, we present the pertinent results to the analysis of the two language courses that we were able to follow. In the next, we bring some additional information that could be collected from the other courses.

The language MOOCs that we could access in the second half of 2016 were: Aprende chino para negocios (Learn Chinese for Business) and Español en Línea ELE (Online Spanish). The first, made available on the UNIMOOC platform, was developed by Lun Yu Chinese School, based in Madrid, Spain. It is a school specializing in teaching Mandarin to Spanish speakers. The school offers the MOOC as a kind of introduction to the language, inviting students from its face-to-face courses. In its publicity material, the school claims to follow the criteria of the Confucius Institute for language teaching. The Institute is an official organization dedicated to the dissemination of the language, with insertion in several western universities, such as Unesp, in Brazil, by means of accords and including specific teaching and learning standards, like the Common European Framework of Reference for the Teaching of Languages.

The second course was developed by UNED, the National University of Distance Education of Spain. Created in the early 1970s, UNED is today the largest Spanish university, with more than 250,000 students enrolled. In addition to its 26 undergraduate and 43 master's level courses, UNED has invested heavily in MOOCs and has even created a platform for its massive courses: UNED-COMA (COMA is the acronym in Spanish for *Cursos Online Masivos Abiertos*).

To carry out the analysis of the courses, the research group developed an observation guide based on the criteria established by the Connectivist Learning Theory for the characterization of a MOOC. The guide was composed of questions that could guide the researchers in their observations and reflections on the courses they were participating in. The guide was composed of 16 questions:

1. What was proposed at the beginning of the course was consistent with its final result?
2. Is the course divided into units?
3. If yes, did the units have clear, explicit and well-defined objectives? Add comments.

4. When starting a unit, is it clear what the trainee is expected to develop in terms of knowledge and skills?

5. Do units have suggested or mandatory time? If so, how much? Is there flexibility to finish a unit? If it is not completed within the specified time, is there any loss or penalty for the student?

6. When you move to a new unit, do you feel that you have mastered the contents of the previous unit? Do you feel valued, self-assessed and able to go on?

7. Does the course present a didactic guide that helps to situate the student in the course, indicating the unit in which he/she is, what he/she has already covered and what is still ahead? If so, describe how the guide is presented and what elements make it up.

8. What types of assessment are presented throughout the course?

9. Is the course taught by a single teacher or a group of teachers? In the case of a group, does there seem to be a responsible teacher working with collaborators or do they work independently?

10. Does the course have discussion forums? If so, what are their goals? How do they work?

11. Does the course have PDF or PPT materials? If yes, what kind of content do they have? How are they structured (texts, images, etc.)?

12. Does the course have videos? If yes, what are their goals? Are they teacher talks, or do they involve other types (animations, short films etc.)? What is the average duration? Were there any exceptions (very long or very short)?

13. Was there any kind of videoconference with the teachers or between students? If yes, how did it work?

14. Are there other materials offered by the course to the student?

15. Overall, were you satisfied with the course? What did you like the most and what did you like the least about this experience?

16. Is there any other information you would like to add?

In our analysis, we will not follow the strict sequence of the questions as presented above. Trying to be more dynamic and making our reflections more objective, we will group the questions into categories, as follows: course structure, available materials, interaction, assessments and global perception of learning.

As far as structure is concerned, we will start our analysis by breaking the course down into units. Following the Connectivist principles, it is important for the students to understand the process from the beginning, being able to visualize the course tutorial to facilitate their choices in terms of their learning strategies. The student needs to have control of their learning process, and because of this, the course structure must be transparent. In this sense, the Mandarin course was clearly divided into six lessons, all of them with topics defined a priori. The Spanish course, however, was divided into modules: Presentation, Modules 0 to 4 and Closing, but only in the Presentation there is a description of unit 1. The description is made in Spanish, French, English

and Portuguese. The other units do not have the objectives described. The only way for the student to know what to expect to learn is from the unit title. In neither course there is a requirement or even a stipulation of time to study the content and complete the assignments. The students are totally free to set the speed of their progress. In addition, in both cases, the students have a progress indicator available for them. Whenever you log in, you are referred to where you left off. In the case of the UNIMOOC platform, in the Mandarin course, the system specifies the current lesson with black letters, in contrast to the blue color of previous lessons and those to come. The COMA platform, on the other hand, uses blue to indicate previous titles.

In terms of available materials, the Mandarin course is basically made up of videos. Each lesson is presented through a single video of approximately four minutes, except for the last two, which have seven and eight minutes, respectively, in which one of the teachers who takes turns in the presentation of the subjects gives explanations in Spanish and pronounces the words and expressions in the target language. Additionally, next to each video there is a "More content" button. By clicking on it, students are referred to a Google Docs link where they have access to a text file with vocabulary. It is a table with four columns: the Chinese ideogram, the corresponding writing in western characters, the grammatical category and the meaning. In the same way, the Spanish course is presented mainly by videos. Some are reproduced from Power Point presentations offering grammatical explanations with the teacher's voice in the background. In general, the videos are eight to nine minutes. The longest is one of the first, which explains the structure of a film review, lasting around fifteen minutes. There are images and texts used in a harmonic way, without excesses in any part. In addition to the videos, the Spanish course also has audio material in MP3 format.
Interaction was practically nonexistent in both cases.

This is surprising considering that the theory underlying MOOCs emphasizes students' socialization and self-organization. For Connectivism, digital media should serve as a bridge between learners. Learning takes place when the networks are created, with nodes provided with databases, applications and other technological tools, but for learning to occur human touch and interaction must be added. In the case of the Mandarin course, there is only one forum for solving doubts, where students can resort to the organizing team to help. If students ask no questions, however, there is no direct contact with teachers at any time. Even less with other students. At no time is this interaction possible. The Spanish course is not very different. Here, however, not even a forum is available for the students to ask help from teachers. The only interaction that occurs is the delivery of the final work, which is sent to the team, who will evaluate and give feedback.

As far as assessments are concerned, the Mandarin course presents some multiple-choice questions at the end of each of the six lessons. After the lessons are completed, there is a test with five questions, also in multiple-choice format. When answering at least four correctly (80%), the trainee obtains a free certificate issued by the University of Alicante, a partner of the Lun Yu School. In the Spanish course, there are multiple

choice exercises interspersed with writing exercises. All are submitted, but only the final task receives feedback from the team responsible for the course. This is a generic feedback, in which the evaluator congratulates the student and thanks for the participation. The course offers certificate only with payment.

Finally, we come to the last point, which concerns the global perception of learning. The observations made by the assistant researchers, first individually and later discussed in the meeting, as already mentioned, lead to the conclusion that the courses did not completely respond to what had been proposed. In the case of the Mandarin course, the objective of the course would be "to provide basic knowledge of Mandarin, which will allow to establish business in China". Considering the complexity of Mandarin, a distant language to speakers of Spanish and Portuguese, it would be naïve to expect that a ten-hour course would be enough to achieve this objective, although very modest. Even with such a timid goal, the course delivers less than it promises, reducing itself to explore, superficially, the tonal structure of the Chinese pronunciation, aspects related to greetings and courtesy, business card etiquette and numerals. Hardly would anyone be able to make an expressive import/export agreement with so few elements. Obviously, the course will help those who have the support of an interpreter, allowing them to express small courtesies in the target language, which will certainly be well received by the hosts, but the language skills developed do not exceed this modest limit.

The data collected in relation to the Spanish course showed similar results. Despite presenting itself as a course based on the exploration of the cinematographic review genre, little was worked within this focus. Only the first modules explore the topic. The others turn to grammar, pure and simple. Still, the assistant researchers concluded that, from a grammatical perspective, the course gave an account of presenting, discussing and teaching some rudiments of Spanish. What most attracted the attention, in a positive way, was the fact that the MOOC was clearly designed for students who already had some notion of Spanish, being suitable for students of level A2, according to the European Common Framework. In this sense, in the introduction of the course itself, a link was made available so that the interested party could take a level test to know in advance if he would be able to follow up. In any case, with almost no interaction and an unfulfilled work proposal, the course does not seem to achieve the objectives it proposes.

So far, we have been able to get a clear view of how the two language courses presented as MOOCs by their creators should work and how they really worked. Two different language courses were evaluated, available on different platforms and by different institutions. In the next section, we will present ten other courses. We will not delve deeply into these MOOCs. The intent is simply to demonstrate the variety of existing courses, the platforms and institutions that make them available, and to make some general comments that can help us gain insight into how these MOOCs have worked and what may still be missing so that they become adequate to a Connectivist perspective.

From Quixote to blended learning: variations on the same theme

As we mentioned from the beginning, we found only two language courses available in the second half of 2016 and none of them on the more traditional platforms, *Coursera* and *edX*. We then explored other courses, according to the criteria already mentioned. Courses in Portuguese or Spanish were chosen as shown in Table 1.

Table 1 – Courses, universities and platforms

COURSE	UNIVERSITY	PLATFORM
Educación de calidad para todos. Equidad, inclusión y atención a la diversidad	Universidad Autónoma de Madrid.	edX
El desafío de Innovar en la Educación Superior	Universidad de Chile	Coursera
Ensino Híbrido: Personalização e Tecnologia na Educação	USP e UNESP	Coursera
Explorando os recursos educacionais da Khan Academy	Fundação Lemann, Instituto Península, Instituto Natura, Ismart	Coursera
Fundamentos do Google para o Ensino	Fundação Lemann, Foreducation & Google for Education	Coursera
Hacia una práctica constructivista en el aula	Pontificia Universidad Católica de Chile	Coursera
Introducción al Aprendizaje Basado en Problemas	Universidad Zaragoza	Miriada X
La España de El Quijote	Universidad Autónoma de Madrid	edX
Técnicas Cuantitativas y Cualitativas para la Investigación	Universitat Politècnica de València – UPV	edX
Tecnologías para la educación	Universitat Politècnica de València – UPV	edX

Source: Own elaboration.

The courses are listed alphabetically. Five of them were made available on the *Coursera* platform and four on *edX*. A final course was found on the Miríada X platform, a Spanish platform linked to Universia, the largest network of Ibero-American universities, based in Spain and linked to Banco Santander.

In general, the criticisms made to these ten courses were the same as those made to the first two. The exception is the organization of the *edX* platform courses, which present some of the elements proposed by the Connectivist Learning Theory, such as explicit objectives, defined deadlines, syllabus, and data that students need to organize

their trajectory and prepare their study strategies. This, however, does not seem to be the merit of the institutions, but of the platform, which submits the courses to a standardized structure.

The biggest criticisms, again, fell on little interaction and lack of feedback. In general, students are not placed in situations where they must interact. Even teachers do not often interact. There are courses, for example, where there is a list of responsible teachers and only one or two appear in the videos. The students never get to know them. Not even in the course presentation video they show up. In some circumstances, however, the expected presence of teachers does not promote the desired motivation: monotonal readings from Power Point presentations filled with text or extensive readings in front of the cameras. This simple transposition from face-to-face to the virtual context bores the students and can be one of the reasons that lead to the high attrition rates found in this type of course. According to a survey conducted in 2013, attrition numbers are so high that only 2% to 14% of students concluded the courses (PERNA et al., 2013).

The frequent evaluations of activities with the concise "Very good!" or "Congratulations, keep it up!" – when there is feedback – is another obstacle to retain the students. There is no stimulating debate. The teachers do evaluate their student tasks, but because they have thousands of students, they either end up by restricting evaluation to a minimum or use peer-to-peer (p2p) evaluation, with little guidance on how to review each other's activities. The result is that feedback is reduced to the emotional level, avoiding content or technical aspects.

One of the assistant researchers in the group, who collaborated in the participant observations, due to lack of interaction, no feedback and the continuous requirements to fill out satisfaction questionnaires in one of the courses, stated at the last meeting of the research group to discuss the data collected, that as a student, she did not feel important: "It even seems like they were studying us". Considering the very strong marketing bias one perceives in some of the courses, it is possible, that the objective is to advertise their graduate courses online, paid in euros or dollars.

There are, however, good points. Videos with dialogues or dramatizations gave a more playful tone to the contents and managed to capture students' attention. Teachers who know how to dialogue with the camera without being stuck in presentations or texts were also positively evaluated. Another aspect found in some courses was flexibility in the final assignment. The possibility of exploring different media and tools to produce the final project, with a generative theme, and a format free from monolithic molds seemed to yield interesting materials, giving the students creative freedom.

In the next section, based on the positive and negative aspects we found, and considering the principles of Connectivism, we present our final considerations, bringing some elements that we consider important for a language course in the MOOC format.

Looking for new roads

Taking the same sequence from our analysis, we will begin this section by taking up our considerations about the desirable structure of a MOOC from the Connectivist Learning Theory in contrast to what was found in our research. So, regarding the structure of MOOCs, we need to emphasize the need for them to be transparent. A course needs to have its overall goal clearly defined from the beginning, as well as the specific objectives of each unit, which need to be visible and comprehensible by the students. A student cannot start a unit without a full view of its objective, because this is the only way to comply with the principle of self-organization, according to Siemens, for whom students can only organize themselves if the learning goals are clear. It is important to remember that self-organization is a fundamental element in Connectivism. We do not learn alone. We learn when we form networks with learning tools and with other people and these connections become meaningful only if we know very clearly where we are at any moment and where we need to go, as object-oriented human beings.

Considering available materials and interaction, we believe the most important result from this investigation is the finding that the way teachers present themselves in videos is critical. No matter how well a video is designed and produced, it will lead to poor results if it does not reflect, at least in part, human interaction. The teacher needs to be able to interact with the camera, talk to the people on the other side. He must also refrain from long and tedious readings or from hiding behind presentations with abundant texts. He can use dialogues with colleagues, simulating – or actually doing – brief interviews, always trying to set the right tone, without excessive formality or extreme informality. Creative role plays to demonstrate some point or content are often welcomed. Also, videos should never be too long. If a topic is too long, it is more productive to break it down into several small videos. Salman Khan, for example, founder of Khan Academy, suggests that an exhibition video should not exceed ten minutes (KHAN, 2013). Another useful suggestion is that these videos have a transcript in PDF, so that the student can consult a specific point later, whenever necessary, without having to watch the video all over again. If possible subtitles should be offered in different languages, considering that a MOOCs may reach the whole world.

As far as assessment is concerned, it is important to respect the principle of student autonomy. Testing restricted to multiple-choice items contribute very little to learning. Fewer evaluations, allowing for students to express themselves and use their creativity, tend to be more welcomed and yield better results. In general, considering the courses we analyzed, the assistant researchers felt more at ease in those where they could freely use any tool to carry out some type of production that involved their personal perception regarding the content studied. In one course, for example, students were given the freedom to design a digital artifact, to use Siemens terminology, to synthesize their understanding of one of the topics covered. The artifacts were sent to peers in a p2p system for evaluation. Videos, animations and other objects were made using different resources available to students. They were given complete freedom. By

producing their own materials and evaluating their peers' materials, students deepened and consolidated a more profound learning process. Making room for the students to choose their own readings and relate them to the content covered in the course also proved to be a positive point.

Finally, it is important to note that in none of the MOOCs analyzed we found those aspects, which are regarded as basic by Siemens in this type of course, such as weekly synchronous sessions with invited speakers and facilitators, daily email newsletters, use of RSS to follow course participants' blogs; emphasis on social systems as effective means of self-organization for students, etc. It seems that many advances are still needed for the reality of MOOCs to meet the theory.

We described in this paper the results of our investigation and offered some suggestions on how a MOOC for language teaching should be designed to be significant and efficient considering the concepts that guide Connectivist Learning Theory. There are still several gaps that need to be filled, including: (1) more suggestions for creating such courses; (2) clear definition of the elements that are essential for the proper functioning of the courses, with meaningful results for the students; (3) design of a metaMOOC, i. e. a MOOC that discusses and presents the necessary suggestions for designing a Massive Online Open Course, oriented to the teaching of languages.

- *RESUMO: Com o aumento da oferta na rede mundial de computadores dos chamados Cursos Online Massivos e Abertos, mais conhecidos por seu acrônimo em inglês MOOCs (Massive Online Open Courses), cria-se a necessidade de que se pesquise de maneira mais aprofundada sua dinâmica, sua pedagogia e sua estrutura a fim de que se possa estabelecer um juízo sobre as possibilidades de uso efetivo desta nova ferramenta para o ensino de línguas estrangeiras. Neste trabalho, situado no âmbito das pesquisas em CALL (Computer Assisted Language Learning), partimos da perspectiva da Teoria Conectivista, estabelecida, sobretudo, a partir dos estudos de Siemens e Downes, para analisar dois cursos de línguas estrangeiras na modalidade MOOC, oferecidos em diferentes plataformas digitais: o curso de espanhol para estrangeiros oferecido pela UNED em plataforma própria e o curso de mandarim para hispanofalantes veiculado pela plataforma UNIMOOC. Em ambos os casos, evidencia-se que os cursos não cumprem o que prometem, apresentando falhas no processo de ensino, sobretudo no que tange à necessidade de criar espaços propícios à interação. Este trabalho inscreve-se em uma proposta de pesquisa mais ampla, cujo objetivo é estabelecer critérios adequados para a criação de MOOCs para o ensino de línguas estrangeiras.*

- *PALAVRAS-CHAVE: Ensino de línguas estrangeiras. CALL. MOOCs. Conectivismo.*

REFERENCES

BATES, T. **Comparing xMOOCs and cMOOCs:** philosophy and practice. 2014. Available in: <http://www.tonybates.ca/2014/10/13/comparing-xmoocs-and-cmoocs-philosophy-and-practice/>. Accessed on: 8 may. 2016.

CAPRA, F. **A Teia da Vida**. São Paulo: Cultrix, 2006.

CHAFKIN, M. **Udacity's Sebastian Thrun, Godfather Of Free Online Education, Changes Course**. 2013. Available in: <https://www.fastcompany.com/3021473/udacity-sebastian-thrun-uphill-climb>. Accessed on: 20 apr. 2017.

FLICK, U. **Introdução à Pesquisa Qualitativa.** Porto Alegre: Artmed, 2009.

GARCÍA ARETIO, L. Los MOOC están muy vivos. Respuestas a algunas preguntas. In: **RIED - Revista Iberoamericana de Educación a Distancia**, v. 20:1, p.09-27, 2017.

_____. MOOC: ¿tsunami, revolución o moda pasajera? In: **RIED – Revista Iberoamericana de Educación a Distancia**, v.18:1, p.9-21, 2015.

KHAN, S. **Um mundo, uma escola**. Rio de Janeiro: Intrínseca, 2013.

KOP, R.; HILL, A. Connectivism: learning theory of the future or vestige of the past? **The International Review of Research in Open and Distributed Learning**, v.9, n.3, p.1-13, 2008.

MARTÍN-MONJE, E.; BÁRCENA, E. (Eds.). **Language MOOCs**: providing learning, transcending boundaries. Berlin, New York: Mouton de Gruyter, 2015.

PAPPANO, L. The Year on the MOOC. In: **New York Times**. 2012. Available in: <http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html>. Accessed on: 20 apr. 2017.

PERNA, L.; RUBY, A.; BORUCH, R.; WANG, N.; SCULL, J.; EVANS, C.; AHMAD, S. The life cycle of a million MOOC users. In: **MOOC Research Initiative Conference**. 2013.

SIEMENS, G. **Learning Ecology, Communities, and Networks:** extending the classroom. 2003. Available in: <http://www.elearnspace.org/Articles/learning_communities.htm>. Accessed on: 15 mar. 2016.

_____. **Knowing Knowledge**. 2006. Available in: <http://www.nodosele.com/editorial>. Accessed on: 8 feb. 2011.

SIEMENS, G.; MCAULEY, A.; STEWART, B.; CORMIER, D. **The MOOC Model for Digital Practice**. 2010. Available in: <http://www.elearnspace.org/Articles/MOOC_Final.pdf>. Accessed on: 10 oct. 2012.

EARLY RECOGNITION OF SPELLING PROBLEMS IN STORIES INVENTED BY TWO NEWLY LITERATE PUPILS: WHEN AND HOW THEY HAPPEN

Eduardo CALIL [*]
Luísa Álvares PEREIRA [**]

- ABSTRACT: The acquisition of spelling competence is a complex process, involving lexical and grammatical questions. Research, however, almost always places the spelling from an autonomous point of view and disconnected from the other components of writing. In this text, we present the relevance of the Ramos System that captures students in an ecological situation of text production in pairs, allowing access to the processes for solving orthographic problems. Collaborative writing also grants access to comments made by students during the process of textual linearization. Our study focuses on the recognition of spelling problems (SP) and the comments made regarding such problems by two 2nd grade students during the production of six invented stories. More than a quantitative analysis of the types of SP identified in the product, we were interested in making a qualitative and fine analysis of oral recognitions of SP, particularly those SP anticipated by the writers. Our results indicate that: i. Recognition motivates comments that are not always related to the identified SP; ii. Recognition and comments are related to the orthographic contents taught in the classroom; iii. Some of the recognized SP involve the articulation between different linguistic levels. These aspects can contribute for the comprehension of orthographic learning in didactic situations provided by collaborative writing.

- KEYWORDS: Classroom. Text production. Spelling. Erasure. Metalinguistic. Learning. Collaborative writing.

Introduction

The path taken by the writers from the first texts onward, until they become autonomous as text producers, is long and complex, depending, above all, on a systematic teaching and reflective activity over their own writing (KELLOGG, 2008; PEREIRA, 2008; PEREIRA; BARBEIRO, 2010). In the beginning of this process, the

---

[*] Federal University of Alagoas (UFAL), Education Center, Maceio - Alagoas - Brazil. eduardocalil@hotmail.com

[**] University of Aveiro, Aveiro - Portugal. Department of Education and Psychology. lpereira@ua.pt

writer has to deal, on the one hand, with a notation system which involves graphic and phonetic representations and its graphic and orthographic distinctions (TOLCHINSKY, 2003; MOREIRA; PONTECORVO, 1996) and confront the relations between rules, norms and irregularities (NUNES; BRYANT, 2014, ZORZI, 1998, PINTO, 1998). On the other hand, this writing sub-system depends on its own text and transcription generation process (BEREITER; SCARDAMALIA, 1987; BERNINGER et al., 1994), which demands knowledge regarding orthographic, semantic, syntactic, morphologic, punctuation and accentuation aspects, all of them tied with the articulation and concatenation between words and sentences in the written text.

Many studies about the acquisition of orthography adopt quantitative methodologies that analyse the orthographic understanding of the child from the analysis of pseudo- or invented words (CASSAR; TREIMAN, 1997; REGO; BUARQUE, 1999), spelled words, short phrases (ROSA; NUNES, 2010), short texts (NOBILE; BARRERA, 2009), known texts that were rewritten (MOREIRA, 1996), or evaluation tests (NUNES; BRYANT; BINDMAN, 2006). Generally, these and other studies adopt different instruments of data collection such as dictations (CASSAR; TREIMAN, 1997; ZORZI, 1998), interviews (DE GAULMYN; LUIS, 1997; MORIN, 2005) or simply collections of texts written in the classroom (CARRAHER, 1985; CAGLIARI, 1989).

However, as Chiss and David (2011b) argue, if writing is a cognitive activity characterised by a series of parallel operations (HAYES, 1996), then the learning of orthography should also be further investigated during processes of textual production. The apprentice writer, while producing a text, simultaneously resorts to multiple information relating to the content, form and function of what is being written, seeking to solve diverse and heterogeneous problems. Their action as a writer is framed by their linguistic knowledge and cognitive capacities such as long term and working memories (KELLOGG, 2001; MCCUTCHEN, 2000; FAYOL; MIRET, 2005), which are activated during the writing process.

Spelling problems (SP), especially in novice writers, run parallel to other components of the writing process, with textual and linguistic questions, making it impossible to write a text that simply reflects on spelling problems. It is also not possible to anticipate at which point of the text the student will have doubts about the way a certain word is written, or in how a spelling problem would relate to other sub-systems or linguistic levels (for example, orthography and the use of the upper-case letter, the division of a word at the end of the line [translineation], grammatical or lexical-semantic knowledge to write homophone words). Thus, to better understand the development of textual production in novice writers and the relation with its inherent components, it is legitimate to value real writing situations, and offer mechanisms to identify those components during the text in course.

In this study, we shall approach that issue, discussing the recognition of spelling problems by a couple of newly literate students, during the writing processes of fictional narratives.

Acquisition of orthography and metalinguistic reflections

Since the 1990s, studies in the acquisition of orthography began to analyse explanations, justifications and comments of novice writers, using interviews or exercises of text production as a methodological instrument. The work of the Linguistics of Writing and Acquisition Research Group (LÉA) is representative of such studies and has influenced several other investigations.[1]

Among those investigations, we may highlight studies based on the "metagraphic interview" technique.[2] This technique was adapted and used by David (2001, 2003), Morin (2005) and Barbeiro (2007). Those studies adopted differentiated methodological procedures, which, nevertheless, allowed analyses of verbalizations of orthographic questions from young writers. David (2001), for example, showed some typologies of what he named "meta-graphical dialogues" established from interactive sequences with the participation of two or three children, in which one of them revises the other's text, highlighting spelling problems, in the presence and, sometimes, with the participation of, the investigator. The author describes types of components that would, generically, characterise the comments made during the interactive situation: justification x revision, agreement x disagreement, facts x comments, explicit participation x implicit participation, decision x non-decision.

In another type of experimental situation, based on the technique of the individual interview, Morin (2005) presented a list of words in French with specific spelling difficulties and asked 7-year-old students to write them down, questioning them afterwards about the written production. His findings confirmed former studies, demonstrating a variety of metagraphic comments related to phonological and morphological aspects and visual memory. They also indicated the importance and interference which formal instruction had in the way of thinking of these novice writers.

Similar results were also obtained by Barbeiro (2007), who proposed a collaborative writing assignment, in groups of 3, with 4 different levels of education (2nd, 4th, 6th and 8th grade), where each student of the group should write their own text (a fictional report), from a collective oral construction. This is one of the few studies which analysed the recognition of spelling problems that occurred during the production of written texts in real time, in conditions that were closer to the didactic practices of the classroom. For this reason, in the typology of identified spelling problems, the author considered the use of upper-case and lower-case lettering, translineation (hyphenation) and accentuation. Furthermore, Barbeiro proposed 4 categories for the analysis of the types of episodes: manifestations of auto-correction, hetero-correction, indication of

---

[1] Take a look at, for example, the work of Bousquet et al. (1999), Jaffré (1995), Chiss and David (2011a), Morin and Nootens (2013), Fijalkow, J. e Fijalkow, E (1991) and De Gaulmyn and Luis (1997).

[2] The metagraphic interviews conducted by the researchers and inspired by Piaget's clinical interviews are centred on spelling. In those interviews, according to Brissaud and Cogis (2011, p.45), "it is asked that the students comment, from the text they just wrote, on some of the graphemes they produced. The student, reflecting upon their choices, is then asked to explain the linguistic materials they chose and the reasons behind the choices. As with the words metalinguistic and metacognition, the term metagraphic explains the reflexive return, in effect in the graphemes".

the orthographic form in advance, help request from colleagues to solve a spelling difficulty (BARBEIRO, 2007, p.117).

That typology seems to be more interesting than the one proposed by David (2001) and Morin (2005), mainly because it considers orthographic forms "by the anticipation of occasional difficulties by colleagues" (BARBEIRO, 2007, p.118). In his results, this type of episode was not the objective of the analysis, which was centred on the quantitative aspects of the inaccuracies (faults in transcription, oral transcription, non-observation of morphological and phonological rules, lexical problems, accentuation, lower/upper-case lettering, graphic unity, translineation) and on the types of behaviour of the students (if the student self-corrected himself/herself, if they corrected the other student, if he/she anticipated a difficulty another student might have, if they requested the help of another student to solve a difficulty). Another result obtained in that study concerns what Barbeiro called "justification", namely, "verbal exchanges between elements of the group, to seek the adoption of a solution. [...] argumentation relative to spelling difficulties, to justify indications, answers to requests, self- or hetero-corrections" (BARBEIRO, 2007, p.120). Only 25 occurrences of "justification" were identified, related to the different spelling problems. The use of audio recording for data collection did not allow us to know exactly what was being written while the students were speaking. As we shall observe below, the author also did not explore the arguments used by the different writers and how their "justifications" differed.

S25: à noite/ vá/ "à"/ now its "á with heich"/ isn't it?/ "à" noite[3]

S27: "existe" noite?

(BARBEIRO, 2007, p.121)

According to Barbeiro (2007), this is one of the 6 registered episodes in which the students sought to differentiate the uses of 'à' (to the) and 'há' (there is). Even though it indicates the importance of the shared action between the students regarding metalinguistic reflection (")justifications"), there was no concern from the researcher to differentiate the types of arguments in relation to the students' level of education, neither did their methodology allow correlating the oral identification of the SP with the moment of its inscription (that is, whether the comment was made before, during, or after writing a certain word).

Despite the importance of the comments[4] made by the students, to understand the way they observe, think and solve (or not) certain spelling problems, we still do not have studies dedicated to the analysis of the arguments used, their relations with the text in course and the didactic practices to which the students are subject.

---

[3] The transcription provided by Barbeiro (2007) is not immediately understood by the reader. Here, the student is demonstrating doubt about how to graph "à", in the syntagma "à noite". They make an affirmative question saying that the "á" is with an "h". The utterance from the other student brings the information that refers to the semantic value of "há" ("there is"). However, it is not clear whether the student wrote the "a" with an acute or grave accent.

[4] We understand the term "justification", due to its characteristic semantic weight, not to characterise what the writer spontaneously speaks to their colleague. We propose the term "comment" (CALIL, 2016) given its comprehensive semantic scope, which can be explanatory, justificative, or a brief doubt or observation about the linguistic or textual element that is being written.

From a microgenetic perspective, within the field of study of Textual Genetics (FABRE, 1990; BORÉ, 2010; DOQUET, 2011), within an enunciative approach (CALIL, 2012a, 2013), we propose to analyse the comments made by young writers during the manuscript in course. Our analysis unit is the dialogue text (DT) co-enunciatively established during the paired writing process. We identified the DT from the emergence of a recognition and return from the writers in relation to a determined textual element, namely, everything that might potentially belong to the material composition of a text, be it the choice of the character's name, of an upper-case letter, of an accentuation mark or just a line to separate or divide a word. In the text in course, that recognition and return is not operating at every instance, but when it occurs on a certain textual element, it attaches to it the status of "object" or, as we have been calling it (CALIL, 2016), transforming the textual element into a "textual object". What the writers say regarding a recognised textual object (TO) is treated as a "comment" that refers to this object. The enunciated comment may contribute to the maintenance, modification or erasure of the referred TO, changing the text in course and, at the same time, tracing its textual genesis. It is for this reason that the relation between the recognition of textual objects and the comments that relate to them is treated as a "commented oral erasure" (CALIL, 2016). The commented oral erasures (COE) represent points of tension in the flux of writing, relative to the inscription and linearization[5] of one term or another, involving different linguistic and textual levels (pragmatic, graphic, lexical, semantic, orthographic, syntactic, of punctuation). The comments related to the recognised TO expose arguments of different values, as already indicated in other studies (CALIL, 2013, 2016).

In this study, we shall identify the COE of orthographic textual objects (spelling problems) and analyse the comments that refer to them. In our theoretical and methodological proposition we shall thus consider the student's spontaneous speech in the classroom's co-enunciative and ecological[6] conditions and the process of inscription and linearization of the text in course.

Scope of the spelling problems using the Ramos System

It is not a simple challenge to obtain information about the way novice writers think during real situations of textual production in a classroom context. Those who

---

[5] As was defined in Calil (2016), in line with the genetic perspective which is the scope of this study, these two terms are directly related to the moment when the text is being registered on the sheet of paper (letters, lines, strokes, drawings). The "linearization" concerns the graphic materiality of what the students have orally discussed to write, that is, the moment when a textual element (graphic, orthographic, lexical, semantic, syntactic, textual) is concatenated and articulated to another textual element in the syntagmatic chain of the text in course. Accordingly, we shall assume that linearization results from the dialogue, which necessarily involves the articulation and concatenation between the textual elements inscribed on the sheet of paper, whereas the inscription of a graphic mark may occur with no relation to the linearity of the text in course.

[6] For the reader unfamiliar with the specificity of our field of study, it is important to specify that this term refers to the preservation of the classroom's daily and interactive characteristics. This is guaranteed by the way the data is registered with the Ramos System, as shall be described ahead.

have or have had professional experience as teachers know how much the classroom is a dynamic, interactive, and dialogical space, especially when dealing with 7-year old students, whose experience as producers of text is still very incipient and the doubts about multiple aspects of writing (spelling, line limits and margins, orthography, separation of words, translineation, accentuation, punctuation, among various others) are still at the surface. Adding to those difficulties, it is not enough to have access to what the student thinks, the way they think what they think, but also to what they are thinking *while* writing.

Considering the importance of collaborative writing to the metalinguistic activity of the student (DAIUTE; DALTON, 1993, SWAIN; LAPKIN, 1998, STORCH, 2013), we elected the proposition of textual production in pairs as a privileged object of study to access the spontaneous way the student thinks. To this end, we accompanied the development of a didactic project, conducted by a teacher from Portugal, of the 2nd grade of elementary school, which involved the textual production of invented stories. We applied the same methodological design, with an ethnographic and ecologic orientation, which had already been adopted in previous studies (CALIL, 2008, 2009), asking Brazilian students, grouped in pairs, to textually produce a single text. Using the Ramos System,[7] 6 proposals for the production of invented stories within a paired group (B and L, female, 7 years old) were recorded. In each proposal, one student would alternately dictate (dictator), while the other would write (writer). This way, each one of the students wrote 3 stories and dictated the other 3 stories.[8]

Furthermore, we collected a set of complementary materials (the school's curricular proposal, an interview with the teacher, didactic materials adopted and used, students' notebooks, questionnaires to parents, photos of the school and of the classroom) so that we would be able to have the characterisation, contextualisation, and description of the school's daily reality and the educational contents that were valued in the didactic practices.

The material collected offers, on the one hand, information directly related to the writing process: the real-time writing process[9] and the product of this process (school manuscript). On the other hand, it offers information about the context and didactic practice of the classroom. We were able to observe at least two aspects: 1. What was written and spoken by the students during the text in course; 2. What was taught by

---

[7] The Ramos System, developed by the Laboratório do Manuscrito Escolar (CALIL, 2016), allows for a multimodal capture of the writing process in real time (in a classroom). Through the synchronisation of the filmed record of the interaction between the students, the tracing of the pen over the paper sheet and the audio register of the dialogue between participants, during the moment of the manuscript in course, the Ramos System provides a large quantity of information that grants access, in real time, to what the students say while combining, inscribing and linearizing the graphic material of the invented story.

[8] Despite this instruction, exchanging pens was not forbidden. The teacher was attentive in order to direct the attention between the students and the division of the task in pairs, avoiding that a single student would be the sole responsible for creating and writing the invented story. It is worth noting that, in order to preserve the ecological characteristics of the international dynamic between teacher and students, all other researchers and auxiliaries would leave the classroom during the execution of the task. This favoured the preservation of the interaction between teacher and students.

[9] The Ramos System generates a synchronised film of the writing process of each participating pair.

the teacher during the weeks leading up to the accomplishment of the proposals of text production. In this article, we shall take the first aspect to be the privileged object of analysis, considering:

- In the school manuscript (textual product), what was inscribed and linearized:
a. Words written orthographically.
b. Written words with spelling problems, but with no erasure marks.[10]
c. Marked words or letters, indicating the students' recognition of spelling problems.
- The manuscript in course (textual process), what happened during the inscription and linearization:
d. Tension points[11] related to the spelling problems recognised by the students, regardless of whether those problems were erased or not, or if they occurred only orally, but with the correct inscription of the word.

If we had only the textual product as an object of study, the student's recognition of the SP could be indicated with an erasure, but we would not know what the student was thinking while erasing. If what the students talked about during the moment they were writing had only been registered in audio, as in Barbeiro (2007), we would not know exactly what was being written while they were speaking. From an empirical point of view, the comments made by the students do not seem to us to be equivalent, during a metagraphic interview conducted by the researcher and the spontaneous comments, enunciated unpredictably, between the students, *during* the production of a text. With the synchronized film, we do not only have a dimension of the simultaneity between what was spontaneously spoken and what was inscribed, but we are also able to know, in real time, where the students were looking or pointing at, what the teacher was doing, with whom she/he was interacting at the time the dyad students were writing, and what she/he was saying to her/his students.

Thus, there are innumerable possibilities of interpretation. Regarding our question, we can have an erased SP, but with no explicitation of what the student thought while recognising and erasing the SP indicated in the manuscript. We can have an inscribed SP with no erasure, but that generated reflections between the students during the moment of linearization of a word. We can also have SP that was recognised through what the teacher or another student said. And, finally, we can have a SP that is erased

---

[10] As shown in the pioneer works by Fabre (1990) and many other researchers that followed, erasure in the school manuscripts marks the recursiveness of writing and indicates metalinguistic reflections which are made effective by the writer. From the point of view of the textual product, it is only possible to identify recognitions of orthographic problems by the writer if they made some type of erasure in the text.

[11] The identification of those 'tension points' (CALIL, 2012a) considers the relation between the written record, in the time and space of inscription in the paper sheet, and the interactive and dialogical record made face to face, involving overlapping turns of speech, oral reformulations, body movements, facial expressions, eye direction, the teacher's interventions and interactions with other colleagues.

or not, accompanied by comments, identifying, justifying or making explicit the SP or the reason why the erasure was made.

Differently from studies on the acquisition of orthography, some of them quoted in our introduction, which elect as object of study dictations of words, pseudo-words, evaluation tests, sentences or texts, identifying and describing the errors, inaccuracies or orthographic deviations of the students, our double object (product and process) requires another type of interpretative approach to spelling problems.

A first observation refers to what we shall treat as a spelling 'problem'. For us, 'problems' are not necessarily 'errors', 'deviations', 'transgressions', 'incorrections' or 'difficulties'. In relation to the product, we consider a 'problem' to be any inscriptions that differ from the orthographic convention; an occurrence generally analysed by most studies on the acquisition of orthography, which take as object of study what the student effectively wrote. Meanwhile, in relation to the process, 'problems' are identified from the recognitions made by the students of orthographic textual objects and the comments that topicalize aspects relating to the way a word is spelled or how a certain sound is orthographically represented, even if its inscription onto the sheet of paper was not made effective or if it was inscribed correctly and without difficulties, a dimension rarely considered by studies in acquisition of orthography. Given the spontaneous nature of the students' speech while interacting face to face, sometimes these comments present arguments that contain repetitions, hesitations or ironic intonations, and they can be intersected by pauses or interruptions and accompanied by gestures, taking the pen and the sheet of paper, body movements and facial expressions complementary to what is being said.

In our analysis, this double dimension constituted by the process and the product in the interpretation of the 'spelling problem' shall be treated in an articulated and complementary manner, with the SP being intercepted from that multi-modal perspective.

A large quantity of studies present a broad range of spelling problems that were faced by students soon after they understood the alphabetic principle (CAGLIARI, 1989; ZORZI, 1998; CARRAHER, 1985, among others).[12] In general, we can basically place them into 2 groups, composed of different sub-groups:

- Lexical SP relative to how the word is presented in the dictionary:
  - Support in orality
  - Homophonic representations
  - M before B/P
  - Mute/sound exchange
  - Nasalization

---

[12] These and other studies in the areas of Linguistics and Psychology propose various, sometimes conflicting, classifications for the spelling problems in the Portuguese language. Here, we list some of them, in a general way, so that the reader may have an idea of the complexity of the problems faced by recently-literate writers while producing their texts.

- ○ Accentuation
- ○ Segmentation
- ○ Translineation
- • Grammatical SP linked to the variations of a word in relation to its context of occurrence:
- ○ Markings of male and female gender
- ○ Subjective concordance
- ○ Verbal concordance
- ○ Grammatical class
- ○ Morphological derivations
- ○ Verbal variations (conjugation, tense and verbal person)
- ○ Use of upper-case associated to the punctuation subsystem.

We do not intend to make a quantitative analysis of these types of SP, nor elect beforehand a specific type of SP to be analysed, such as those discussed in David (2008) or by Morin (2005) by selecting the problems according to specific criteria (frequency, mute letter, digraph, among others). Here, this list is used to situate the reader in relation to some of the possible SP recognised in the literature. The literature also indicates what are the most probable spelling problems the students encountered in their first texts. Notwithstanding the fact that we know that double consonants such as the "ss" or the "rr" or the digraphs "nh" and "lh" will be problematized by the students shortly after understanding the alphabetic principle (NUNES; BRYANT, 2014), the fact that the students are writing, in real time, an invented story does not allow us to predict the types of spelling problems that will occur and be recognised by the students. It is also not possible to know which orthographic aspects will be recognised by the students and which will not. However, the synchronised film allows us to identify if the inscription of a SP was preceded by any sort of comment from the students or if the comment occurred after the effective inscription of a certain graphic or orthographic form. In other words, we establish as a point of reference for the identification of spelling problems recognised in advance (SP-RA) the comments uttered in the moments that immediately preceded their graphic recording onto the sheet of paper. Inversely, we can also identify the spelling problems recognised subsequently (SP-RS): those that were identified by the students and commented on only after its graphic recording was performed.

In line with the dynamic of the dialogally-produced manuscript in course, which is subject to the unpredictability of spontaneous and co-enunciative speech, to the associative relations established during dialogue (CALIL, 2012b, 2016) and to the materiality of the linguistic elements concatenated and inscribed linearly in the sheet of paper, our case study shall have a qualitative focus, based on a micro-genetic analysis of the orthographic COE, in order to describe the students' anticipation of SP and the way these SP are solved (or not)..

Results and discussion

In the 6 school manuscripts written by B and L, a total of 634 words were written and we identified 157 SP. On average, 1 SP for every 4 written words. The graph below shows this relation per process.

Graph 1 – Relation between words x SP per textual process.
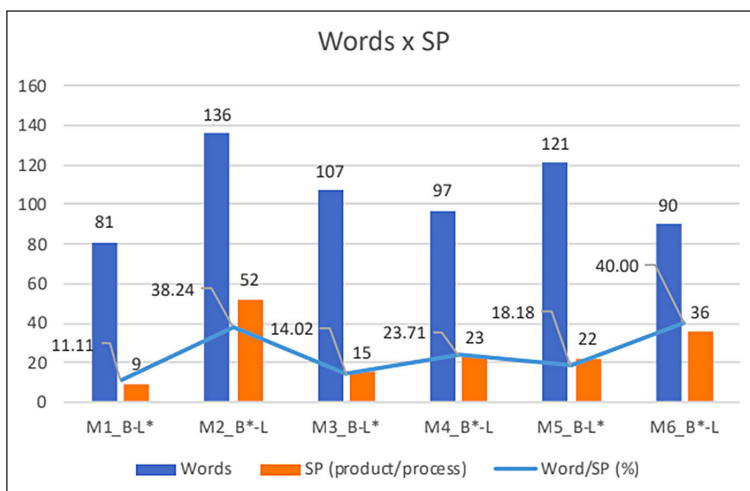
The difference between the quantity of words and the identified SP in each process and its respective manuscript seems to relate to the student that was responsible for writing. In the processes in which L wrote[13] and B dictated, we have fewer SP: Manuscript 1 (M1) = 11%, Manuscript 3 (M3) = 14%, and Manuscript 5 (M5) = 18%. In the manuscripts written by B, the SP were greater in number: Manuscript 2 (M2) = 38%, Manuscript 4 (M4) = 23% and Manuscript 6 (M6) = 40%. This difference suggests that L would have greater orthographic knowledge than B, producing, therefore, a lesser number of words inscribed with SP. We shall develop this question in another study, since the current analysis centres only on the early recognitions and their relation with the identified SP.

In order to explain and deepen this problematic, we present M2_B*-L, written by B. This was the school manuscript where we found the largest quantity of written words and also highest occurrence of SP. In order to identify the SP, we established the following codification:[14]

---

[13]   The information referring to who wrote what is indicated by an asterisk beside the first letter of the female student's name.

[14]   In this codification we do not separate the lexical SP from the grammatical SP. For example, SP48, in line 15, indicates a grammatical SP: "do dinosauros" [incorrect plural]. SP49 indicates a lexical SP: "dinosauros" [there is a missing "s"

- Numerals inside the geometric shapes: indicate all the identified SP not only in the product (school manuscript), but also in the process (the film of the text in course; this means that inscribed and non-inscribed SP were considered, namely when the word was written correctly, but recognised orally, from the comments made by the students.
- Rectangle: marking the SP identified in the product.
  - Red rectangle: SP with a visible erasure in the product.
  - Green rectangle: non-erased SP or with an erasure that is not visible[15] in the product.
- Circle: marking an SP identified in the process.
- Arrow to the right (⌒): SP recognised in advance.
- Arrow to the left (⌒): SP recognised subsequently.
- Orange arrow: SP commented by B.
- Blue arrow: SP commented by L.
- Numerals in L and B: indicate the quantity of SP recognised by the students.

Figure 1 – Manuscript 2 'The Snow White and the three dinosaurs' (06/02/2015), with identification of the SP.
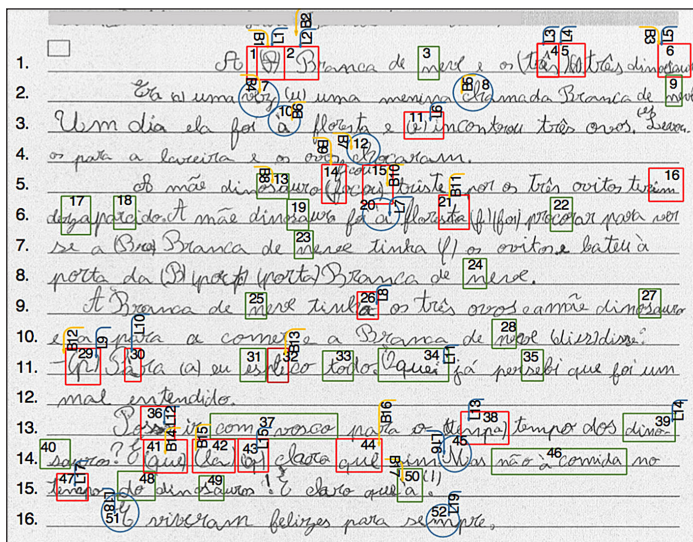
Source: School Manuscript Lab (LAME).

in the word "dinossauros"].

15 According to Textual Genetics, the "written" erasure may be defined as any modification or alteration relative to already linearized textual elements. Our methodological drawing, by synchronising the produced text and the text in course, allows for a retrieval of the visible erasures (those that are graphically marked on the paper, as for example, smudges, overwritten letters/words, strokes/lines, bars, etc.) and the erasures that are not visible (those that are not graphically marked, a result of adding a letter or punctuation mark, for example). In order to observe the relation between the SP and the erasure, we indicate with red rectangles the SP that had visible erasures and with green rectangles the SP that were not erased or to which were added erasures, often not visible in the product.

- In this manuscript, we have 16 lines and 136 words, we identified and numbered the occurrence of 52 SP, classified as follows:
- Erased SP (red rectangles): 21 occurrences
  - SP1, SP2, SP4, SP5, SP6, SP11, SP14, SP15, SP16, SP21, SP26, SP29, SP30, SP32, SP36, SP38, SP41, SP42, SP43, SP44, SP47.
- Non-erased or non-visible SP (green rectangles): 23 occurrences
  - SP3, SP9, SP13, SP17. SP18, SP19, SP22, SP23, SP24, SP25, SP27, SP28, SP31, SP33, SP34, SP35, SP37, SP39, SP40, SP46, SP48, SP49, SP50.
- SP that were commented on, but not erased (blue circles): 8 occurrences
  - SP7, SP8, SP10, SP12, SP20, SP45, SP51, SP52.
- SP recognised by B: 17 occurrences
  - SP-RA (orange arrow to the right): 6 occurrences
    - B1, B3, B4, B7, B8, B17.
  - SP-RS (orange arrow to the left): 11 occurrences
    - B2, B5, B6, B9, B10, B11, B12, B13, B14, B15, B16.
- SP recognised by L: 19 occurrences
  - SP-RA (blue arrow to the right): 2 occurrences
    - L16, L18.
  - SP-RS (blue arrow to the right): 17 occurrences
    - L1, L2, L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L17, L19.

These results provide interesting information about the produced SP and the students' ability to recognise them. There was a balanced distribution between the erased and, thus, to some extent, recognised SP (21 occurrences) and those that were not erased or did not present visible erasures (23 occurrences). Through this manuscript production process, we also identified commented SPs that were not erased in any way.

We were also able to observe that there is a strong relation between erased and commented SP. From the 21 erased SP, only SP16 was not accompanied by a comment. This suggests that the erasure, in the production of a text in pairs, strongly favours the oral enunciation of some form of metalinguistic reflection. Inversely, we highlight that only 3 of the 23 non-erased or non-visible SP received comments: SP34, SP39, SP50.

Another relevant aspect to be emphasised concerns the SP-RA and SP-RS. Both students recognised approximate quantities of SP. Student B recognised 17 SP, while student L made 19 recognitions. However, student B, responsible for writing the story, anticipated 6 of the 8 SP recognised in advance, while student L, responsible for dictating the story, made 17 subsequent recognitions. One of our working hypotheses is that such difference between the students would be related to the level of knowledge and information each one of them possesses, as well as to the role of dictator and writer taken on in the task proposed by the teacher. That would be another issue to be explored in future studies.

In order to better understand the anticipated and subsequent recognitions of spelling problems and the comments related to them, we shall present a microgenetic and enunciative[16] analysis of 3 words that were inscribed in the manuscript in course, in which their linearization underwent the incidence of erasures and/or were accompanied by orthographic comments enunciated by the students. We chose the SP related to the word "branca" (white), line 1 (SP1, SP2), the word "dinossauro" (dinosaur), line 1 (SP6) and line 5 (SP13) and to the word "há" (to have or exist), line 15 (SP50). We have chosen these SP given the different ways of correlation between the erasures and the comments.

During the inscription of "branca" a visible erasure and four comments occurred (SP-RA B1, SP-RS B2, SP-RS L1, SP-RS L2). The word "dinossauro" was inscribed 5 times throughout the school manuscript. In the first occurrence (SP6) there was an erasure, accompanied by two comments (SP-RA B3, SP-RS L5). In the second (SP13), there was no erasure, but there was an anticipated comment (SP-RA B8). In the other three occurrences of this same word there occurred neither erasures nor comments. In the word "há" (SP50) there is no erasure mark in the product, even though an anticipated comment (SP-RA B17) was identified in the process.

SP with RA in the representation of the syllable "bran" of the proper noun "Branca de Neve"

The two first SP recognised in this process were also the first SP that were erased in the manuscript.

Figure 2 – Stroke across the letter 'B' of the word "Branca" (line 1, SP1, SP2).
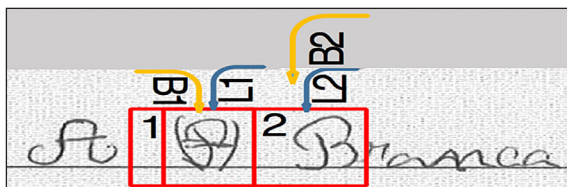
Source: School Manuscript Lab (LAME).

Indeed, the identification of the erasure of the letter 'B' does not explain why the student made such an erasure. Any assertion as to why or what the motive was for the student to make the erasure is mere speculation. The manuscript in itself bears no evidence to this. Through this identification, it is not possible to know whether the student faced a spelling problem or if it was a graphic difficulty, relative to the incomplete

---

16 This type of analysis involves the dialogic and genetic character of the manuscript in course, occurring over the tension points identified in the textual flux.

tracing of the upper-case letter 'b'. Even if we suppose it to be an SP relative to the use of upper-case letters, we still can say nothing about how the student interpreted it. In a similar way, a researcher basing her/himself only on the way the word "branca" was inscribed might suppose, unwittingly, that there was an erasure over the letter "n". There was not. When the student wrote it, she made the first "leg" of the letter "n", lifted the pen from the paper sheet and when she came back to writing, started inscribing the letter "n" once more, moving the pen over the initial tracing of the word. This way of writing the letter "n", recorded by the Ramos System, isn't even characteristic of a graphic erasure, which might indicate a hesitation about the shape of the letter. There was also no form of comment (which is the object of our study) during the moment that letter was inscribed. For this reason, we do not mark this point of the text in course as associated with the representation of the nasalization or with any other type of SP. This is a good example of how a researcher might make erroneous inferences by accessing only the textual product.

The synchronised film generated by the Ramos System gives another dimension to the written erasure that was produced in "branca". From our theoretical and methodological input, we can say that in this tension point of the writing process, as well as the visible erasure, we have the occurrence of comments that relate to the word, characterising what we have been defining as a commented oral erasure (CALIL, 2016). As we shall demonstrate, the oral erasure evidences that the TO was brought about by a spelling problem orally recognised in advance by student B, before being inscribed. This occurrence happened right after the article "A" was inscribed, seconds before beginning the inscription of the name "Branca", which would constitute the story's title. We may observe in the dialogue below that the comments generated pertain to the recognition of 3 problems directly related with orthography:

- Problem with the orthographic representation of the consonant group "br" at the beginning of "branca".
- Problem with nasalization, orthographically represented by "an".
- Problem with the use of the upper-case letter for the proper noun.

There is still a problem on a graphic level, affecting the interpretation made by one of the students, as we shall also see next.

Figure 3 – L pointing to the letter 'B' of her name;
B looking to where L is pointing at.

Source: School Manuscript Lab (LAME).

DT1: "Branca" (00:28:24:06 – 00:29:16:00[17])

213. L: (Asking B to continue writing) Come on! (B is finishing writing the letter 'a' [A] in the title and L reading) ...aaaa... (Then, dictating) ..…...Bran-ca.

214. B*: (Stops writing and is repeating 'branca' in a doubtful tone) *Branca...* (Babbling and murmuring) *...bran... bran...*[B1]

215. L: (Repeating, with an emphasis on the 'bran') *...Braaan-ca ::* (L, wearing the badge where her name and surname is written, points to the letter 'B', the first letter of her surname). *You want to look to hear ...*[L1a]

216. B*: (Whispering.) *...bran... ca...* (B turns to L and looks at where L was pointing, her badge. B looks at the letter 'B', which was being pointed at by L. B goes back to the paper sheet and continues writing, beginning the upper-case letter 'B' [B]. Halfway while tracing the line for the upper-case letter 'B', B interrupts and says in a tone of disapproval) *Ei...* (Erasing the incomplete capital letter 'B' [P̶] she had started writing.) *I wrote it with a capital.* (Making a parenthesis, as the teacher showed, to indicate that the letter 'B' was erased [(P̶)])[B2a]

217. L: *Alright then.* (Making with her index finger the letter's trace across the table.) [L1b]
218. B*: (looking at L) *It is lower-case.* [B2b]
219. L: Yeah it is. (Realising her mistake) *No. No. It is upper-case, as well (laughing). It is also… with upper-case. You write it with a capital letter.* [L2] (B is finishing making the parenthesis and tracing again the letter 'B' in upper-case [B]. L attentively accompanies the tracing made by B) A... bran.... ca...
220. B*: (Writing the rest of the word 'branca', lifting the pen at the beginning of the letter "n" e putting it once more over the paper sheet to keep tracing the word.) Bran... [ran] ...ca [ca]...
221. L: ...ca...

In this DT1, which lasted less than 60 seconds, B's hesitation in turn 214, which is marked by her doubtful tone while repeating the word 'branca' (white), anticipates the SP that refers to the way the sound [brã] of the first syllable of the word should be graphically represented. The way B stammers, murmurs and repeats the syllable 'bran', accompanied by her facial expression, indicates that she recognises problems related to the orthographic representation of that syllable. At the same time that her enunciation (B1) anticipates and recognises the SP to be inscribed, she also expresses doubt about how 'bran' is written, probably an outcome of the consonant group 'br' and the nasalization.

In turn 215 we have the occurrence of L1. This enunciation from L indicated how the student interpreted B's earlier enunciation and action. She understood B's hesitation, not as a doubt relating to the encounter of consonants or to how the nasalization is represented, but as a problem related to the way the capital letter 'B' is inscribed. The evidence of L's interpretation is in the almost enigmatic enunciation L1a ('you want to look to hear…'), along with the gesture pointing to her own badge (Figure 3), pointing at the first letter of her surname, a capital 'B'.

In the video, there is no oral or gestural indication that B was asking about the way the letter 'B' was written. Furthermore, it is highly unlikely that B would not know how to inscribe the first letter of her own name. The description accompanying turn 216 describes B's gesture and movement, indicating that the student began writing the capital letter 'B' without great hesitation and no comment.

However, before concluding the tracing of the letter, B interrupts her gesture, making, in two turns of speech, two simple and inter-related comments (B2a and B2b), with the same argumentative value. Still in turn 216, she recognises as an "error" having started writing the letter in its upper-case form, erasing the upper part of the letter she had started writing. While she is doing it, she comments in a tone of disapproval: 'I wrote with a capital letter' (B2a). L reacts, insisting on the graphic question of the capital letter 'B', drawing an imaginary stroke on the table (turn 217, L1b). In turn 218, B once again denies that 'Branca' should be written with a capital letter, firmly stating:

'It is lower-case' (B2b). B's misconception by thinking that the 'B' in 'Branca de Neve' is written in lower-case, producing the erasure that stems from this interpretation, is corrected by L (L2), in turn 219; 'No. No. It is upper-case, as well. (Laughing.) It is also with... with upper-case. It is written with a capital.'

We have in this DT1 a tension point around the way the syllable "Bran" is written, which we consider to be a TO anticipated and recognised by B. Over this TO concur simultaneously, at least, three spelling problems and one graphic problem. Strictly speaking, in B1 (turn 214) we have an anticipation related to the problem of orthographic representation of the sound [brã], even though it isn't possible to know whether B questions herself about the orthographic representation of the consonant group 'br' or the orthographic representation of the phoneme /ã/. The replica contained in L1a and L1b about the tracing of the letter 'B' was not directly related to the B's SP-RA. In the same way, the SP relative to the use of upper-case (B2a and B2b) was also not related to that SP-RA. The recognition of that SP (upper-case) occurred after the beginning of the inscription of the letter 'B', generating its erasure and being solved by the comment marked in L2.

The arguments presented by these 4 comments (B1, B2, L1, L2) are a consequence of the TO being inscribed and linearized (the name of the character). However, the difference in value between those arguments (where the themes are the consonant group or nasalization and the graphic form of the letter 'B' or use of upper-casing) evidences the parallel functioning of different problems faced during the writing process, imposing simultaneously and in a concatenated way several questions for the recently literate writer to solve.

In relation to the problematics of the use of upper-case in names, it is worth observing the students' incomplete orthographic knowledge. This knowledge is based on the visual and graphic aspect of the word 'Branca', and not on the knowledge regarding the grammatical class that determines its correct form. In all of the six occurrences of the proper noun 'Branca de Neve' (Snow White), the word 'branca' was inscribed with a capital letter, while the word 'neve' (snow) was inscribed in lower-case.

SP with RA in the representation of the phoneme /s/ in 'dinossauro' (dinosaur)

Two other SP recognised in advance by B refer to the representation of the phoneme /s/ in the word 'dinossauro' (dinosaur). In the six inscriptions present in the manuscript, the digraph 'ss' for the phoneme /s/ in 'dinossauro' was represented by just one 's'. In its first occurrence we identify SP6, which was anticipated by a comment (B3), followed by the inscription of an erasure of substitution and by a subsequent comment (L5).

Figure 4 – Erasure over the grapheme 'c' and substitution
by 's' in the word 'dinossauros' (line 1, SP6).

Source: School Manuscript Lab (LAME).

In the first inscription of the word "dinossauro" (Figure 4), SP6 occurred. In this SP we can easily identify an erasure of overlap, indicating the competition between the grapheme 'c' and the grapheme 's' for the orthographic representation of the digraph 'ss'. If such an erasure were analysed only through the product we would not have clear evidence about which grapheme was inscribed first. Even though we may suppose that there was the initial inscription of the grapheme 'c', so that the distinction between the phonemes /k/ and /s/ in the syllables 'ca' and 'sa' may be more easily observed by the child, we do not exactly know what was thought by the student when she wrote it, what was the relation she established between the 'c' and the 's' and what led her to observe the difference, thus avoiding representing the phoneme /s/ with the grapheme 'c' in this position.

In SP13, line 5, we have the inscription of the same word 'dinossauro', however it did not produce an erasure, even though it was accompanied by an other comment indicating the spelling recognition in advance (B8) of this same type of SP.

Figure 5 – Spelling of 'dinosauro' for the term 'dinossauro' (line 5, SP13).

Source: School Manuscript Lab (LAME).

In the second register of the word 'dinossauro' (Figure 5), just like in the other four inscriptions of this term (SP19, SP27, SP39, SP49), the representation of the digraph 'ss' with just the grapheme 's' is continued, with no erasure mark that might suggest doubt or hesitation in spelling.

A study by Monteiro (1999) suggested that the contextual rules involving the cases of representation of the grapheme 'ss' for the phoneme /s/ are initially represented by the child with a single letter. It should be noted, however, that problems of representation related to the phoneme /s/ go far beyond these contextual rules, since their homophonic character involves multiple representations. The complexity required for the mastery

of the orthographic representations of this phoneme is evident in the student's own manuscript. We either have the correct use, such as in for example, 'três' (line 1), 'vez' (line 2), 'dezaparcido' (line 6), 'disse' (line 10),[18] 'posso' (line 13),[19] 'felizes' (line 16), or an incorrect representation of the phoneme /s/: 'dezaparcido' (line 6), 'esplico' (line 11), 'persebi' (line 11).

In relation to the first two occurrences of the term 'dinossauro', the respective dialogue texts show the anticipation of the spelling problem related to the way the phoneme /s/ is represented, during the linearization of the syllable "ssau".

Figure 6 – B writing the letter 'c' for the 'ssau' syllable.

Source: School Manuscript Lab (LAME).

DT2: 00:30:06:08 – 00:30:23:21

254. B*: Dinossauros... (Writing) ...di [di]...
255. L: ...nooo...
256: B*: ...no [no]... *...ssauu... o cê-só...* (tracing the letter 'c' [c]) (L looking at what B is writing and speaking along with her) *...ssaaaauu...* [B3] (Interrupting the tracing of the letter 'c', indicating doubt about how to write the 'ssau' syllable)
257. L: (Seeing that B wrote the letter 'c') *It is not with that one*[L5].

---

258. B*: (Overwriting the letter 's' on the letter 'c' and making the rest of the syllable 'ssau' ...dino... [e̶ sau]) ...ssau...
259. L: ...ro...
260. B*: ...ro [ro]...
261. L: The dinossauros.

[...]
DT3: 00:40:05:00 - 00:40:28:19

414. B*: (Finishing writing "the mother") ...dinossauro... (Writing 'dinossauro'.) di [di]... no [no]...
415. L: Di-no-ssau-ro.
416. B*: Di-no... *'ce' - 'sapo'... ssau!*[B8] (Writing) Sau [sau]... ro [ro]... the mother dinossauro... (Turning to L) And the mother dinossauro... it is written.

Both DT bring similar comments for the anticipation of the same type of SP about the same word. In the first SP (SP6) we have comment B3: 'o ce-sô… …ssaaauuu…'. In the second SP (SP13 B8), we have: 'ce-sapo...sau'. These enunciations from B suggest that she is analysing and trying to solve the orthographic representation of that phoneme, which has an intense and complex concurrence in the Portuguese language system, involving the graphemes 's', 'c', 'ç', 'x' and the digraphs 'ss', 'sc', 'sç', 'xc', and 'xs'.

L's brief comment in L5 (turn 257) eliminates B's hesitation, assuring her that the correct form is not with the letter 'c'. Her comment "It is not with that one," even though it does not verbalise which is the correct grapheme, indicates, on the one hand, some metalinguistic knowledge about the different graphemes used to represent the phoneme /s/ and, on the other hand, shows her acting as a reviser of the manuscript in course, contributing to generate the erasure of the 's' over the 'c', at 00:30:17:17. We observe that the revising role assumed by L is associated with the shared attention and directed by the visual aid (eye direction) of what is being written. This aspect seems to have an important role in the shared writing of students of this age group.

Even though the inscription of the word 'dinossauro', with a single 's', was kept in all the occurrences of this word, and that the SPs, in which these graphemes and digraphs used to represent the phoneme /s/ concur, are widely recognised by studies on the acquisition of orthography (NUNES; BRYANT, 2014; ROSA, 2004, MONTEIRO, 1999), the two brief comments made by B are revealing of the way she thinks and solves this tension point. Firstly, as she had already done when faced with the SP in "bran", B anticipates aloud the SP, as if she was speaking and recognising a difficulty to herself. At the same time, such anticipation is accompanied by an attempt of resolution, which is revealed by her spontaneous speech: turn 256, '...no [no]… …ssauu… o cê-só…' and turn 416 'Di-no… 'ce' - 'sapo'… ssau.'.

By highlighting the syllable, breaking it down and associating the first letter to the word "sapo" (frog) (B8), B retrieves from her long-term memory part of the complexity involving the representation of the phoneme /s/, making explicit, in turn, content that was lectured in the classroom. In the months prior to the proposal of textual production (October, November, and December), the teacher had emphasised in the didactic work the teaching content relative to simple syllabic families. Among them, the syllables of the 'ca' and 'sa' family are intensely emphasised, as attested by many of the exercises registered in the school notebooks of the students, which were intensely worked on.

Figure 7 – School exercises involving the syllabic families of 'sa' and 'ca'-

Source: School Manuscript Lab (LAME).

The homophonic relation between the graphemes 'c', 'ç', 's', the digraph 'ss' and the phoneme /s/ is explicitly highlighted in these school tasks, presenting the respective syllabic families highlighted, sometimes in red, in the words "**sa**bonete", "**si**no", "**se**te", "**sa**ia", "**sa**nta", "**sa**pato", "**sa**po", "ta**ça**", "len**ço**", "a**çú**car", "**ce**noura", "bi**ci**cleta" (soap, bell, seven, skirt, holy, shoe, frog, bowl, tissue, sugar, carrot, bicycle). Even though this type of switch is common at this moment of childhood development, it seems to be difficult to deny the interference of the teaching content valued by the

teaching practise and the didactic material had on student B's way of thinking. By emphasising the occurrences of 's', 'c' and 'ç' in different "syllabic families", either by the identification of the initial letter, or by its occurrences in the middle of words, their homophonic relations are evidenced. B anticipated the SP while writing the word "dinossauro", using as a resource to solve the problem the emphasis of the first letter of the word "sapo", which was repeatedly emphasised in her school exercises and, probably, in her teacher's speech. To us, more important than describing the student's predictable 'error', is responding when the students identify the SP along the text in course and the way they deal with the linguistic information they have received in order to solve the identified spelling problems. B, by anticipating that type of SP, expresses her doubt in the representation of the phoneme /s/ in this position, indicating a meta-phonological recognition, which is still incipient, since the grapheme 'c' before the phoneme /a/ never receives the phonic value /s/. B's metalinguistic doubt, in turn, makes L verbalise a meta-phonological and meta-graphic analysis, differentiating the graphemes 'c' and 's' in the representation of the phoneme /s/.

SP with RA in the representation of the phoneme /a/ in "há".

As we know, the spelling problems of students in this phase of the process of alphabetization are many and diverse. And, most likely, the problems related with the homophonic relations are the most frequent, as indicated by Zorzi (1998), Nunes and Bryant (2014) and other researchers. The appropriation of these relations by the young writer seems to depend on different interrelated aspects:

- Information received through social interaction (teacher, parents, classmates, didactic material).
- Frequency of occurrence in the student's reading material.
- Cognitive ability related to long-term memory and the students' working memory in order to understand and retrieve the received linguistic information (KELLOGG 2001, MCCUTCHEN, 2000).

These aspects seem to occur in the early recognition that shall be discussed below.

As indicated in Figure 1, this school manuscript was produced on 6 February 2015. On 26 January, the teacher worked with the students on the difference between 'ah', 'há' and 'á', as shown in Figure 8.

Figure 8 – Class exercise about the differentiation of 'ah', 'há' and 'à'.
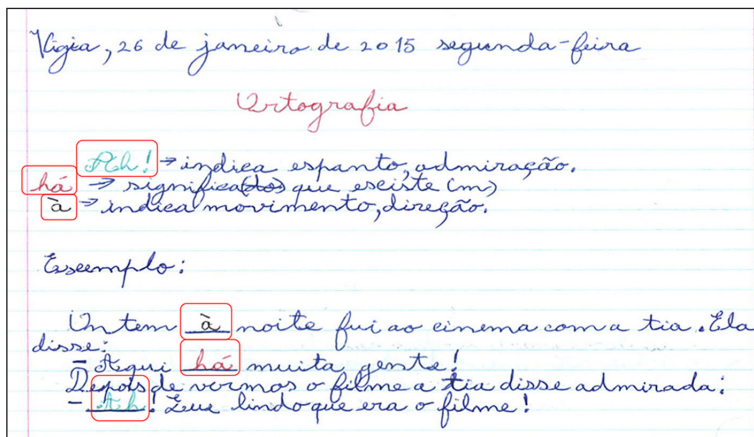
This orthographic[20] task concerned 3 forms of manifestation of the phoneme /a/, present in 3 grammatical categories: 'ah' (exclamatory interjection), 'há' (3rd person singular of the present tense of the indicative of the verb 'haver' (to be/to have), and 'à' (contraction).

In the story 'Snow White and the three dinosaurs,' produced 10 days after this task was completed in the classroom, the orthographic form 'à' was inscribed in six points of the manuscript:

1st. Line 3, SP10: 'Um dia ela foi à[B6] floresta.' ('One day she went to the forest')

2nd. Line 6, SP20: 'A mãe dinossauro foi à[L7] floresta...' ('the mother dinosaur went to the forest)

3rd. Line 7-8: '...e bateu à / porta da Branca de Neve.' ('and knocked at / the door of Snow White)

4th. Line 11, SP30: 'Pàra[L10], eu explico tudo.' ('Stop, I will explain everything')

5th. Line 14-15, SP46: 'Mas não à comida no / tempo dos dinossauros!' ('But there is no food in the / time of the dinosaurs')

6th. Line 15, SP50: 'É claro que à[B17].' ('Of course, there is')

There are no visible erasure marks in any of these six inscriptions of the accentuated grapheme 'à', which would also lead us to suppose that the students did not recognise

---

20   It is not the objective of this study to analyse or discuss the school exercises that were proposed to the students. From an ethnographic and ecological point of view, the presentation of these documents has the function of describing the didactic context, so that it enhances access to the received information, correlating them to what is said and written during the text in course.

any orthographic problem upon their inscription. In the first three inscriptions, located in lines 3, 6, and 7, the use of the 'a' with a contraction accent is orthographically graphed, a suprising fact, given that the correct use of the contraction does not seem to be common in texts by Brazilian students of this age. In Brazilian Portuguese, the most likely construction for those enunciations would be: 'ela foi na floresta' (she went in the forest), 'ela foi para a floresta' (she went to the forest) or 'ela bateu na porta' (she knocked on the door). At the 4th point the occurrence of 'à' in the first 'a' for 'a' ('to') is surprising. All the comments identified in SP10 B6, SP20 L7 and SP30 L10, respectively, 1st, 2nd, and 4th points, were subsequent and for this reason we shall not analyse them in this study.

In the 5th point, even though there is a SP (SP46), there was neither erasure nor comment, which indicates that the students did not recognise it as a SP. In the last inscription of 'à' (6th point, line 15, SP50) there was the early recognition B17, which we shall discuss next.

Figure 9 – Inscription of 'à' (SP50) for the verbal form 'há' (haver - to be/ to have) in the phrase 'É claro que há' (Of course there is) (line 15).
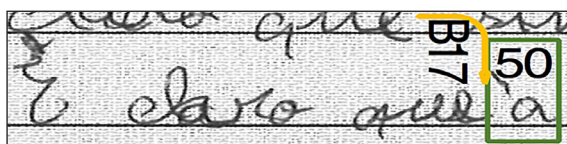
Source: School Manuscript Lab (LAME).

Figure 10 – Moment when B gestured on the table,tracing the grapheme "a" with a grave accent.
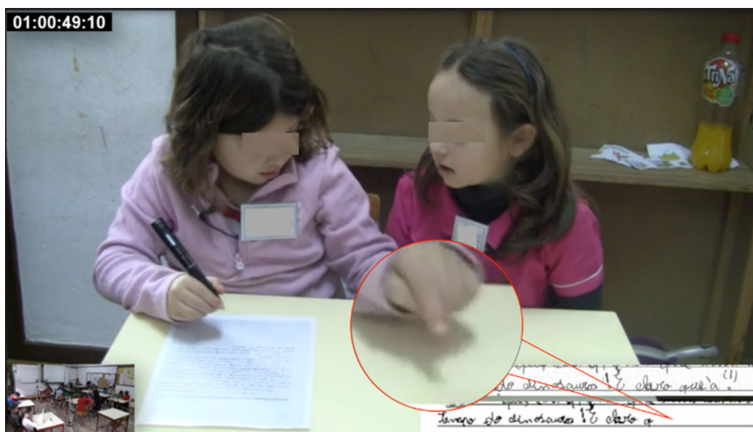
Source: School Manuscript Lab (LAME).

728. L: (B finishing writing 'É claro', in line 15. L dictating the continuation 'que há'.) ...que...

729. B*: ...que [q]...

730. L: ...háááá.

731. B*: ...*háá?* (Stops writing, turning to L with a doubtful expression and making a gesture with the index finger, tracing over the table the letter 'a' and the grave accent; speaking in a low voice) ...*de 'à à à'...?* [B17a]

732. L: (L, not seeing the gesture made by B on the table, and not seeming to have heard her question; L dictating and repeating what she said before.) É claro que há. Que há! (B makes the gesture once more, drawing with her index finger the letter 'a' on the table. L is not looking ate the gesture made by B, repeating yet again what is to be written.) É claro que há.

733. B*: (turning to the page and whispering about whether it has an accent.) *Should I put an accent after?* [B17b]

734. L: (Not hearing and dictating again) 'É claro que há' (Impatiently.) 'Come on! Write!'

735. B*: (Writing again.) It is [u]... Of course...

736. L: (Emphatically speaking.) Que há!

737. B*: Que há. (B writing [e].) Que há [a]...

738. L: (Rereading) É claro que há. (B inscribing the grave accent above the 'a' [à].)

739. B*: (Rereading.) Que há.

740. L: (Indicating where to put the period.) Now, you put...

741. B*: Exclamation?

742. L: No. Full stop. (Indicating on the sheet where to inscribe the period)

As mentioned above, the students, especially writer B, faced with the SP she recognised during the manuscript in course, is responding to what has been taught in class. Here, she seeks to differentiate the different grammatical forms that have a representation corresponding to the phoneme /a/. The gestures made over the table during turns 731 and 732, more than the actual enunciations that accompanied those gestures (B17a and B17b), clearly evidence it. Even though the differentiation between the forms of representation of the phoneme /a/ involves lexical and grammatical knowledge, we assume that, in the initial development of writing, the search for differentiation happens visually, retrieving visual-graphic information stored in the long-term memory and offered by the literary social context. In the case of these students, this context is strongly represented in the didactic practises and school exercises performed. The school fulfills its role by giving the information and emphasising the spelling problems that arise from the multiple representations (in this case, homophony) of phonemes. What is most

relevant to us is not whether the student understood or not the differences between the different ways of representation. What is relevant is the way how she, by anticipating the SP and recognising the TO, not only recovers the requirement to use a grave accent, by remembering it, as indicated by the gesture made over the table, but also the fact that she places in relation its homophonic representations. Such sensitivity to the SP will certainly lead her to learn the linguistic categories associated to the phoneme /a/ in its multiple and distinct orthographic occurrences, in the different linguistic levels (vowel, verb, article, pronoun, interjection, contraction).

The same 'há' in SP46 was not recognised as a spelling problem by the student.

It is worth noting that the fact that SP46 – also written as 'à' just a few minutes before, but not recognised as a SP – and the fact that SP50 addresses the same SP, but was recognised as a SP — illustrates the linguistic complexity involved in the process of learning orthography. Strictly speaking, in the representation of the phoneme /a/ in each of them, we have the simultaneous occurrence of two SP, related to the phenomenon of homophony,[21] which circumscribes it. On the one hand, the SP involving the representation of the silent letter 'h', which requires, in addition to the knowledge of its morphemes in the composition of verbal forms, the knowledge of the radical 'hav' from the verb 'haver', both, radical and morpheme, presenting great irregularity. On the other hand, the SP related to the orthographic accentuation of the verbal form 'há', a tonic monosyllable, where the acute accent does not mark any phonic difference, thus being also "mute". In none of the two inscriptions did the students observe or try to inscribe the grapheme 'h' which, as indicated by Nunes; Bryant (2014) is a late acquisition. However, the grave accent, present in the Portuguese language only in contractions, is over-generalised, that is, the grapheme 'à' served as much to represent the two SP related to the verb 'haver', in its verbal form 'há', as it was used, unpredictably, for the recording of 'pàra' (4th point, line 11, SP30). The recurrence of the same graphic form six times in a row may be related to the working memory activated during the text in course. Student B, by anticipating the SP50 related to the phoneme /a/, would resort to the previously written graphic form 'à'. Even though it was incorrectly used, there is an initial metalinguistic reflection from the student, even if only based on the graphic aspect of the contraction, and not in its function in relation to the linearization of the sentence 'É claro que há!' ('Of course there is!').

Somehow, the school exercise that differentiated 'à', 'há' and 'ah' seems to have had interfered in the student's metalinguistic activity. We may question whether the teaching of that spelling problem would be appropriate for this level of education (2nd year of Basic education in Portugal). But perhaps such a question might not be pertinent, since during the text in course the words being enunciated and linearized by the writers cannot be controlled. What strikes us as relevant from a learning perspective is that the student recognised a spelling problem in that point of the text in course. Even if

---

[21] Analyses of the interference of homophony in the generation of unpredictable spelling problems are also present in Felipeto (2013), Calil and Felipeto (2008) and Calil (2007).

that recognition is intermittent and limited to the accentuation mark, it indicated the student's cognitive ability to retrieve received information and her metalinguistic ability to relate the different orthographic representations for the same phoneme.

Conclusion

This case study clearly shows that recently literate students bear different strategies and information relative to orthography. They respond to the recognised SP during the text in course according to the linguistic knowledge and cognitive resources they possess. This indicates that, despite this being one of the first written texts produced by the students, they already have relative autonomy for the resolution of those problems. The metalinguistic activity related to the recognised spelling problems justifies the necessity of valuing such didactic situations involving collaborative writing in pairs. On the one hand, the didactic explanation of specific orthographic content has great importance for the learner of orthography. On the other hand, the didactic practise of production, particularly in pairs, seems to contribute to the educational content to be recognised and articulated in the text under construction.

As is characteristic of students in this stage of education, spelling problems are great in number and the identified difficulties are related to what the literature has already been describing. In the six manuscripts that were produced, we found, on average, one SP for every three words written, involving a significant diversity of problems. In the case of the invented story 'Snow White and the three dinosaurs', 136 words were written, and 52 SP occurred. Considering the 21 erased SP, and the 8 SP that were not erased but commented on, we can state that the students recognised little more than half of the SP. The presence of these SP in the manuscript indicates the knowledge level they possess and the way they resort to the information received to solve the identified SP.

The adopted methodological design, based on the theoretical foundation offered by Textual Genetics, articulated the process (text in course) with the product (written text), while respecting the ecological conditions of the classroom. The synchronised film generated by the Ramos System favoured an ethnographic, qualitative and microgenetic analysis, the main characteristic of which is in the association between the identified SP, the erasures inscribed in the sheet of paper relating to those SP and the comments enunciated by the female students. From this procedure, the paired writing of a single text stands out, favouring dialogue and strengthening the metalinguistic activity, whether in the form of comments that anticipate SP, or in the form of comments that seek solutions for the SP that were recognised after being inscribed, aiming to ensure continuity in the writing of the text.

Regarding this aspect, we found interesting behavioural characteristics in the pair of students studied, which would have to be developed in future studies involving other pairs, and even with other writing processes but the same pair, in the same classroom

conditions. We observe that SP were predominantly anticipated by the writer, while the subsequent recognition SP came from the observations made by the dictator, taking on the role of reviser of what was being inscribed and linearized. Regarding this aspect, the visual follow-up, namely, accompanying what was written with the eyes, seems to be of paramount importance. Another characteristic that may have an important impact for didactic practise and the assessment of students' knowledge within this age range refers to the fact that, on the one hand, the recognitions in advance expressed doubts, as in the case of the inscription of 'bran' in the word 'branca', of 'ssa' in the word 'dinossauro' and of 'há' for the verb 'haver'. On the other hand, the subsequent SP regnition made by L provided answers and information for the B's doubts, as for example, explaining that 'bran' starts with a 'b', that the 'b' in 'Branca de Neve' is upper case or that the grapheme 'c' is not equivalent to the grapheme 's'.

We were also able to point out two other aspects that differ greatly from other studies on the acquisition of orthography. The first refers to the presence of SP with comments, but that did not have a visible erasure mark, as was the case with 'dinosauro' (SP13 B8) and 'à/há' (SP50 B17). The second, which is not explored in this study, refers to the fact that even words that were written correctly (with no SP) and no erasures had their inscription accompanied by orthographic comments, such as 'vez' (SP7 B4), 'chamada" (SP8 B5), 'à floresta' (SP10 B6; SP20 L7), 'chocaram' (SP12 B7), 'Mas' (SP45 L16), 'E' (SP51 L18) and 'sempre' (SP52 L19).

This analysis also allowed intersecting the comments made by the students and the instruments and activities developed by the teacher and recorded in the students' notebook, evidencing that the comments they produced contain metalinguistic reflections about the taught content.

Taking these aspects into consideration, and perhaps contributing to other work on this subject, we may begin to elaborate an answer for the question stated on our title.

- Despite not knowing when the child will interrupt the textual flux when faced with a SP, since it is not possible to predict beforehand what words will emerge during the text in course, the anticipation of a SP seems to coincide with the difficulties faced from the understanding of the alphabetic principle, namely, we can expect students to recognise in advance a SP *when* these SP refer to vocal or consonant digraphs, to the accentuation or use of mute letters.

- At this still initial stage, the way *how* recognition emerges seems to be intermittent, i.e., in the present case, it does not occur in every digraph, accentuated letter or mute letter. However, its occurrence may indicate the educational content to which the students are exposed. This challenges pre-conceived ideas that the teaching of grammar has no effect on a student's textual production.

- It is difficult to know *why* the early recognition of a specific SP happens instead of another, given the cognitive and subjective dimension of each writer. However, even though we cannot respond to that just yet, we might argue that

they emerge because the student is given the opportunity to produce texts early in their school years.

These three points justify the importance of the professional gestures of the teacher, in this pathway of alphabetization, in that they contribute to students' progression in the written production in a reflective way, creating situations for them to comment, reason, and explain through their writing or that of their colleagues. This *démarche didactique* of constant questioning of linguistic phenomena shall lead them to a double confrontation: a progressive resolution of problems encountered in the text in course, among them spelling problems, and to the mastery of that system, which is undoubtedly very complex, with a learning process that shall take years to be concluded.

Acknowledgements

CALIL, E.; PEREIRA, L. Reconhecimento antecipado de problemas ortográficos em escreventes novatos: quando e como acontecem. **Alfa**, São Paulo, v.62, n.1, p.87-120, 2018.

▪ *RESUMO: A aprendizagem da ortográfica constitui um processo complexo, envolvendo questões lexicais e gramaticais. Muitos estudos sobre essa aprendizagem tratam os problemas ortográficos de modo independente e separado da produção textual. Neste estudo defendemos a importância de se analisar a aprendizagem da ortografia a partir da perspectiva proposta pela Genética Textual, colocando em destaque a gênese do processo de escritura e criação textual. Apresentamos o Sistema Ramos, metodologia de investigação que registra o processo de escritura em tempo e espaço real da sala de aula. Esse Sistema oferece informações multimodais (fala, escrita, gestualidade) sobre o que alunos, em duplas, reconhecem como problemas ortográficos (PO) e os comentários espontâneos feitos quando estão escrevendo o texto. Este estudo analisa justamente o momento em que aconteceram esses reconhecimentos e os comentários de duas alunas no 2º ano de escolaridade, durante a produção de seis histórias inventadas. Mais do que uma análise quantitativa dos tipos de PO identificados no produto, apresentamos uma análise enunciativa e microgenética de*

reconhecimentos de PO e seus comentários, particularmente aqueles PO antecipados pelas escreventes. Os resultados indicam: i. Reconhecimentos ensejam comentários nem sempre relacionados ao PO identificado; ii. Reconhecimentos e comentários estão relacionados aos conteúdos ortográficos ensinados em sala de aula; iii. Alguns PO reconhecidos envolvem a articulação de diferentes níveis linguísticos. Esses aspectos podem contribuir para a compreensão da aprendizagem da ortografia em situações didáticas propiciadas pela escrita colaborativa a dois.

▪ *PALAVRAS-CHAVE: Sala de aula. Produção textual. Ortografia. Rasura. Metalinguístico. Aprendizagem. Escrita colaborativa.*

REFERENCES

BARBEIRO, L. F. Episódios ortográficos na escrita colaborativa. Anais do **XXII Encontro Nacional da Associação Portuguesa de Linguística**, Lisboa, APL, p.111-125, 2007.

BEREITER, C.; SCARDAMALIA, M. **The psychology of written composition**. Hilsdale. NJ. Lawrence Erlbaum Associates, 1987.

BERNINGER, V.; CARTWRIGHT, A. C.; YATES, C. M.; SWANSON, H. L.; ABBOTT, R. D. Developmental skills related to writing and reading acquisition in the intermediate grades: Shared and unique variance. **Reading and Writing,** Dordrecht, v.6, p.161-196, 1994.

BORÉ, C. **Modalités de la fiction dans l'écriture scolaire.** Paris: Éditions l'Harmattan, 2010.

BOUSQUET, S.; COGIS, D.; DUCARD, D.; MASSONNET, J.; JAFFRE, J.-P. Acquisition de l'orthographe et modes cognitifs. **Revue française de pédagogie**, Paris, v.126, p.23-37, 1999.

BRISSAUD, C.; COGIS, D. **Comment enseigner l´orthographe aujourd´hui?** Paris: Hatier, 2011.

CAGLIARI, L. C. **Alfabetização e linguística**. São Paulo: Scipione, 1989.

CALIL, E. D'efeitos d'a(língua): o fenômeno linguístico "homortográfico. In: CALIL, E. (Org.). **Trilhas da escrita:** autoria, leitura e ensino. São Paulo: Cortez, 2007. p.77-118.

_____. **Escutar o invisível:** escritura & poesia na sala de aula. São Paulo: Editora da Unesp, 2008.

_____. **Autoria:** a criança e a escrita de histórias inventadas. Londrina: Eduel, 2009.

_____. Dialogues between two pupils during the process of writing a fictional story: Verbal erasures and their forms of representation. In: COOREN, F.; LÉTOURNEAU, A. (Eds.). **(Re)presentations and Dialogue**. Amsterdam: John Benjamins Publishing Company, 2012a. p.325-341.

_____. Do amarelo ao quem tem fé: séries associativas na escritura de um poema em sala de aula. In: DEL RÉ, A.; ROMERO, M. (Orgs.). **Na Língua do outro:** estudos interdisciplinares em aquisição de linguagens. São Paulo: Cultura Acadêmica, 2012b. p.257-276.

_____. Dialogisme, hasard et rature orale. Analyse génétique de la création d un texte par des élèves de 6 ans. In: CALIL, E.; BORÉ, C. (Éds.). **L'école, l'écriture et la création:** Études franco-brésiliennes. Louvain-la-neuve: L'Harmattan-Academia, 2013. p.157-188.

_____. Writing, memory and association: newly literate students and their poetry creation processes. **Recherches Textuelles**, Lorraine, n.13, p.105-121, 2016.

CALIL, E.; FELIPETO, C. A singularidade do erro ortográfico nas manifestações d'Alíngua. **Estilos da Clínica**, v.Xlll, n.25, p.118-137, 2008.

CARRAHER, T. N. Explorações sobre o desenvolvimento da ortografia em português. **Psicologia:** teoria e pesquisa, v.1, n.3, p.269-285, set./dez. 1985.

CASSAR, M.; TREIMAN, R. The beginnings of orthographic knowledge: children's knowledge of double letters in words. **Journal of Educational Psychology**, v.89, n.4, p.631-644, 1997.

CHISS, J-L.; DAVID, J. Orthographe: état des recherches et approches didactiques, **Le français aujourd'hui**, Paris, n.5, p.215-220, 2011a.

_____. L'orthographe du français et son apprentissage. **Le français aujourd'hui**, Paris, n.5, p.233-244, 2011b.

DAIUTE, C.; DALTON, B. Collaboration between children learning to write: Can novices be masters? **Cognition and Instruction**, New York, n.10, p.281-333, 1993.

DAVID, J. Typologie des procédures métagraphiques produites en dyades entre 5 et 8 ans. L'exemple de la morphologie du nombre. In: DE GAULMYN, M. M.; BOUCHARD, R.; RABATEL, A. (Eds.). **Le processus rédactionnel. Écrire à Plusieurs voix.** Paris: l'Harmanttan, 2001. p.281-292.

_____. Les procédures orthographiques dans les productions écrites de jeunes enfants. **Revue des sciences de l'éducation**, Paris, v.XXIX, n.1, p.137-158, 2003.

_____. Les explications métagraphiques appliquées aux premières écritures enfantines, **Pratiques** [En ligne], p.139-140, 2008, mis en ligne le 15 décembre 2008,

consulté le 30 septembre 2016. Available in: <http://pratiques.revues.org/1230>. Access in: 30 jul. 2017.

DE GAULMYN, M.-M.; LUIS, M.-H. Genèse des représentations métalinguistiques de la langue écrite. **Linx**, Paris, v.27, n.2, p.107-113, 1997.

DOQUET, C. **L'écriture débutant:** pratiques scripturales à l'école élémentaire. Rennes: Presses Universitaires de Rennes, 2011.

FABRE, C. **Les brouillons d'écoliers ou l'entrée dans l'écriture.** Grenoble: Ceditel/ L'Atelier du Texte, 1990.

FAYOL, M.; MIRET, A. Écrire, orthographier et rédiger des texte. Writing, spelling, and composing. **Psychologie française**, Paris, n.50, p.391-402, 2005.

FELIPETO, C. 'À Dieu petit crapaud': quand l'homonymie produit des désordres et des ratures dans un processus d'ecriture en collaboration. In: BORÉ, C.; CALIL, E. (Org.). **L'école, l'écriture et la création.** Paris: Ed. Bruylant-Academia, 2013. p.189-202.

FIJALKOW, J.; FIJALKOW, E. L'écriture inventée au cycle des apprentissages. **Dossiers de l'Éducation**, Paris, n.18, p.125-167, 1991.

HAYES, J. R. A new framework for understanding cognition and affect in writing. In: LEVY, C. M.; RANSDELL, S. (Eds.). **The science of writing**: theories, methods, individual differences and applications. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1996. p.1-27.

JAFFRÉ, J. P. Compétence orthographique et acquisition. In: DUCARD, D.; HONVAULT, R.; JAFFRÉ, J. P. (Eds.). **L'orthographe en trois dimensions.** Paris: Nathan, 1995. p. 94-158.

KELLOGG, R. T. Competition for working memory among writing processes. **The American Journal of Psychology**, New York, n.114, p.175-191, 2001.

_____. Training writing skills: a cognitive developmental perspective. **Journal of writing research**, New York, v.1, n.1, p.1-26, 2008.

MCCUTCHEN, D. Knowledge, processing, and working memory: Implications for a theory of writing. **Educational Psychologist**, New York, n.35, p.13-23, 2000.

MONTEIRO, A. M. L. "Sebra – ssono – pessado – asado". O uso do "S" sob a ótica daquele que aprende. In: MORAIS, A. G. de (Org.). **O aprendizado da ortografia**. Belo Horizonte: Autêntica, 1999. p.43-60.

MOREIRA, N. R. (Org.). **Chapeuzinho Vermelho aprende a escrever:** estudos psicolinguísticos comparativos em três línguas. São Paulo: Ática, 1996.

MOREIRA, N. R.; PONTECORVO, C. Chapeuzinho/Cappuccetto: as variações gráficas e a norma ortográfica. In: MOREIRA, N. R. (Org.). **Chapeuzinho Vermelho aprende**

a escrever: estudos psicolinguísticos comparativos em três línguas. São Paulo: Ática, 1996. p.78-122.

MORIN, M-Fr. Declared Knowledge of beginning writers. **L1 – Educational Studies in Language and Literature**, Amsterdam, n.5, p.385-401, 2005.

MORIN, M-F.; NOOTENS, P. Étude des procédures verbalisées en lecture et en écriture chez des forts et faibles orthographier au début du primaire. **Repères**, Paris, n.47, p.83-107, 2013.

NOBILE, G. G.; BARRERA, S. D. Análise de erros ortográficos em alunos do ensino público fundamental que apresentam dificuldades na escrita. **Psicologia em Revista**, v.15, n.2, p.36-55, 2009.

NUNES, T.; BRYANT, P. **Leitura e Ortografia:** além dos primeiros passos. Porto Alegre: Penso, 2014.

NUNES, T.; BRYANT, P.; BINDMAN, M. The effects of learning to spell on children's awareness of morphology. **Reading and Writing**, v.19, n.7, p.767-887, 2006.

PEREIRA, L. Á. **Escrever com as crianças. Como fazer bons leitores e escritores.** Porto: Porto Editora, 2008.

PEREIRA, L. Á.; BARBEIRO, L. F. A revisão textual acompanhada como estratégia de ensino da produção escrita. In: LUNA, M. J. D.; SPINILLO, A. G.; RODRIGUES, S. G. (Eds.). **Leitura e Produção de Texto.** Recife: Editora Universitária da UFPE, 2010. p.51-80.

PINTO, M. da G. L. C. A ortografia e a escrita em crianças portuguesas nos primeiros anos de escolaridade. In: _____. **Saber viver a linguagem. Um desafio aos problemas de literacia.** Porto: Porto Editora, 1998. p.139-193.

REGO, L. L. B.; BUARQUE, L. L. Algumas fontes de dificuldade na aprendizagem de regras ortográficas. In: MORAIS, A. G. (Ed.). **O aprendizado da ortografia**. Belo Horizonte: Autêntica, 1999. p.21-41.

ROSA, J. Morphological awareness and the Spelling of Homophone Forms in European Portuguese, **Lidil**, n.30, p.133-146, 2004. Available in: <http://lidil.revues.org/873>. Access in: 30 jul. 2017.

ROSA, J.; NUNES, T. Pode-se melhorar a escrita das vogais indistintas pelas crianças? **Educar em Revista**, v.38, p.113-127, 2010.

STORCH, N. **Collaborative Writing in L2 Classrooms.** Bristol (UK): Multilingual Masters, 2013.

SWAIN, M.; LAPKIN, S. Interaction and Second Language Learning: two adolescent french immersion students working together. **The Modern Language Journal**, Montreal, v.82, n.3, p.320-337, 1998.

TOLCHINSKY, L. **The cradle of culture and what children know about writing and numbers before being taught.** Mahwah, NJ: Lawrence Erlbaum, 2003.

ZORZI, J. L. **Aprender a escrever:** a apropriação do sistema ortográfico. Porto Alegre: Artes Médicas, 1998.

CHARACTERIZATION OF TEMPORAL COMPLEMENTARITY: FUNDAMENTALS FOR MULTI-DOCUMENT SUMMARIZATION

Jackson Wilke da Cruz SOUZA*
Ariani Di FELIPPO**

- ABSTRACT: Complementarity is a usual multi-document phenomenon that commonly occurs among news texts about the same event. From a set of sentence pairs (in Portuguese) manually annotated with CST (Cross-Document Structure Theory) relations (Historical background and Follow-up) that make explicit the temporal complementary among the sentences, we identified a potential set of linguistic attributes of such complementary. Using Machine Learning algorithms, we evaluate the capacity of the attributes to discriminate between Historical background and Follow-up. JRip learned a small set of rules with high accuracy. Based on a set of 5 rules, the classifier discriminates the CST relations with 80% of accuracy. According to the rules, the *occurrence of temporal expression in sentence 2* is the most discriminative feature in the task. As a contribution, the JRip classifier can improve the performance of the CST-discourse parsers for Brazilian Portuguese

- KEYWORDS: Linguistic description. Complementarity. CST. Multi-document Summarization. Natural Language Processing.

Introduction

The access and availability of digital information have been grown very rapidly. According to the projections of Taufer (2013), digital information will amount to 44 zettabytes by 2020. Some Natural Language Processing (NLP) subareas seek to develop computational applications capable of handling this tremendous amount of data.

One of these subareas is Multi-Document Summarization (MDS), which goal is to automate the production of a summary given a group of texts on the same topic compiled from different sources (MANI, 2001). The vast majority of the MDS work relies on the production of extractive summaries (or extracts), which are composed of sentences taken exactly as they appear in the source texts. Such summaries tend to

* Federal University of São Carlos (UFSCar) / Interinstitutional Nucleus of Computational Linguistics (NILC), São Carlos - São Paulo - Brasil. jackcruzsouza@gmail.com

** Federal University of São Carlos (UFSCar) / Interinstitutional Nucleus of Computational Linguistics (NILC), São Carlos - São Paulo - Brasil. Department of Letters (DL). arianidf@gmail.com

be *informative*, since they convey the central content of the collection to the point of replacing the reading of the source texts, and *generic,* i.e., they are aimed at a broad readership community (KUMAR; SALIM; RAZA, 2012).

The multi-document summaries have been generated in three stages: (i) *analysis* (i.e., interpretation of the source collection to obtain an internal representation of its content); (ii) *transformation* (i.e., the main stage of the summarization model; it takes the internal representation of the source texts to produce the summary internal representation), and (iii) *synthesis* (i.e., the summary internal representation is linguistically realized into the final summary) (SPARCK-JONES, 1993; MANI, 2001).

According to the amount and level of linguistic knowledge, MDS can be shallow or deep (MANI, 2001). A shallow MDS method uses little or no linguistic knowledge; it usually performs a statistical analysis of source texts in order to produce extracts. Shallow methods/systems usually require low-cost processing and achieve higher robustness and scalability. However, the extracts generated by these methods tend to suffer from lack of text coherence and cohesion, and from low informativeness. Deep MDS approaches, in turn, use sophisticated linguistic knowledge codified in grammars, semantic repositories and discourse models, and thus they have a restricted application and a high cost of development. On the other hand, the deep MDS paradigm generates extracts with less linguistic problems and also *abstracts*.

In order to produce informative and generic extracts, it is necessary to select the most important sentences of the collections, avoiding redundancy and contradiction, and keeping complementarity among them. Redundancy, contradiction and complementarity are some of the so-called multi-document phenomena whose identification and treatment are important for the linguistic quality and informativeness of the multi-document extracts. We illustrate the three phenomena with the sentence pairs (S1 and S2) in (1), (2), and (3) (with the original Portuguese sentences in parenthesis).

(1) *Redundancy*
S1: The margin of error is plus or minus 2 percentage points. (*A margem de erro é de dois pontos percentuais, para mais ou para menos.*)
S2: The margin of error is 2 percentage points. (*A margem de erro é de 2 pontos porcentuais.*)
(2) *Complementarity*
S1: For Japan, the reported magnitude 6.8 is considered "strong". *(No caso do Japão, a magnitude apontada de 6,8 é considerada "forte".)*
S2: In Niigata, after an earthquake of same magnitude (6.8) in October 2004, 65 people were killed and more than 3,000 injured. *(Em Niigata, um terremoto em outubro de 2004, também de magnitude 6,8, matou 65 pessoas e deixou mais de 3.000 feridos.)*
(3) *Contradiction*
S1: Both José Maria Eymael (PSDC - Christian Social Democratic Party) and Rui Pimenta (PCO - Workers' Cause Party) have not reached 1% of voting

intentions. *(José Maria Eymael, do PSDC, e Rui Pimenta, do PCO, não chegaram a obter 1% das intenções de voto.)*

S2: Candidates José Maria Eymael (PSDC) and Ruy Pimenta (PCO) did not score. *(Os candidatos José Maria Eymael (PSDC) e Ruy Pimenta (PCO) não pontuaram.)*

The sentences in Example (1) are redundant because the main content is very similar. In Example (2), the sentences are complementary, since they share some information (*6.8 magnitude earthquakes in Japan*), but S2 provides additional information not presented in S1. Specifically, S1 provides historical information on *a 6.8 magnitude earthquake that struck a region of Japan in 2004*. Finally, in Example (3), S1 and S2 contradict each other because, in S2, the candidates did not receive any vote, and, in S1, they received some votes, but the number of votes did not add up to 1% of the intentions.

Specifically, the identification of such phenomena during the texts analysis is very important because: (i) the most redundant sentences of the collection convey its main content and thus must be included in the summary; (ii) relevant and complementary sentences should also be included in the summary, and (iii) redundant or contradictory information should not be selected for the summary. In order to do that, linguistic descriptions of the multi-document phenomena are essential, since they provide clues to be followed by the MDS methods. In this paper, we focus on the (temporal) complementarity, since redundancy (e.g., HATZIVASSILOGLOU et al., 2001; NEWMAN et al., 2004; HENDRICKX et al., 2009; SOUZA; DI-FELIPPO; PARDO, 2013) and contradiction (e.g., CONDORAVDI et al., 2003; MARNEFFE; RAFFERTY; MANNING, 2008; MARNEFFE, 2012) are the phenomena that have been investigated more extensively in the literature.

In Section 2, we describe the multi-document relations of the CST (*Cross-Document Structure Theory)* model (RADEV; JING; BUDZIKOWSKA, 2000) that codify complementarity and the main methods of identifying the CST relations. In Section 3, we present the *corpus* used in this work and then the selection of temporal complementary pairs of sentences from the *corpus*. Section 4 describes the linguistic characteristics of the temporal complementarity and the translation of such characteristics into attributes or features that can be used to automatically identify the CST relations. In Section 5, we describe the linguistic description of the *corpus* that is necessary to evaluate the potential of the attributes to detect the temporal complementarity. Finally, in Section 6, we present the evaluation results regarding the potential of the attributes to detect the CST relations of temporal complementarity and some final remarks.

Related studies

Two sentences from different texts on the same topic can be related to each other in a number of ways (MANI, 2001). The analysis of the relationships among such sentences

(i.e., multi-document or intertextual analysis) has been extensively investigated during the last decades in the NLP field. An example of NLP application that benefits from such analysis is Multi-Document Summarization (MDS), which aims at generating a unique summary from the content of several source texts. The investigation of cross-document relationships provided a set of rhetorical relations identified among sentences from topically related documents. These rhetorical relations are based on the CST model (RADEV; JING; BUDZIKOWSKA, 2000).

The CST model allows the connection (in pairs) of semantically related textual units (e.g., sentences) from documents on the same topic. Originally, it was proposed a set of 24 CST relations (Table 1).

Table 1 – Original set of CST relations.

Identity	Modality	Judgment
Equivalence	Attribution	Fulfillment
Translation	Summary	Description
Subsumption	Follow-up	Reader profile
Contradiction	Elaboration	Contrast
Historical background	Indirect speech	Parallel
Cross-reference	Refinement	Generalization
Citation	Agreement	Change of perspective

Source: Radev (2000).

Some authors have refined the original CST relations, proposing more compact sets (e.g., ZHANG; OTTERBACHER; REDEV, 2003; MAZIERO, 2012; MAZIERO; JORGE; PARDO, 2014). For Brazilian Portuguese, the original set was reduced to 14 relations and they were organized into two groups (MAZIERO, 2012; MAZIERO; JORGE; PARDO, 2014): (i) content relations (Identity, Elaboration, Equivalence, Contradiction, Summary, Subsumption, Overlap, Historical background and Follow-up), and (ii) form relations (Attribution, Citation, Modality, Indirect Speech and Translation). The content relations, in particular, codify the following multi-document phenomena: redundancy, complementarity and contradiction.

A number of papers have addressed the benefits of CST for MDS. The study proposed by Zhang, Blair-Goldensohn and Radev et al. (2002) was the first to consider multi-document structural relationships, codified by the CST relations, to generate a summary. Specifically, the authors first use MEAD (RADEV; JING; BUDZIKOWSKA, 2000; RADEV et al., 2003), a summarization system based on *cluster centroids*[1], to rank the source sentences and produce an initial extract. Then the low-salience

---

[1] In general, the analysis of the source texts in cluster and centroid-based MDS methods consists in grouping sentences that are highly similar to each other. Thus, the clusters with similar sentences represent the "topics" of the collection. Each cluster is represented by a centroid, i.e., a set of statistically significant words. The cluster centroids are used to

sentences ranked by MEAD are replaced by sentences that have more CST relations in the collection, which tend to be more informative.

Jorge and Pardo (2010), focusing on Brazilian Portuguese, also apply CST in MDS. In this work, they rank sentences according to the number of CST relations they have in the collection. More recently, Cardoso (2014), also using a *corpus* in Brazilian Portuguese, developed a MDS method that integrates CST, Rethorical Structure Theory (RST) (MANN; THOMPSON, 1987), and subtopics to model the summarization process. All the mentioned research works deal with manually annotated *corpora*, but a CST annotation is an expensive and time-consuming task, since it requires a highly trained team of experts capable of producing a considerable amount of data.

In order to solve this problem, there have been efforts put forth to automatically identify CST relations in texts using Machine Learning (ML)[2] techniques. Zhang et al. (2003) focused on the detection of six CST relations across source sentences. The developed classifier[3] was able to efficiently identify unrelated sentence pairs, but showed poor performance in classifying the type of relations.

Miyabe, Takamura and Okomura (2008) attempted to detect the Equivalence and Transition relations. The method of the authors to automatically identify these relations necessarily starts with the detection of Equivalence and then Transition (i.e., relation that occurs between sentences with the same information, differing by numerical values; this relation would be similar to the CST Contradiction relation).

In Zahri and Fukumoto (2011), the identification of Identity, Paraphrase (similar relation to CST Equivalence), Subsumption, Overlap, and Elaboration is a stage of a summarization application. In such MDS system, the authors used the headlines of the documents to extract sentences with salient terms from the source texts using the statistical model. Then they assigned rhetorical relations among those sentences that were learned by a ML algorithm. Finally, they ranked the sentences by measuring their relative importance within the source collection through the method called PageRank[4] (ERKAN et al., 2004), and selected the most salient sentences to compose the extract. According to the authors, the combination of PageRank along with rhetorical relations among sentences helps to avoid the generation of extracts with redundant information.

Kumar, Salim and Raza (2012) applied *linguistic* (e.g., verbal similarity between two sentences), *structural* (e.g., length sentence), and *statistic* (e.g., word overlap) features

---

[2] identify the sentences in each cluster that are most similar to the centroid. Thus, the system selects the sentence that is most relevant to each cluster. For more details on cluster and centroid, see Jurafsky and Martin (2009).

[2] Machine learning is a field of Artificial Intelligence that aims at exploring the study and construction of algorithms that can learn from and make predictions on data.

[3] In Machine Learning, a classifier is an algorithm that takes a set of parameters (or features) that characterize objects (or instances) and uses them to determine the type (or class) of each object. Using a training set containing a list of instances with known classifications, since each class is described by a set of attributes or features, the classifier decides how the parameters ought to be weighed and combined in order to separate the various classes of instances. Then, the weights determined in the training set are applied to a set of instances that do not have known classes in order to determine what their classes are likely to be (MITCHELL, 1997).

[4] PageRank is an algorithm used by Google Search to rank websites in their search engine results (https://en.wikipedia.org/wiki/PageRank).

for the automatic identification of four CST relations: Identity, Overlap, Subsumption, and Description. To evaluate the method, the authors used a dataset taken from CSTBank (RADEV; OTTERBACHER; ZHANG, 2004), a multi-document *corpus* of English news articles whose sentences were annotated with CST relationships. Specifically, Kumar et al. selected 476 sentence pairs for training and 206 sentence pairs for testing. In order to identify the CST relations, they used three different ML techniques, and the results showed that Identity is the easiest relation to be detected (f-measure > 90%). The authors point out that this result may be related to the high lexical similarity between the sentences linked by Identity, which facilitates the automatic identification of the relation.

For Brazilian Portuguese, there are a few studies on automatic detection of CST relations (MAZIERO, 2012; MAZIERO; JORGE; PARDO, 2014; SOUZA; DI-FELIPPO; PARDO, 2012, 2013).

Maziero (2012) and Maziero, Jorge and Pardo (2014) developed the multi-document discourse parser called CSTParser, which applies the most popular set of similarity features to detect a CST relation between two sentences: (i) difference in number of words between the sentences (S1–S2); (ii) percentage of words in common between S1 and S2; (iii) position of S1 in the text (0 – beginning: first three sentences; 1 – middle; 2 – end: last three sentences); (iv) position of S2 in text (the same as above); (v) number of words in the longest common substring between S1 and S2; (vi) difference in the number of nouns between S1 and S2 (common and proper nouns); (vii) difference in the number of adverbs between S1 and S2; (viii) difference in the number of adjectives between S1 and S2; (ix) difference in the number of verbs between S1 and S2; (x) difference in the number of proper nouns between S1 and S2; (xi) difference in the number of numerals between S1 and S2, and (xii) number of possible synonyms in common in S1 and S2. The CSTParser achieved an overall accuracy of 68,13%, which is the average accuracy obtained by a classifier for Overlap, Subsumption, Elaboration, Equivalence, Historical background and Follow-up, and by hand-crafted rules for Identity, (explicit) Contradiction, Attribution, Indirect Speech, and Translation[5]. According to the authors, this overall accuracy is a good result given the subjectivity of the cross-document analysis.

Souza, Di-Felippo and Pardo (2012, 2013) focused on the automatic detection of the CST relations that represent redundancy (Identity, Equivalence, Summary, Subsumption and Overlap) and the types of redundancy (i.e., total, partial, or null) codified by these relations (2012). To investigate the formal characteristics of redundancy, the authors used CSTNews (CARDOSO et al., 2011), a multi-document *corpus* of journalistic texts that were annotated at sentence level with CST relations. Besides sentence position in source texts, the authors applied the following attributes: (i) word overlap,[6] (ii) noun and verb overlap, (iii) morphosyntactic pattern (e.g., noun+preposition+noun) overlap,

---

(iv) subject overlap (i.e., occurrence or identical subjects) (v) main verb overlap, (v) object (direct or indirect) overlap, (vii) part-of-speech tag overlap, and (viii) occurrence of synonyms. Using PART (WITTEN; FRANK, 1998) and J48 (QUINLAN, 1993) ML algorithms, Souza, Di-Felippo and Pardo (2013) showed that the classifier based on all attributes identifies the redundancy types (total, partial and null) with 97.7% of accuracy, and the CST relations with 62.2% of accuracy. The second best classifier uses only one attribute (noun overlap) and achieves 91.1% of accuracy for the types and 60% for the CST relations.

Based on this review of the literature, we observed that several works have been investigated the automatic detection of CST relations, especially those relations that represent redundancy. However, the CST relations that refer to complementarity (i.e., Follow-up, Historical background and Elaboration) have been automatically detected using attributes that characterize redundancy, such as those investigated by Souza, Di-Felippo and Pardo (2012, 2013), which are not specific of complementarity.

CSTNews *corpus* and complementarity

In order to describe the complementarity phenomenon, we used CSTNews, a reference multi-document *corpus* in Brazilian Portuguese (CARDOSO et al., 2011). The CSTNews *corpus* is composed of 50 clusters (or collections), totaling 140 source texts, 2,088 sentences and 47,240 words. The clusters are organized into the following categories: *world* (14), *politics* (11), *daily news* (13), *science* (1), *money* (1), and *sports* (10). The source texts were compiled from the following online news agencies *Folha de São Paulo, Estadão, O Globo, Jornal do Brasil,* and *Gazeta do Povo*.

Each cluster is composed of: (i) two (2) to three (3) sources texts; (ii) mono-document summaries; (iii) six (6) multi-document abstracts and six (6) manual multi-document extracts; (iv) one (1) automatic multi-document extract; (v) manual CST annotation among source texts; (vi) manual annotation of temporal expression of each source texts; (vii) part-of-speech tagging and syntactic parsing of each source text; (viii) semantic annotation of nouns and verbs with their corresponding Princeton WordNet[7] synsets (FELLBAUM, 1998); (ix) manual annotation of aspects (e.g., *who, what, where, when,* etc.) of one (1) source text; (x) annotation of each source document with RST[8] relations (MANN; THOMPSON, 1987), and (xi) manual subtopic segmentation of each source text.

---

two variations of *word overlap*, i.e., *noun overlap* and *verb overlap*, to compute similarity based on the number of nouns and verbs shared by two sentences.

[7] WordNet is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives, and adverbs are organized into synonym sets (or synsets) (e.g., {car, auto, automobile, machine, motorcar}), each representing one underlying lexical concept. Different relations link the synonym sets, such as antonymy, hyponymy/hypernymy, meronymy/holonymy, entailment, and cause. More details on Princeton WordNet are available at <https://wordnet.princeton.edu/>.

[8] The RST model allows the analysis of text coherence. Such analysis consists of verifying if the elementary discourse units (EDUs), which are the minimal building blocks of a discourse structure, are interconnected. Each EDU in a

Based on the CST typology proposed by Maziero et al. (2012), we selected a set of sentence pairs from CSTNews annotated with the following CST relations: Follow-up, Historical background and Elaboration. In order to select the pairs, we used the CSTNews online interface[9]. According to Maziero et al. (2012), Historical background and Follow-up are content relations that codify two different types of temporal complementarity, as illustrated in examples (i) and (ii) of Table 2, respectively.

Table 2 – Examples of temporal complementarity.

TEMPORAL COMPLEMENTARITY	SENTENCE PAIR
(i) S2 presents some historical background about the event described in S1 (S1←S2); the historical event is the focus of S2 (*Historical background*)	S1: A plane crash in Bukavu, a city in eastern Democratic Republic of the Congo (DRC) killed 17 people on Thursday, said a United Nations spokesman on Friday. (*Um **acidente aéreo na localidade de Bukavu, no leste da República Democrática do Congo (RDC)**, matou 17 pessoas na quinta-feira, informou nesta sexta-feira um porta-voz das Nações Unidas*.) S2: Air accidents are frequent in Congo, where 51 private companies operate oldplanes built in the former Soviet Union. (***Acidentes aéreos** são frequentes **no Congo**, onde 51 companhias privadas operam com aviões antigos principalmente fabricados na antiga União Soviética*.)
(ii) S2 presents additional information which has happened since S1 (S1←S2); the events in S1 and S2 are related and have a relatively short period of time between them (*Follow-up*)	S1: **The secondary runway at São Paulo's Congonhas Airport opened** at 6 am, only for departures. (*A pista auxiliar de Congonhas abriu às 6h, apenas para decolagens.*) S2: **São Paulo's Congonhas Airport only opened** for landings at 8:50 am. (*Congonhas só abriu para pousos às 8h50.*

Source: Own elaboration.

---

relation is classified as nuclei (i.e., more important propositions) or satellites (i.e., complementary information). When coherent, the units of a text are linked to each other by rhetorical relations (also known as coherence or discourse relations). Relations with one nucleus and one satellite are said to be mononuclear relations. When all the interconnected units are nuclei, we have a multinuclear relation.

9 <http://nilc.icmc.usp.br/CSTNews/>.

We illustrate the complementarity in (i) with a pair of sentences from distinct news reporting a plane crash in Congo. Specifically, the sentences present common content (in bold), and S2 provides historical information (underlined) about the other, which is the frequent occurrence of plane crashes in Congo (since the commercial carriers operate with old aircrafts). According to the typology of Maziero et al. (2012), Historical Background is the CST relation that represents the complementarity illustrated in (i), characterizing as temporal; in this case, the temporal complementarity is related to the frequent aspect of the main event described in S1.

To illustrate the complementarity in (ii), we provide a pair of sentences compiled from distinct news reporting delays and cancellations of flights at Congonhas Airport due to bad weather. The temporal complementarity relation between S1 and S2 occurs because they share some information (*time which Congonhas Airport opened*), and the event reported in S2 followed the event described in S1 after a short period. According to the typology of Maziero, Jorge e Prado (2014), this type of temporal complementarity is codified by Follow-up.

Elaboration, unlike the relations illustrated in Table 2, does not involve temporal information. The example in Table 3 illustrates this non-temporal complementarity.

Table 3 – Example of non-temporal complementarity.

Non-temporal complementarity	Sentence pair
S2 details some information present in S1 (S1←S2); S2 does not repeat information present in S1 and the additional information present in S2 is the focus of S2. (*Elaboration*)	S1: Although the improvement project is approved, the **work schedule** has not been released yet. (*Apesar da definição, o cronograma da obra não foi divulgado.*)
	S2: The **work schedule** depends on final studies being carried out by Infraero. (*O cronograma da obra depende de estudos finais que estão sendo realizados pela Infraero*

Source: Own elaboration.

The sentences in Table 3 were extracted from news reporting a renovation project of the Congonhas Airport. We observed that S1 and S2 share some content (the project schedule), and S2 details some information present in S1. The additional information in S2 is the focus of the sentence and consists of the reason why the project timeline has not been published (it depends on final studies being carried out by Infraero – Brazil's national airport authority).

Table 4 shows the total number of sentence pairs annotated with the CST relations that express the different types of complementarity in CSTNews.

Table 4 – The statistics of complementary in the CSTNews *corpus*.

COMPLEMENTARITY	CST RELATION	QT. OF PAIRS	TOTAL
Non-temporal	Elaboration	343	343
Temporal	Historical background	77	370
	Follow-up	293	
			713

Source: Own elaboration.

In Table 4, we see that non-temporal complementarity, expressed by Elaboration, occurs in 343 sentence pairs. We also see that temporal complementarity occurs in 370 sentence pairs, with 293 cases of Follow-up and 77 cases of Historical background. Thus, there are 713 sentence pairs with complementarity in CSTNews.

To date, we manually analyzed 45 pairs of each CST relations that codify temporal complementarity, totaling a *subcorpus* with 90 pairs. The manual analysis of the *subcorpus* allowed us to detect linguistic characteristics of temporal complementarity, represented by Historical background and Follow-up. In the next section, we describe these linguistic characteristics and present their conversion into machine-tractable attributes or features.

Identifying machine tractable attributes

Two complementary sentences extracted from different news texts reporting the same event are relatively similar, which can be seen in the examples that illustrate the complementarity phenomenon. Such similarity underlies the very definition of the CST relations, since Zhang and Radev (2005) state that such relations only occur between semantically related sentences. Based on the degree of similarity, the classes of content CST relations in the Maziero, Jorge e Prado (2014) typology can be organized into the following hierarchy: redundancy > complementarity > contradiction. Thus, complementarity is a multi-document phenomenon that involves intermediate level of similarity. It is not known, however, if the different temporal complementarities, expressed by Historical background and Follow-up CST relations, present the same level of similarity. Thus, we decided to investigate the potential of redundancy as a distinctive feature. In order to make it possible, we have selected the most efficient attributes that have been used in the literature to automatically detect similarity among sentences.

According to Hatzivassiloglou et al. (2001), Newman et al. (2004) and Souza, Di-Felippo and Pardo (2012, 2013), there are three efficient features to capture similarity or redundancy between two sentences, which are: (i) noun overlap, (ii) sentence position (in the source text), and (iii) subtopic overlap. Noun overlap is an

efficient measure since most of the frequent words that are content words in texts or sentences are nouns. As we mentioned, *noun overlap* (Nol) is a version of the traditional measure *word overlap.* Calculation of Nol is shown in (4). The result value of *noun overlap* varies between zero and one. If the value is one or close to one, it means high similarity, but, if the value is zero or close to zero, it means the opposite, i.e., low similarity between sentences.

$$(4) \quad \text{Nol (S1, S2)} = \frac{\text{\# Common nouns (S1,S2)}}{\text{\# Nouns (S1) + \# Nouns (S2)}}$$

Concerning (ii), Souza Di-Felippo and Pardo (2012, 2013) state that sentence position indicates similarity because news texts commonly follow the inverted-pyramid structure. Such structure illustrates how information is organized in the following blocks of decreasing relevance: (i) lead, i.e., the main information of the news text; it corresponds to the first paragraph; (ii) body, i.e., paragraph(s) that follow(s) the lead, illustrating the main information of the text, and (iii) closer, i.e., the final paragraph which reinforces or debunks the lead (LAGE, 2002).

Thus, if two sentences occupy close positions in their correspondent source texts, they tend to convey similar information. If the distance between the positions they occupy in the texts is long, the content of the sentences tend to be not so similar. Under this hypothesis, Souza, Di-Felippo and Pardo (2012) proposed the *sentence distance* attribute, described in (5). For example, given a sentence pair from a cluster *x* (S1 and S2), where S1 is the Sentence 6 from the Text 1 and S2 is the Sentence 4 from the Text 2, the distance value between them is equal to 2 (positions). In this paper, the distance value is divided by the longest distance between two sentences identified in the *subcorpus*, which normalizes the feature by document length. Thus, the equation described in (5) generates a distance value between zero and one. If the value is zero or close to zero, it means high similarity between sentences; if the value is one or close to one, it means the opposite, i.e., low similarity.

$$(5) \quad \text{Distance (S1, S2)} = \frac{\text{\# Distance between S1 and S2}}{\text{\# Longest distance in } subcorpus}$$

The attribute (iii), *subtopic overlap*, is a refinement of the sentence distance feature. In the inverted pyramid structure, the lead is the main information (or topic) of a piece of news and the details about the lead are the subtopics, which are directly or indirectly linked to the topic according to thematic progression (KOCH, 2009). Due to the inverted pyramid structure, sentence position indicates content similarity. However, the inverted pyramid is just a writing guideline and thus a piece of news might present relatively different structure. Consequently, it is not always possible to capture redundancy based

on sentence position. Thus, subtopic overlap might be a relevant feature to identity redundancy since it is independent of the position occupied by the sentences. Moreover, once *subtopic overlap* is a semantic feature, we say that it is richer or more informative than noun overlap, which is based on word forms. In order to capture similarity based on subtopic, we proposed the *subtopic overlap* attribute (SubTol). As a binary feature, the SubTol values can be "yes" or "no", indicating whether a sentence pair conveys the same subtopic or not.

Besides similarity or redundancy, temporal complementarity, expressed by the Historical background and Follow-up CST relations, involves the occurrence of temporal marks, such as simple adverbs and expressions.

The Example (6a) illustrates Historical background. In (6a), we observe the occurrence of temporal expressions only in S1, i.e., *on Thursday* (*na quinta-feira*) and *on Friday* (*nesta sexta*). In Example (6b), the two temporal expressions *on Thursday night* (*na noite desta quinta-feira*) and *since last 13th (July)* (*desde o último dia 13*) occur in S1, and the only expression *in 1996* (*em 1996*) occurs in S2.

(6a) S1: A plane crash in Bukavu, in the Eastern Democratic Republic of the Congo (DRC) killed 17 people on Thursday, said a United Nations spokesman on Friday. (*Um acidente aéreo na localidade de Bukavu, no leste da República Democrática do Congo (RDC), matou 17 pessoas na quinta-feira, informou nesta sexta-feira um porta-voz das Nações Unidas.*)

S2: Air accidents are frequent in Congo, where 51 private companies operate elderly planes built in the former Soviet Union. (*Acidentes aéreos são frequentes no Congo, onde 51 companhias privadas operam com aviões antigos principalmente fabricados na antiga União Soviética.*)

(6b) S1: TAM airlines confirmed on Thursday night that the right thrust reverser was deactivated since last 13th (July). (*A TAM confirmou, na noite desta quinta-feira, que o Airbus da TAM estava com o reverso do lado direito desligado, desde o último dia 13.*) S2: In 1996, a failure in the reverser caused an accident with a TAM Fokker-100, occurred seconds after takeoff from Congonhas Airport. (*Em 1996, uma falha no reverso foi a causa do acidente com o Fokker-100 da TAM, ocorrido segundos depois da decolagem, também em Congonhas.*)

The sentence pair in Example (7a) illustrates the Follow-up relation. Follow-up involves the occurrence of temporal expression in both sentences of the pair. The expression *at 8:50 a.m.* (às 8h50) in S2 indicates the complementary event, which occurred after the event described in S1 (*the secondary runway opened at 6 am)*. In Example (7b), the temporal *during this Sunday, day 6 (durante este domingo, dia 6)* occur in S1 and the adverb *after (depois)* in S2.

(7a) s1: The secondary runway at São Paulo's Congonhas Airport opened <u>at 6 am</u> only for departures. (*A pista auxiliar de Congonhas abriu às 6h, apenas para decolagens*.) S2: São Paulo's Congonhas Airport opened for landings only <u>at 8:50 am</u>. (*Congonhas só abriu para pousos às 8h50*.)

(7b) <u>S1: During this Sunday</u>, <u>day 6</u>, bloody fights occurred (*Durante este domingo, dia 6, foram travadas lutas sangrentas*.)

S2:   The Israeli offensive was launched <u>after</u> a string of Hezbollah attacks on Sunday that caused the biggest casualties for Israel in the four weeks of the conflict. (*A ofensiva israelense foi lançada <u>depois</u> de uma sequência de ataques do Hezbollah no domingo que causou as maiores baixas para Israel nas quatro semanas do conflito*.)

Thus, to use these linguistic clues for the automatic detection of temporal complementarity, we proposed four (4) binary attributes: occurrence of temporal expression in S1 (TES1), occurrence of ET in S2 (TES2), occurrence of adverb in S1 (ADVS1) and occurrence of adverb in S2 (ADVS2).

Table 5 synthesizes the total set of seven (7) attributes used in this research.

Table 5 – Temporal complementarity atributes.

ATTRIBUTE	DESCRIPTION	ACRONYM
Noun overlap	It captures redundancy based on the number of nouns the sentences of a pair have in common.	NoI
Distance	It captures redundancy based on the distance between the positions/locations occupied by the sentences of a pair in their correspondent source text	Distance
Subtopic overlap	It captures redundancy based on the subtopic overlap between the sentences of a pair.	SubToI
Occurrence of temporal expression in S1	It captures temporal complementarity by the occurrence of temporal expression in Sentence 1 of a pair	TES1
Occurrence of temporal expression in S2	It captures temporal complementarity by the occurrence of temporal expression in Sentence 2 of a pair	TES2
Occurrence of adverb in S1	It captures temporal complementarity by the occurrence of adverbs in Sentence 1 of a pair	ADVS1
Occurrence of adverb in S2	It captures temporal complementarity by the occurrence of adverb in Sentence 1 of a pair	ADVS2

Source: Own elaboration.

Next, we describe the linguistic characterization of the 90 sentence pairs to evaluate the potential of the attributes.

Linguistic description of the *corpus*

The characterization process of the 90 sentence pairs for the automatic evaluation of the features consisted in describing the sentences in a way that we could calculate the seven features or attributes for each pair. Specifically, the sentence description was based on previous annotations of the CSTNews *corpus*.

In order to calculate NoI, Distance, ADVS1 and ADVS2 attributes, we identified the nouns and adverbs that constitute each sentence and the position of the sentences in their source texts. The nouns and adverbs and also the sentence position were compiled from the syntactic annotation of CSTNews, which was automatically performed by the parser[10] PALAVRAS (BICK, 2000). We manually revised the automatic annotation of CSTNews, in order to reduce *noise*[11] (i.e., tagging errors) and *silence* (i.e., tagging omissions) produced by the parser. Table 6 illustrates the syntactic annotation[12] of the sentence *São Paulo's Congonhas Airport only opened for landings at 8:50 am* (*Congonhas só abriu para pouso às 8h50*). The annotation in Table 6 includes the position the sentence ("s4") and the following nouns, *Congonhas* (*pos*[13]*=np*") and *pousos* (landings) (*pos="n"*). For the linguistic description, we selected the values of the attribute *lemma* of nouns and adverbs. For describing the sentence in Table 6, for example, we selected *pouso* (landing) and *Congonhas*.

Table 6 – Example of syntactic annotation of CSTNews.

```
</s><s id="s4" text="Congonhas só abriu para pousos, às 8h50.">
    <terminals>
        <t id="1" word="Congonhas" lemma="Congonhas" pos="np"/>
        <t id="2" word="só" lemma="só" pos="adv"/>
        <t id="3" word="abriu" lemma="abrir" pos="v-fin" morph="PS 3S IND VFIN"/>
        <t id="4" word="para" lemma="para" pos="prp"/>
        <t id="5" word="pousos" lemma="pouso" pos="n"/>
        <t id="6" word="," lemma="--" pos="pu"/>
        <t id="7" word="a" lemma="a" pos="prp"/>
        <t id="8" word="as" lemma="o" pos="art"/>
        <t id="9" word="8h50" lemma="8h50" pos="n"/>
        <t id="10" word="." lemma="--" pos="pu"/>
    </terminals>
</s>
```

Source: Own elaboration.

---

[10] Computational tool that can analyze and identify the sentence constituents and their syntactic functions (CARROL, 2004).

[11] This is the case of *8:50,* which was wrongly annotated as noun.

[12] Each sentence (*s*) is annotated with two attributes: *id*, i.e., position of the sentence in the text, and *text*, i.e., the sentence itself of the *corpus*. The constitutive elements of *s* (words or expressions and punctuation symbols) are called *terminals* (or *tokens*). Each of them is described by four attributes: *id* (i.e., position in sentence), *word* (i.e., occurrence of the word or expression), *lemma* (i.e., canonical form) and *pos* (i.e., part-of-speech or word category).

[13] POS stands for *Part-of-Speech*.

We selected the subtopics of the previous annotation of CSTNews described in Cardoso et al. (2011). Tables 7 and 8 show the source texts of the sentences that illustrate the Historical background relation in Table 2. Based on the annotation[14], we may see, for example, that S1 from Table 2 is the S1 in Document 2 (Table 7) and its content conveys the subtopic labelled as "*the crash*", while S2 from Table 2 is the S4 in Document 1 (Table 8) and its content coveys a different subtopic, labeled as "*history*".

Table 7 – Example of subtopic annotation in CSTNews (Document 2, *Cluster* 1).

S1: A plane crash in Bukavu, in the Eastern Democratic Republic of the Congo (DRC) killed 17 people on Thursday, said a United Nations spokesman on Friday *(Um acidente aéreo na localidade de Bukavu, no leste da República Democrática do Congo (RDC), matou 17 pessoas na quinta-feira, informou nesta sexta-feira um porta-voz das Nações Unidas)*
S2: The victims of the accident were 14 passengers and three crew members. *(As vítimas do acidente foram 14 passageiros e três membros da tripulação.)*
S3: Everyone died when the plane, hampered by the bad weather, failed to reach the runway and crashed in a forest that was 15 kilometers from the airport in Bukavu. *(Todos morreram quando o avião, prejudicado pelo mau tempo, não conseguiu chegar à pista de aterrissagem e caiu numa floresta a 15 quilômetros do aeroporto de Bukavu.)*
S4: The plane exploded and caught fire, said the UN spokesman in Kinshasa, Jean-Tobias Okala. *(Segundo fontes aeroportuárias, os membros da tripulação eram de nacionalidade russa.)*
S5: The plane exploded and caught fire, added UN spokesman in Kinshasa, Jean-Tobias Okala. *(O avião explodiu e se incendiou, acrescentou o porta-voz da ONU em Kinshasa, Jean-Tobias Okala.)*
S6: "There were no survivors", said Okala. *"Não houve sobreviventes", disse Okala.)*
<t LABEL="the crash" TOP= "1")>

S7: The spokesman said the plane, a Soviet Antonov-28, of Ukrainian manufacturing and under ownership of the Trasept Congo, a Congolese company, also took a mineral load *(O porta-voz informou que o avião, um Soviet Antonov-28 de fabricação ucraniana e propriedade de uma companhia congolesa, a Trasept Congo, também levava uma carga de minerais.)*
<t LABEL= "detail about the aircraft" TOP= "3">

Source: Own elaboration.

---

[14] The annotation of a subtopic includes two attributes: LABEL, i.e., a brief description of the subtopic, and TOP, i.e., a sequential number of annotation in the cluster (CARDOSO et al., 2012).

Table 8 – Example of subtopic annotation in CSTNews (Document 1, *Cluster* 1).

S1: At least 17 people died after a passenger plane crashed in Democratic Republic of Congo. *(Ao menos 17 pessoas morreram após a queda de um avião de passageiros na República Democrática do Congo.)*   S2: According to a spokesman from UN, the plane was trying to land at the airport in Bukavu in the middle of a storm. *(Segundo uma porta-voz da ONU, o avião, de fabricação russa, estava tentando aterrissar no aeroporto de Bukavu em meio a uma tempestade.)*   S3: The aircraft hit a mountain and crashed into a forest about 15 Km from the runway end. *(A aeronave se chocou com uma montanha e caiu, em chamas, sobre uma floresta a 15 quilômetros de distância da pista do aeroporto.)*   **<t LABEL= "the crash" TOP= "1">**
S4: Air accidents are frequent in Congo, where 51 private companies operate elderly planes built in the former Soviet Union. *(Acidentes aéreos são freqüentes no Congo, onde 51 companhias privadas operam com aviões antigos principalmente fabricados na antiga União Soviética.)*   **<t LABEL= "historical background" TOP= "2">**
S5: The crashed airplane, operated by Air Traset airline, carried 14 passengers and three crew members *(O avião acidentado, operado pela Air Traset, levava 14 passageiros e três tripulantes.)*   S6: The airplane was a flight from the mining town Lugushwa, 130 kilometers away from its destination, Bukavu, *(Ele havia saído da cidade mineira de Lugushwa em direção a Bukavu, numa distância de 130 quilômetros.)*   **<t LABEL= "the crashed aircraft" TOP= "1">**
S7: Aircraft are used extensively for transport in Democratic Republic of Congo, a huge country where there are few paved road. *(Aviões são usados extensivamente para transporte na República Democrática do Congo, um vasto país no qual há poucas estradas pavimentadas.)*   S8: In March, the European Union banned all Congolese airlines from operating in Europe. *(Em março, a União Européia proibiu quase todas as companhias aéreas do Congo de operar na Europa.)*   S9: Only one kept the permission. *(Apenas uma manteve a permissão.)*   S10: In June, the International Air Transport Association also included Congo in a group of several African countries it classed as an "embarrassment" to the industry. *(Em junho, a Associação Internacional de Transporte Aéreo incluiu o Congo num grupo de vários países africanos que classificou como "uma vergonha" para o setor.)*   **<t LABEL= "historical background" TOP= "2">**

Source: Own elaboration.

We selected the temporal expressions (TE) from the *corpus* annotation described in Menezes-Filho and Pardo (2011). In order to identify and classify the TEs of the source sentences from CSTNews, the authors used the typology of Baptista, Hagège and Mamede (2008). Such typology organizes TEs into 4 classes or types: (i) *calendar time* (i.e., the TE corresponds to a unique anchoring of the process onto the timeline), (ii) *duration* (i.e., the TE does not anchor the process onto the timeline) (e.g., *A reunião durará 2 horas*) (*The meeting will last 2 hours*), (iii) *frequency* (i.e., the TE relates the process to the timeline through multiple anchoring instances) (e.g., *Ocorrerá entre os dias 29 e 31 de julho*) (*It will take place between 29 and 31 July*), and (iv) *generic* (i.e., the expression does not anchor any event onto the timeline) (e.g., *Eu gosto do mês de julho*) (*I like July*). The calendar time class is further structured in: (i) *hour* (e.g., *Ele chegou às 9h30m*) (*He arrived at 9:30am*), (ii) *date* (e.g., *Em 7 de julho foi registrado 52.77 graus Celsius*) (O*n July 7th, 52.77 degrees Celsius were recorded*), and (iii) *interval* (i.e., TE involving two explicit dates (e.g., *A reunião durará entre 1 e 2 horas*) (*The meeting will last from 1 to 2 hours*). Finally, the *date* type is also classified based on the temporal reference of the TE and/or its indeterminacy regarding its anchoring on the timeline. In this sense, the following subtypes are distinguished: (i) *enunciation* (e.g., *Ele partiu em março*) (*He departed in March*)), (ii) *textual element*, somewhere in the text (e.g., *Um acidente no dia anterior*) (*A car accident the day before*), and (iii) *absolute dates*, directly computable from the TE (e.g., *Ele foi lançado em maio de 2009*) (*It was released in May 2009*). Based on such typology, the expression às 6h (*at 6am*) in (8) (see case (i) in Table 2) was annotated as *calendar time* of the subtype *hour*.

(8) A pista auxiliar de Congonhas abriu **<TE TYPE="CALENDAR_TIME" SUBTYPE="HOUR">**às 6h **</TE>**, apenas para decolagens

We organized the linguistic information necessary to calculate the 7 attributes in an single *xls* file, which is here illustrated in two Tables (9 and 10) for matter of space. In Table 9, we exemplify the characteristics of the sentences showed in Table 2 that are needed for calculating the numeric attributes, i.e., Nol and Distance.

Table 9 – Pre-processing of the *corpus* to calculate the numeric attributes.

CORPUS		LINGUISTIC DESCRIPTION	
PAIR	CST RELATION	NOUN	SENTENCE POSITION
1	Historical background	*acidente, localidade, Bukavu, leste, República Democrática do Congo, RDC, pessoa, porta-voz, Nações Unidas, quinta-feira, sexta-feira* (accident, city, Bukavu, eastern, Democratic Republic of the Congo, RDC, people, spokesman, United Nations, Thursday, Friday)	1
		acidente, Congo, companhia, avião, União Soviética (accident, Congo, company, plane, Soviet Union)	4
2	Follow-up	*pista, Congonhas, decolagem* (runway, Congonhas, departure)	6
		Congonhas, pouso (Congonhas, landing)	4

Source: Own elaboration.

In Table 10, we exemplify the description of the sentences that is needed for calculating the binary attributes (i.e., SubTol, TES1, TES2, ADVS1, and ADVS2). "X" indicates that the linguistic feature does not occur in the sentence.

Table 10 – Pre-processing of the *corpus* to calculate the binary attributes.

Corpus		Linguistic description		
Pair	CST Relation	SUBTOPIC	TEMPORAL EXPRESSION	ADVERB
1	Historical background	1	calendar/date/enunciation	X
		2	X	X
2	Follow-up	1	calendar/hour	X
		1	calendar/hour	X

Source: Own elaboration.

After describing each sentence of the *subcorpus*, we manually calculated the attributes for each sentence pair. The values resulting from the attributes specification of the 90 sentence pairs in our *subcorpus* were also organized in an *xls* file, as shown in Table 11.

Table 11 – Two sentence pairs and their correspondent attributes and values.

	Corpus			*Attributes*				
Pair	CST Relation	Nol	Distance	SubTol	TES1	TES2	ADVS1	ADVS2
1	Historical background	0.133	3	no	yes	no	no	no
2	Follow-up	0.4	2	yes	yes	yes	no	no

Source: Own elaboration.

In Table 11, the pair 1 is characterized by the attribute-value pair Nol=0.125, which means that the sentences have little information content in common. The value 0.125 for Nol results from the fact that among the 16 different nouns contained in S1 (i.e., *acidente, localidade, Bukavu, leste, República Democrática do Congo, RDC, pessoa, porta-voz, Nações Unidas, quinta-feira, sexta-feira*) and the 5 nouns contained in S (i.e., *acidente, Congo, companhia, avião, União Soviética*), only *acidente* occurs in both. Concerning Distance, we normalized[15] the values due to the different sizes of the source texts. Thus, the initial value 3 obtained for pair 1 in Table 11 was divided by the highest value of Distance observed in the *subcorpus* (i.e. 34), resulting in 0.088, the normalized value. Moreover, the sentence pair 1 do not share the same subtopic, which is indicated by the value "no" for the SubTol attribute. We also observe in Table 11 that there is no occurrence of adverbs in the pair 1, but temporal expressions, on the other hand, occur in S1 of the pair 1 (*na quinta-feira* and *nesta sexta-feira*).

Once the attributes for each sentence pair of the *subcorpus* were computed, we evaluated the potential of the attributes for automatic distinguishing Historical background from Follow-up.

Evaluation of the attributes

The attributes were automatically evaluated based on ML algorithms available in Weka (*Waikato Environment for Knowledge Analysis*) (HALL et al., 2009), an open source software toolkit from the University of Waikato (New Zealand). The toolkit supports both supervised and unsupervised ML algorithms from various Artificial Intelligence approaches.

In this paper, we used the supervised inductive learning approach. In such ML task, an algorithm learns from a training set (or *corpus*) whose classes (i.e., label to be learned) of the examples are known. In general, each example of a training data is a pair consisting of an input object (i.e., a sentence pair and its attributes) and a desired

---

15 The normalization process aims at reducing the chances of data becoming inconsistent. Thus, a normalized attribute takes values in the range [0, 1].

output value or class (i.e., the CST relation of the pair). A supervised learning algorithm analyzes the training set and produces a *classifier* that should be able to predict the correct classes of new and unlabeled examples.

Specifically, we applied the *10-fold cross validation*[16] technique, which gets more realistic estimates of the error rates for classification, since our dataset is relatively small. To evaluate the attributes, we applied *precision, recall* and *f-measure*, i.e., metrics most commonly used to quantify the performance of NLP techniques (MITCHELL, 1997). *Precision* is the proportion of the instances (i.e., sentence pairs) returned by the classifier that were correct (i.e., instances classified as belonging to the correct class). *Recall* is the proportion of all possible correct instances that were returned. In other words, recall indicates the fraction of pairs annotated by human experts as belonging to a class *x* that were also correctly identified by the classifier. Finally, *f-measure* is the harmonic mean of the previous two measures. In this paper, we only used general accuracy for evaluating the results.

Although there are different ML paradigms, i.e., connectionist, mathematical (or probabilistic), and symbolic, we focused on symbolic algorithms, since they produce rules that can be easily interpreted and verified by human experts. Nonetheless, we have also tested other ML techniques from other paradigms, for comparison purposes only.

To identify the CST relations that codify complementarity, we tested the well-known connectionist method called *Multi-Layer Perception* (MLP), with the default Weka configurations. We achieved 82.2% of general accuracy. Among the several mathematical or probabilistic methods in Weka, we run Naïve-Bayes and SMO. Naive-bayes achieved the same accuracy of MLP (i.e., 82.2%), while SMO obtained 80% of accuracy.

Among the symbolic algorithms, we tried the same set used in previously related works, such as those of Maziero (2012), and Souza, Di-Felippo and Pardo (2012, 2013). Thus, we specifically tried One-R (or *One Rule*) (HOLTE, 1993), PART (WITTEN; FRANK, 1998), JRip (COHEN, 1995) e J48 (QUINLAN, 1993). One-R is probably the simplest symbolic algorithm, since it uses only the most discriminative feature or attribute to produce a single set of rules over this feature. JRip and PART are basic algorithms. They examine the classes from training data in growing size and generate an initial set of logical rules for the class. Such rules commonly combine two or more attributes. J48 builds decision trees from a set of training data. The most common way to build a decision tree is by top-down partitioning, and thus the most discriminative (or more generic) attribute corresponds to the topmost decision node and all other descendant nodes are less discriminative attributes.

J48 and PART built the biggest set of rules (13 and 8, respectively) with similar accuracy: both obtained 81% of general accuracy. One-R, as we mentioned, produces a single rule based on single attribute. In this case, the One-R algorithm selected TES2 as

---

[16] In k-fold cross-validation, the *corpus* is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single one is retained for test, and the remaining (k 1) subsamples are used as training data. The process is repeated k times, with each of the k subsamples used once as the test data. The results are averaged over all the runs.

the most important attribute and, as usual, it surprisingly produced a very good result, obtaining 80% of accuracy. JRip, in its turn, generated the smallest set of rules (only 5) with general accuracy of 79%.

Thus, JRip learned a small set of rules with a very good level of accuracy. Such combination (i.e., manageable rule set and high accuracy among the symbolic approaches) makes the choice of JRip perfectly adequate for our purposes. Table 12 presents the 5 rules of JRip, which are followed by the number of instances (pairs) correctly classified and incorrectly classified, and the precision of the rule, given by the number of correctly classified instances over all the instances classified by that rule.

Table 12 – JRip logic rules for identifying temporal complementarity.

Rules	Correct	Incorrect	Precision
1. If TES2= hour, then *Follow-up*	12	0	100%
2. Elseif TES2= no, then *Follow-up*	30	8	78.9%
3. Elseif NoI>=0.315 e TES1=date-absolute, then *Follow-up*	4	0	100%
4. Elseif TES2=hour_date-enunciation, then *Follow-up*	3	0	100%
5. Elseif *Historical background*	41	4	91.1

Source: Own elaboration.

Based on these rules, one can say that the TES2, NoI and TES1 features define the pairs annotated as Follow-up in CSTNews. Among them, TES2 is the most discriminative attribute, since three of the total five rules are based on it. For *feature selection*[17], we applied the InfoGainAttributeEval algorithm, also available at Weka. This algorithm also indicated the relevance of TES2. In addition, one can say that, if none of the 4 first rules are applied, the default class or (i.e., CST relation) is Historical Background, which is given by rule 5.

It is also interesting to notice how productive the rules are. For instance, rules 1 and 2 deal with many more cases than rules 3 and 4, which is natural to happen due to the way the ML process chooses the attributes to start the rules. Thus, we can say that the classifier might still achieve good results by using only the two first rules (1 and 2) for Follow-up and the last default rule (5) for Historical background. In Table 13, we have the JRip confusion matrix, through which we verify in more details how the classifier is dealing with each class or relation. Each column of the matrix represents the instances (sentence pairs) in a predicted class (relation), while each row represents the instances in an actual class.

---

[17]   Feature selection is the process of selecting a subset of relevant features. The aim of this process is to improve the performance of the classifiers. The central premise when using a feature selection technique is that the data contains features that are either redundant or irrelevant, and, removing them can reduce the processing time and generate simpler models.

Table 13 – Confusion matrix of JRip algorithm.

Class Test	*Follow-up* (45)	*Historical background* (45)
Follow-up	35	10
Historical background	9	36

Source: Own elaboration.

It can be observed from the results in Table 13 that, from the total of 45 Follow-up pairs, the rules of JRip correctly classified 35 of them, and, from the total of 45 Historical background pairs, the algorithm correctly identified 36 of them. Based on this performance, we conclude that JRip correctly classified the pairs from both classes in a very similar way.

It is important to say that such results are only indicative of the features that characterize temporal complementarity and the corresponding CST relations, as well as the discriminative power of the attributes. We say that because our *subcorpus*, from which the classifiers were learned and tested, is very small. Apart from that, it is important to say that this is a pioneer study on the characterization of temporal complementarity as a linguistic phenomenon in MDS.

Future studies may include creating a testing *corpus*, by selecting new sentence pairs annotated as Follow-up and Historical background from CSTNews. Consequently, the *subcorpus* of 90 pairs analyzed here could be used as training data only, and the resulting classifier could be evaluated in a different set of sentence pairs. We expect that future efforts will be put in linguistic description of non-temporal complementarity and the corresponding CST relation (Elaboration).

Acknowledgments

- *RESUMO: A complementaridade é um fenômeno multidocumento comumente observado entre notícias que versam sobre um mesmo evento. A partir de um corpus em português composto*

por um conjunto de pares de sentenças manualmente anotadas com as relações da Cross-Document Structure Theory (CST) que explicitam a complementaridade temporal (Historical background e Follow-up), identificou-se um conjunto potencial de atributos linguísticos desse tipo de complementaridade. Por meio de algoritmos de Aprendizado de Máquina, testou-se o potencial dos atributos em distinguir as referidas relações. O classificador simbólico gerado pelo algoritmo JRip obteve o melhor desempenho ao se considerar a precisão e o tamanho reduzido do conjunto de regras. Somente com base em 5 regras, tal classificador identificou Follow-up e Historical background com precisão aproximada de 80%. Ademais, as regras do classificador indicam que o atributo ocorrência de expressão temporal na sentença 2 é o mais relevante para a tarefa. Como contribuição, salienta-se que o classificador aqui gerado pode ser utilizado nos analisadores discursivos multidocumento para o português do Brasil que são baseados na CST.

▪ *PALAVRAS-CHAVE: Descrição linguística. Complementaridade. CST. Sumarização Multidocumento. Processamento Automático de Língua Natural.*

REFERENCES

ALEIXO, P.; PARDO, T. A. S. CSTTool: um parser multidocumento automático para o Português do Brasil. In: IV Workshop on MSc Dissertation and PhD Thesis in Artificial Intelligence – WTDIA. 2008. **Proceedings**… Salvador/Brasil.

BAPTISTA, J.; HAGÈGE, C.; MAMEDE, N. Proposta de anotação e normalização de expressões temporais da categoria TEMPO para o HAREM II. In: Encontros do Segundo HAREM, 2008. **Actes**... p.1-24.

BICK, E. **The parsing system "PALAVRAS":** Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. 2000. 412 f. Thesis (PhD) - Aarhus University, Denmark University Press, 2000.

CARDOSO, P. C. F. **Exploração de métodos de sumarização automática multidocumento com base em conhecimento semântico-discursivo**. 2014. 182 f. Tese (Doutorado em Ciências da Computação) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2014.

CARDOSO, P. C. F.; MAZIERO, E. G.; JORGE, M. L. C.; SENO, E. M. R.; DI FELIPPO, A.; RINO, L. H. M.; NUNES, M. G. V.; PARDO, T. A. S. CSTNews – A discourse-annotated *corpus* for single and multi-document summarization of news texts in brazilian portuguese. In: 3rd RST Brazilian Meeting, 2011. **Proceedings**… Cuiabá/Brazil. p.88-105.

CARROL, J. Parsing. In: MITKOV, R. (Ed.). **The Oxford handbook of computational linguistics.** Oxford/USA. Ed. Oxford University Express, 2004. p.233-248.

COHEN, W. Fast effective rule induction. In: 12th International Conference on International Conference on Machine Learning, 1995. **Proceedings**... California/USA. p.115-123.

CONDORAVDI, C.; CROUCH, D.; DE PAIVA, V.; STOLLE, R.; BOBROW, D. G. Entailment, intensionality and text understanding. In: HLT-NAACL 2003 workshop on Text meaning, 9., 2003. **Proceedings**... Edmonton/Canada: Association for Computational Linguistics, 2003. p.38-45.

ERKAN, G.; RADEV, D. R. LexPageRank: prestige in multi-document text summarization. In: Empirical Methods in Natural Language, 2004. **Proceedings**… Barcelona/Spain. p.365-371.

FELLBAUM, C. **WordNet:** an electronic lexical database. California: Ed. MIT Press, 1998.

HALL, M.; FRANK, E., HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v.11, Issue 1, 2009.

HATZIVASSILOGLOU, J. L.; KLAVANS, J. L.; HOLCOMBE, M.; BARZILAY, R.; MCKEOWN, K. Simfinder: a flexible clustering tool for summarization. In: NAACL Workshop on Automatic Summarization, 2001. **Proccedings**…Pittsburgh/USA. p.1-9.

HENDRICKX, I.; DAELEMANS, W.; MARSI, E.; KRAHMER, E. Reducing redundancy in multi-document summarization using lexical semantic similarity. In: 2009 Workshop on Language Generation and Summarisation, 2009. **Proceedings**... Suntec/Singapore: Association for Computational Linguistics, 2009. p.63-66.

HOLTE, R. C. Very simple classification rules perform well on most commonly used datasets. **Machine Learing**, Ed. Springer, v.11, n.1, p.63-90, 1993.

JORGE, M. L. C.; PARDO, T. A. S. Experiments with CST-based Multidocument Summarization. In: ACL Workshop TextGraphs-5: Graph-based Methods for Natural Language Processing, 2010. **Proceedings**… Uppsala/Sweden. p.74-82.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing:** an introduction to natural language processing, computational linguistics, and speech recognition. v.2. Englewood: Ed. Prentice Hall, 2009.

KOCH, I. G. V. **Introdução à linguística textual:** trajetória e grandes temas. 2. ed. São Paulo: Contexto, 2009.

KUMAR, Y. J.; SALIM, N.; RAZA, B. Cross-document structural relationship identification using supervised machine learning. **Applied Soft Computing**, v.12, n.10, p.3124-3131, oct. 2012.

LAGE, N. **Estrutura da notícia**. São Paulo: Ed. Ática, 2002.

MARNEFFE, M-C DE. **What's that supposed to mean? Modeling the pragmatic meaning of utterances**. 2012. 178 f. Thesis (PhD in Linguistics) – Department of Linguistics, Stanford University, Stanford, 2012.

MARNEFFE, M-C DE.; RAFFERTY, A. N.; MANNING, C. D. Finding contradictions in text. In: Annual meeting of the ACL, 46., 2008. **Proceedings**... Columbus/USA, p.1039-1047.

MANI, I. **Automatic Summarization.** Amsterdam/Netherlands: Ed. John Benjamins Publishing Company, 2001.

MANN, W. C.; THOMPSON, S. A. **Rhetorical structure theory**: a theory of text organization. California/USA: Ed. University of Southern California – Information Sciences Institute, 1987. p.87-190.

MAZIERO, E. G. **Identificação automática de relações multidocumento**. 2012. 117 f. Dissertação (Mestrado em Ciências da Computação) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2012.

MAZIERO, E. G.; JORGE, M. L. R. C.; PARDO, T. A. S. Revisiting Cross-document Structure Theory for multi-document discourse parsing. **Information Processing & Management**, v.50, n.2, p.297-314, 2014.

MENEZES FILHO, L. A.; PARDO, T. A. S. Detecção de Expressões Temporais no Contexto de Sumarização Automática. In: 2nd STIL Student Workshop on Information and Human Language Technology, 2011. **Proceedings**… Cuiabá/Brasil. p.1-3.

MITCHELL, T. M. **Machine learning**. v.45. Burr Ridge, IL: McGraw Hill, 1997.

MIYABE, Y.; TAKAMURA, H.; OKUMURA, M. Identifying cross-document relations between sentences. In: 3rd International Joint Conference on Natural Language, 2008. **Proceedings**… Hyderabad/India. p.141-148.

NEWMAN, E.; DOMN, W.; STOKES, N.; CARTHY, J.; DUNNION, J. Comparing redundancy removal techniques for multi-document summarization. In: Starting AI researchers' symposium, 2004. **Proceedings**…Valencia/Spain. p.223-228.

QUINLAN, J. **Programs for machine learning.** San Mateo/USA: Ed. Morgan Kaufmann Publishers, 1993.

RADEV, D. A common theory of information fusion from multiple text sources step one: cross-document structure. In: 1st SIGdial workshop on Discourse and dialogue, 10., 2000. **Proceedings**… Hong Kong/China: Association for Computational Linguistics, 2000. p.74-83.

RADEV, D. R.; JING, H.; BUDZIKOWSKA, M. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation and user studies. In:

ANLP/NAACL Workshop on Automatic Summarization, 2000. **Proceedings...** Seatle/ WA: North American Association for Computational Linguistics, 2000. p.21-29.

RADEV, D. R.; TEUFEL, S.; SAGGION, H.; LAM, W.; BLITZER, J.; QI, H.; CELEBI, A.; LIU, D.; DRABEK, E. Evaluation challenges in large-scale multi-document summarization: the mead project. In: 41st Annual Meeting of the Association for Computational, 2003. **Proceedings**… Sapporo/Japan: Association for Computational Linguistics, 2003. p.375-382.

RADEV, D.; OTTERBACHER, J.; ZHANG, Z. CSTNank: A corpus for the study of Cross-document Structural Relationship. In: **Proceedings fo Fourth International Conference on Language Resources and Evaluation**. Lisboa, 2004.

SOUZA, J. W. C. **Descrição linguística da complementaridade para a Sumarização Automática Multidocumento.** 2015. 105 f. Dissertação (Mestrado em Linguística) – Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2015.

SOUZA, J. W. C.; DI-FELIPPO, A.; PARDO, T. A. S. **Investigação de métodos de identificação de redundância para Sumarização Automática Multidocumento**. Série de Relatórios do NILC (NILC-TR-12): São Carlos/Brasil. 30 f. 2012.

_____. Identificação da redundância na Sumarização Automática Multidocumento: explorando métodos superficiais. In: 3rd Student Workshop on Information and Human Language Technology (TILic), 2013. **Proceedings**… Fortaleza/Brasil. p.1-3.

SPARCK JONES, K. What might be in a summary? **Information Retrieval**. v.93, p.9-26, 1993.

TAUFER, P. Massa de informações digitais pode ser usada em benefício da população. **Jornal da Globo**, 26 dez. 2013. Disponível em: <http://g1.globo.com/jornal-da-globo/ noticia/2013/12/massa-de-informacoes-digitais-pode-ser-usada-em-beneficio-da-populacao.html>. Acesso em: 02 fev. 2015.

WITTEN, I. H.; FRANK, E. **Generating accurate rule sets without global optimization.** Working paper series, 1998.

ZAHRI, N. A. H. B.; FUKUMOTO, F. Multi-document summarization using link analysis based on rhetorical relations between sentences. In: 12th International Conference on Computational Linguistics and Intelligent Text Processing, 2., 2011. **Proceedings**… Tokyo/Japan. p.328-338.

ZHANG, Z.; RADEV, D. Combining labeled and unlabeled data for learning cross-document structural relationships. In: **Natural Language Processing** – I JCNLP. Springer, p.32-41, 2005.

ZHANG, Z.; OTTERBACHER, J.; REDEV, D. R. Learning cross-document structural relationships using boosting. In: International Conference on Information and Knowledge Management, 2003. **Proceedings**… Las Vegas/USA. p.124-130.

ZHANG, Z.; BLAIR-GOLDENSOHN, S.; RADEV, D. R. Towards CST-enhanced summarization. In: Innovative applications of artificial intelligence conference, 2002. **Proceedings**… Edmonton/Canada. p.439-446.

MORPHOLOGICAL COMPLEXITY AND LEXICAL PROCESSING COSTS

Alina VILLALVA*
Carina PINTO**

- ABSTRACT: Although many words are formed by more than one morphological constituent, not all of them are complex words. In the framework of morphological analysis, the term 'complex word' usually sets apart words formed by a root, a stem index and inflectional affixes, from words formed by derivation, modification or compounding. This distinction is quite simplistic since all words display a certain degree of complexity. In the literature, there are abundant claims that morphological structure plays an important role in word processing, but the level of morphological complexity is never taken into account. In this paper, we will try to contribute to the discussion of the role of morphological structure for written word processing, namely by taking into consideration the level of morphological complexity of a particular set of Portuguese derived words. We will look at the results of a priming experiment involving a lexical decision task on three sets of derivatives in *-oso*: the first set is formed by compositional structures; in the second, we have included words that display an allomorph of the suffix (i.e. *–oso ~ –uoso*); and, in the third set, we gathered words that make use of an allomorphic base. The results of this experiment confirm that derived word processing is sensitive to the morphological structure of the word and they also show that compositional structures involve lower processing costs. Hence, these results allow us to claim that the degree of morphological complexity of complex words needs to be considered for the study of written word processing.

- KEYWORDS: Morphological complexity. Visual processing. Derivation.

Languages are complex systems, formed by complex modules that accommodate complex domains.[1] The complexity of one of these modules, the lexicon, must be evaluated at three distinct levels: the first level concerns the assignment of a complexity index to each lexical unit;[2] the second level deals with the calculation of the complexity

---

* University of Lisbon (UL), Faculty of Letters, Linguistics Center, Lisbon - Portugal. Department of General and Romance Linguistics. alinavillalva@campus.ul.pt

** University of Lisbon (UL), Faculty of Letters, Linguistics Center, Lisbon - Portugal. pintocarinaal@gmail.com

1 Gong and Coupé (2011) present the state of the art on language complexity with regard to other complex systems. See also the discussion of linguistic complexity in Mufwene (2012).

2 We assume here, like Villava and Silvestre (2015), that the lexicon contains several kinds of lexical units, namely roots, affixes, words and lexicalised phrases.

of words; and the third level concerns the evaluation of the degree of complexity of lexical relationships (e.g. words that share the same affix; words that make use of competing affixes). These three levels of the analysis of lexical complexity are of great relevance for a better understanding of word processing and lexical access.

In this paper, we intend to contribute to the discussion of the second level by looking at complexity contrasts of a particular type of European Portuguese[3] derived words (i.e. adjectives in -*oso*). More specifically, we wish to find out if higher processing costs are somehow correlated to a higher index of word complexity, but in order to confirm or refute this claim, we need to know (i) how to assign a complexity index to a word and (ii) how to measure word processing costs.

The framework that will sustain our morphological analysis is based on the characterization of lexical units presented in Villalva and Silvestre (2015) and in the typology of complex word structures drawn by Villalva (2000, 2008). Experimental evaluation and the study of processing will be mainly based in Taft and Forster (1975), Taft (1979, 1994), Rastle et al. (2000) and Rastle et al. (2004).

Complexity and linguistic complexity

The concept of complexity is often invoked in the analysis of many domains of knowledge, but it is often used in an unprincipled way, which can lead to unpredictable and unsystematic interpretations. Simon's (1962) work probably sets the beginning of the theoretical debate in this field. He argues that complexity is a characteristic of systems that take a hierarchical form: complex systems are formed by subsystems, which in turn have their own subsystems, with no limit established for this chain. It can thus be inferred that, regardless of their specific content, hierarchical systems are always complex systems.

Since languages are hierarchical systems, formed by hierarchical modules that accommodate several hierarchical domains, it can be concluded that linguistic systems and all their subsystems are complex systems. However, applying the concept of complexity to each subject of knowledge raises particular questions, and its application to the particular domain of linguistics is still quite limited.

According to Gong and Coupé (2011, p.370), languages can be seen as complex adaptive systems, because they involve a significant number of units and modules that generate a structural complexity at various levels. Languages are complex adaptive systems because their agents - the speakers - have the ability to change the system itself (cf. STEELS, 1997, 2000; BECKNER et al., 2009).

Another claim on linguistic complexity is offered by Bane (2008, p.69):

---

[3]   Considering that the formation of adjectives in -*oso* does not present significant differences in the varieties of contemporary Portuguese, we postulate that the results that we obtained with European Portuguese speakers are not exclusive to this variety of Portuguese.

i. Linguistic complexity remains constant over time;
ii. The grammar of a given language is as complex as the grammar of any other language.

These are interesting claims because they allows us to assume that linguistic variation does not affect the degree of complexity of any language, although it may shape it differently. For instance, though the 'richness' or 'poverty' of inflectional morphological systems is often invoked in correlation with syntactic processes, we can assume that the lower complexity of one subsystem will be compensated by a greater degree of complexity of another subsystem. However, to the best of our knowledge, there is no way of measuring 'richness' or, in other words, of measuring the complexity of any of the components of any of the linguistic systems.

Equally interesting is the contribution by Mufwene (2012), who describes the concept of complexity in the field of linguistics, considering the following issues:

i. Unit complexity (e.g. the phonetic inventory size) and rules of each linguistic subsystem - Mufwene calls it **bit complexity**;
ii. **Interactive complexity**, which regards the relationship between units and rules, within each module, and also the relationship between different modules.

This is an important consideration because it suggests that the complexity of units differs from the complexity of their interaction and it further claims that the complexity of each subsystem also differs from the complexity of the system globally considered.

A combination of these standpoints allows us to set a hypothesis on linguistic complexity based on the following assumptions:

1. It is mandatory to delimit a stable window over a linguistic system, which can be achieved if the selected sample is consistent;
2. Linguistic systems have a similar degree of complexity, but the complexity of their subsystems is variable; therefore, each assessment should select a subsystem and analyse it independently;
3. Considering that the degree of complexity of a language is based on two vectors (the intrinsic value of each unit and the value of its interaction with other units in a structure), the assessment should be directed to only one of these domains.

Assuming that the analysis of word processing data can bring relevant information to the understanding of the complexity of morphological structures, we will present, in this paper, the results of an experimental study, considering complex denominal adjectives derived by suffixation in -*oso*. This study was carried out on a homogeneous sample of European Portuguese speakers.

Previous attempts to quantify word complexity deal with the description of linguistic structures and their function; the predictability of a given word sequence in speech;

the structural regularities of languages or the size of the phonetic inventory; among others (see GONG; COUPÉ, 2011), but they have produced vague measures, making the measurement of linguistic complexity too subjective. Not much work has been done in the domain of morphological complexity either, and most of the existing research is about inflection[4]. In this context, the research on derived morphological structures complexity is pioneer.

On the other hand, the discussion on complexity makes it possible to understand that the distinction between simple words and complex words, that is usually considered in the domain of morphological analysis, is as important as it is simplistic. In fact, this division merely allows to oppose words consisting of a root and its thematic and inflectional specifiers (if required by the root) and words that combine one or more affixes (derivational or modifiers) or two or more roots. It goes without saying that this is a relevant distinction: the set of simple words is an important set in the lexicon of any given language; and the identification of the set of complex words is equally important, since it provides the evidence required for the description of all the word formation processes of the language (past or contemporary). The problem that the discrimination between simple and complex words does not solve is that not all simple words are equally simple; neither all complex words are equally complex. As far as we know, no lexicological source of information allows us to obtain the set of Portuguese simple words. It is also impossible to gather the set of complex words globally. It is possible to generate subsets of complex words formed by a given prefix or suffix, although the results will need to be filtered, since the search is produced on the basis of a spelling criterion (e.g. words that end in *ity*) and not of a morphological criterion (e.g. words that contain the suffix *-ity*). In the case of simple words, no orthographic strategy allows to produce any results. So, we will now explore the kinds of problems that measuring the complexity of simple and complex words may arise.

The complexity of simple words

The basic issue, in this domain, is to determine if a similar morphological complexity index can be assigned to all the roots that occur in simple words, or if these roots belong to different classes of morphological complexity. We are persuaded that simple word roots are not all alike, but we need to identify which factors will substantiate those differences.

One of these factors is probably related to the identification of the word class of roots. Villalva and Silvestre (2015) establish a basic distinction between roots that can occur in simple words (cf. [*tóxic*]o) and those that cannot (cf. [*aqu*]ífero). The former are classified as inherently intransitive predicates (though in other contexts they may

---

4   The research on morphological complexity has been mainly developed by the Surrey Morphology Group (cf. <http://www.smg.surrey.ac.uk/>), within the scope of a project carried out between 2009 and 2015 (see **Morphological complexity:** typology as a tool for delineating cognitive organization).

occur as transitive predicates (cf. *neuro*[*tóxic*]*o*)), as complements (cf. [*toxic*]*idade*) or as modifiers (cf. [*tóxic*]*o-dependente*). The latter, usually called neoclassical roots because they are Latin and Ancient Greek borrowings, are classified as inherently transitive predicates because they can never occur in simple words: these roots occur as complements in some types of derivatives (cf. [*aqu*]*oso*) and they also occur in morphological compounds, as its head (cf. *aqui*[*fer*]*o*) or as its complement (cf. [*aqu*]*ífero*).

In this section, we will just consider the first type of roots, that is, those that can occur in simple words.[5] Although there is no reliable information available, we believe that the amount of simple words that belong to a single word class is quite considerable. Words such as *perna* 'leg', *grosso/a* 'thick' ou *pedir* 'ask' belong unequivocally to the domain of nouns, adjectives and verbs, respectively. The roots that are part of these words will therefore have a single word class specification:

(1) [*pern*]$_{NR}$ *a*   'leg'
    [*gross*]$_{ADJR}$ *o/a*   'thick'
    [*ped*]$_{VR}$ *ir*   'to ask'

There is, however, a set of roots that is equally or even more important than the previous one – it is the set of roots that occur in different simple words, as in the following examples:

(2) a.   [*murch*]$_{ADJR}$ *o/a*   'faded'
         [*murch*]$_{VR}$ *ar*   'to fade'
    b.   [*danç*]$_{NR}$ *a*   'dance'
         [*danç*]$_{VR}$ *ar*   'to dance'
    c.   [*velh*]$_{ADJR}$ *o*   'old'
         [*velh*]$_{NR}$ *o*   'old person'
    d.   [*sec*]$_{NR}$ *a*   'drought'
         [*sec*]$_{NR}$ *o*   'dry'
         [*sec*]$_{ADJR}$ *o/a*   'dry'
         [*sec*]$_{VT}$ *ar*   'to dry'

These roots have not the same lexical status[6]: some will be simple lexical units, owning an unambiguous categorical specification (e.g. [*murch*]$_{ADJR}$; [*danç*]$_{NR}$; [*sec*]$_{ADJR}$), while others are underspecified roots (e.g. [*velh*]$_{[+N]R}$); and the remaining are the output of conversion processes (cf. [*murch*]$_{VS}$; [*danç*]$_{VR}$; [*sec*]$_{NR}$; [*sec*]$_{VR}$). This is not the focus of our discussion here – we merely wish to find out if a different status regarding word class membership has consequences for the assessment of word complexity. For

---

[5]   Transitive roots will be mentioned in the following section, when we will discuss morphological compounds.

[6]   Cf. Villalva (2013).

the purpose of the present study, we will assume that the least degree of complexity belongs to the unambiguous intransitive roots (cf. [*pern*]$_{NR}$), and that the highest degree of complexity lies in the roots that were formed by conversion (cf. [*murch*]$_{VR}$, [*danç*]$_{VR}$, [*sec*]$_{NR}$, [*sec*]$_{VR}$). For this reason, the formation of the lists that we have used for the experimental work, included two series of simple words, but they all contain inherently intransitive and unequivocally noun roots (e.g. *veneno* 'poison', *mentira* 'lie', *luxo* 'luxury', *conflito* 'conflict').

On the other hand, some roots present a unique form, whatever structure they occur in, but others do not. The existence of alternating forms may be related to morphophonological issues (cf. 3a) or may arise from lexical idiosyncratic circumstances, such as the introduction of neoclassical loans (cf. 3b):

(3) a. *cão*   'dog'
         canil   'kennel'
     b. *veia*   'vein'
         venoso   'venous'

The 1st and 2nd series of data in our experiment are formed by roots that have no alternating forms. Roots that have alternating forms are in the 3rd series.

The complexity of complex words

The evaluation of the complexity of complex words raises the difficulty level referred above even further, since, as we have already mentioned, there is no work done in this domain for Portuguese or other languages. There are many open research lines, such as the comparative weight of affixation vs. compounding[7], of prefixation vs. suffixation (or other affixation types where available); the comparison of configurations involving different levels of embedding and various kinds of interaction between prefixation, suffixation and compounding; and also the evaluation of the productivity of different word formation processes. However, to the present study, only questions concerning the status of inherently transitive roots and allomorphic derivational suffixes will be taken into account.

Inherently transitive roots have been introduced in the Portuguese lexicon since the 18th century, along with the development of scientific and technical terminologies. Words that carry them often arrive as loans, for instance from French (e.g. *termómetro* 'thermometer'), though these roots have been borrowed from Latin and Ancient Greek. Consider the words in (4). The root [*pedr*] in (4a) is semantically equivalent to the root of Latin origin [*petr*], which occurs in (4b) and to the root of Greek origin [*lit*] in

---

[7]   Only morphological compounding (which involves the sequencing of roots, not words) is included here. Morphosyntactic and syntactic compounding are not morphological word formation processes.

(4c). Only the first one appears in a simple word (e.g. *pedra* 'stone') that is associated to a semantic value accessible to native speakers. It is assumed that the semantics of the words containing this root are the most transparent and that the processing of their derivatives is compositional. As for the other two roots (cf. 4b and 4c), none of them can be known outside the context of the complex words in which they occur. The meaning of these complex words is often non-compositional.

(4) a. *pedra*   'stone'
        pedreira   'stone quarry'
     b. *pétreo*   'stony'
        petrificar   'to petrify'
     c. *litografia*   'lithography'
        megalítico   'megalithic'

In our experiment, the 3rd series of data contains variants of roots that can only occur in complex words (e.g. *aquoso* 'watery' / *água* 'water'; *medroso* 'fearful' / *medo* 'fear').

Alternation can also affect affixes. In the specific case of *-oso*, there is an allomorph (*-uoso*) that is unpredictable in contemporary Portuguese. The 2nd series of data is formed by lexicalized words that contain this allomorph (cf. *conflituoso* 'conflicting'; *luxuoso* 'luxurious').

This could be a reliable index for calculating the morphological complexity of complex words. So, our initial hypothesis is that lexicalised words, either because they contain an allomorph of the root (3rd series) or an allomorph of the suffix (2nd series), reveal higher processing costs than compositional words.

Assessment of morphological complexity

The degree of complexity of a word is more than a simple arithmetic sum of the complexity indexes associated to each of its constituents – probably, its assessment requires an algorithmic function that we are far from being able to devise. There are, therefore, two lines of work worth exploring: on one hand, we need to identify what can determine the complexity index of each of the morphological constituent and the algorithm that can allow us to calculate the complexity index of each word; and, on the other hand, we need to find tools that will allow the validation of the stipulations advanced by linguistic analysis.

In the following sections we will explore the possibility of finding a validation tool based on the analysis of the morphological and lexical processing data obtained experimentally.

Morphological complexity and processing

Several experimental studies have already allowed claiming that the morphological structure plays a role in visual word recognition and lexical access. These studies were developed within a wide range of experimental paradigms, as reviewed by McQueen and Cutler (1997). A more recent state of the art can be found in Pinto (2017).

Priming experiments have helped to identify the features that are relevant for lexical activation, as well as their role. The results presented by Frost, Kugler, Deutsch and Forster (2005) and by Velan and Frost (2011) suggest the existence of two hierarchically organized processes:

1. The morpho-orthographic stage characterizes the earlier phases of word recognition - the process is activated when the visual input (the written word) is complex and compositional - in this phase the recognition of forms (base and affix) is activated;
2. The morpho-semantic stage is activated later - at this stage, the previously recognized forms are semantically interpreted.

Other studies (cf. TAFT; FORSTER, 1975; TAFT, 1979, 1994) suggest that all morphological constituents are stored in the mental lexicon as independent lexical inputs and the access to meaning is obtained by decoding the meaning of each individual unit. Morphological priming studies, such as Laudanna and Burani (1995) or Järvikivi et al. (2006), have shown that words containing affixes with various allomorphs have a longer latency response. Probably, the visual processing of morphologically complex words can be affected by different properties of morphological constituents, namely the existence of allomorphs and the productivity of each affix.

Much of the evidence that morphological structure plays a role in the visual processing of complex words is compatible with both a global word processing model and an autonomous processing of word constituents. Nevertheless, it is still unclear when and how morphological analysis comes into play.

Experimental assessment

As a contribution to the discussion on morphological complexity, we have designed and applied four experiments: a simple lexical decision task and three morphological priming tests. By performing these two types of experimental procedures, we intended to find out if the introduction of a prime corresponding to the base form of a derived target facilitates the visual processing of these complex words or not. A positive response may indicate that morphological analysis is available and facilitates the understanding of derivatives.

On the other hand, performing the same priming test with three different exposure intervals will allow us to verify whether morphological processing is performed at a more initial stage or at a later phase of visual processing. We assume here that morphological processing occurs at various stages of word processing, depending on the complexity of the constituents of complex words.

Finally, we have used three series of words derived by the same suffix: the 1st series gathers compositional structures, while the other two include derivatives that present an allomorph of one of the constituents. The hypothesis behind this choice is that words with a compositional structure are accessed more quickly than words with a structure disturbed by the occurrence of allomorphs.

Morphological Data

All the words that we have tested are formed by –*oso*, which is a denominal adjective-forming suffix. We took into account the number of syllables: simple words have two or three syllables (e.g. *veneno* 'poison', *luxo* 'luxury', *medo* 'fear') and the derivatives have four or five syllables (e.g. *venenoso* 'poisonous', *luxuoso* 'luxurious', *medroso* 'fearful'). The frequency of use in European Portuguese was equally checked with CRPC[8], which allowed us to select high frequency words, although derivatives always have a relatively low frequency rate, as an inherent feature.

Finally, and taking into account the morphological structure of the derivatives, we have set three series of ten words, which correspond to the following conditions:

1st derivatives with a compositional structure (e.g. *venenoso* 'poisonous') - the base is a noun root (*venen*- 'poison'), the suffix is –*oso;* the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by the expression $X]_{NR}$ *oso* = 'that has $X']_{ADJ}$ (*venenoso* 'poisonous' = 'that has *venen(o)* 'poison');

2nd derivatives with an allomorph of the suffix (e.g. *luxuoso* 'luxurious') - the base is a noun root (*lux*- 'luxury'), the suffix allomorph is –*uoso*; the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by the expression $X]_{NR}$ *oso* = 'that has $X']_{ADJ}$ (*luxuoso* 'luxurious' = 'that has *lux(o)* 'luxury');

3rd derivatives with an allomorphy of the base (e.g. *medroso* 'fearful') - the base is an allomorph of a noun root (*medr*- ~ *med*- 'fear'), the suffix is –*oso*; the derivative is an adjective (presented in the masculine singular form) that can be paraphrased by an expression $X]_{NR}$ *oso* = 'that has $X']_{ADJ}$ (*medroso* 'fearful' = 'that has *medo* 'fear').

---

[8]   CRPC is an electronic contemporary Portuguese Corpus that includes more than 1.6 million words. It is available at <www.clul.ul.pt/pt/recursos/183-reference-corpus-of-contemporary-portuguese-crpc>.

Methodology

All these experiments were carried out with Portuguese subjects. The sample gathered healthy individuals, college students, from Lisbon and Leiria. This study has obtained a favourable opinion from the National Commission of Data Protection (Authorization nº 7788/2013).

In order to exclude cognitive-linguistic changes, the following exclusion criteria were applied to the sample:

1) Cerebral Vascular Accident;
2) Epilepsy;
3) Cranio-Encephalic Trauma;
4) Major depression / schizophrenia diagnosed by a medical specialist;
5) Uncorrected visual changes;
6) Uncorrected auditory changes;
7) Changes in written language characterized in DSM IV;[9]
8) Serious medical illness that could lead to the appearance of linguistic alterations;
9) Drug Addiction / Alcoholism;
10) Bilingualism.

In total, we have collected data from 116 subjects. Their informed consent was obtained and all subjects were submitted to an assessment of oral and written language changes. None of the subjects had any deviant behaviour.

Procedure

For the experiments reported in this paper we have used three different prime exposure times: 50 milliseconds (=ms), 100 ms and 150 ms. Experiments were built using the software E-Prime® 2.0. Visual stimuli were presented on a Compaq Presario® computer. The following table shows the distribution of subjects,[10] by experience:

Table 1 – Distribution of subjects by four experiments

Prime presentation during 50ms	32
Prime presentation during 100ms	30
Prime presentation during 150ms	27
Lexical decision	27

Source: Own elaboration.

---

[9] DSM IV (*Diagnostic and Statistical Manual of Mental Disorders*) is a manual from the American Psychiatry Association that is used to diagnose mental illnesses.

[10] The collected sample is a convenience sample. The data collection was done with 32 subjects by experience, but data treatment led to an exclusion of some subjects for several reasons, such as percentage of errors in the responses.

The lexical decision test included an experimental list formed by thirty words presented as a target in the priming test (e.g. *venenoso* 'poisonous', *luxuoso* 'luxurious', *medroso* 'fearful'). The other three tests included a series of ten pairs of words: the prime is the base word and the target is the derivative (e.g. *veneno* 'poison' – *venenoso* 'poisonous'; *luxo* 'luxury' – *luxuoso* 'luxurious'; *medo* 'fear' – *medroso* 'fearful'). We have also used pseudo-words as fillers. These pseudo-words were obtained by a systematic procedure of replacement of syllables of the target words (e.g. *bexigoso* – *goxiboso*). The fillers are used, canonically, to motivate the lexical decision and conceal the purpose of the test, thus fulfilling the role of distractors.

All the experiments started with six training items (i.e. words that are not part of the test). Both the training items and the experimental items (i.e. test words) appeared in the centre of the screen, in black on a white background, using Times New Roman font size 18, and lowercase. They were preceded by a mask (+), which was displayed in the same position for 500ms, serving as a fixation point. In the case of the morphological priming experiment, the prime (50 ms, 100 ms or 150 ms) appeared immediately after the mask, instantaneously followed by the target word, which was available on the screen until the subjects made the lexical decision, using the computer keyboard.

Results

For the data analysis, we have used the SPSS® software version 20. As usual in similar cases, we have carried out the cleaning of the output data. The outliers were dealt taking into account the following criteria:

1) exclusion of wrong answers;
2) exclusion of values over 2000 ms;
3) replacement of disparate values greater than 10,000 ms and smaller than 250 ms by the mean of the subject in the condition;
4) replacement of values above the mean in ± 2,5 standard deviations by the mean of the subject in the condition.

This cleaning aimed to reinforce the quality of the answers that might be threatened by errors of execution or by an inherent variability to the sample elements.

The following table shows the percentage of excluded data, taking into account the above criteria:

Table 2 – Percentage of deleted data

	Prime exposure time	Wrong answers	Disparate values	Values > mean ± 2,5 SD	Values >2000ms
Lexical decision		4,50%	0 %	2,33%	5,04%
Morphological priming	50 ms	6,50%	0%	2,67%	0,44%
	100 ms	2,00%	0,11%	2,15%	4,08%
	150 ms	7,41%	0,27%	2,40%	6,41%

Source: Own elaboration.

Normality tests were also performed -they did not reveal a normal distribution. This is an expected circumstance, since the observation focuses on reaction times - the data collected has a limit on the left, that is, there is always the limit zero (0) that prevents symmetry. Thus, non-parametric statistical tests have been used to analyse the data.

Assessment of the role of the morphological condition

This experiment aimed to evaluate the costs of word visual processing associated to three morphological conditions: regular derivation, derivation with suffix allomorphy and derivation with base allomorphy. Table 3 presents the results by task: lexical decision; priming (50 ms) plus lexical decision; priming (100 ms) plus lexical decision; and priming (150 ms) plus lexical decision.

Table 3 – Descriptive data from the four experiments

Prime exposition		Compositional structure	Suffix allomorphy	Base allomorphy
Lexical Decision	**Mean (SD)**	932,77 (317,93)	1007,66 (356,45)	993,13 (322,84)
	Median	854,00	893,00	896,00
	1st quartile	697,50	727,50	738,00
	3rd quartile	1099,00	1270,00	1208,00
50 ms	**Mean (SD)**	782 (220,21)	908 (310,37)	861 (229,96)
	Median	713,00	795,00	812,50
	1st quartile	662,50	688,00	710,00
	3rd quartile	901,00	1075,50	970,00

Prime exposition		Compositional structure	Suffix allomorphy	Base allomorphy
100 ms	**Mean (SD)**	879 (282,86)	950 (321,40)	920 (272,15)
	Median	802,00	861,00	856,00
	1st quartile	682,00	722,00	731,00
	3rd quartile	994,00	1178,00	1028,00
150 ms	**Mean (SD)**	875 (328,80)	970 (342,13)	921 (290,58)
	Median	761,50	886,00	858,00
	1st quartile	659,00	709,00	704,50
	3rd quartile	1019,00	1176,00	1032,00

Source: Own elaboration.

These four experiments present fairly homogeneous results, and the condition 'compositional structure' (= CS) is always different from the other two. Graphic 1 shows that there are statistically relevant differences between reaction times (marked by braces), in the following cases:

1. Lexical decision experiment
   compositional structure – suffix allomorphy (U=24367,5; p=0,034)
   compositional structure – base allomorphy (U=27062,5; p=0,018)

2. Morphological priming experiment– 50ms
   compositional structure - suffix allomorphy (U=34023,5; p=0,000)
   compositional structure - base allomorphy (U=34122; p=0,000)

3. Morphological priming experiment – 100ms
   compositional structure - suffix allomorphy (U=34308; p=0,009)
   compositional structure - base allomorphy (U=36190; p=0,008)

4. Morphological priming experiment – 150ms
   compositional structure - suffix allomorphy (U=23949,5; p=0,000)
   compositional structure - base allomorphy (U=25324; p=0,004)

Graphic 1 – Significant differences in the different experiments
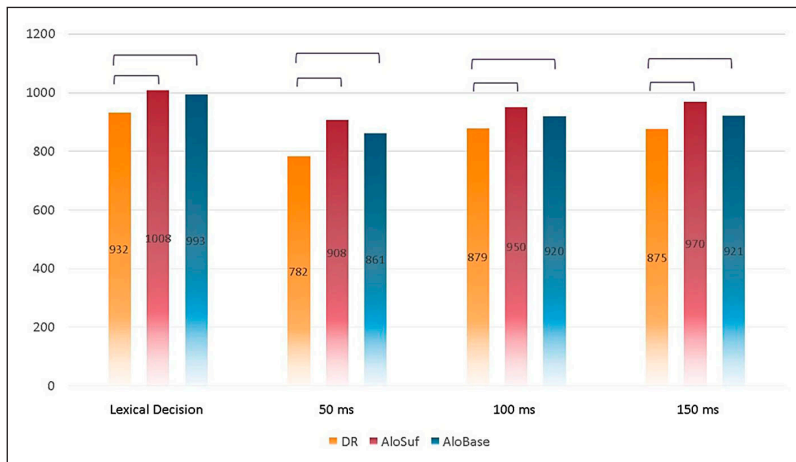
Source: Own elaboration.

These results show that compositional structures are always processed in less time and that, conversely, structures that have a suffix allomorph are those that require longer processing time.

Assessment of the role of priming

The following results evaluate the processing costs due to the existence or non-existence of priming and the different exposure time to prime. Results are presented by condition.

1. Compositional structure *(e.g. orgulho 'proud'/orgulhoso ' proud')*
   There are statistically significant differences in the following reaction times:
   1. Prime exposure of 50 ms – prime exposure of 100ms ($U=34644$; $p=0,000$)
   2. Prime exposure of 50 ms – prime exposure of 150 ms ($U=34196,5$; $p=0,005$).
   3. Prime exposure of 50 ms – lexical decision ($U=28155,50$; $p=0,000$).

Graphic 2 – Significant differences in the condition 'compositional structure'
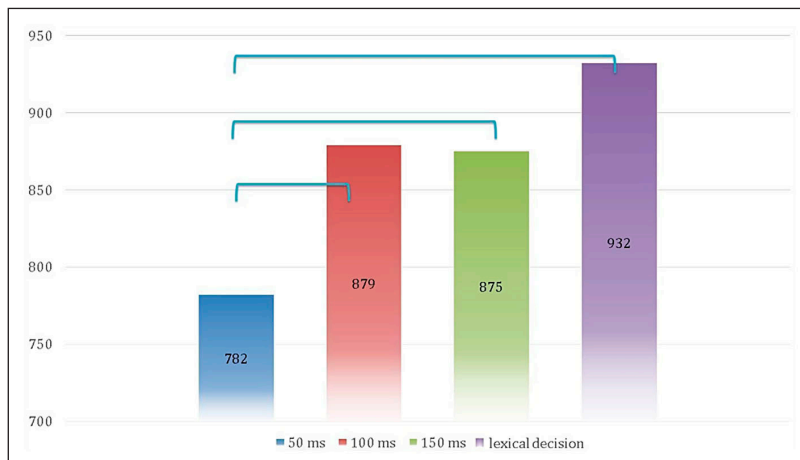
2. Suffix allomorphy (e.g. *defeito* 'defect' / *defeituoso* 'defective')

In this condition there are significant differences in the reaction times, except when the exposure time of 50 ms is contrasted with the exposure time of 100 ms:

1. Prime exposure of 50 ms – prime exposure of 150 ms (U=30093; p=0,034).
2. Prime exposure of 50 ms – lexical decision (U=26072,5; p=0,002).

Graphic 3 – Significant differences in the condition 'derivation with suffix allomorphy'
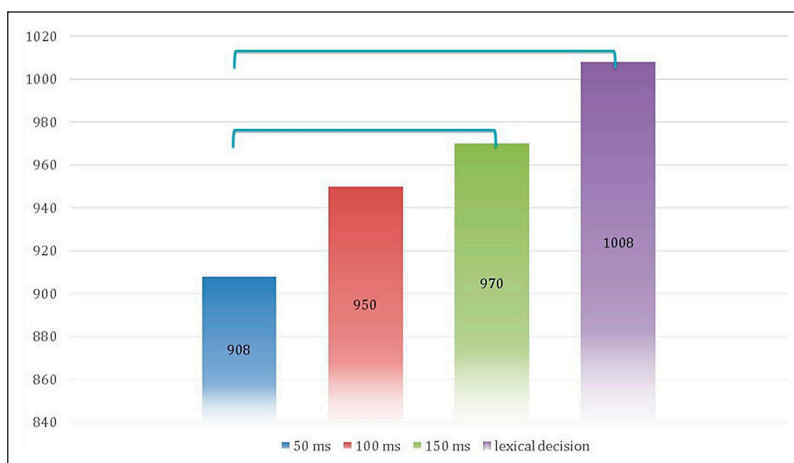
3. Base allomorphy (e.g. *lume* 'fire' / luminoso'bright')

This condition repeats the trends previously found. Significant differences in reaction times are obtained in the following cases:

1. Prime exposure of 50 ms – prime exposure of 100 ms (U=36610; p=0,007).
2. Prime exposure of 50 ms – prime esposure of 150 ms (U=30360; p=0,03).
3. Prime exposure of 50 ms – lexical decision (U=66474; p=0,038).

Graphic 4 – Significant differences in the condition 'base allomorphy'
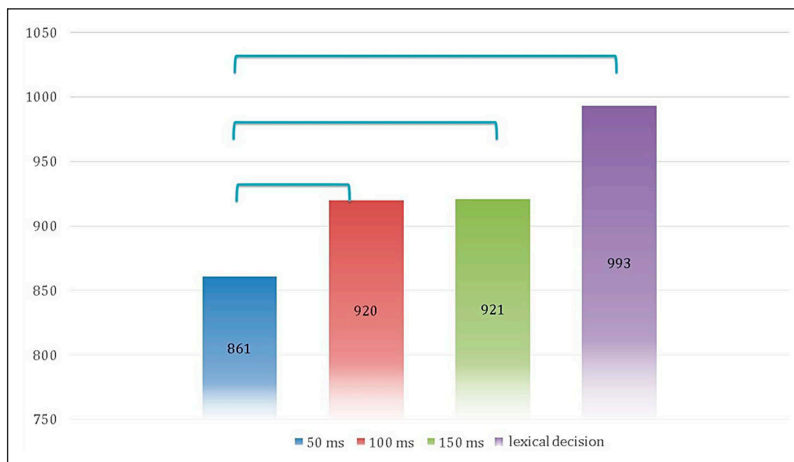
Source: Own elaboration.

Discussion

In the first experiment, which evaluated visual processing costs associated with three morphological conditions (regular derivation, derivation with suffix allomorphy and derivation with base allomorphy), our results indicate that compositional structures are processed in less time than lexicalised words, either by an allomorphy of the base or the suffix. We have also found out that visual processing of structures with suffix allomorphy requires a greater effort than all the other, which can be due to a smaller salience of affixes (perhaps semantic, perhaps formal, but the data do not allow to draw safe conclusions) with respect to roots.

The second set of experiments, which evaluated the costs of visual processing due to the existence or non-existence of priming and different prime exposure times, confirmed that reaction time decreases when there is priming, which means that when morphological analysis is induced, processing costs decrease. On the other hand, the contrast of prime exposure times (50ms, 100ms and 150ms) has also generated some

interesting results, since the reaction time increases with the increase of prime exposure time, irrespective of the morphological condition.

The combination of all these observations allows us to conclude that the processing of derived words reading is facilitated when the presentation of their base forms induces morphological analysis, and it also suggests that the effect is more visible in an initial phase of visual word processing.

Final remarks

This study was driven by a discussion on the nature of morphological structures and on the fragility of current knowledge about its degree of complexity. The opposition between simple and complex words captures a basic distinction between words that are not formed by word formation processes and words that are derived, modified, or compounds. However, it can be easily demonstrated that not all simple words are equally simple and that not all complex words are complex in the same way.

From the point of view of morphological analysis, it is relatively easy to set complexity indexes to words constituents, and to proceed to the calculation of the complexity of each word, but this is merely the output of theoretical stipulations. This is why we decided to look at morphological processing data as a means to validate our linguistic assumptions. We have selected a particular word formation process, the formation of –*oso* derivatives, contrasting compositional structures (cf. *venenoso* 'venomous') to lexicalised structures, with suffix allomorphy (cf. *luxuoso* 'luxurious') and base allomorphy (cf. *medroso* 'fearful').

The results obtained in the four experiments allow us to draw the hypothesis that visual processing of complex words derived by suffixation in Portuguese is sensitive to properties of its morphological structure. They also reveal that compositional structures have lower processing costs than structures disturbed by factors such as suffix allomorphy or base allomorphy. Additionally, we found out that 'suffix allomorphy' has higher processing costs than 'base allomorphy'. This finding is somehow counterintuitive, but it may indicate that words that integrate base allomorphs (cf. *arenoso* 'sandy') are closer to lexicalization (that leads to a direct lexical access) than words that include suffix allomorphs (cf. *luxuoso* 'luxurious'). The latter, which are heaviest for word processing, seem to remain analysable, but require an additional effort to ensure that the suffix is duly recognized.

Another aspect that arises from the results analysis concerns the priming role. In fact, since reaction time decreases with the existence of an exposure to a prime, it may be assumed that processing of derived words always involves morphological analysis tasks, since it is facilitated by the exposure to a base form. On the other hand, reaction time increases in direct proportion of the increase of exposure time to a prime, which seems to indicate that the facilitation factor offered by the presence of a prime is more relevant in an earlier time window than in the later stages of processing.

In sum, the results that we have obtained corroborate the hypothesis that complex compositional words, which contain inherently intransitive roots and the canonical form of a suffix (in this case, -*oso*), have lower processing costs and therefore a lower complexity index than lexicalized complex words. However, the second hypothesis initially considered was contradicted by the results that indicate that lexicalization triggered by suffix allomorphy is more onerous for processing than lexicalization triggered by root allomorphy. In other words, when the subjects identify an inherently transitive root, they give up word morphological analysis, and they process it as a simple word; when the root is positively identified as an intransitive root, although the suffix corresponds to an allomorph with a random distribution, the attempt to analyse the word morphologically lasts a bit longer until that process is abandoned. Presumably, then, the index of complexity of derived words is sensitive to the nature of the process of lexicalization.

- *RESUMO: Embora muitas palavras sejam formadas por mais de um constituinte morfológico, nem todas são habitualmente consideradas palavras complexas. No quadro da análise morfológica do Português, o conceito de 'palavra complexa' divide as palavras formadas por um radical, um constituinte temático e, eventualmente, um ou dois sufixos de flexão, das palavras formadas por estes mesmos constituintes e ainda os que são trazidos pelos processos de derivação, modificação ou composição. Esta distinção é redutora porque todas as palavras contêm algum grau de complexidade, mas não há instrumentos de análise que permitam medi-la. Procuraremos contribuir para a discussão da avaliação da complexidade das palavras com base em dados do processamento da leitura. A literatura apresenta diversas descrições que mostram que a estrutura morfológica desempenha um papel importante no processamento visual. Neste trabalho procuraremos encontrar novas evidências, testando hipóteses relacionadas com a composicionalidade das palavras. Usamos os métodos de priming morfológico e decisão lexical sobre três conjuntos de derivados em –oso: o primeiro é formado por estruturas composicionais, o segundo é constituído por palavras onde ocorre um alomorfe do sufixo e o terceiro contém palavras onde ocorre um alomorfe da forma de base. Os resultados obtidos confirmam que o processamento das palavras derivadas é sensível à sua estrutura morfológica e mostram também que as estruturas composicionais envolvem menores custos de processamento. Estas evidências permitem-nos propor critérios a ter em consideração na avaliação da complexidade das palavras.*

- *PALAVRAS-CHAVE: Complexidade morfológica. Processamento visual. Acesso lexical. Derivação.*

REFERENCES

BANE, M. Quantifying and measuring morphological complexity. In: CHANG, C. B.; HAYNIE, H. J. (Eds.). **Proceedings of the 26thWest Coast Conference on Formal Linguistics**, p.69-76, 2008.

BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHISTIANSEN, M. H.; CROFT, W.; ELLIS, N.; SCHOENEMANN, T. Language Is a Complex Adaptive System: Position Paper. **Language Learning**, n.59, p.1-26, 2009.

FROST, R.; KUGLER, T.; DEUTSCH, A.; FORSTER, K. Orthographic structure versus morphological structure: principles of lexical organization in a given language. **Journal of Experimental Psychology**, n.31, p.1293-1326, 2005.

GONG, T.; COUPÉ, C. A report on the workshop on complexity in language: developmental and evolutionary perspectives. **Biolinguistics**, n.4, p.370-380, 2011.

JÄRVIKIVI, J.; BERTRAM, R.; NIEMI, J. Affixal salience and the processing of derivational morphology: the role of suffix allomorphy. **Language and Cognitive Processes**, n.21, p.394-432, 2006.

LAUDANNA, A.; BURANI, C. Distributional properties of derivational affixes: implications for processing. In: FELDMAN, L. (Org.). **Morphological Aspects of Language Processing**. Hillsdale, NJ: Erlbaum, 1995. p.345-364.

MCQUEEN, J.; CUTLER, A. Cognitive processes in speech perception. In: HARDCASTLE, W.; LAVER, J.; GIBBON, F. (Eds.). **The Handbook of Phonetic Sciences**. Oxford: Blackwell, 1997. p.489-520.

MUFWENE, S. S. The emergence of complexity in languages: an evolutionary perspective. In: MASSIP-BONET, A.; BASTARDAS-BOADA, A. (Eds.). **Complexity Perspectives on Language, Communication and Society**, p.197-281, 2012.

PINTO, C. **O papel da morfologia no processamento da palavra escrita**. 2017. 259 f. Dissertação (Doutoramento em Voz, Linguagem e Comunicação) – Faculdade de Letras da Universidade de Lisboa, Lisboa, 2017.

RASTLE, K.; DAVIS, M.; MARSLEN-WILSON, W.; TYLER, L. Morphological and semantic effects in visual word recognition: a time course study. **Language & Cognitive Processes**, n.15, p.507-537, 2000.

RASTLE, K.; DAVIS, M.; NEW, B. The broth in my brother's brothel: morpho-ortografic segmentation in visual word recognition. **Psychonomic Bulletin & Review**, n.11, p.1090-1098, 2004.

SIMON, H. A. The architecture of complexity. **Proceedings of the American Philosophical Society**, n.106, p.467-482, 1962.

STEELS, L. The Synthetic Modeling of Language Origins. **Evolution of Communication Journal**, n.1, p.1-3, 1997.

STEELS, L. Language as a Complex Adaptive System. In: SCHOENAUER, M.; DEB, K.; RUDOLPH, G.; YAO, X.; LUTTON, E.; MERELO, J. J.; SCHWFEL, H.-P. (Eds.). **Proceedings of the 6th International Conference on Parallel Problem Solving from Nature**, n.1917, p.17-26, 2000.

TAFT, M. Recognition of affixed words and the word-frequency effect. **Memory and Cognition**, n.7, p.263-272, 1979.

_____. Interactive-activation as a framework for understanding morphological processing. **Language and Cognitive Processes**, n.9, p.271-294, 1994.

TAFT, M.; FORSTER, K. Lexical storage and retrieval of prefixed words. **Journal of Verbal Learning and Verbal Behavior**, n.14, p.638-647, 1975.

VELAN, H.; FROST, R. Words with and without internal structure: what determines the nature of orthographic and morphological processing? **Cognition**, n.118, p.141-156, 2011.

VILLALVA, A. **Estruturas morfológicas. Unidades e hierarquias nas palavras do Português**. Lisboa: FCG, 2000.

_____. **Morfologia do Português**. Lisboa: Universidade Aberta, 2008.

_____. Bare morphology. **Revista de Estudos Linguísticos da Universidade do Porto**, v.8, p.121-141, 2013.

VILLALVA, A.; SILVESTRE, J. **Introdução ao Estudo do Léxico. Descrição e Análise do Português**. Petrópolis: Vozes, 2015.

CONJUGATION OF IRREGULAR VERBS BY NATIVE BRAZILIAN PORTUGUESE SPEAKING CHILDREN: A STUDY IN THE LIGHT OF LEXICAL PHONOLOGY AND MORPHOLOGY

Tamires Pereira Duarte GOULART[*]
Carmen Lúcia Barreto MATZENAUER[**]

- ABSTRACT: This study investigates the acquisition of irregular verbs in Brazilian Portuguese (BP) when conjugated in the Indicative Present, Subjunctive Present and Simple Past tenses by native BP speaking children. It aims at describing and analyzing the interaction between Morphology and Phonology which is found in the inflection of irregular verbs, in the light of Lexical Phonology and Morphology (LPM). Data show that linguistic phenomena of morphophonological origin, which are highly complex, can be considered processes of late acquisition by Brazilian children. LPM leads to the understanding that verb regularization, one of the manifestations of irregular conjugations, can be explained by the non-correspondence between the phonological and morphological levels of the language.

- KEYWORDS: morphophonological acquisition; irregular verbs; Lexical Phonology and Morphology

Preliminary Considerations

In the process of language acquisition, from the input they receive, children must abstract the grammar of the language, that is, the units that constitute each component of the linguistic system, such as phonology, morphology and syntax, as well as the relations among them. Relations can occur either between units within the same component or between units of different components. The latter implies interfaces between the components that make up the grammar of the language.

Considering relations between the components of the grammar, this study focuses on the interface that is established between Phonology and Morphology and aims at the behavior, in the phase of language acquisition, of the adjunction of inflectional

* Catholic University of Pelotas (UCPEL), Postgraduate Program of Letters, Pelotas – Rio Grande do Sul – Brazil. tamirespdgoulart@gmail.com

** Catholic University of Pelotas (UCPEL), Postgraduate Program of Letters, Pelotas – Rio Grande do Sul – Brazil. carmen.matzenauer@gmail.com

morphemes to irregular verb forms of Brazilian Portuguese (BP) in cases with alternations in the phonological plane of the language.

The aim of this paper is to discuss the use of forms of the Indicative Present, Simple Past and Subjunctive Present, which, in verbs classified as 'irregular', show alternation of consonant phonemes. Examples of such verbs are *fazer*, *trazer* and *medir* ('to do', 'to bring' and 'to measure'), which, in the previously mentioned tenses, show the following alternations: [s] alternates with [z], as in *faço*, *fazes* (*fa[s]o*, *fa[z]es*); [g] alternates with [z] and [s], as in *trago*, *trazes*, *trouxe* (*tra[g]o*, *tra[z]es*, *trou[s]e*); [s] alternates with [d], as in *meço*, *medes* (*me[s]o*, *me[d]es*).

Research data, which was collected by an instrument that elicited the use of forms in the verb tenses under study, were analyzed and formalized with the support of Lexical Phonology and Morphology, since it is a theory that predicts the interaction between Morphology and Phonology, pointing out relations between the morphological structure of a word and the phonological rules that apply to it.

Irregular verbs and consonantal phonological alternations

The morphological component of Brazilian Portuguese allows regular and irregular verbs. For their characterization, it should initially be pointed out, following Câmara Jr. (1970), that a verbal is composed of a theme (T), formed by a root (R) and a theme vowel (TV), besides inflectional suffixes (FS), which can be of mood and tense (SMT) and number and person (SNP). Considering this structure, verbs are said to be regular when they keep an unchanged root and follow the general pattern in the adjunction of inflectional morphemes; it is the case of the verb *amar* ('to love') in all its conjugations: *amo, amava, amarei, amasse.*

Differently, irregular verbs, the focus of this paper, are those in which there is a change in the root when they are conjugated. Examples are the verb *dizer* ('to say'), and its inflections *digo* and *dizes* (*di[g]o*, *di[z]es*), and verbs that deviate from the paradigm imposed by the grammar of the language, thus, showing changes in inflectional morphemes. The verb *saber* (*to know*), whose conjugation in the Indicative Present has the forms *sei*, *sabes*, *sabe*, is an example of this type of irregularity. The verb form "*sei*" moves away from the paradigm of the first person of the singular, conjugated in the Indicative Present, that usually manifests with the personal-number suffix -o, as in *canto*, *danço*, *escrevo*, *faço*, *estudo*.

In irregular verbs, either due to changes in the root or because of escape to the paradigm, alternations may occur. In the functioning of the language, the phenomenon of alternation is characterized by replacing a phoneme by another and there are two types of alternations: vowel alternation, which implies the exchange of vowels, for example, [o]vo ~ [ɔ]vos, and the consonantal alternation, which is understood as the replacement of a consonantal phoneme with another one. Examples are consonantal changes at the end of roots of conjugated verbs, as in *fazer ~ faço ~ fazes* (*fa[z]er*, *fa[s]o*, *fa[z]es); trazer ~ trouxe ~ traz* (*tra[z]er*, *trou[s]e*, *tra[s]*).

The previously described consonantal alternation seems to be highly complex for BP speakers. As a result, it is argued that its acquisition can be considered a process of late nature in the acquisition of the BP irregular verb system of BP that extends, at least, to the age of 10.

The process of regularization of irregular verbs

The tendency towards regularization and at less marked in the functioning of the grammar of languages is recognized as part of the evolutionary process of linguistic systems, as well as of the process of language acquisition. In this context, the regularization of verb forms is inserted.

Therefore, there is regularization of irregular verbs when the speaker produces an irregular verb in agreement with the pattern of the regular verb. Examples are the production of "*fazi*" instead of "*fiz*" ('did') and "*trazi ~ trazei*" instead of "*trouxe*" ('brought').

Lexical Phonology and Morphology (LPM)

Lexical Phonology and Morphology (LPM) is a theory that enables researchers to look at language while considering the relationship between morphology and phonology. As a result, generalizations and patterns of morphophonological occurrences can be singularly captured.

Linguistic systems show cases of interface among the components of the language. So, neither in the phonological scope nor in the morphological one alone, would it be possible to explain them. An example, in this context, is the phenomenon under study, which puts into analysis the production of irregular verbs that exhibit phonological alternations in their conjugations motivated by the adjunction of morphology units.

In the 1980s, Lexical Phonology gained prominence. The pioneers were Kiparsky (1982, 1985) and Mohanan (1982), but many phonologists quickly adopted it because it possesses high level of theoretical explanation, besides being responsible for making morphology return to the setting of language studies (SCHWINDT, 2006).

LPM is dedicated to look at the lexicon not only as a keeper of structures, but also as a set of phonological rules that communicate with morphological ones. One of its greatest contributions is the understanding that the lexicon of the language is organized in a series of levels or strata which are responsible for the occurrences of these rules.

Within each stratum, both morphological rules of word formation and phonological processes are applied pairwise. Strata are arranged so as to reflect the ordering of the word formation processes, as shown in Figure 1.

Figure 1 – Model proposed by Kiparsky (1982):
representation of the lexicon structure.
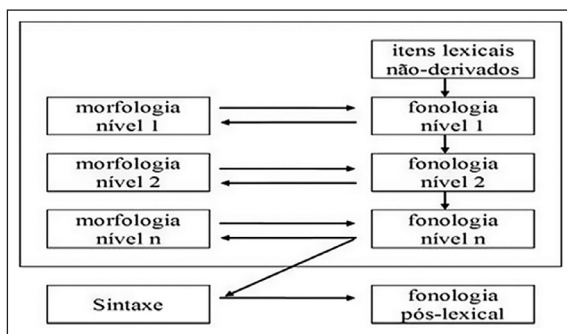
Source: Lee (1992).

The model shown in Figure (1), which was introduced by Kiparsky (1982), considers the entry of the lexicon by the phonology module. From that point on, relations between phonological and morphological strata begin to be established. Finally, the word leaves the lexical level through phonology and goes directly to the syntax, at the post-lexical level.

Lee (1995) emphasizes that the components of phonology and morphology are mixed, so that relevant phonological rules apply to the output of any morphological rule, determining entry to another phonological rule, and so on.

In the analysis of Portuguese, Bisol (2010) proposes the division of lexicon into two levels, i. e., the root and the word, identified as Level 1 and Level 2, respectively. This structure, shown in Figure (2), is the most used one in most LPM proposals, although it is the option of each language to divide lexicon into different levels.

Figure 2 – Model suggested by Bisol (2010) for Portuguese.
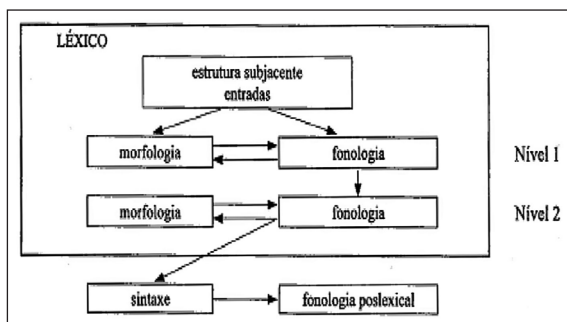
Source: Bisol (2010, p. 92).

Bisol's proposal (2010) introduces the LPM model with two possibilities from the underlying structure. Input can happen either through the morphology module or through the phonology module whereas the output of the word from the lexical module to the post-lexical module is provided directly by phonology and follows the path to the syntax. In this level, it undergoes post-lexical phonological rules.

In order to analyze the data of this study, its results are formalized with the support of LPM, considering essentially Lee's proposal (1995) for BP. On this basis, this study advocates that the process of regularization of verb forms in the speech of 6-9-year-old Brazilian children is due to the lack of correspondence between Phonology strata and Morphology strata.

Methodology

The study was carried out with a group of 6-9-year-old native BP speakers – from Rio Grande do Sul (RS) state, Brazil, – who speak a dialect variety that includes the use of the personal pronoun 'tu' ('you'). It focused on the production of irregular verbs and aimed to identify whether the process of regularization of such verbs can still be found in this age group. When it occurred, the study investigated the motivation of the phenomenon within the assumptions of Lexical Morphology and Phonology.

The corpus consisted of data on eight native BP speaking children. All were literate and monolingual and belonged to one of the age groups listed in Table 1. In addition, children's ages were linked to their schooling level. Table 1 shows the selection of informants regarding schooling level and age[1].

Table 1 – Subjects' age groups and schooling levels.

Age Group 1 (FE 1):	06 years old	2 subjects	2nd year - Elementary school I
Age Group 2 (FE 2):	07 years old	2 subjects	3rd year - Elementary School I
Age Group 3 (FE 3):	08 years old	2 subjects	4th year - Elementary school I
Age Group 4 (FE 4):	09 years old	2 subjects	5th year - Elementary school I

Source: Goulart (2015).

In each group of selected ages, children were subdivided with respect to sex, i. e., a girl and a boy composed each age group, in order to be able to establish relations between the gender and possible peculiarities and/or similarities in the process of acquisition of the irregular verb morphology of BP.

---

[1] Because this paper is a part of a Master's thesis, it has only some data of the research corpus and the description of only one of the instruments created for data collection.

To collect data, an instrument, which focused on the production of irregular verb morphemes in the 1st and 2nd singular persons of the Indicative Present, Simple Past and Subjunctive Present, was created. The instrument consisted of sentences with gaps to be completed by the informants with appropriate forms of the verb highlighted in red, as shown in Figure 3.

All sentences led to the production of the expected verb tense. The test was presented on slides, on the screen of a laptop (model shown in Figure 3). Two types of slides were exhibited: one that highlights the verb to be conjugated and another one with sentences whose gaps had to be completed with the inflected verb.

Figure 3 – Model of Instrument I[2].
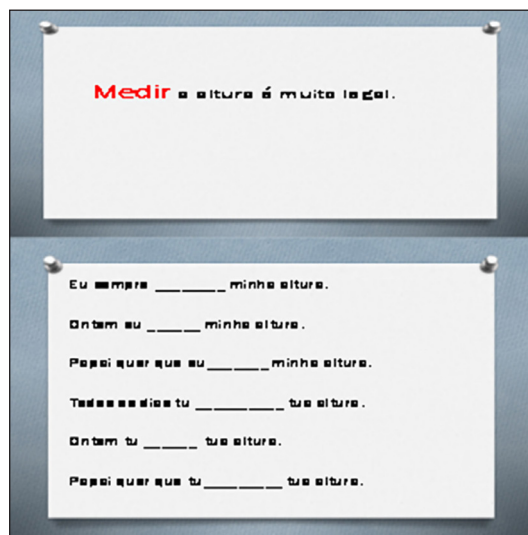
Source: Goulart (2015).

Initially, the researcher showed the first slide, containing the sentence with the verb to be used. She asked the child to read the sentence and then, moving to the next slide, asked the child to read and complete the sentences. Throughout the application, the researcher tried to interact constantly with the informants. Preliminarily, a familiarization test, similar to that of the instrument, had been applied, but with sentences containing only regular verbs.

---

2   "*To measure height is cool*".
    I always ___ my height.
    I ___ my height yesterday.
    Daddy wants me to___ my height.
    You___ your height everyday.
    You___ your height yesterday.
    Daddy wants you to___ your height.

The instrument also included slides that aimed at distracting the children, i. e., they had sentences with regular verbs. In addition, it should be emphasized that the disposition of sentences in the test was randomly done to avoid the sequence of verb conjugation in the order of grammatical persons, such as: *eu faço, tu fazes; que eu faça, que tu faças* ('I do', 'you do'; 'me to do', 'you to do'). When, in the ordering of sentences, the sequence of pronouns *eu - tu* ('I - you') appeared, the verb tense was alternated.

The instrument had 10 slides which displayed irregular verbs. Each one had 06 different sentences with gaps to be filled. Thus, each informant had to fill 60 spaces with the oral use of irregular verbs. Regular verbs, whose presence in the instrument only occurred as a distraction from the focus of the research, totaled 48 regular verb productions, distributed in 08 slides. Table 2 shows the irregular verbs selected for the research proposed in this paper.

Table 2 – Irregular verbs and their alternations analyzed by the Master's thesis.

Irregular Verbs		Consonantal Phonological Alternation
Dizer	('to say')	[z], [g], [s]
Fazer	('to do')	[z], [s]
Satisfazer	('to satisfy')	[z], [s]
Trazer	('to bring')	[z], [g], [s]
Poder	('can')	[d], [s]
Ter [3]	('to have')	[], [v]
Perder	('to lose')	[d], [k]
Medir	('to measure')	[d], [s]
Ouvir	('to hear')	[v], [s]
Pedir	('to ask')	[d], [s]

Source: Goulart (2015).

In irregular verbs, whether by root change or by fugue to the paradigm, phonological alternations can occur (shown in Table 2). They are characterized by replacing a phoneme with another one. The principle of phonological alternations is maintained by this study, since one can observe the replacement of a consonant with another one. For example, in *medir* '(to measure'), one of the possibilities of conjugation is the replacement of consonant "d" with consonant "s". This fact occurs in the root of the verb and, due to this fact, a phenomenon of morphological order is also manifested. In view of this fact, LPM was brought to the analysis since it is a theory that can explain linguistic processes that deal with units of phonology and morphology at the same time.

---

[3] Consonant alternations related to the verb "ter" are of a differentiated nature in this list of verbs, since they refer to the inflection paradigm, whereas alternations recorded for the other verbs occur in the root of the word.

According to Câmara Júnior (1970), consonantal alternation is constituted by variation of the root, which contributes to express grammatical notions of tense, mood and person that are primarily represented by suffixes. Souza, Silva and Koch (2009) state that it is this type of irregularity that allows to distinguish deviant morphological "patterns", since one of the characteristics of regular verbs is root immutability.

Results and Data Analysis

In order to carry out data description, descriptive tables with information about every informant were outlined. It should be emphasized that, at that moment, standard and nonstandard forms, according to the BP irregular verb conjugation, were considered. Therefore, the focus is to identify nonstandard forms, which are manifested through the regularization process and do not show consonantal alternations which are necessary for the inflections of irregular verbs, especially in the Indicative Present and Subjunctive Present.

Thus, it should be pointed out that the consonantal alternation corresponds to the form which, in this paper, is referred to as Standard (P)[4], whereas its non-occurrence, which corresponds to the production of the regularized verb form, is interpreted as Nonstandard (NP). When NP productions are shown in the tables, the phonetic transcription of the consonantal segment is also identified. It evidences the use of the verb form in disagreement with the target language, that is, it implies the escape of the morphophonological alternation displayed by the verb conjugation which is considered standard[5].

Informant 1

Informant 1/ FE 1/ F1		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	NP fa[z]o	NP satisfa[z]o	NP tra[z]o	P	NP per[d]o	P	NP mɛ[d]o	NP ou[v]o	NP pɛ[d]o
	2nd person	P	P	P	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	P	NP tra[z]i	P	P	P	P	P	P
	2nd person	P	P	P	NP tra[z]eu	P	P	P	P	P	P
Subjunctive Present	1st person	P	NP fa[z]a	NP Satisfa[z]a	P	P	NP per[d]a	P	NP mi[d]a	NP ou[v]a	NP pɛ[d]a
	2nd person	P	P	NP Satisfa[z]a	P	P	NP per[d]a	P	NP mi[d]a	NP ou[v]a	NP Pe[d]a vogal média baixa

---

Informant 1, 06 years of age, female, presented regularizations in the conjugations of the irregular verbs "*fazer, satisfazer, trazer, perder, medir, ouvir* and *pedir*". Regarding the verb "*fazer*" ('to do') and its derivative "*satisfazer*" ('to satisfy'), the informant produced regularized morphemes, which were faithful to the root of the verb *faz-*; as a result, conjugations *fa[z]o, fa[z]a, satisfa[z]o* and *satisfa[z]a* were observed. Thus, the girl did not perform the irregular conjugation, which would occur for the production of alternations with [s], found in *fa[s]o, fa[s]a, fi[s]*, both in the Indicative and Subjunctive.

Regularizations, which also maintain faithfulness to the base-morpheme of the verb, such as *tra[z]o, tra[z]i, tra[z]eu*, were some of the conjugations that this informant carried out. Thus, the /z/ phoneme alternated neither with [g] nor with [s] in the Indicative mood, but alternated with [g] in the Subjunctive mood. Because the Subjunctive mood is less frequent, the standard use may be interpreted as the result of an unanalyzed form. Construction of the forms *per[d]o* and *per[d]a*, rather than *per[k]o* and *per[k]a*, was also found.

The expected alternation for the verb "*ouvir*" ('to hear'), of form [v] with [s], was not performed by the child. On the contrary, what prevailed was the regularization and faithfulness to the base form of the root *ouv-*. Such phenomenon was observed in *ou[v]o* and *ou[v]a*, examples of the informant's production.

Informant 2

Informant 2/ FE1/ M1		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	P	NP satisf[eito]	P	P	NP per[d]o	P	NP mi[d]o	NP Ou[v]o	NP pɛ[d]o
	2nd person	P	P	P	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	P	P	P	P	P	P	P	P
	2nd person	P	P	P	P	P	P	P	P	P	P
Subjunctive Present	1st person	P	P	P	P	P	NP per[d]a	P	NP mi[d]a	NP ou[v]a	P
	2nd person	P	P	P	P	P	NP per[d]a	P	NP mi[d]a	NP o[v]a	P

Informant 2, 06 years of age, male, conjugated the verbs "*dizer, fazer, trazer, poder* and *ter*" in agreement with the standard grammar of the language, alternating consonantal phonemes. However, in relation to the verbs "*satisfazer, perder, medir, ouvir* and *pedir*", the child produced regularized conjugations.

Concerning conjugations of the verb *"perder"* ('to lose'), *per[d]o* and *per[d]a* were produced. They are forms which were regularized by the child in both Indicative and Subjunctive. As for the verb *"ouvir"* ('to hear'), the forms *ou[v]o, ou[v]a* and *o[v]a* were constructed. For the verb *"medir"* ('to measure'), Informant 2 used *mi[d]o* and *mi[d]a*, with vowel reduction [e → i]. With regard to the verb "*pedir*" ('to ask'), the child operated with the characteristic alternations of the Subjunctive Present and produced *pɛ[d]o* in the Indicative Present, a regularization that underwent the process of vowel alternation, but without effecting the consonantal alternation referring to /d/ for [s].

Informants 3, 4, 5, 6, 7 and 8, whose data are shown below, produced regularized verb forms that show the same nature as those previously mentioned.

Informant 3

Informant 3/ FE2/ F1		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	P	NP satisfa[z]o	NP tra[z]	P	NP per[d]o	P	NP mɛ[d]o	NP ou[v]o	NP pɛ[d]o
	2nd person	P	P	P	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	NP satisfa[z]i	P	P	P	P	P	P	P
	2nd person	P	P	NP satisfa[z]esse	P	P	P	P	P	P	P
Subjunctive Present	1st person	P	P	NP satisfa[z]a	NP tra[z]a	P	NP per[d]a	P	NP mɛ[d]a	NP ou[v]a	NP pɛ[d]a
	2nd person	P	P	P	NP tra[z]a	P	NP per[d]a	P	NP mi[d]a	NP ou[v]a	NP pɛ[d]a

Informant 3, 07 years of age, female, made regularizations for the verbs "*satisfazer, trazer, poder, perder, medir, ouvir* and *pedir*". The verbs "*dizer, fazer* and *ter*" were the only ones that did not undergo this phenomenon.

Informant 4

Informant 4/ FE2/ M1		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	P	NP satisfa[z]o	NP tra[z]o	P	NP per[d]o	P	P	NP Ou[v]o	P
	2nd person	P	P	P	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	P	P	P	P	P	P	P	P
	2nd person	P	P	P	P	P	P	P	P	P	P
Subjunctive Present	1st person	P	P	NP satisfa[z]a	NP tra[z]a	NP po[d]a	NP per[d]a	P	NP mi[d]a	NP ou[v]a	NP pi[d]a
	2nd person	P	P	NP satisfa[z]a	NP tra[z]a	P	NP per[d]a	P	NP me[d]a	NP ou[v]a	NP pe[d]a

Informant 4, 07 years of age, male, had the regularization of verbal morphemes for the verbs "*satisfazer, trazer, poder, perder, medir, ouvir* and *pedir*" in his productions.

Informant 5

Informant 5/ FE3/ F		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	P	P	P	P	P	P	P	NP o[v]o	P
	2nd person	P	P	P	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	P	P	P	P	P	P	P	P
	2nd person	P	P	P	P	P	P	P	P	P	P
Subjunctive Present	1st person	P	P	P	P	P	P	P	NP mɛ[d]a	NP o[v]a	NP pi[d]a
	2nd person	P	P	P	P	P	P	P	NP me[d]a	NP ou[v]a	NP pe[d]a

Informant 5, 08 years of age, female, did not apply the morphophonological rules that the irregular conjugations of the verbs "*medir, ouvir* and *pedir*" require.

Informant 6

Informant 6/ FE3/ M		dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Indicative Present	1st person	P	NP fa[z]o	NP satisfa[z]o	P	P	NP per[d]o	P	NP me[d]o	NP ou[v]o	NP pɛ[d]o
	2nd person	P	P	NP satisfa[z]eu	P	P	P	P	P	P	P
Indicative Simple Past	1st person	P	P	NP satisfa[z]i	P	P	P	P	P	P	P
	2nd person	P	P	NP satisfa[z]esse	NP tra[z]i	P	P	P	P	P	P
Subjunctive Present	1st person	P	NP fa[z]a	NP satisfa[z]a	P	P	P	P	NP mi[d]a	NP ou[v]a	NP pɛ[d]a
	2nd person	P	NP fa[z]a	NP satisfa[z]a	P	P	P	P	NP me[d]a	NP ou[v]e	NP pi[d]a

Informant 6, 08 years of age, male, only conjugated the verbs "*dizer, poder* and *ter*" in agreement with the language pattern.

As for the other verbs of the study, Informant 6 distanced himself from the pattern in the production of some inflections and produced them in regularized form, without the use of the morphophonological rule. Data on Informant 6, as well as data on the other children, indicate that the child, who is still in the phase of acquisition of the morphophonology of the language, reveals floating morphological characteristics by comparison with the standard grammar of the target language. These are features that may either show up properly or may be far from the pattern.

Informant 7

Informant 7/ FE4/ F			dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Instrument I	Indicative Present	1st person	P	P	P	P	P	P	P	P	P	P
		2nd person	P	P	P	P	P	P	P	P	P	P
	Indicative Simple Past	1st person	P	P	P	P	P	P	P	P	P	P
		2nd person	P	P	P	P	P	P	P	P	P	P
	Subjunctive Present	1st person	P	P	P	P	P	P	P	NP mi[d]a	P	P
		2nd person	P	P	P	P	P	P	P	NP mɛ[d]a	P	P

Informant 7, 09 years of age, female, used regularized verb conjugations for the verb *"medir"* ('to measure') in the Present Subjunctive; both forms *mi[d]a* and *mɛ[d]a* show that, although the child's grammar is almost reaching the target system of the standard language, the instability generated by the complexity of the acquisition of BP

morphology is still present at the age of 09. They are conjugations that are produced in regularized form, especially in the inflections in the Subjunctive mood of the verb system.

Informant 8

Informant 8/ FE4/ M			dizer	fazer	satisfazer	trazer	poder	perder	ter	medir	ouvir	pedir
Instrument I	Indicative Present	1st person	P	P	NP satisfa[z]o	P	P	P	P	NP mɛ[d]o	NP o[v]o	NP pɛ[d]a
		2nd person	P	P	P	P	P	P	P	P	P	P
	Indicative Simple Past	1st person	P	P	P	P	P	P	P	P	P	P
		2nd person	P	P	P	P	P	P	P	P	P	P
	Subjunctive Present	1st person	P	P	NP satisfa[z]a	P	P	P	P	NP mi[d]a	NP o[v]a	P
		2nd person	P	P	NP satisfa[z]a	P	P	P	P	NP mɛ[d]a	NP o[v]a	P

Informant 8, 09 years of age, male, produced inflections whose roots did not comply with the rules of consonantal alternations required by the irregular Portuguese verb pattern for the verbs *satisfazer, medir, ouvir* and *pedir*.

Like the other children in this study, this informant regularized some conjugations – that pertain to the standard of PB regular verbs –, guided by the root of the irregular verb in the infinitive. The forms *satisfa[z]o, satisfa[z]a, m[ɛd]o, m[id]a, m[ɛd]a, o[v]o, o[v]a* and *p[ɛd]a* are examples of this regularization.

Data in the light of Lexical Phonology and Morphology

All informants performed some verb inflection in disagreement with the standard norm of conjugation of the irregular verb category. Table 3 shows how often regularizations of each verb under investigation were produced by the children.

Table 3 – Estimates of the regularization process.

FORMS OF NONSTANDARD (NP) PRODUCTION		
Verb	**The use of verb forms**	
	Instrument I	
	Number of occurrences	**Production**
DIZER	100% production in agreement with P (standard)	
FAZER	02	fa[z]o
	03	fa[z]a
SATISFAZER	05	satisfa[z]o
	01	satisfa[z]ei
	01	satisfa[z]eu
	01	satisf[eito]
	02	satisfa[z]i
	03	satisfa[z]esse
	11	satisfa[z]a
TRAZER	05	tra[z]o
	04	tra[z]i
	02	tra[z]eu
	01	tra[z]
	01	tra[z]ei
	07	tra[z]a
PODER	01	po[d]a
PERDER	10	per[d]o
	16	per[d]a
TER	100% production in agreement with P (standard)	
MEDIR	06	mɛ[d]o
	01	me[d]o
	03	mi[d]o
	19	mi[d]a
	05	mɛ[d]a
	06	me[d]a
OUVIR	07	ou[v]o
	02	ou[z]o
	04	o[v]o
	18	ou[v]a
	03	ou[z]a
	05	o[v]a
PEDIR	06	pɛ[d]o
	01	pi[d]o
	05	pe[d]a
	07	pɛ[d]a
	05	pi[d]a

Source: Goulart (2015).

Data show that the most recurrent regularizations in the informants' conjugations were forms *per[d]o, per[d]a* (verb - *perder*), *m[id]a* (verb - *medir*), *p[ɛd]a, p[id]a* (verb - *pedir*), *ou[v]a* (verb - *ouvir*) and *satisfa[z]a* (verb –*satisfazer*). The Subjunctive mood (Subjunctive Present) has greater instability as to its inflection, in high percentage by comparison with productions in the Indicative Present. This fact is believed to be related to the complexity that involves the acquisition of the morphophonology of BP and to the fact that it is the tense and the verb mood (Subjunctive Present) that is considered marked in the language. The more marked the verb mood in the language, the greater the possibility of variation in relation to the standard form. The Subjunctive mood is also used in subordinate sentences; it implies the construction of a period composed of subordination and requires great syntactic complexity to the point that, in the oral use of the language, it is seldom used, even by adults.

The Indicative Simple Past had low index of regularization, which is attributed to the fact that this is a tense that demands few consonant changes in its inflections in the Indicative mood. Moreover, its use does not require syntactic complexity either, since it can either be employed in simple periods or composed of coordination, which is often used in the language.

Regarding the Indicative Present, data can be considered relevant, since they are based on previous studies (YAVAS; CAMPOS, 1988; SANTOS; SCARPA, 2003) which consider this tense as one of the first to be acquired by native BP speakers. Thus, it may be inferred that, although it is a tense which is early used by the subjects, especially in the regular class of verbs, the irregular verb forms of the Indicative Present that are carried out with consonantal alternations are complex for children in the acquisition phase, a fact that contributes to their late emergence[6].

Difference in non-standard production between boys and girls was below 1%, a fact that leads to the understanding that verb regularization is not related to the subjects' gender. In view of it, it is evident that the use of morphophonological events in the conjugation of irregular verbs is found in the linguistic development of all children in the sample, regardless of the gender, unlike other language acquisition processes in which girls and boys may present different behavior, such as the acquisition of the segmental inventory of language phonology.

The corpus under analysis also reveals that there is gradual development of irregular verb acquisition in relation to age, which is also superimposed on the subjects' schooling level. The older the group (and the higher the schooling level), the more the process of producing different forms from the standard (NP) decreases.

This relationship proposes that irregular verb forms of BP are acquired as the child's exposure to the language system increases. So, the older the subjects, the lower the regularization rate they produce.

---

[6]   Data described in this paper also point out that the phenomenon of verb regularization occurs at higher percentage with verbs of the third conjugation (-ir). The determining factor that explains these occurrences was not investigated in detail, but it is emphasized that instruments have been developed to investigate this influence, as well as to highlight the possible differences that exist in irregular verb acquisition, when dealing with different verb moods.

The frequency of use of verbs was a factor that was of particular relevance to this analysis. In this context, data show that non-consonantal alternation occurs, in a prevalent way, when dealing with the less frequent verbs in the informant' linguistic input, as Andersen (2008) had already pointed out. He stated that the verbs "*ter, poder, dizer* and *fazer*" are among the ten most frequent BP verbs. In fact, in this research, these were the verbs that least suffered the process of verb regularization. Besides, both verbs "*ter* and *dizer*" were not inflected any time in a nonstandard way, whereas the verbs "*poder* and *fazer*" were targets of a very low number of regularized productions.

In accordance with Lexical Phonology and Morphology, there are at least three explanations for the process that affects NP conjugations, that is, regularized conjugations of verb forms, evidenced by this paper:

a) the assumption that the subjects have not yet fully acquired the Verbal Morphology of the language and, because of it, present a blockade at the morphological level;
b) the assumption that the lack of alternations in the conjugation of irregular roots is motivated by a phonological gap, directly linked to the phonological level;
c) the assumption that the child has already acquired both the Verbal Morphology of BP and its Phonology and, thus, has regularized verbal morphemes due to the motivation required by the morphological relationship that, in the lexical component, requires interaction between Morphology and Phonology.

Firstly, regarding (b): it does not seem to account for verb forms produced differently from the standard, since the segmental and syllabic inventory of the child with typical development and most rules that determine its functioning in the language are acquired until the age of 05 (LAZZAROTTO-VOLCÃO, 2009).

Considering (a), it seems to be discarded, because data collected by this study show that the informants have already acquired the Verbal Morphology, since they master, in accordance with the language pattern, the morphemes of tense and mood and of number and person, that is, all verb inflections, including the verbal paradigm. Examples of this fact are the inflections *satisfa[z]o* and *per[d]a*, which reveal that informants retain, in the different forms of the pattern, the tense-indicating morphemes, as well as those of number and person. As a result, it could be understood that Verbal Morphology in all age ranges under study has already been acquired. It seems, therefore, that inflectional morphemes of the class of verbs, corresponding to the categories under investigation by this study, have already been acquired. In addition, all children, in all age groups under study, used the forms of the verb *ter* ('to have') at 100% adequacy.

The alternative proposed in (c) is able to explain the phenomenon of nonstandard occurrences, demonstrating, through LPM (Figure 3), that regularizations are motivated by the lack of correlation that must be maintained between Morphology and Phonology so that the process of alternations in irregular verbs follows the pattern of the language. The phonological rule of alternation is not applied to the conjugation of some verbs, because

there is no adequate match between morphological and phonological components and, therefore, the morphophonological interaction is not carried out.

Figure 4 shows lack of match between phonological and morphological components, based on Kiparsky's proposal (1982, 1985), in order to understand the forms produced by the subjects in this research.

Figure 4 – Proposal for nonstandard forms, based on Kiparsky (1982, 1985).
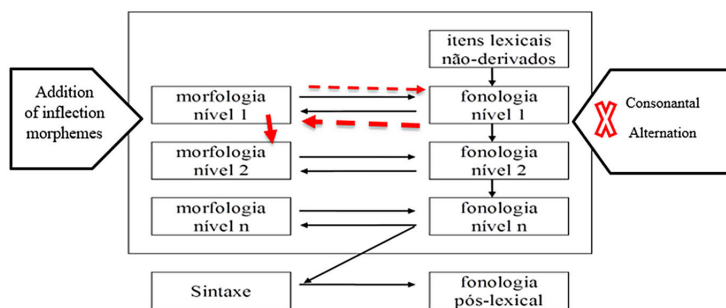
Source: Goulart (2015), based on Kiparsky's Model (1982).

In the theoretical model of LPM, proposed by Lee (1995)[7] for Brazilian Portuguese, irregular flexion is at level 1 (α), while regular flexion manifests at level 2 (β).

Verb productions carried out by children in this study indicate that, at the age of 06, subjects already operate with irregular forms, since they adequately produce the consonant phonological alternations of certain verbs, especially those whose frequency is high in the linguistic input.

However, data also point out that, in all age groups under study, children still regularize irregular verb forms, thus evidencing the late acquisition of morphophonological processes of the BP verb system, especially of less frequent verbs (the case previously emphasized) and the less frequent verb tense: Subjunctive Present. Simultaneous use of irregular forms for some verbs and non-use for others lead to the interpretation that the acquisition of these verbs requires incorporation into the child's grammar:

I) of regular inflectional verbal morphemes (mood and tense, number and person);
II) of irregular verbal morphemes (irregularities due to the paradigm);
III) of alternations of irregular roots (irregularities due to root).

Facts in (II) and (III) are irregular; therefore, they are unpredictable and have to be learned by the child with the use (the frequency) of the language. Besides, they

---

7   According to Lee (1995), at level 1 (also identified as (α)), there are irregular inflections, the Derivation and the Composition I; at level 2 (β), productive formation of PB occurs, together with regular verb inflections of the language; in the post-lexical level, there is Composition II. For further details, see Lee (1995).

require interaction between Morphology and Phonology. For this reason, the frequency of the use of verbs in the language is a conditional factor of the acquisition of irregular verbs, i. e., the more frequent the use of an irregular verb, the more easily its structure will be acquired.

Based on these arguments, the proposal in (c) is the one that seems to be most adequate to explain the corpus of this study, taking into account, above all, the morphophonological nature of the process, which is essential for the standard irregular BP. Because it is complex in language, it contributes to the late acquisition of verb forms whose root involves consonantal alternations in their conjugations.

According to Lee's proposal (2005), in the analysis of data described by this paper, non-communication of morphological and phonological components of the language occurs only when considering irregular verb forms, which are inflected in the child's speech in a regularized way. Then, they go to level 2 (β) without undergoing rules of irregular inflection and, at level 2 (β), they are submitted to the treatment given to regular verbs. This division of verb, irregular and regular inflection in BP at different levels occurs in compliance with the Elsewhere Condition Principle, in which more restricted rules apply before the most general ones, that is, irregular inflection, which is the most marked and most restricted one, occurs before regular inflection, whose rules are less marked and more universal, not only in BP, but in languages of the world.

In the passage to level 2 (β), the verb that did not undergo irregular inflection (because, despite submitting to morphological rules, it did not reach the consonantal changes at the phonological level[8] due to lack of correlation between the components) suffers regular inflection exclusive of level 2 (β), preserving itself in a regular and faithful way to the root of the verb, which is a morpheme that is in the deep lexicon, integrating the underlying representations of the speakers of the language.

According to Lee (1995), it is also at level 2 (β) that the productive formation of BP occurs, together with regular verb inflections of the language. It highlights the reason why regular verbs are understood as the productive class of Portuguese and, therefore, of unmarked nature in the language. In this sense, regularization in the process of verb acquisition shows that children operate in accordance with the unmarked pattern. It becomes the most natural phenomenon, when they have not yet acquired the irregularities of the linguistic system.

In addition, this fact may also show that, when speakers produce regularizations (nonstandard forms), they are dealing with a single internalized grammar, unlike studies (LORANDI, 2006; AVELEDO, 2006) which postulate two operating grammars in the verb system: one is responsible for the conjugation of regular verbs and the other one is in charge of the conjugation of irregular verbs. Regularization, in these researches, is understood as being an inflection established by the grammar of regular verbs and, consequently, maintained in the subjects' speech with faithfulness to the root of the verb.

---

[8]   The locus of the absence of consonantal alternation, as part of the process of acquisition of irregular verb forms, is represented in Figure 4 (on the right).

In the light of the Lexical Theory, the process of non-standard forms was observed to occur through the relations established between the components and strata that constitute the grammar of the language. Thus, the regularization phenomenon is situated in the field that establishes the interchange between the components to form the BP vocabulary and, because of it, it is a process that operates with a single grammar, capable of explaining both the standard and the non-standard inflection. It seeks for linguistic parameters for events that mark the development of language acquisition, specifically the acquisition of BP irregular verbs.

Final considerations

In the verb acquisition phase, the child operates with at least two possibilities of inflection of irregular verbs: either in agreement with the target of the language, performing phonological alternations in line with the pattern; or in disagreement with the target, carrying out the phenomenon that is understood in the literature as regularization (LORANDI, 2006). Inflection that occurs without the proper alternations of irregular conjugation maintains the root of the verb in the subject's speech and, therefore, inflects the irregular class in accordance with the conjugation of regular verbs.

Taking into account that Brazilian studies carried out up to the present moment attest that the phenomenon of verb regularization just occurs until the age of 5 years, data collected by this research disagree, since it shows evidence that children over the age of 5 years produce regularized verb forms in the conjugation of certain irregular verbs.

Therefore, this study reaffirms the importance of the schooling level, a variable linked to age, to the stabilization of BP irregular inflection. In addition, the analysis concluded that:

a) the corpus of this study shows, through the conjugations under analysis, that the subjects under investigation have already got the morphology of the language, since they produce the morphemes of mood and tense, of number and person and the markers of the BP verb paradigm;

b) regularizations in the conjugation of irregular verbs in the view of this study, which is based on LPM, have morphological nature in the BP system, because they are associated with relations conditioned by the mismatch between the phonological and the morphological components of the language;

c) because it implies a morphophonological phenomenon, acquisition of the irregular verb system is late and may extend to at least 9 years of age, in the case of native BP speakers; in this study, the variable age, linked to the schooling level, points out that the greater the child's linguistic experience, the greater the irregular conjugations s/he produces, considering the standard form of the language;

d) verbs – and their inflections – that are maintained at high frequency in the language are acquired earlier than those that show low frequency.

Under the assumptions of Lexical Phonology and Morphology, the use of verb forms which are different from the standard, without the presence of consonantal alternations, can be explained by the non-articulation between the modules of Morphology and Phonology of the language. Lack of matching between these components in the construction of certain irregular conjugations, especially in the first person of the Indicative Present and in the 1st and 2nd persons of the Subjunctive Present, points to the high complexity that permeates the acquisition of the BP irregular verb class. Considering that the subjects of this study already operate with this relation and produce consonantal alternations, mainly for the high frequency verbs in the linguistic input, this study understands that they are still in phase of acquisition of the morphophonology of the language due to the coexistence of standard (P) and nonstandard forms (NP) in their productions.

With this analysis, we have evidence of the complexity of morphophonological phenomena and, in this fact, there is a justification for the late stage of its acquisition: the acquisition of morphology of verbs is complex due to the different affixes that it involves. Besides, difficulty gradually increases when tenses are in the subjunctive mood (from the syntactic and semantic point of view, the use of this verb mood implies a demanding operation in the relations between clauses and meanings). Complexity becomes even greater when there is interaction between the morphological and phonological levels of the language, thus, requiring vowel and consonantal alternations in the process of verb inflection.

Results achieved by this study therefore confirm that, when phenomena of morphological nature are considered, the process of language acquisition by children can extend to the age of 9 years or more, including the use of irregular verbs.

- *RESUMO: Este estudo investiga o processo de aquisição de verbos irregulares do Português Brasileiro (PB), quando conjugados nos tempos Presente do Indicativo, Presente do Subjuntivo e Pretérito Perfeito do Indicativo, por crianças falantes nativas da língua, com o objetivo de descrever e analisar, sob o enfoque da Fonologia e Morfologia Lexical (Lexical Phonology and Morphology - LPM), a relação morfofonológica presente nas flexões dos verbos irregulares. Os dados revelam que os fenômenos linguísticos de origem morfofonológica são de alta complexidade para os falantes e podem ser considerados como processos de aquisição tardia pelas crianças brasileiras. Por meio da LPM compreendemos que a regularização verbal, uma das manifestações das conjugações irregulares, pode ser explicada pela não-correspondência entre os níveis fonológico e morfológico da língua.*

- *PALAVRAS-CHAVE: Aquisição morfofonológica. Verbos irregulars. Fonologia e Morfologia Lexical.*

REFERENCES

ANDERSEN, E. M. L. Representações Lexicais Subjacentes: verbos e léxico inicial. **ReVEL**, v.6, n.1, p.1-31, 2008.

AVELEDO, F. El processamento de verbos regulares e irregulares en el español infantil: mecanismo dual vs. Conexionismo. **Boletín de Linguística**, v.18, n.26, p.5-32, jul. 2006.

BASILIO, M. **Teoria Lexical**. 7. ed. São Paulo: Afiliada, 2004.

BISOL, L. **Introdução e estudos de fonologia do português brasileiro**. 5. ed. Porto Alegre: EDIPUCRS, 2010.

CÂMARA JÚNIOR, J. M. **Uma forma verbal portuguesa:** estudo estilístico-gramatical. Rio de Janeiro: Jornal do Comércio – Rodrigues & Cia, 1975.

_____. **Estrutura da Língua Portuguesa**. 38. ed. Rio de Janeiro: Vozes, 1970.

CASTILHO, A. T. de. **Nova Gramática do Português Brasileiro**. São Paulo: Contexto, 2012.

GOULART, T. P. D. **A Produção de Formas Verbais Irregulares por Crianças Falantes do Português Brasileiro (PB)**. 2015. 147 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade Católica de Pelotas, Pelotas, 2015.

KIPARSKY, P. Lexical morphology and phonology. In: YANG, S. (Ed.). **Linguistic in the morning Calm**. Seoul: Hanshin Publishing Co, 1982.

_____. **Some consequences of lexical phonology**. Phonology yearbook 2. London: Cambridge University Press, 1985.

LAZZAROTTO-VOLCÃO, C. **Modelo Padrão de Aquisição de Contrastes:** uma proposta de avaliação e classificação dos Desvios Fonológicos. 2009. 218 f. Tese (Doutorado em Letras) – Universidade Católica de Pelotas, Pelotas, 2009.

LEE, S. Fonologia Lexical do Português. **Cadernos de Estudos Linguísticos**, Campinas, v.23, p.103-120, 1992.

_____. **Morfologia e Fonologia Lexical do Português do Brasil**. 1995. 201 f. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1995.

_____. Fonologia Lexical – Modelos e Princípios. **Letras de Hoje**, Porto Alegre, v.31, n.2, p.129-137, 1996.

LORANDI, A. **Formas Morfológicas na Gramática Infantil:** um estudo à luz da Teoria da Otimidade. 2006. 185 f. Dissertação (Mestrado em Linguística Aplicada) – Pontifícia Universidade Católica, Porto Alegre, 2006.

MOHANAN, K. P. **Lexical Phonology**. 1982. 218 f. Doctoral dissertation – Massachusetts Institute of Technology, Cambridge, 1982.

SANTOS, R. S.; SCARPA, E. M. A aquisição da morfologia verbal e sua relação com o acento primário. **Letras de Hoje**, Porto Alegre: EDIPUC, v.38, n.4, p.249-260, 2003.

SCHWINDT, L. C. A relação entre Morfologia e Fonologia na História dos Estudos Linguísticos. In: MARTINS, E. S.; CANO, W. M.; MORAES FILHO, W. B. (Org.). **Léxico e morfofonologia:** perspectivas e análises. Uberlândia: EDUFU, 2006.

SOUZA & SILVA, M. C. P.; KOCH, I. V. **Linguística Aplicada ao Português:** Morfologia. 17. ed. São Paulo: Cortez, 2009.

WETZELS, W. L. Mid Vowel Neutralization in Brazilian Portuguese. **Cadernos de Estudos Linguísticos**, n.23, p.19-55, 1992.

YAVAS, F.; CAMPOS, J. Aquisição da Morfologia Verbal do Português como L1 e L2. **Letras de Hoje**, Porto Alegre: PUCRS, n.74, p.81-95, 1988.

THE ACQUISITION OF PRE-TONIC VOWELS IN BRAZILIAN PORTUGUESE

Graziela Pigatto BOHN[*]
Raquel Santana SANTOS[**]

▪ ABSTRACT: This article discusses acquisition of pre-tonic vowels in Brazilian Portuguese by 3 monolingual children acquiring the *paulista* dialect, aged between 1;4 and 3;5, and its relation to the acquisition of stressed vowels. Based on Miranda (2013), we start out from the supposition that the acquisition of pre-tonic vowels is subject to the instability of this position, and segments affected by phonological processes take longer to be acquired. The children's productions show that high pre-tonic vowels are acquired in contrast with mid vowels, (/i,o/ and /e,u/), and that the pre-tonic /o/ is acquired before /e/. We analyse these results based on Contrastive Hierarchy Theory (DRESHER, 2009), according to which the lexical representation of segments is specific to each language and only contrastive and active features must be present in the representation. We propose that the acquisition of pre-tonic vowels follows the Principle of Maximum Contrast: because of the instability found in this position, segments must be maximally contrastive; that is, they must contrast in place and height. The pre-tonic /e/, being more unstable (cf. CALLOU; MORAES; LEITE, 2002; VIEGAS, 2001; YACOVENCO, 1993), is the last one to be acquired, bringing with it the pretonic /u/.

▪ KEYWORDS: Phonological acquisition. Phonological contrast. Pretonic vowel.

Introduction

The goal of this study is to discuss the acquisition of vowels in the pre-tonic position in Brazilian Portuguese (BP). Previous studies (RANGEL, 2002; BONILHA, 2004; VOGELEY, 2011) suggest that vowel systems emerge as proposed by Jakobson (1968[1941]): first low vowels, followed by high and then mid vowels. These proposals assume that features and their organization are universal, leading to predictions that do not always correspond to facts. For example, if features are innate and their hierarchical organization and acquisition pattern are universal, (low > high > mid), then the prediction

* Catholic University of Santos (UNISANTOS), Santos - São Paulo - Brazil. Center for Education and Communication Sciences. grazielabohn@gmail.com

** University of São Paulo (USP), São Paulo - São Paulo - Brazil. Department of Linguistics. raquelss@usp.br

of a theory such as Feature Geometry (CLEMENTS; HUME, 1995), for example, is that the mid front and back vowels emerge at the same time. As facts do not always corroborate this prediction, a lot of analyzes assume sub-stages to capture the asymmetries found in the data. Furthermore, it is also possible to find in the data not only asymmetries but different paths of acquisition, and this type of result calls into question the description given by theories that argue for a universal pattern of acquisition. Most studies of vowel acquisition in BP either focus only on the acquisition of stressed vowels, or consider all word positions and analyze stress as an independent variable that may (or may not) affect the acquisition process (RANGEL, 2002; MIRANDA; MATZENAUER, 2009). Conversely, Bohn (2015) analyzes the acquisition of stressed and pre-tonic vowels separately and finds that the order of acquisition in these two positions is different. Following Miranda (2013), Bohn argues that the order of acquisition of mid vowels is different in the pre-tonic position because these vowels behave differently in this position, and children are sensitive to this difference. Based on studies of vowel harmony affecting pre-tonic vowels in adult speech, which indicate that this process affects only the pre-tonic /e/ (CALLOU; MORAES; LEITE, 2002; YACOVENCO, 1993; VIEGAS, 2001), Bohn assumes that, because /o/ is more stable than /e/ in this position, it is acquired first.

Our hypothesis is that not only is stability relevant in the emergence and acquisition of segments, but also that the order of acquisition follows a Principle of Contrast, which governs the emergence of segments, following the assumptions of Contrastive Hierarchy Theory (hereinafter CHT, DRESHER, 2009).

In order to argue in favor of this analysis, this paper is organized as follows: the first section reviews previous studies on the acquisition of vowels in BP; the second section presents the assumptions of CHT and an analysis on BP phonology within this theoretical framework. The third section provides a summary of Bohn (2015)'s work on vowel acquisition in PB based on CHT. In the fourth and fifth sections we present the methodology and results found in the study. The sixth and seventh sections present a general discussion of the results and the final considerations, respectively.

Vowel Acquisition - previous studies

Studies of the acquisition of the phonological inventory date back to the last century. Jakobson (1968[1941]) was the first to propose an order of acquisition for segments, and he proposed that the acquisition process is guided by maximum contrasts and its order is universal. With respect to vowels, the first contrast is between consonants and vowels (the consonant being the most closed one - the labial stop /p/ - *vs* the most open vowel - /a/). The next contrast is between the low and high vowels (/a/ vs /i/). The following contrast can occur either with respect to frontness (front /i/ vs. /u/) or height (high vs. mid - /i/ vs. /e/). Many studies, however, have drawn attention to the amount of variability found in children's data, calling into question the universal analyzes (eg,

VIHMAN et al., 1986; MACKEN, 1979; PYE et al., 1987, among others), but still today many studies seek to find the Jakobsonian proposed order.

Rangel (2002) was the first to analyze the acquisition of vowels in BP. Based on Feature Geometry (CLEMENTS; HUME, 1995), and making use of two corpora, a cross-sectional study of 72 children and a longitudinal study of three children (from 1:0 to 1;11 years old), Rangel proposes that the child begins the acquisition of PB vowels with all places of articulation specified while acquiring the aperture nodes gradually, arguing in favor of the following order of acquisition: 1st stage (/a> i> u/), 2nd stage (/e>o/), and 3rd stage (/ɔ>ɛ/). This proposal describes exactly the acquisition path suggested by Jakobson, but it is not enough to explain why vowels which are distinguished only by place of articulation are not acquired simultaneously. Also, in Rangel's analysis, the coronal vowel precedes the acquisition of the labial in the first and second stages, but in the third stage we find the labial vowel being acquired before the coronal vowel. In order to explain these differences, Rangel proposes that vowels are acquired in sub-stages.

Also, although Rangel proposes a single order of acquisition for the cross-sectional and longitudinal studies, the results found for each of these studies show differences in how contrasts are established, as shown in Table 1 below.

Table 1 – Order of vowel acquisition.

Cross-sectional:	a > i, u > e, o > ɔ > ɛ
Child Gabi:	a, i > u > e, o > ɔ > ɛ
Child Ana:[1]	a, i, u > o > e, ɛ, ɔ

Source: Rangel (2002).

With regard to the unstressed position, although Rangel takes into account all the vowels produced in the word and classifies them according to stress, she only does so to verify whether the syllable stress would be a facilitator favoring the correct production of vowels. Rangel observes, however, that stress is not always a strong factor for the correct production of vowels: in the cross-sectional study, for example, the correct production of the vowel /i/ was in the pos-tonic syllable, while for /e/ and /o/ the most favorable position was the stressed syllable. In the longitudinal study, the pre-tonic syllable was the most favorable for both /e/ and /o/. In her study, stress was also a factor shown as relevant for the neutralization of /e/ and /o/ in the final pos-tonic position. Based on these results, Rangel assumes a pos-tonic vowel system consisting of three vowels up to the age of 1;5 and a pos-tonic vowel system consisting of five vowels /a, e, i, o, u/ from the age of 1;6 on.

---

[1]   This is based on the Phonological inventory shown for this child on Table 6 (RANGEL, 2002, p.111), but we hightligh that, in Table 7 (p.122), the order presented is different: /a, e, i, o, u/ > /ɛ, ɔ/.

Another proposal for the acquisition of BP vowels comes from Miranda and Matzenauer (2009)'s analysis based on Lee (2008)'s model for the BP vowels. According to Lee, BP vowels are specified with three features: [high], [low] and [ATR]. Miranda and Matzenauer base their analysis on Rangel (2002)'s results but disconsider the difference in the period of acquisition between the mid-low vowels /ɛ/ and /ɔ/ (only one month. Consequently, Miranda and Matzenauer argue in favor of three stages in the acquisition of BP vowels (/a, i, u/ > /e, o/ > /ɛ, ɔ/), eliminating the sub-stages proposed by Rangel. It must be highlighted, however, that despite distinguishing three vowel sub-systems in PB by different height levels (stressed position with four levels, pre-tonic position with three and pos-tonic position with two), Miranda and Matzenauer do not look into the acquisition of these sub-systems separately, similarly to Rangel.

One other study is Bonilha (2004), which analyzes data from one child acquiring PB. According to her, the child acquires BP vowels in three stages: 1st (/ a, e, i, o, u /), 2nd (/ɛ/), and 3rd (/ɔ/), all before the age of 1;9. However, Bonilha (2004) herself points out that her analysis is based on a small number of data. Working within the Optimality Framework, she argues that the vowel inventory of the first stage indicates that there has been a demotion of markedness constraints, which results in the hierarchy as in (1):

(1) H1 = Fidelity >> {* [dorsal], * [labial], * [coronal], * [+ open1], * [+open2], * [+open3], *[-open1], *[-open2], *[-open3], * [+ sonorant], * [+ aproximant], * [+ vocoid], * [+ voiced], * [+ continuant],
However, this hierarchy of constraints would also allow the acquisition of the mid-low vowels at this moment. Therefore, the author suggests that a conjuction of constraints drives the process as in (2):

(2) H1 = {* [+open3] & * [labial]] (seg), [[* +open 3] & * [coronal]] (seg)} >> Fidelity >> {* [dorsal], * [labial], *[coronal], *[+open1], *[+open2], *[+open3], *[-open1], *[-open2], *[-open3], *[+sonorant], *[approximate], * [+ vocoid], * [+ voiced], * [+ continuant]}

Once demoted below the faithfullness constraints, the constraint conjunction would be deconstructed. The difference between the stages of acquisition of /ɛ/ and /ɔ/ would result from the different moments of the demotion of [[* [+ open3] & * [coronal]] (seg) and [* [+ open3] & * [labial]] (seg). With the demotion of [[* [+ open3] & * [labial]] (sec), the child acquires /ɔ/, but does not yet acquire /ɛ/. What draws our attention is the different order of acquisition of the vowels in relation to the other studies (as for Bonilha five vowels would be acquired at the same time). However, as Bonilha only analyzes data from one child, it is not possible to know what type of analysis would be done if variability were to be found.

Subsequent studies specifically focused on the acquisition of pre-tonic vowels. Matzenauer (2009), based on Lee (2008), focuses on the acquisition of these vowels to verify not only the order of emergence of the segments in this position but also the child's sensitivity to phonological processes affecting pre-tonic vowels. Matzenauer's

analysis is cross-sectional and is based on data from 72 children aged 1;2 to 2;1, divided into 12 age groups. Based on Rangel (2002)'s and Miranda and Matzenauer (2007)'s proposals, Matzenauer claims that, in the pre-tonic position, acquisition takes place in two stages only, since the mid-low vowels /ɛ/ and /ɔ/, acquired in the third stage, are not licensed in this position. In the first stage the low vowel /a/, and the high vowels /i, u/ are acquired, resulting from a maximum contrast of height; in the second stage, the mid-high vowels /e, o/. In spite of arguing in favor of stages, Matzenauer points out that mid-high vowels emerge almost simultaneously to high vowels in the pre-tonic position and finds that their first productions occur mainly in reduplications. According to Matzenauer, although the emergence of pre-tonic vowels occurs early, their acquisition is not immediate. These vowels are subject to many phonological processes, in particular the mid-high vowels, which tend to be realized as a high vowel.

With a view to understanding the lexical representation of pre-tonic vowels, especially the mid-high ones, Miranda (2013) analyzes data from one child acquiring BP. Her analysis investigates two hypotheses: (i) the segments are stored in the lexical representations as discrete and autonomous entities, regardless of the syllabic position they occupy; (ii) there is a difference in how vowels are acquired based on the syllabic position they occupy. Miranda observes that between ages 1;8 and 2;4, pre-tonic mid vowels are either produced as high or mid. According to her, the fact that there is no such variation in the stressed position in this child's data reinforces the fact that the pre-tonic position is a place of instability in the acquisition. This confirms the hypothesis that the child is sensitive to the syllabic position a vowel occupies. For Miranda, there seems to be no strong opposition between mid and high in the pre-tonic position, which provides space for experimentation for the child. Thus, the segments are acquired based on how they operate in the system, not as simple units of a set.

Vogeley (2011) also assumes that the acquisition reflects the variation found in the adult language. She analyzes pre-tonic data from the Recife dialect, which presents mid-low vowels in this position, and hypothesizes that the emergence of vowels in the pre-tonic position in this dialect is different from other dialects (VOGELEY; HORA, 2008).[2] Based on a qualitative analysis, the order proposed by Vogeley for the dialect of Recife is: 1st stage (/a, i, u/), 2nd stage (/ɛ, e/), 3rd stage (/o, ɔ/). With regard to the the acquisition of pre-tonic vowels, her results show the acquisition of the mid-low vowels before the mid-high ones, with the front vowel being acquired before the back vowel, with the following order: 1st stage (/a, i, u/), 2nd stage (/ɛ/), 3rd stage (/ɔ/). Vogeley proposes an underlying representation for the vowel system of the Recife dialect which licenses the mid-low vowels in the pre-tonic position, but points out that this does not mean that there should be a contrast between /ɛ, ɔ/ and /e, o/ in this position, as the mid-high vowels are still produced in this dialect by vowel harmony when followed by another mid-high vowel, as is the case of [sor.ˈve.te] 'sorvete'/ 'ice cream' and

---

[2] In this analysis, conducted with a small number of data (the four children had between 9 and 49 tokens each) the author took into account all vowels produced, i.e. stressed and unstressed.

[se.ˈbo.la] 'cebola'/ 'onion', for example. In this sense, Vogeley (2011) argues in favor of the existence of an underlying representation of pre-tonic vowels in PB where these vowels are underspecified in terms of height and later specified according to the dialect. According to Feature Geometry Theory (CLEMENTS; HUME, 1995), this means that [open$_3$] is phonologically deactivated in the underlying representation and is only specified when inserted in a dialect - for the Southern dialect, known as gaucho, there is activation of [-open$_3$], and for the Recifense dialect, [+open$_3$] in the pre-tonic position.

None of the studies described so far take into account the variability oftentimes found in children's data. Bohn (2015) attempts to account for this variability. As the author develops her analysis by assuming CHT, we show her results after presenting the main characteristics of this theory in the following section.

Contrastive Feature Hierarchy and Brazilian Portuguese

As we saw in the previous section, phonological proposals that assume universal features or hierarchies cannot account for the variability oftentimes found in children's data or how systems with the same phonological inventories may be subject to different phonological processes. CHT proposes a formal, uniform model which accounts for variability both among systems and among individuals speaking the same language (cf. DRESHER, 2009).

Dresher (2015, p.165, emphasis added) proposes that

> [...] the phonological component of grammar computes features, but these features are not innate. Rather, they are created by the learner as part of the acquisition of phonology. Further, Universal Grammar (UG) requires that these features be organized into contrastive feature hierarchies that reflect phonological activity and the contrasts in the lexical inventory.

That is, for Dresher, what is universal is the structure and the hierarchical organization of features, and not features themselves. It is the concept of a hierarchy of features that is inherent in the phonological grammar, not its substance. And it is through an innate capacity to perceive acoustic correlates that features are identified and hierarchized.

According to CHT, the phonological representations of segments are based on phonological contrasts and on the phonological processes affecting them. In fact, the idea of contrast has been fundamental in the linguistic analyzes of phonological grammars in order to determine the phonemes of a phonological inventory and its internal organization. However, Dresher stresses that the initial proposals of contrastive hierarchies (such as that of CHERRY; HALLE; JAKOBSON, 1952; HALLE, 1971) did not establish clear criteria for the determination and arrangement of contrastive features in a branched and hierarchical structure. CHT is intended to rescue the notion

of contrast and hierarchy in a model of phonological grammar, accounting for the lack of consensus in the previous proposals. For Dresher, if contrastive hierarchies vary from language to language, there needs to be a strategy to determine which features are contrastive in each system. Dresher assumes that the phonological grammar computes only features that are active in the system and dispenses with redundant ones, adopting the Contrastive Hypothesis formulated by Hall (2007, p.87), according to which, "The phonological component of a language L operates only on those features which are necessary to distinguish the phonemes of L from one another."

As for the organization of features, Dresher (2009) proposes that it should be done following the *Successive Division Algorithm,* hereinafter SDA, which determines that all segments of the inventory must receive a representation that distinguishes each from all the others. The algorithm consists of the following steps:

(3) a) Begin with no specification: assume all sounds are allophones of a single undifferentiated phoneme
     b) If the set is found to consist of more than one contrasting member, select a feature and divide the set into as many subsets as the feature allows for
     c) Repeat step (b) in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member. (DRESHER, 2009, p.16)

As only the concept of a hierarchical contrastive structure and the division algorithm are innate, and there is no pre-established order for the organization of features, nothing prevents the same set of phonemes from having different organizations. For an inventory such as /a, i, u/, for example, we can have [high] as the first contrast, dividing the set into two groups: /i, u/ being specified with [+ high] and /a/ with [-high]; second, we divide the [+high] vowels – which are still uncontrasted – into [+ rounded] (/u/) and [-rounded] (/i/). In this representation, the [high] feature contrasts all segments, while [rounded] is relevant only for /i/ and /u/ - cf. (4a). This same inventory, in another language, could have a different hierarchical order of features, starting with [rounded]. In this second hierarchy, all segments of the inventory receive the [rounded] contrast and only /a/ and /i/ will contrast in terms of [high] - cf. (4b).

Figure 4 – a) Hierarchy starting with [high], b) Hierarchy starting with [round]
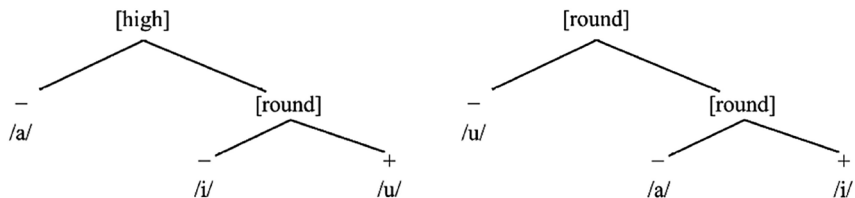
These different configurations capture facts such as why there are languages in which processes affect /i, u/ but not /a/, while in others, processes affect /i, a/ but not /u/.

We would like to point out, however, that Dresher does not explain what would trigger the SDA to be applied. That is, Dresher's proposal is based on the fact that the features must be contrastive; the SDA is an algorithm that divides an allophonic inventory based on contrastive features; but there is no mention of a principle that governs this application, as noted in Dresher (2015, p.172, emphasis added):

> At some point the learner discovers a contrast between a [low] vowel /A/ and a non-low vowel /I/ (the symbols are for convenience only). [...] Consider, for example, the first contrast in the vowel system, that between low and non-low vowels. How does the learner know to include [ə] with the low vowels rather than with the non-low vowels? Part of the answer must be that [ə] alternates with [a] as its ATR counterpart. This suggests that even at the first step in the idealized acquisition sequence there must be a certain amount of look-ahead that makes use of information about contrasts later on in the sequence.

In addition to the concept of contrast, another assumption of CHT is that the hierarchical order must also be determined by the way the segments operate in the phonological systems of the language, unlike the hierarchical universal ordering based on articulatory motivations adopted by Feature Geometry (CLEMENTS; HUME, 1995), for example. This means that the variability between the different contrastive hierarchies reflects how each system is affected by phonological phenomena. The choice between the two possible hierarchies for the same inventory /a, i, u/ shown above will therefore depend on the relevance of each feature in the phonological processes of the language. In the second proposed hierarchy, in which /u/ is specified only for roundness, we could say that in this system, /u/ is affected by processes in which only roundness is relevant; that is, despite being a high vowel, the height feature is not active in the processes that affect this vowel, and, therefore, it is not part of its phonological representation.[3]

Lee (2008) proposes a contrastive hierarchy for vowels in BP, according to CHT, cf. (5):

---

[3]   Analyses of this type can be found in Zhang (1996), Dresher (2009) and Mackenzie (2009).

Figure 5 – Contrastive Hierarchy of PB vowels proposed by Lee (2008)
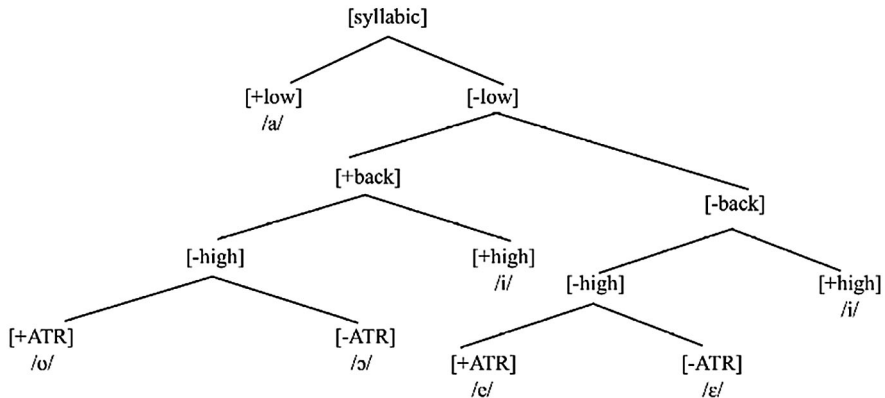
Bohn (2015) points out that Lee's proposal accounts for the phonological processes affecting BP vowels described by Bisol (1996), including vowel elision, which, according to Bisol, affects only /a/ in the Southern dialects; Lee's hierarchy, however, does not account for the elision process as it occurs in the Paulista dialect, where it affects all unstressed back vowels /a, u, o/ (cf. NOGUEIRA, 2007 and SANTOS, 2007). Bohn's proposed hierarchy captures the process as it occurs in the Paulista dialect:(6)):

Figure 6 – Contrastive Hierarchy of PB
vowels proposed by Bohn (2015)
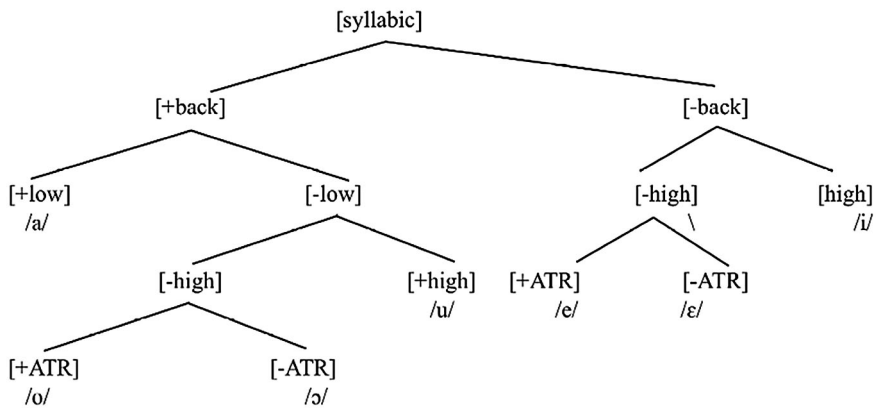
We understand that the existence of two hierarchies for the same language is not a problem for CHT; it is also what explains variability within a language: one phonological inventory with different organizations. And it is through how phonological processes function in the language that we can realize the different organization of inventories.

The implementation of CHT in acquisition raises an important question for both theory and data: is it possible for children to build different hierarchies?

CHT and phonological acquisition

In her study, Bohn (2015) shows that three children exposed to the Paulista dialect of BP follow different paths when acquiring the vowel system of the language. By analysing longitudinal data, Bohn reconstructed the contrastive hierarchies taking into account the order in which vowels became frequent in the speech of each child (following the methodology proposed by Ingram, 1989) and the substitutions found in the data. If, for instance, in an initial stage only one vowel is frequent in a child's speech, the contrast is between vowels *vs.* non-vowels, represented by Bohn with the [syllabic] feature. If another vowel becomes frequent, a second contrast is specified (which characterizes another stage[4]), and so on, until all vowels have enough contrasts that distinguish them from all the others. The existence of two or more vowels under the same contrast can be evidenced in the substitutions found in children's productions. For example, the productions ['azɐ] for /'xɔza/ 'rosa'/'pink' (1;9), ['podʊ] for /sa'patu/ 'sapato'/ 'shoe' (1;8), and [putɐ] for /'pɔrta/ 'porta'/'door' (2;0) for child L. indicate that /a/, /o/, /ɔ/ and /u/ share a contrastive feature and, as the hierarchy is being constructed, these vowels should contrast with each other. The advantage of this model is to explain the variability found in several longitudinal acquisition studies and also attested to in Bohn's data: while one child begins the hierarchy with a contrast in height (Child A.), the others (Children Am. and L.) begin with a contrast in place and follow the same order in the acquisition of contrastive features despite presenting a different order of vowel acquisition. As seen in (7) below, all children begin by contrasting /a/ with all the other vowels. However, child A. almost concomitantly acquires /o/, whereas children Am. and L. acquire /e/ and /i/, respectively:

(7) a.   Child A .:   / a, o /> / i, e, u /> / ɔ, ɛ /   (see representation in (5))
     b.   Child Am .:/ a /> / e /> / o /> / i /> / ɔ, u, ɛ / (see representation in (6))
     c.   Child L .:   / a /> / i /> / e, o /> / ɔ, ɛ /> u   (see representation in (6))

As mentioned, CHT uses the phonological processes affecting segments as clues for the organization of the features in the hierarchy. Thus, the child does not build classes (or nodes, in terms of Feature Geometry) that group vowels that are not subject to the

---

4   New stages can be due to the acquisition of vowels or to the reorganization of features in the hierarchy – cf. Bohn (2015).

same phonological process. For example, a child acquiring BP will not use [a, e, i] allophonically as there are no phonological processes affecting these three vowels in PB.

In sum, in discussing the acquisition of the vowel system in PB from the CHT perspective, Bohn can deal with the variability found among learners, which is not possible within other theoretical frameworks.

Methodology

The *corpora* of this study consist of data from 3 children (A., Am., L.) acquiring the Paulista dialect of BP (data from Santos, 2005). The children were recorded in their natural environment every week for 30 minutes. The amount of data collected in the stressed and pre-tonic positions for each child and the period of data collection are shown in Table 2.

Table 1 – Number of data analysed per child distributed according to stress position.

Child:	Age	Stressed	Pre-tonic	Total
A.	1;0 – 2;11	3.761	3.009	6.770
Am.	1;02 – 3;05	1.540	890	2.430
L.	1;04 - 2;11	2.633	1.511	4.144

Source: Adapted from Bohn (2015, p. 108).

As we have seen, CHT assumes that children can have the same phonological inventory, but build different structures to contrast phonemes. One methodological consequence of this assumption is that cross-sectional data should not be analyzed within this framework, as children in the same age range can have the same phonological inventories but may organize them into different structures. Only by looking into longitudinal data and observing the emergence of phonemes and substitutions is it possible to infer what hierarchies children are building. For this reason, the data of these three children are analyzed separately in this study.

The word productions to be considered in the analysis is also a decision that has important consequences for the results. For example, it is not possible to observe substitutions in words created by the children, as there is no target word for comparison. Therefore, only words that are part of the lexicon of the language were considered in the study.

Productions were considered correct taking into account the adult language. However, cases that show a variation in the adult form were not considered substitutions (e.g. in [ka.'de] ~ [ke.'de] for /ka'de/ '*cadê*'/'where', in which the production of [e] was not considered as a substitution for /a/ as both forms are also attested in adult speech). The data were phonetically transcribed and, in cases of doubt, the F1 and F2

frequencies of the vowel were used as parameters for the final transcription.[5] In addition, special attention was paid to productions that could result from phonological processes. The literature on the acquisition of phonological rules in the adult language is still scarce in BP (but cf. SILVA, 2008; SANTOS, 2007), but less sparse on phonological processes which are typical of children's speech. However, these rules and processes may conceal the acquisition of phonemes. Our decision, then, was not to work with productions that could be the result of a phonological process of the adult language. For example, the production of [si'gu.la] for /se'guɾa/ '*segura*'/'to hold-imperative' (Am. 2;4) may be either an evidence of /i, e/ being acquired (in this stage, as we will see in Table 4, the vowel /e/ is still not used very often by this child), or the result of a process of vowel harmony. Other examples are cases of syllabic reduplication, such as [te.'ta.ta.ta] for /ga'veta/ '*gaveta*' / 'drawer' (Am. 2;10), a process which is typical of language acquisition. These types of data could bias the results, so they were not considered in the analysis.

Finally, an important methodological decision is the use of types or tokens. It is not uncommon in the literature of phonological acquisition to see that children succeed in producing a segment correctly in a specific word, whereas in other words the same segment is constantly being replaced. This may be due to being a segment that appears in the child's name, the name of a close person or even a favorite toy, that is, a word which is produced frequently. If an analysis is conducted by *tokens* and one word occurs constantly, the results could wrongly suggest the acquisition of a phoneme. For this reason, a number of current analyzes consider that the child should present a minimum percentage of correct productions in different types. In this study, the data were extracted and organized according to the methodology proposed by Ingram (1981, 1989). According to this methodology, the analysis is guided by phonetic types. In the case of variability in production, Ingram's methodology sets up criteria for deciding which phonetic type should be used in the analysis, described in (8) below (1989, p. 204):

(8) a) If a phonetic *type* occurs in a majority of the phonetic tokens, select it.

b) If there are three or more phonetic types, select the one that shares the most segments with the others.

c) If there are two phonetic types, select the one that is not pronounced correctly.

d) If none of the above work, select the first phonetic type listed.

---

[5] Because the vocal tract of the child is still developing, and variation in formants may be found, Bohn uses the child's own productions as a formant parameter to classify the vowel productions, and not the values found in the literature. Hence, the frequencies of two productions of each stressed and pre-tonic oral vowel clearly produced by each child were measured, in each session, in order to obtain the means of the formants of each vowel for each child at that session. To illustrate this procedure, we have here the production [pega] for /pɛga/ *pega* ´catch-imperative´of L. (verb in the imperative, second person singular) at age 1;9. The mean of F1 and F2 of vowel [e] in this production presented the following values: 831.31 Hz and 1701.12 Hz, respectively. The approximation of these values with the average obtained for vowel /e/ in the stressed position at that age for this child (F1 = 892.09 Hz and F2 = 1745.30 Hz) allowed Bohn to classify the production as a mid-high front vowel and not as a mid-low front vowel (which, for this child at this age, presented the following values: F1 = 1028.86 Hz and F2 = 2218.56 Hz).

Sounds produced by children are then classified into three frequency categories (marginal, in use, frequent), which vary according to the size of each sample. Table 3 shows the frequency criteria for vowels. When the child has a sample of 1 to 25 types, a certain sound must occur four times or more to be considered *frequent* in their speech; if it occurs two or three times, it will be considered *in use*, and if it occurs only once, it will be considered *marginal*. As the number of phonetic *types* increases in a sample, so does the number of times the sound should appear.

Table 2 – Phonetic types and frequency criteria.

Number of phonetic types	Marginal (low)	In use (medium)	Frequent (high)
1 – 25	1	2, 3	4 or more
26 – 37	2	3, 4	5 or more
38 – 67	2	3 – 5	6 or more
68 – 87	3	4 – 6	7 or more
88 – 112	3	4 – 7	8 or more
113 or more	4	5 - 8	9 or more

Source: According to Fee (1991, p.351).

This first step generates the phonetic inventory of the child. In the second step of this methodology, the substitutions are computed, providing evidence of which phonological contrast has already been acquired and which segments are still being used allophonically.

With the information in steps 1 and 2, the children's phonological inventory is organized. According to this method, a sound will be considered acquired in the child's phonological system when (i) it is classified as *frequent*; or (ii) it is classified as *in use* but also appears as a substitute for another sound. Ingram points out that there must be consistency in the classification of a sound as *frequent* in the samples. That is, the classification as *frequent* must be maintained in all subsequent samples, which indicates that this methodology must only be applied to longitudinal data.

Results

The results by *types* brought in this section are reported in Bohn (2015), but will be arranged differently in order to facilitate our discussion.

Table 4 below shows the order of acquisition of vowels, in stressed and pre-tonic position, for each child; Tables 5 and 6 present the substitutions produced by each child for vowels in stressed position; and Tables 7 and 8, the substitutions produced by each child for the vowels in the pre-tonic position.

Table 3 – Order of acquisition of stressed and pre-tonic vowels for each child.

Stressed		A.		Am.		L.	
		Pre-tonic	Stressed	Pre-tonic	Stressed	Pre-tonic	
Age	1;1						
	1;2						
	1;3						
	1;4					/a/	
	1;5						
	1;6	/a, o/	/a/			/i/	/a/
	1;7		/o/				
	1;8	/i, e, u/	/i/			/e, o/	
	1;9						
	1;10	/ɔ/		/a/		/ɔ, ɛ/	
	1;11	/ɛ/		/e/	/a/	/u/	/o, i/
	2;0		/e/				/u/
	2;1		/u/				
	2;2						
	2;3			/o/			
	2;4			/i/			
	2;5						
	2;6						
	2;7						/e/
	2;8				/o/		
	2;9						
	2;10						
	2;11				/i/		
	3;0			/ɔ, u/			
	3;1			/ɛ/			
	3;2				/e/		

Source: Adapted from Bohn (2015, p.111-137).

Table 4 – Phoneme substitutions in stressed position for each child (1;4 to 2;3).

	phonemes	1;1	1;2	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0	2;1	2;2	2;3
A	/a/															
	/i/															
	/u/															
	/e/															
	/o/															[ɔ](1)
	/ɛ/							[e](1)	[e](2)	[e](4)	[e](1)	[e](1)	[e](1)	[e](4)		
	/ɔ/							[o](1)	[o](3)	[o](2)			[o](2)	[o](2)	[o](6)	[o](2)
	Number of types	0	1	4	4	7	12	6	61	111	193	206	247	297	381	298
Am	/a/															
	/i/															
	/u/															
	/e/															
	/o/													[u](1)		
	/ɛ/									[a](1)						
	/ɔ/															[o](1)
	Number of types	-	0	0	1	1	0	4	4	3	7	10	28	32	36	30
L	/a/							[o](1)						[o](1)		
	/i/													[e](1)		
	/u/															
	/e/													[i](2)		
	/o/								[u](1)	[u](2)	[u](2)			[u](2)		
	/ɛ/								[e](2) [i](1)	[e](2)			[e](4)	[e](3)	[e](2) [i](1)	[e](2) [i](1)
	/ɔ/						[a](1)	[o](1)	[o](1)	[a](2) [o](1)		[o] (1)	[o](2) [u](2)	[o](2) [a](1)	[o](2)	
	Number of types	-	-	-	22	13	16	28	52	124	101	182	210	231	205	148

Source: Adapted from Bohn (2015, p.111-137).

Table 5 – Phoneme substitutions in stressed position for each child (2;4 to 3;5).

	phoneme	2;4	2;5	2;6	2;7	2;8	2;9	2;10	2;11	3;0	3;1	3;2	3;3	3;4	3;5
A	/a/														
	/i/														
	/u/														
	/e/														
	/o/			[ɔ](1)	[ɔ](1)			[ɔ](1)							
	/ɛ/	[e](2)	[e](1)	[e](2)	[e](1)	[e](1)		[e](2)	[e](2)						
	/ɔ/		[o](3)	[o](8)	[o](4)	[o](1)	[o](3)	[o](9)							
	Number of types	296	312	268	311	230	43	382	91	-	-	-	-	-	-
Am	/a/														
	/i/														
	/u/														
	/e/														
	/o/	[ɔ](1)													
	/ɛ/				[a](1)	[a](1)		[a](1) [e](1)	[a](1)						
	/ɔ/	[o](1)	[o](2)				[u](1)								
	Number of types	55	53	67	57	69	53	111	41	129	154	176	133	117	169
L	/a/														
	/i/														
	/u/														
	/e/														
	/o/			[u](1)											
	/ɛ/	[e](1)													
	/ɔ/	[o](1) [a](1)		[o](1)											
	Number of types	224	258	48	167	160	155	219	70						

Source: Adapted from Bohn (2015, p.111-137).

Table 6 – Phoneme substitutions in pre-tonic position for each child (1;4 to 2;4).

	phoneme	1;1	1;2	1;3	1;4	1;5	1;6	1;7	1;8	1;9	1;10	1;11	2;0	2;1	2;2	2;3	2;4
A	/a/									[o](1)							
	/i/									[e](1)							
	/u/																[o](1)
	/e/									[i](4)	[i](4)	[i](5)	[i](2)	[i](3)		[i](2) [ε](2)	[i](3) [ε](2)
	/o/									[u](5)	[u](4)			[u](1)	[u](2)	[u](3)	[u](2)
	Number of types	0	2	4	1	6	11	5	28	77	145	117	165	386	272	201	298
Am	/a/																
	/i/																
	/u/																
	/e/																
	/o/										[u](1)			[u](1)			
	Number of types	0	0	0	1	5	0	5	4	2	6	12	15	21	16	11	25
L	/a/											[e](1)					
	/i/											[e](1)	[u](1)				
	/u/									[ɔ, i]		[i]	[e,i]				
	/e/									[i](1)		[i](6)	[i](8) [a](1)	[i](10)	[i](8)	[i](6)	[i](11)
	/o/										[u](1)			[u](3)	[u](6)	[u](7)	[u](6)
	Number of types	-	-	-	11	3	22	20	32	45	56	90	109	102	97	102	138

Source: Adapted from Bohn (2015, p.111-137).

Table 7 – Phoneme substitutions in pre-tonic position for each child (2;5 to 3;5).

	phoneme	2;5	2;6	2;7	2;8	2;9	2;10	2;11	3;0	3;1	3;2	3;3	3;4	3;5
A	/a/													
	/i/													
	/u/													
	/e/	[i](4) [ɛ](1)		[i](1)										
	/o/	[u](1)			[u](1)									
	Number of types	225	214	229	169	40	338	76	-	-	-	-	-	-
Am	/a/			[u](1) [e](1)										
	/i/													
	/u/													
	/e/						[u](1)				[i](1)			
	/o/			[u](1)										
	Number of types	19	34	19	27	28	77	14	67	85	109	81	87	120
L	/a/													
	/i/													
	/u/													
	/e/	[i](20)	[i](6)	[i](22)	[i](3)	[i](1)	[i](11)	[i](2)	-	-	-	-	-	-
	/o/	[u](12)	[u](1)	[u](15)	[u](9)	[u](7)	[u](2)	[u](3)	-	-	-	-	-	-
	Number of types	202	51	166	81	49	94	41	-	-	-	-	-	-

Source: Adapted from Bohn (2015, p.111-137).

Tables 5 to 8 also show the amount of data per type. For example, at 2;5, A. produced 225 different words (cf. Table 8). Of these 225 words, 5 words showed substitutions: /o/ was produced as [u] in one word, and /e/ was produced as [i] in four other words. As can be seen, children can show substitutions for target phonemes even after a phoneme has been acquired, but the number of substitutions decreases over time, at the same time as the number of different lexical items being produced increases.[6]

Tables 7 and 8 show that children produce pre-tonic syllables in a very early period: A. has some productions as early as 1;2, and begins a more productive stage at 1;9. L.

---

[6] An item that may have been frequently substituted in a session may fail to be so at a later time. For example, child A. produces [ko.'ku.sa] for /ba'gũ.sa/ 'bagunça'/ 'mess' at 1;9 (as shown by the substitution of [o] for /a/ in Table 5). Observe, however, that this was the only time such substitution occurred.

already produces pre-tonic syllables at 1;4 (when data started being collected) and they become more productive at 1;11. In the case of Am., some productions start at 1;5, but note that Am.'s samples are always much smaller than the other two children's, only increasing significantly after age 3;0.

As can be seen in Table 4, A. acquires pre-tonic vowels in the same order as the stressed vowels (stressed: a, o > i, e, u > ɔ > ɛ; pre-tonic: a > o > i > e > u), with the only difference being time of acquisition, with pre-tonic vowels being acquired after the stressed vowels, except for /a/ and /i/, which are acquired simultaneously in the stressed and pre-tonic positions. In (9) we show examples of substitutions in pre-tonic position by child A.

(9) /a/ [ko'kũsɐ]   /ba'gũsa/   *bagunça* 'mess'   A. 1;09;19
    /e/ [si'a]   /se'ɾa/   *será* 'maybe'   A. 2;04;26
        [tɛ'iɲa]   /estɾe'liɲa/   *estrelinha* 'little star' A. 2;03;17
    /i/ [be'gadʊ]   /obɾi'gadu/   *obrigado* 'thank you'   A. 1;09;11
    /o/ [ku'sa]   /ko'saɾ/   *coçar* 'to itch'   A. 1;10;12
    /u/ [ko'ɛkɐ]   /ku'ɛka/   *cueca* ´underpants´   A. 2;04;10

The instances of /a, i, u/ substitutions are very systematic: one lexical item each. On the other hand, /e, o/ show more substitutions than correct productions. All such cases involve their counterparts with the same place of articulation ([u] for /o/ and [i] for /e/). Correct productions for 'fechou' /fe'ʃow/ and 'pesado' /pe'zadu/ occur before /e/ is acquired, but A. also produces a number of words in which /e/ is substituted by a high vowel (e.g. /ʒo'gaɾ/ → [ʒu'ga] 'jogar'/ 'play', /to'maɾ/ → [tu'ma] 'tomar' / 'take', /se'ɾa/ → [si'a] 'será'/ 'maybe', /pe'gaɾ/ → [pi'ga] 'pegar' / 'pick up'). According to Bohn (2015), this alternation may be due to the fact that the child is already trying to work with the processes that affect the pre-tonic position, especially vowel harmony (as in [mi.'ni.nʊ] for /me'nino/ 'menino'/'boy') and free context neutralization (as in [tu.'ma.tʃɪ] for /to'mate/ 'tomate'/'tomato'). According to these two processes, /e, i/ are produced as [i] and /o, u/ as [u]. Note that it is exactly the type of substitution that child A. does: the high vowel replaces the mid-high vowel. Bohn's analysis is corroborated by other substitutions found in the data: [ko'ɛkɐ] for /ku'ɛka/ 'cueca'/'underpants', [be'gadʊ] for /o.bɾi'gado/ 'obrigado'/'thank you'*,* [ko'kũsɐ] for /ba'gũsa/ 'bagunça'/ 'mess', for example, are produced with a low or high pre-tonic vowel in order to harmonize to the stressed vowel (see productions in (9) above). The only word that does not follow the same pattern is /pi.ɾu'litu/ '*pirulito*'/'lolypoppy', produced as [pi.i'i.tu]. Our hypothesis is that the child is applying here a process of harmony: having adjacent high vowels on both sides, A. produces the high front vowel instead of the high back vowel. And in 'estrelinha', A. makes use of [ɛ], a vowel that does not exist in the pre-tonic subsystem, without any apparent reason (such as an assimilation to an adjacent vowel), but maintains the same place of articulation as the correct vowel /e/.

Two facts draw our attention in Am. data: vowel /u/ is not acquired before 3;5, the last age analyzed for this child. In addition, the order of acquisition is different in the two positions (stressed: a > e > o > i > u > ɔ > ɛ; pre-tonic: a > o > i > e). The same time difference between the acquisition of the stressed and the pre-tonic found in child A. is also found in child Am. As shown in the data quantity row, Am. has a very small vocabulary (whereas child A. produces more than 100 different words at 1;10 and child L. produces this quantity at 2;0, Am. only reaches this quantity at 3;2. However, Am. is the child who makes the least substitutions. Am. replaces /a/ by [u] and [e], each one in a type. Vowel /o/ is replaced by [u] in some words, and /e/ is replaced by [i] in a single word. In (10) we have the pre-tonic substitutions found in child Am.'s data.

(10) /a/   [mu'ʃo]   /desmã'ʃa/   *desmanchar* ´to dismantle´   Am. 2;07;08
         [be'igɐ]   /ba'xiga/   *barriga* 'belly'   Am. 2;07;25
      /e/   [niku'la]   /selu'la/   *celular* ´cell phone´   Am. 2;10

It is worth noting the substitution of /a/ for [u, e], vowels with different places of articulation, at the same age (2;7), and after /a/ being considered frequent in her speech. However, note that in *barriga* 'belly', we have /a/-raising, partially assimilating to the stressed vowel. In *desmanchar* ´to dismantle´, our hypothesis is that the labiality of /m/ has affected vowel /a/ - specifically, the [labial] feature was assimilated by the vowel.

Finally, child L. also acquires pre-tonic vowels in a different order when compared to the stressed vowels (stress: a > i > e, o > ɛ, ɔ > u; pre-tonic: a > i, o > u > e). Similarly to child A., L. acquires, within the period analyzed, all the vowels that fill this position in BP, and similarly to child Am., the order of acquisition in stressed position is different from pre-tonic position. For both Am. and L., the last vowel to be acquired is /e/. Examples of production by L. are found in (11).

(11) /a/   [se'peʷ]   /ʃa'pɛw/   *chapéu* 'hat'   L. 1;11;23
      /u/   [se'miʷ]   /su'miw/   *sumiu* 'disappeared'   L. 2;0;28
         [si'biʷ]   /su'biw/   *subiu* 'climb-past'   L. 2;0;28
      /e/   [a.le'fa.di] /ele'fãte/   *elefante* 'elephant'   L. 1;11;16
         [i.'ʃey]   /fe'ʃey/   *fechei* 'closed'   L. 1;11;02
      /i/   [bũ:'ka]   /bi'ka/   *brincar* 'to play'   L. 2;0;14
         [de'gɛʷ]   /mi'gɛw/   *Miguel* (proper name)   L. 1;11;23
      /o/   [bu'to]   /bo'tow/   *botou* 'put-past'   L. 2;7;8

Table 9 below summarizes the order of acquisition of the segments in the stressed and pre-tonic positions for the three children in this study:

Table 8 – Summary of the order of vowel acquisition

Child	Context	Order of acquisition
A.	Stressed	a, o > i, e, u > ɔ > ɛ
	Pre-tonic	a > o > i > e > u
Am.	Stressed	a > e > o > i > u, ɔ > ɛ
	Pre-tonic	a > o > i > e
L.	Stressed	a > i > e, o > ɛ, ɔ > u
	Pre-tonic	a > i, o > u > e

Source: Adapted from Bohn (2015, p.111-137).

Discussion

Jakobson (1968 [1941]) proposes that vowel acquisition starts with the low vowel contrasting with the high vowels (/a, i, u/). Rangel (2002)'s and Miranda and Matzenauer (2009)'s analysis of stressed vowels indicate the same order of acquisition of stressed and pre-tonic vowels. Bonilha (2004) points out that, in the first stage, the child analyzed in her study had already acquired /e, i, o, u/. The studies on the acquisition of pre-tonic vowels, specifically, draw our attention to phonological processes that cause instability of the vowels in these positions (eg. MATZENAUER, 2009; MIRANDA, 2013; VOGELEY, 2011), but also point to a single acquisition path: low and high vowels, followed by mid vowels (high or low, depending on the dialect).

As could be seen, none of the three children analyzed in our study followed this pattern, either because they have inverted the order of acquisition (for example, child Am. acquired the mid high vowels before the high vowels) or because different vowels were acquired simultaneously, as was the case of child A., who has acquired /i, u/ at the same time as /e/).

In order to explain these results, the first hypothesis to be considered is that children build different hierarchies for stressed and pre-tonic vowels. The main consequence of this hypothesis is that it cannot be said that the pre-tonic system is a subsystem of the stressed vowels, generated by neutralization of the mid vowels - the most accepted analysis for the description of BP vowels (cf. CÂMARA JR., 1977; WETZELS, 1992, among numerous others). Following this hypothesis, it should be expected to find different orders of acquisiton for stressed and pre-tonic vowels and different substitutions, as feature hierarchies could be organized differently. However, although for two children (Am. and L.) the order was different, we found that most substitution patterns are those that also appear in the stressed position.

The alternative hypothesis is that there is a single structure constructed by the child, regardless of the position the vowel occupies in the word. Tables 10, 11 and 12 indicate which vowels had already been acquired in stressed position and what the feature hierarchy for stressed vowels was when pre-tonic vowels were being acquired (according to Bohn, 2015).

Table 9 – Acquisition of stressed and pre-tonic vowels and hierarchical structure – A.

A.	Pre-tonic:	Stressed	Hierarchy (stressed vowels)	
1;6	a	a, o	**Stage 1:**	[low] [non-low]
1;7	o	a, o	**Stage 1:**	[low] [non-low]
1;8	i	a, o, e, i, u	**Stage 2:**	[low] [non-low] [back] [non-back] [high] [non-high]
2;0	e	a, o, e, i, u, ɔ, ɛ	**Stage 3:**	[low] [non-low] [back] [non-back] [high] [non-high] [ATR] [non-ATR]
2;1	u	a, o, e, i, u, ɔ, ɛ	**Stage 3:**	[low] [non-low] [back] [non-back] [high] [non-high] [ATR] [non-ATR]

Source: Adapted from Bohn (2015, p.152-162).

Table 10 – Acquistion of stressed and pre-tonic
vowels and hierarchical structure – Am.[7]

Am.	Pre-tonic	Stressed	Hierarchy (stressed vowels)	
1;11	a	a, e	**Stage 2:**	[back] [non-back]
2;8	o	a, e, o, i	**Stage 4:**	[back] [non-back] [low] [non-low] [high] [non-high]
2;11	i	a, e, o, i	**Stage 4:**	[back] [non-back] [low] [non-low] [high] [non-high]
3;2	e	a, e, o, i, u, ɛ, ɔ	**Stage 5:**	[back] [non-back] [low] [non-low] [high] [non-high] [ATR] [non-ATR]

Source: Adapted from Bohn (2015, p.162-173).

---

7   Up to 3; 5, the vowel /u/ had not been frequently produced in the pre-tonic position by Am.

Table 11 – Acquistion of stressed and pre-tonic
vowels and hierarchical structure – L.

L.	Pre-tonic	Stressed:	Hierarchy (stressed vowels)		
1;6	a	a, i	**Stage 2:**	[back] [non-back]	
1;11	o,i	a, i, e, o, ɔ, ɛ, u	**Stage 6:**	[back] [non-back]	
				[low] [non-low]	
				[high] [non-high]	
				[ATR] [non-ATR]	
2;0	u	a, i, e, o, ɔ, ɛ, u	**Stage 8:**	[back] [non-back]	
				[low] [non-low]	
				[high] [non-high]	
				[ATR] [non-ATR]	
2;7	e	a, i, e, o, ɔ, ɛ, u	**Stage 8:**	[back] [non-back]	
				[low] [non-low]	
				[high] [non-high]	
				[ATR] [non-ATR]	

Source: Adapted from Bohn (2015, p.174-190).

As can be seen in Tables 10, 11 and 12, when pre-tonic vowels emerged, children already had the same vowel in the stressed position. Therefore, the structure they had built for the stressed position at the moment of acquisition of the pre-tonic also captures the acquired pre-tonic vowel. For example, when child Am. acquires the pre-tonic /o/ at 2;8, a hierarchical structure for stressed /o/ with the feature [back, low, high] had already been constructed and allows this vowel's production in pre-tonic position. We did not find any cases in the children's data in which the acquired pre-tonic vowel could not be produced taking into account the hierarchical structure of the stressed vowels at that moment of acquisition.

The hypothesis that there is a single hierarchy that captures the vowel neutralization of BP, but leads to several questions in the context of language acquisition: why do vowels in the pre-tonic position appear later and not at the same time as the stressed vowels? If there is a single hierarchy, how to explain the different order of acquisition often attested to in the data? And finally, what governs the order of acquisition of pre-tonic vowels?

As we have seen, from the first moment of analysis, at 1;4, children were already able to produce pre-tonic syllables, an aspect already pointed out by Santos (2005, 2007), which means that at an early stage they have the syllabic position where unstressed vowels are produced. However, studies of various aspects of phonological acquisition also point to the higher prosodic salience of stressed syllables. The studies in Lamprecht (2004) show that the stressed syllable is a facilitator for the correct production of the segments (vowel and consonant), and that complex syllables (either with branching rhymes or branching onsets) are first correctly produced in the stressed position. In addition, we can observe that the stressed syllables are the locus of greater stability for

the production of vowels, as has already been suggested by Rangel (2002). Miranda, in her study, points to what is at the heart of the CHT: "This [mid vowels being stable in stressed position and unstable in pre-tonic position] indicate that segments are acquired based on how they behave in the system and not as a single unit of a set." (MIRANDA, 2013, p.94)[8].

In other words, Miranda claims that the only reason for the later acquisition of the vowels in pre-tonic position is vowel instability in this position, a result of the processes that affect these vowels, a position shared by Bohn (2015).

Let us see how Miranda's proposal would work. According to Wetzels (1992), BP has phonological processes that affect only vowels in unstressed syllables (neutralization in the pre-tonic position, neutralization of non-final pos-tonic vowels, neutralization of the final pos-tonic vowels, and vowel harmony); processes that affect only vowels in stressed syllables (datilyc lowering, spondaic lowering, mid-vowel lowering in verb roots, neutralization by occlusive adjustment in coda); and a process that affects vowels regardless of their stress (nasal neutralization).

Wetzels (1992) suggests a nasal neutralization process that explains why, in nasal-closed syllables, only mid-high vowels emerge (e.g /ˈõn.xa/ '*honra*'/ˈhonor' and /ˈɔ.ɾa/ ' 'hora'/'hour'). However, children have no evidence that this is a process of neutralization (between low and mid-high vowels), and nothing would prevent them from processing this information as a gap in the system.

Cases affecting stressed vowels are also instigating from the acquisition point of view. Coda occlusive neutralization shows cases in which syllables closed by a nonsonorant stop are always produced as mid-low vowels (e.g [ˈsɛk.so] 'sexo'/ 'sex', [iˈnɔks] '*inox*' / 'stainless steel'). However, children have no evidence that there is neutralization between mid vowels in this position. On the contrary, children are exposed to the fact that these vowels, when in pre-tonic syllables, become mid-high, following the rule of neutralization of the pre-tonic position ([sekˈsista] 'sexista'/ 'sexist', [i.nok.siˈda.vew] 'inoxidável' / 'stainless steel-adj'). The spondaic lowering process shows that in words with an exceptional stress pattern (paroxitons with a heavy final syllable) the mid vowel is always produced as mid-low ([reˈpɔɾteɾ] 'repórter'/ 'reporter', [ˈsɛ.zar] '*César*' / 'Ceasar'). However, this rule has many exceptions (for example, the rule is blocked when the syllabic weight is generated by a inflectional suffix ([ˈtes.til] 'textil' / 'textile') and children are exposed only to the neutralization of the pre-tonic vowel (eg [xepoɾˈtaʒẽj] 'reportagem' / 'news report', [seˈzaɾja] 'cesárea' / 'cesarean (c-cut)'). Datylic lowering shows that the mid vowels in proparoxytone position are always produced as mid-low ([eskeˈleto] 'esqueleto' / 'skeleton' > [eskeˈlɛ.tiko] 'esquelético' / 'skeletal'). Finally, verbal root lowering captures the fact that in verbal forms the mid-low vowel always occurs, whereas in the noun form both a mid-low and a mid-high vowel can occur ([ˈfoɾ.sa] 'força' /

---

8   "Isso [vogais médias são estáveis na posição tônica e instáveis na posição pretônica] parece apontar para a ideia de que os segmentos são adquiridos com base no funcionamento do sistema e não como simples unidade de um conjunto." (MIRANDA, 2013, p.94).

'force-noun' vs. ['fɔɾ.sa] 'força' / 'force-verb-imp', [a'pelo] 'apelo' / 'appeal-noun' vs. [a'pɛlo] 'apelo' / 'appeal-verb- 1sg. present'.

Common to almost all of the processes described above (except for the dactylic lowering and verbal lowering) is the fact that the child is not exposed to alternate forms, but to a single form. It is only by comparing several vocabulary items that one can infer that there is a phonological process being applied - which would explain the absence of a given phoneme in one position. In short, in the stressed position, children have no evidence that vowel neutralization occurs.

The phonological processes in unstressed syllables behave differently. In all of them the child has clues that there is a process of neutralization either because it is an optional process (such as Vowel Harmony - e.g. p[e]rigo ~ p[i]rigo 'perigo' / 'danger'), or because derived words show an alternation (e.g. *p[ɛ]le* 'pele'/ 'skin' > p[e]lado 'pelado'/'naked´, for the neutralization of the pre-tonic; ['kaxʊ] 'carro'/'car' > [kaxose'ɾia] 'carroceria'/ 'wagon', for the neutralization of the final unstressed vowel; núm[e]ro ~ núm[i]ro 'número' / 'number' for non-final pos-tonic neutralization). Thus, it can be said that the vowel system is more stable in the stressed position than in the pre-tonic position. The alternations are clues, for the children, that it is one vowel system, but it might take longer to establish which vowel is the underlying representation in unstressed position when compared to the stressed position, where there is no alternation. In short, at the same time that the phonological processes provide evidence for the pre-tonic and its relation to the stressed system, the alternations also make the acquisiton of the vowels in this position more difficult and consequently take more time.

Considering the data in this study, we have shown that for all the children analyzed, vowel /a/ was the first to be acquired. In adult speech, in the pre-tonic position, vowel /a/ does not undergo neutralization, it is the most stable of all the vowels in this position; its early acquisition, then, confirms Miranda (2013)'s and Bohn (2015)'s analyses. However, the next vowels to be acquired are all subject to neutralization in adult speech: /ɛ, e/ and /ɔ, o / by neutralization of the pre-tonic position, /e, i/ and /o, u/ by vowel harmony. Regarding these vowels, Miranda (2013) follows Jakobson (1968 [1941]), for whom the acquisition will follow the same pattern as in the stressed position: initially the high vowels and then the mid-vowels. Bohn (2015)'s results do not follow this order, however (cf. Table 9). Bohn points out that, in adult language, phonological processes do not affect front and back vowels the same way, and assumes that the child is sensitive to differences in how this phonological process is applied in adult speech. She exemplies her reasoning with vowel harmony: according to Callou, Moraes and Leite (2002), Viegas (2001), Yacovenco (1993) and Casagrande (2004), vowel harmony affects vowel /e/ much more often than /o/, which leads these authors to propose that only /e/ undergoes the process in BP, and that the cases in which /o/ is raised to [u] are phonetically motivated. As /e/ undergoes more processes than /o/, it is a more unstable vowel, and that is why, according to Bohn, this phoneme would be the last to be acquired in the pre-tonic position.

However, if Bohn's analysis explains why /e/ is acquired later, it does not explain why there is an asymmetry in acquisition: why does the child acquire /i, o/ and not /i, u/? That is, if we assume instability affects acquisition, with no changes in the hierarchical structure of the vowels and no other factor influencing the process, we should expect children to follow the path proposed by Miranda: a> ... > o > e (where ... indicates a possible ordering between /i/ and /u/ depending on one being more unstable than the other). However, that is not what Bohn's results show. For the three children, in the pre-tonic position, after the low vowel /a/ children acquired a high and mid vowel (simultaneously or with little temporal difference).

According to Jakobson's proposal, vowel acquisition depends on contrast: low > high > mid. CHT proposes that the internal structure of the segments is built following feature contrasts (DRESHER, 2009), but it does not explain how contrasts are acquired. Bohn (2015) applies CHT to observe if this theory can explain the acquisition data and proposes that it is possible to have evidence of the construction of the hierarchies through the substitutions found in the data: if the vowels [o, u, a] alternate, it is because at that moment they share the same feature and do not contrast. Our proposal here is that the child deals with two variables: a principle of maximum contrast (inspired by the notion of contrast proposed by CHT) and the segments instability in the input (as pointed out by Miranda 2013, among others). According to CHT, only contrastive features are acquired. We propose that there should be a Principle of Contrast that governs the order of acquisition of segments. Looking at our data, this means that the child, when acquiring the pre-tonic position, selects, within the hierarchy, which has already been constructed, which segments are more contrastive. For example, child L., at age 1;6, has only vowel /a/ in the hierarchy (evidenced in the stressed position), and it is this only vowel that it also produced in the pre-tonic position. At 1;10, L. also starts producing /i, o/ in pre-tonic position when she already produces /a, i, e, o/ in stressed position. The question we pose is: why does the child select /i/ and not /e/? The first part of our answer to this question concerns the Contrast Principle in its maximum version (which we will call the Principle of Maximum Contrast): if the child selects /o, e/ or /i, u/, these pairs of segments are contrastive, but not maximally contrastive. Observe that /i, u/ are maximally contrastive with respect to /a/ - they show a maximum contrast of height and place -, but they contrast in place only, as they are both high vowels.[9] If the child selects /o, e/ we have the same contrast problem: they contrast with respect to place only, and a smaller contrast is established in relation to /a/ (as now, instead of high *vs.* low vowels, we have a mid-high *vs.* low vowel).

On the other hand, if the child selects /e, u/ or / i, o/, the contrast is now greater: in both cases, vowels contrast with respect to place and height. The child would then have three levels of height in the pre-tonic position: /a/ (low), /o/ (mid-high) and /i/ (high). We emphasize that this does not mean that the child has created an asymmetric system, because the system is the same as in the stressed position, where the seven

---

[9] One might argue that these vowels also contrast phonetically in roundness. However, roundness is not contrastive in BP.

vowels have already been acquired. In the pre-tonic position, there are three groups of vowel: /a; i, o; e, u/, regardless of whether the hierarchy of vowel phonemes starts with place (as is the case of Am. and L.) or with height (as is the case of A.).

In summary, the data on the acquisition of pre-tonic vowels indicate that the child treats pre-tonic vowels and stressed vowels the same way; that is, they have the same internal feature hierarchy. Still, the acquisition order is different. Our proposal is that this order of acquisition in the pre-tonic position is affected by the instability of some vowels and by a principle of maximum contrast. The notion of contrast assumed by CHT is responsible for the selection of contrastive features. The Contrast Principle in its maximum version governs the order of acquisition: given the contrastive features, the segments tend to emerge by contrasting maximally (in the greatest number of features).

Finally, one might wonder why the Principle of Maximum Contrast is not applied in the acquisition of the stressed vowels. As we have seen, in the stressed position we have the emergence of vowels of the same height: for A., /i, u/ are acquired at the same time, L. acquires /e, o/ at the same time; Am. acquires /o/ and soon after /e/. The difference between the two systems is that the stressed system is more stable, with vowel contrasts more easily to be distinguished. Thus, the Principle of Contrast would apply, but without the need to be in the greatest possible number of contrasts (Principle of Contrast Maximum).

Final considerations

In this article, we discuss the acquisition of vowels taking into account the fact that different children have different orders of acquisition of the segments. As we have seen, theories that assume a universal hierarchy cannot account for this variability. We assume, then, that while the notion of hierarchy is innate, the organization of feature in the hierarchy is constructed - as proposed by CHT (DRESHER, 2009).

Our objective is to discuss whether the acquisition of pre-tonic vowels is different from the acquisition of stressed vowels. To this end, we revised Bohn (2015)'s data. The author assumes Miranda (2013), who claims that the acquisition of the segments in the pre-tonic position is subject to the instability affecting these vowels. Thus, vowel /e/, for example, would be acquired after /o/ because /e/ is more subject to phonological processes in this position. We argue that instability contributes to the organization of vowels, but it should not be the only explanation, as, if taken to its ultimate consequences, it would predict that the emergence of vowels in the pre-tonic position would be: /a/ > /i, u/ > /e, o/. However, the data from the three children analyzed indicate that the high vowels are not acquired simultaneously, but combined with the mid-vowels (initially /i, o/ and then /e, u/). Thus, we propose that in addition to the instability, there is a Principle of Maximum Contrast that guides the acquisition of pre-tonic vowels in BP. According to this principle, segments with less prosodic salience, such as the pre-tonic subsystem discussed in this study, should be maximally contrastive to be acquired. In this sense, the order of acquisition is not governed by the notion of

contrasts, but by maximum contrasts, which, for pre-tonic vowels in PB, result from the opposition between both height and place. In addition, the results indicate that the child is sensitive to the phonological processes very early as it influences the acquisition of the segmental layer. Since vowel /e/ is very unstable because of the processes it undergoes, the child will first acquire the mid vowel /o/. The Principle of Maximum Contrast will direct the child to acquire a vowel which maximally contrasts with /o/, that is /i/. Our hypothesis is that this principle only applies in contexts of neutralization or less prosodic saliency of the segments, which explains why we do not find its effects in the stress position. A possible verification of this hypothesis may come from studies in which the acquisition of other unstressed positions are conducted (i.e. the non-final pos-tonic and the final pos-tonic).

Acknowledgments

BOHN, G.; SANTOS, R. A aquisição de vogais pretônicas em Português Brasileiro. **Alfa**, São Paulo, v.62, n.1, p.191-221, 2018.

▪ *RESUMO: O artigo discute a aquisição de vogais pretônicas em português brasileiro, por 3 crianças monolíngues adquirindo o dialeto paulista, com idade entre 1;4 e 3;5, e sua relação com a aquisição das vogais no ambiente tônico. Com base em Miranda (2013), partimos do pressuposto de que a aquisição das vogais pretônicas está sujeita à instabilidade desse subsistema, e, portanto, segmentos afetados por processos fonológicos seriam adquiridos mais tardiamente nessa posição. As produções mostram que as vogais altas pretônicas são adquiridas em contraste com as vogais médias, (/i,o/ e /e,u/), sendo a pretônica /o/ adquirida antes de /e/. Analisamos nossos resultados à luz da Hierarquia Contrastiva de Traços (DRESHER, 2009), para a qual a representação lexical dos segmentos é específica de cada língua, trazendo somente os traços contrastivos e ativos em processos fonológicos naquele sistema, e propomos que a aquisição da pauta pretônica é regida por um Princípio de Contraste Máximo: devido a instabilidade dessa posição, os segmentos devem ser maximamente contrastivos, ou seja, por ponto e altura vocálica. A pretônica /e/, por ser a mais instável (cf. CALLOU; MORAES; LEITE, 2002, VIEGAS, 2001 e YACOVENCO, 1993), é a última a ser adquirida, trazendo consigo a pretônica /u/.*

▪ *PALAVRAS-CHAVE: Aquisição fonológica. Contraste fonológico. Vogal pretônica.*

REFERENCES

BISOL, L. Sândi externo: o processo e a variação. In: KATO, M. (Org.). **Gramática do Português Falado**. v. 5. Campinas: Ed. da UNICAMP, 1996. p.55-96.

BOHN, G. P. **Aquisição das vogais tônicas e pretônicas do português brasileiro.** 2015. 218 f. Tese (Doutorado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2015.

BONILHA, G. F. G. **Aquisição fonológica do português brasileiro:** uma abordagem conexionista da Teoria da Otimidade. 2004. 389 f. Tese (Doutorado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2004.

CALLOU, D.; MORAES, J.; LEITE, Y. A elevação das vogais pretônicas no português do Brasil: processo(s) de variação estável. **Letras de Hoje**, Porto Alegre: EDIPUCRS, v.37, n.1, p.9-24, 2002.

CÂMARA Jr., J. M. **Para o estudo da fonêmica portuguesa**. 2. ed. Rio de Janeiro: Padrão, 1977.

CASAGRANDE, G. P. B. **Harmonização vocálica:** análise variacionista em tempo real. 2004. 171 f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2004.

CHERRY, E..; HALLE, M.; JAKOBSON, R. Toward the logical description of languages in their phonemic aspect. **Language**, v.29, p.34-46, 1953.

CLEMENTS, N.; HUME, E. The internal organization of speech sounds. In: GOLDSMITH, J. (Org.). **The handbook of phonological theory**. Oxford: Blackwell, 1995. p.245-306.

DRESHER, B. E. **The Contrastive Hierarchy in Phonology.** Cambridge: Cambridge University Press, 2009.

_____. The arch not the stones: universal feature theory without universal features. **Nordlyd**, University of Tromso, Noruega, n.41, p.165-181, 2015.

FEE, E. J. **Underspecification, Parameters, and the Acquisition of Vowels.** 1991. 489 f. PhD (Dissertation) – University of British Columbia, 1991.

HALL, D. C. **The role and representation of contrast in phonological theory** 2007. 293 f. Tese (Doutorado) – University of Toronto, Toronto, 2007.

HALLE, M. **The sound pattern of Russian:** a linguistcs and acoustical investigation. Second Printing. The Hague: Mouton, 1971[1959].

INGRAM, D. **Procedures for the Phonological Analysis of Children's Language.** Baltimore: University Park Press, 1981.

_____. **First Language Acquisition:** Method, Description, and Explanation. Cambridge: Cambridge University Press, 1989.

JAKOBSON, R. **Child Language, Aphasia and Linguistic Universals.** The Hague: Mouton. 1968[1941].

LAMPRECHT, R.; BONILHA, G.; FREITAS, G.; MATZENAUER, C.; MEZZOMO, C.; OLIVEIRA, C.; RIBAS, L. **Aquisição fonológica do português**: perfil de desenvolvimento e subsidies para terapia. Porto Alegre: Artmed, 2004.

LEE, S. H. Contraste das Vogais no PB. **Portuguese-Brazilian Studies**, v.5, p.201-221, 2008.

MACKEN, M. A. Developmental reorganization of phonology: a hierarchy of basic units of acquisition. **Lingua**, v.49, n.1, p.11-49, 1979.

MACKENZIE, S. **Contrast and Similarity in Consonant Harmony Processes**. 2009. 226 f. Tese (Doutorado em Linguística) – Universidade de Toronto, Toronto, 2009.

MATZENAUER, C. L. B. Sobre as vogais médias pretônicas na aquisição do português brasileiro. **Revista Organon,** Porto Alegre: Universidade Federal do Rio Grande do Sul, n.46, p.71-108, 2009.

MIRANDA, A. R. Insights sobre a representação das vogais pretônicas no Português do Brasil: dados de Desenvolvimento Fonológico e de escrita inicial. **Revista Organon**, Porto Alegre: Universidade Federal do Rio Grande do Sul, v.28, n.54, p.83-100, jan./jun. 2013.

MIRANDA, A. R.; MATZENAUER, A. R. Traços distintivos e a aquisição das vogais do PB. In: HORA, D. (Org.). **Vogais no ponto mais oriental das Américas.** João Pessoa: Idéia/UFPB, 2009.

NOGUEIRA, M. V. **Aspectos segmentais no processo de sândi externo no falar de São Paulo.** 2007. 154 f. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.

PYE, C.; INGRAM, D.; LIST, H. A comparison of initial consonant acquisition in English and Quiché. In: NELSON, K.; KLEECK, A. Van (Org.). **Children's Language.** Hillsdale, NJ: Lawrence Erlbaum Associates. 1987. p.175-190.

RANGEL, G. A. **Aquisição do Sistema Vocálico do Português Brasileiro**. 2002. 170 f. Tese (Doutorado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2002.

SANTOS, R. S. **Projeto de Aquisição do Ritmo em Português Brasileiro**. FAPESP. 2003/13565-4, 2005.

_____. **A aquisição prosódica do português brasileiro de 1 a 3 anos:** padrões de palavra e processos de sândi externo. 2007. Tese (Livre-docência) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.

SILVA, C. C. **Aquisição da regra de assimilação de vozeamento em Português Brasileiro.** 2008. 161 f. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Ciências Humanas e Letras, Universidade de São Paulo, São Paulo, 2008.

VIEGAS, M. C. **O alçamento de vogais médias pretônicas e os itens lexicais**. 2001. 170 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2001.

VIHMAN, M. et al. Phonological development from babbling to speech: common tendencies and individual diferences. **Applied Psycholinguistics,** n.7, p.3-40, 1986.

VOGELEY, A. C. S. **Vogais médias pretônicas:** aquisição e variação. 2011. 236 f. Tese (Doutorado em Linguística) -Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 2011.

VOGELEY, A. C.; HORA, D. Aquisição das vogais médias pretônicas. In: XV Congresso Internacional de la ALFAL, 2008, Montevideo. **Livro de Resúmenes – XV Congresso Internacional de la ALFAL**. Montevideo: Gega s.r.l., 2008. p.150.

WETZELS, L. M. Mid-vowel alternations in the Brazilian Portuguese. **Cadernos de Estudos Linguísticos**, Campinas: Universidade Estadual de Campinas, v.23, p.19-55, 1992.

YACOVENCO L. C. **As vogais médias pretônicas no falar culto carioca**. 1993. 185 f. Dissertação (Mestrado em Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1993.

ZHANG, X. I. **Vowel systems of the Manchu-Tungus languages of China**. 1996. 252 f. Tese (Doutorado em Linguística) - Universidade de Toronto, Toronto, 1996.

SUBJECT INDEX

AUTHOR INDEX

PAPER SUBMISSION INSTRUCTIONS

Alfa: Revista de Linguística

1. Editorial policy

ALFA – Revista de Linguística – the linguistics journal of the São Paulo State University (UNESP), sponsored by the Research Provost, publishes papers authored by professors, researchers, and PhD students in co-authorship with professors and researchers from national and international institutions of learning, teaching and research. Maximum number of co-authors should be 3. Regular issues are organized in a "continuous flow" system, and special issues are edited according to the organizers' convenience. The journal publishes papers, book reviews, interviews and translations with a substantial contribution to any of the various branches of Linguistics.

Contributions in the form of articles should be original and unpublished and cannot be simultaneously submitted for publication in other journal. Only reviews of books published in Brazil in the last 2 years and abroad in the last 4 years should be submitted for publication in the journal. Translations should be preferably of scientific articles and book chapters published within twelve months of submission; interviews should be conducted with researchers with academic prestige acknowledged in Brazil and abroad.

All submissions are read by two anonymous referees. Authors' identities are not revealed to the referees. Should the judgment of the two referees be inconsistent, a third referee will be called in. Once the refereeing process is concluded, the review copies are sent to the author, or to the first author of co-authored papers, with the appropriate instructions.

Papers can be written in one of the following languages: **Portuguese, English, French, Spanish**, or **Italian**. In papers written in **Portuguese**, *TÍTULO, RESUMO*, and *PALAVRAS-CHAVE* should come before the body of the paper, and their English versions, *TITLE, ABSTRACT*, and *KEY-WORDS*, after it. In papers written in any of **the other languages**, the corresponding elements that come before the body of the paper should be written in the same language the paper was written; the corresponding elements that come after the body of the paper should be written in **Portuguese** for papers written in **English** and in **English** for papers written in **French, Spanish**, or **Italian**.

All articles are published in a bilingual format, with English necessarily as the second language version. Authors who submit for approval an article originally in English must, in case it is accepted, provide its version in Portuguese, following the same guidelines indicated for the English language. Only articles with accepted versions in Portuguese (or another chosen language) and English will be

published. If both versions are not accepted, the article will not be published.

The journal editor reserves the right to return a manuscript if it departs from the style requirements. When applicable, a personal letter will be sent to the author, asking for improvements and adaptations.

Authors are responsible for the data and concepts expressed in the paper as well as for the correctness of the references and bibliography.

2. Online submissions

To submit a paper, authors must be registered on the journal's website. To register, create a login name and a password by clicking **Acesso** (Access) on the journal's website. After logging in, fill in the profile by clicking **Editar Perfil** (Profile Editing) and start the submission process by clicking **Autor** (Author) and then **CLIQUE AQUI PARA INICIAR O PROCESSO DE SUBMISSÃO** (Click here to start the submission process). Follow the **five-step submission process** below:

Step1. Confirm the agreement to the Journal Policies (**Condições de submissão**) and the Copyright Terms (**Declaração de Direito Autoral**) by checking the appropriate boxes. Select either **Artigo** (Paper) or **Resenha** (Review paper). Save the form and go to step 2.

Step 2. Enter metadata: first name, last name, e-mail, bio statement, and paper title are obligatory. Save the form and go to step 3.

Step 3. Upload the paper file. Go to step 4.

Step 4. If necessary, upload supplementary files such as appendixes and annexes with research tools, data and tables, which should conform to the ethical standards of assessment, sources of information usually unavailable to readers, and pictures or tables that cannot be inserted into the text itself. Go to step 5.

Step 5. Confirm the submission.

After confirming the submission, authors will receive a confirmation e-mail from the journal editor. After submission, authors can follow the process up, from submission and acceptance, through assessment and final version preparation, to on-line publication.

After submission, articles will be assigned to reviewers by the Editorial Board or special issue editors. The journal's Editorial Board and Editors are responsible for the policy of paper selection, which is available at the link **Sobre a Revista>Processo de Avaliação por Pares** (About the Journal>Peer Review Process).

3. Preparation of manuscripts

3.1. Presentation

Authors should ensure that their electronic copy is compatible with *PC/MSWord*, and use *Times New Roman*, *12-point* size. The page size should be set to A4 (21cm x 29.7cm), and the text body should be one-and-a-half spaced throughout. Leave 3 cm from the top of the page and on the left margin, and 2.0 cm from the bottom of the page and on the right margin. Articles should have a minimum of 15 pages and not exceed **30 pages**, including bibliography, appendixes, and annexes. The text must meet the rules of Portuguese new orthographic agreement, which became mandatory in Brazil from January 2016. Two versions of the paper must be submitted: one containing the name and academic affiliation of

author(s), and one in which all references to the author(s), including citations and bibliographical references are erased.

3.2. Paper format

The format below should be followed:
Title. The title should be centered and set in **bold** CAPITALS at the top of the first page. Runover* titles should be single-spaced.

Author's name: The name of each author follows the title and should be given in full with the surname in CAPITALS and aligned to the right margin, on the third line below the title, with a footnote marked by an asterisk referring to metadata in the following order: acronym and full name of the institution to which author(s) is(are) affiliated, city, state, country, zip code, e-mail.

Abstract. The abstract, which must summarize the contents of the paper (goals, theoretical framework, results, and conclusion), should conform to the following: it should appear on the third line under the name(s) of the author(s), contain at least 150 and at most 200 words, be single-spaced, and, with no indentation for the first line, be preceded by the word ABSTRACT in CAPITALS in the same language of the paper,

Keywords. Each keyword (seven, at most) is followed by a period. They should be preceded by the word KEYWORDS in CAPITALS, and appear two lines below the abstract. The Editorial Board suggests that the keywords should match general concepts of the paper subject domain.

Body of the paper. The body of the paper should be one-and-a-half-spaced throughout. It begins on the third line below the keywords.

Subsection titles. The subsection titles should be typeset in **bold** and aligned to the left margin. They should not be numbered. There should be two one-and-a-half-spaced blank lines before and one one-and-a-half-spaced blank line after each subsection title.

Acknowledgements. Acknowledgements should conform to the subsection title lay-out, and should be preceded by the word **"Acknowledgements"** set in **bold**.

Title in English. For papers written in Portuguese, French, Spanish and Italian, the title in English (with no capitals and no bold) should be placed two blank single-spaced lines after the paper text body.

The abstract in English. For papers written in Portuguese, French, Spanish and Italian, the abstract in English should be typeset in *italics*, be preceded by the word *ABSTRACT*, typeset in *italics* and in CAPITALS, with no indentation for the first line, be single-spaced, and be placed three blank single-spaced lines after the title in English.

The keywords in English. For papers written in Portuguese, French, Spanish, and Italian, the keywords in English should be typeset in *italics*, be preceded by the word *ABSTRACT*, typeset in *italics* and in CAPITALS, and be placed three blank single-spaced lines after the abstract in English.

NOTE: For papers written in English, the title, abstract, and keywords referred to in 7, 8 and 9 above, respectively, should be written in Portuguese.

References. The subtitle **References** should be set in **bold**, with no indentation for the first line, and placed two blank

single-spaced lines after the keywords. The reference list should be single-spaced and ordered alphabetically and chronologically (see 3.3.1 below), placed three blank single-spaced lines after the keywords in English. **12. Bibliography.** The bibliography list, if essential, should come after the reference list. The word "**Bibliography**" should be set in **bold**, with no indentation for the first line, and placed three blank single-spaced lines after the reference list, aligned to the left. It will include all works not mentioned in the paper or in its footnotes.

3.3. Further instructions

3.3.1 Reference guidelines Both reference and bibliography lists should be ordered alphabetically by the last name of the first author. A single space should separate one reference item from the other. The names of the translators must be specified. Examples:

Books

AUTHIER-REVUZ, J. **Palavras incertas:** as não coincidências do dizer. Tradução de Cláudia Pfeiffer et al. Campinas: Ed. da UNICAMP, 1998.
CORACINI, M. J.; BERTOLDO, E. S. (Org.). **O desejo da teoria e a contingência da prática.** Campinas: Mercado das Letras, 2003.
LUCHESI, D. **Sistema, mudança e linguagem**: um percurso na história da linguística moderna. São Paulo: Parábola Editorial, 2004.

Book chapters

PÊCHEUX, M. Ler o arquivo hoje. In: ORLANDI, E. P. (Org.). **Gestos de leitura: da história no discurso**. Tradução de Maria das Graças Lopes Morin do Amaral. Campinas: Ed. da UNICAMP, 1994. p.15-50.

Thesis and dissertations

BITENCOURT, C. M. F. **Pátria, civilização e trabalho**: o ensino nas escolas paulista (1917-1939). 1998. 256 f. Dissertação (Mestrado em História) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1998.

Papers in journals

SCLIAR-CABRAL, L.; RODRIGUES, B. B. Discrepâncias entre a pontuação e as pausas. **Cadernos de Estudos Linguísticos**, Campinas, n.26, p. 63-77, 1994.

Online papers

SOUZA, F. C. Formação de bibliotecários para uma sociedade livre. **Revista de Biblioteconomia e Ciência da Informação**, Florianópolis, n.11, p. 1-13, jun. 2001. Disponível em: <...> Acesso em: 30 jun. 2001.

Newspaper articles

BURKE, P. Misturando os idiomas. **Folha de S. Paulo**, São Paulo, 13 abr. 2003. Mais!, p.3.

EDITORA plagiou traduções de clássicos. **Folha de S. Paulo**, São Paulo, 4 nov. 2007. Ilustrada, p. 6.

Online publications

UNIVERSIDADE ESTADUAL PAULISTA. Coordenadoria Geral de Bibliotecas. Grupo de Trabalho Normalização Documentária da UNESP. Normalização Documentária para a produção científica da UNESP: normas para apresentação de referências. São Paulo, 2003. Disponível em: <...>. Acesso em: 15 jul. 2004.

Paper in edited volumes, conference proceedings, and working papers

MARIN, A. J. Educação continuada. In: CONGRESSO ESTADUAL PAULISTA SOBRE FORMAÇÃO DE EDUCADORES, 1., 1990. **Anais**... São Paulo: UNESP, 1990. p. 114-118.

Films:

Macunaíma. Direção (roteiro e adaptação) de Joaquim Pedro de Andrade. Filmes do Serro/Grupo Filmes/Condor Filmes. Brasil: 1969. Rio de Janeiro: Videofilmes, 1969. Versão restaurada digitalmente, 2004. 1 DVD (105 minutos), color.

Paintings, photographs, illustrations, drawings:

ALMEIDA JÚNIOR. **Caipira picando fumo**. 1893. Óleo sobre tela. 17 cm X 23,5 cm. Pintura pertencente ao acervo da Pinacoteca do Estado de São Paulo.

PICASSO, Pablo. **[Sem título]**. [1948]. 1 gravura. Disponível em: <http://belgaleria.com.br>. Acesso em 19 ago. 2015.

Music CDs (as a unit or tracks)

CALAZANS, T. **Teca Calazans canta Villa Lobos**. Rio de Janeiro: Kuarup Discos, 1999. 1 CD.
CALAZANS, T. Modinha. In: CALAZANS, T. **Teca Calazans canta Villa Lobos**. Rio de Janeiro: Kuarup Discos, 1999. 1 CD.

3.3.2. In-text references and quotations

For references in the text, the surname of the author should be in CAPITALS, enclosed in parentheses; a comma should be placed between the author's last name and year, e.g. (BARBOSA, 1980). If the name of the author is part of the text, only the year is enclosed in parentheses: "Morais (1955) argues..."
Page numbers follow the year and are preceded by "p."; note a comma and a space between year and "p.", and between "p." and the number, e.g. (MUNFORD, 1949, p. 513).
References of the same author with the same year should be distinguished by using lower case letters in alphabetical order, e.g. (PESIDE, 1927a), and (PESIDE, 1927b). For references with one author and up to two co-authors, semi-colons are used to separate the surnames, e.g. (OLIVEIRA; MATEUS; SILVA, 1943); for references with more than two co-authors, the expression "et al." substitutes for the surnames of the co-authors, e.g. (GILLE et al., 1960).

Quotations longer than three text lines should be set in 11-point font size, and set out as a separate paragraph (or paragraphs) on a new line. The paragraph (or paragraphs) should be 4.0 cm from the left margin throughout, without any quotation marks. Quotations shorter than three text lines should be included in double quotation marks in the running text. Quotations from texts in foreign languages must be translated into Portuguese. Published translations should be used whenever possible. The original text should appear in a footnote.

3.3.3. Italics, bold, underlining and quotation marks

Italics: Use italics for foreign words, book titles int the body of the text, or for emphasis.

Bold: Use bold only in the title of the article and in the text headings and subheadings.

Underlining: Avoid using underlining.

Quotation marks: can be used to highlight parts of the major works, such as titles of poems, articles, chapters. The major works should be highlighted in italics, as the statement above; quotation marks must be used in the body of the text for quotations of excerpts of works. Example: A linguística é uma disciplina que "[...] se baseia na observação dos factos e se abstém de propor qualquer escolha entre tais factos, em nome de certos princípios estéticos e morais" (MARTINET, 1972, p.3).

3.3.4. Footnotes

Footnotes should be kept to a minimum and placed at the bottom of the page. The superscript numerals used to refer to a footnote com after any punctuation sign (comma, semicolon, period, question mark, etc.).

3.3.5. Figures

Figures comprise drawings, graphs, charts, maps, diagrams, formulas, models, photographs, x-rays. The identifying caption should be inserted above the figures, centered, preceded by the designation word designative (Chart, Map, Figure etc); if there is more than one, figures must be numbered consecutively in Arabic numerals using the same font and size of the body of the text. Indication of the source and other information necessary for its understanding should appear below the figure. Figures should be submitted as separate files, saved in the program in which they were generated. Maps, photographs and radiographs should also be submitted as separate files, high-resolution (300 dpi). Author(s) are responsible for image copyrights.

3.3.6. Tables and text frames

Tables should be used to present statistical information, and text frames should be used to summarize and organize textual information. The title is inserted above the table, centered, beginning with **Table 1** in bold, followed by a hyphen and the title without emphasis, in the same font and size of the body text; the title of figures should be presented above the figure,

centered, beginning with Figure 1 in bold, followed by a hyphen and the title without emphasis, in the same font and size of the body text. The numbering is consecutive, in Arabic numerals; if you need to specify the data source, it must be placed below the table or the table and the text aligned to the left. Tables should be built with the open side borders and no lines separating columns.

3.3.7. Appendixes and Annexes

When absolutely necessary to the text comprehension, and within the limit of 30 pages, Annexes and / or appendixes, following the subsection style, should be included in the end of the paper, after the references or bibliography.

3.3.8. Review format

The review should contain, at the beginning, the complete reference to the book being reviewed, including number of pages, in Times New Roman, 14 point size, single spaced, no title, no summary, no keywords. The name(s) of the review author(s), in 12 point size, must appear on the third line below the reference of the book reviewed, preceded by "reviewed by [name(s) of author(s)]". Names must be followed by an asterisk referring to a footnote containing the following information: full name and acronym of the institution to which the review author(s) belong; city; state; country; zip code; email. The text of the review should begin on the third line below the name(s) of the author(s) in Times New Roman, 12 point size 12 and spacing 1.5.

Page format is as follows: paper size: A4 (21.0x 29.7 cm); left and top margins 3.0 cm, right and lower 2.0 cm; minimum length of 4 and maximum of 8 pages, including bibliographic references and annexes and/or appendices; indentation: 1.25 cm to mark the beginning of the paragraph; spacing: 1.5.

3.3.9. Translation format

Translated articles are subjected to a peer review process, to decide on the opportunity and the convenience of their publication. They should follow the article format, where applicable. In the second line below the name of the author of the translated text, right-aligned, the name(s) of the translator(s) should appear in the following format: "Translated by [name(s) of the translator(s)]", with an asterisk referring to a footnote with the following information: full name and acronym of the institution to which the translator(s) belong; city; state; country; zip code; email. The translated text must be accompanied with a written authorization of the publisher responsible for the original publication.

3.3.10. Interview format

Interviews are subjected to a peer review process, which decides on the opportunity and the convenience of its publication. The format of the interview is the same required for articles, but the title should contain, besides the general theme, the expression "Interview with [interviewee name]", without emphasis, with an asterisk referring to a footnote containing a brief review of the biography of the

interviewee, which clearly demonstrates her/his scientific relevance. The author(s) of the interview should follow, according to the rules established for articles.

3.3.11. English version

The author(s) of paper accepted for publication in Portuguese, French, Spanish or Italian must provide the English version of the text until the deadline shown in the e-mail notification of acceptance. The standards for citation of authors in the text and the references of the English version are the same as the ones in Portuguese. *Alfa* appoints reviewers to evaluate the English version of the article. The review is restricted to checking the quality of translation, i. e. adequation to the standard norms of English usage for research papers.

In case there are citations of works with an English-language edition, this edition should be used both in the text and in the references. In case there is no English edition, the quoted text should be translated into English, and the text in the original language of the edition used must be included in a footnote.

If the text contains figures scanned from advertisements in newspapers, magazines or similar midia, in Portuguese or another language, the English version of the text must be included in a footnote.

When the text contains examples the understanding of which involves the need to clarify morphosyntactic features, a literal version of them in gloss should be included, followed by the common English translation in single quotation marks. Example:

(1) isso signific-a um aument-o de vencimento-s (D2-SP-360)
    this mean-IND.PRS.3.SG a.M raise-NMLZ of salary-PL
    'this means a raise in salary.'

Conventions for the glosses: *The Leipzig Glossing Rules: conventions for interlinear morpheme-by-morpheme glosses*, edited by the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology (Bernard Comrie, Martin Haspelmath) and the Department of Linguistics at the University of Leipzig (Balthasar Bickel); available in http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

3.3.12. Copyright transfer – publication authorization

Following acceptance of the paper, the AUTHOR AGREES PROMPTLY TO ASSIGN COPYRIGHT to UNESP, which is granted permission to reproduce it and publish it in the journal. The terms "reproduce" and "publish" are understood according to their definitions in the Sections VI and I, respectively, of the Article 5 of the Law 9610/98. The PAPER is permanently available to be used in both print and electronic media, and readers have FREE-OF-CHARGE ACCESS through the Internet to read, download, copy, or print the FULL TEXTS of the PAPERS. This publication authorization is permanent, and UNESP is responsible for keeping the identification of the PAPER AUTHOR(S).

3.3.13. Privacy Policy

All names and addresses appearing in *Alfa* are used exclusively for the stated purposes of the journal and will not be made available for other purposes or to third parties.

Produção Editorial:

FCL - UNESP
Laboratório Editorial