

DIADORIM

19
VOLUME
ESPECIAL



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Diretor Adjunto de Pós-Graduação e Pesquisa

Profa. Dra. Maria Mercedes Riveiro Quintans Sebold

Vice-diretor

Prof. Dr. Pedro Paulo G. Ferreira Catharina

Coordenador do Programa de Letras Vernáculas

Profa. Dra. Angela Beatriz de Carvalho Faria

Substituto Eventual do Coordenador

Prof. Dr. João Antônio de Moraes

Comissão Deliberativa

Representantes Docentes

Língua Portuguesa

Profa. Dra. Regina Souza Gomes

Prof. Dr. Carlos Alexandre Victorio Gonçalves

Profa. Dra. Eliete Figueira Batista da Silveira (suplente)

Literatura Brasileira

Profa. Dra. Anélia Montechiari Pietrani

Prof. Dr. Adauri Silva Bastos

Prof. Dr. Godofredo de Oliveira Neto (suplente)

Literaturas Portuguesa e Africanas

Profa. Dra. Mônica do Nascimento Figueiredo

Profa. Dra. Carmen Lucia Tindó Ribeiro Secco

Profa. Dra. Teresa Cerdeira (suplente de Literatura Portuguesa)

Profa. Dra. Luci Pereira Ruas (suplente de Literatura Africana)

Representantes Discentes

Louise Bastos Corrêa (Doutoranda em Literatura Brasileira)

Victor Augusto Corrêa Azevedo (Doutorando em Língua Portuguesa)

Secretaria do Programa de Pós-Graduação

Maria Goretti Mello e Elizângela Campos

Diretora da Faculdade de Letras

Profa. Dra. Eleonora Ziller Camenietzki

Vice-Diretor

Profa. Dra. Cláudia Fátima Moraes Martins

Diretora Adjunta de Ensino de Graduação

Profa. Dra. Cláudia Fátima Moraes Martins

Diretor Adjunto de Cultura e Extensão

Prof. Dra. Karen Sampaio

Diretor Adjunto de Administração e Finanças

Jorge Luís Fernandes Barbosa

Coordenação de Infraestrutura Acadêmica

Prof. Dra. Christine Nicolaides

Coordenação de Intercâmbio e Internacionalização

Prof. Dra. Danúsia Torres

CENTRO DE LETRAS E ARTES (CLA)

Decania do Centro de Letras e Artes

Decana: Profa. Dra. Flora de Paoli Faria

Vice: Profa. Dra. Cristina Grafanassi Tranjan

Reitor:

Prof. Dr. Roberto Leher

Vice-reitor:

Profa. Dra. Denise Nascimento

Editora Chefe

Violeta Virginia Rodrigues, Universidade Federal do Rio de Janeiro, Brasil

Editoras Adjuntas

Eliete Figueira Batista da Silveira, Universidade Federal do Rio de Janeiro, Brasil

Marcia dos Santos Machado Vieira, Universidade Federal do Rio de Janeiro, Brasil

Editores Assistentes de Literatura

Prof. Dr. Nazir Ahmed Can, Universidade Federal do Rio de Janeiro - Campus Fundão, Brasil

Sofia Maria de Sousa Silva, Universidade Federal do Rio de Janeiro, Brasil

Editoras Assistentes de Língua

Filomena Azevedo Varejao, Universidade Federal do Rio de Janeiro, Brasil

Danielle Kely Gomes, Universidade Federal do Rio de Janeiro, Brasil

Coordenação de Revisão

Ana Paula Victoriano Belchor, Universidade Federal do Rio de Janeiro, Brasil

Equipe de Revisão

Anna Carolina da Costa Avelheda, Universidade Federal do Rio de Janeiro, Brasil

Gesieny Laurett Neves Damasceno, Universidade Federal do Rio de Janeiro, Brasil

Karen Pereira Fernandes de Souza (revisão de resumé) , Universidade Federal do Rio de Janeiro, Brasil

Heloise Vasconcellos Gomes Thompson (revisão de abstracts) , Universidade Federal do Rio de Janeiro, Brasil

Equipe Técnica

Sr. Rafael Andrade, UFRJ, Brasil

Miguel R. Amorim Neto, Universidade Federal do Rio de Janeiro, Brasil

Elir Ferrari, Universidade do Estado do Rio de Janeiro, Brasil

Webdesign e Diagramação

Sr. Rafael Andrade, UFRJ, Brasil

Diadorim: Revista de Estudos Linguísticos e Literários - Volume Especial (2017) - Rio de Janeiro: UFRJ, Programa de Pós-Graduação em Letras Vernáculas, 2017.

Semestral.

ISSN: 1980-2552.



DIADORIM

19

VOLUME ESPECIAL

CONTENTS

Foreword p.(8-10)

Cilene Rodrigues, Fábio Bonfim Duarte, Maria Eugênia L. Duarte

1. Truly Minimal Pronouns p.(11-36)

Anders Holmberg and On-Usa Phimsawat

2. PassiveP and the distinction between eventive, resultative, and stative passives p.(37-71)

João Claudio de Lima Júnior and Marina Rosa Ana Augusto

3. NumP and PossP in dialectal Brazilian Portuguese p.(72-98)

Bruma Karla Pereira

4. On the syntax of subjects in Brazilian Portuguese: using the 'split' pronominal system as the basis for an alternative analysis p.(99-139)

Eloisa Pilati, Rozana R. Naves and Heloisa M. Lima Sales

5. Revisiting Duarte (1995): For a gradient analysis of indeterminate subjects in Brazilian Portuguese p.(140-172)

Juliana Marins, Humberto Soares da Silva and Maria Eugênia L. Duarte

6. A comparative diachronic analysis of Wh-questions in Brazilian and European Portuguese p.(173-196)

Mayara Nicolau de Paula

7. The acceptability of clitic and tonic accusative 3rd person pronouns in written Brazilian Portuguese p.(197-218)

Cândido Samuel Fonseca de Oliveira and Ricardo Machado-Rocha

8. The prosody of elliptical constructions in Brazilian Portuguese: an experimental study p.(219-249)

Lilian Teixeira de Sousa

9. The Morphosyntax of Nominalization in Wayoro (Tupi): a preliminary approach p.(250-275)

Antonia Fernanda de Souza Nogueira

10. Some observations on word order and prosody in Karitiana relative clauses p.(276-304)

Karin Camolese Vivanco

11. Absence of evidence is not evidence of absence: the Pirahã case p.(305-324)

Cilene Rodrigues

12. Exploring agreement displacement from the internal to the external argument in the Tenetehára language (Tupi-Guarani family) p.(325-342)

Quesler Fagundes Camargos

13. Stative morpheme in Shimakonde: an anticausative morpheme? p.(343-376)

Ronaldo Rodrigues de Paula



Foreword

We have the pleasure of presenting this special issue of *Diadorim – Revista de Estudos Linguísticos e Literários do programa de Pós-Graduação em Letras Vernáculas (Faculdade de Letras, Universidade Federal do Rio de Janeiro)*. It contains selected papers from the Intermediate Meeting of the Theory of Grammar Working Group (Encontro Intermediário do GT de Teoria da Gramática), which took place in July 2015 at UFMG – the Federal University of Minas Gerais, with Fabio Bonfim Duarte and Cilene Rodrigues as organizers.

The volume contains thirteen papers. Together they represent an important contribution to the field, as they explore different aspects of syntax and its interfaces. They also represent a variety of languages. Although many papers focus on Portuguese, many other less well-known languages are investigated in the following pages, such as Karitiana, Shimakonde, Tenetehára, Pirahã and Wayoro. Thus, this volume contributes comparative studies on how language computes and expresses formal features, especially those related to PF interface.

We open the issue with a paper by Anders Holmberg (Newcastle University/Cambridge University), one of the invited speakers at the plenary session, who, in collaboration with On-Usa Phimsawat (Burapha University, Thailand), discusses recent developments in their research on null generic pronouns attested in partial pro-drop languages. Their paper shows that the presence of agreement correlates with a restriction to human referents in null inclusive generic pronouns. The seven articles that follow Holmberg & Phimsawat's paper deal with ongoing research on Brazilian Portuguese (BP): Lima Jr. & Augusto's paper proposes an analysis for eventive, resultative and stative passives. Pereira analyzes issues related to number agreement within possessive DPs in a dialect of BP spoken in Minas Gerais. She argues that in this dialect the number feature has been reanalyzed as a person feature, with the result that plural agreement is restricted to phrases located above NumP. Pilati, Naves and Salles investigate the syntax of subjects in Brazilian BP, proposing that the split in the pronominal/inflectional system (1st and 2nd person, inherently definite/referential and 3rd person, underspecified for both features) is the key to accounting for the distribution of 3rd person (spatial/ temporal) locative subjects, both lexical and null, in

BP. Next, Marins, Soares da Silva & Duarte analyze BP's strategies for subject indetermination, showing that they are divided into three different groups, according to a set of features they share with respect to arbitrary and generic reference. Nicolau de Paula brings a diachronic analysis of Wh questions in BP and European Portuguese (EP), based on data collected from theater plays. Her work shows that, contrary to BP, [WhV] is the most frequent order in EP, followed by [WhVS] and [WhSV]. The pattern [WhSV] is triggered by a clefting strategy. Oliveira & Machado-Rocha's paper offers an experimental way of testing the adequacy of using the obsolescent 3rd person accusative clitic and the tonic pronoun, an option available in natural speech in alternation with null objects. Sousa presents an acoustic analysis of sentences containing ellipsis in BP (stripping, gapping and VP ellipsis), concluding that there is a parallel between elision and de-accenting given information, while contrastive information is generally marked intonationally by High or Low-High contour.

The other six papers concentrate on languages from different families. Nogueira's paper discusses the morphosyntax of nominalization in Wayoro, an endangered Tupi language, spoken in the north of Brazil. Vivanco explores the relation between prosody and word-order in Karitiana, a Tupi-Arikém language also spoken in the north of the country. Another language from the north of Brazil, Pirahã, a Mura language, is the subject of Rodrigues's paper, in which she discusses the lack of transparent morphological evidence for self-embedding in Pirahã, suggesting that one should also look at the syntax-semantics interface for this type of evidence. Camargos analyzes agreement in Tenetehára, a language from the Tupi-Guarani family, and proposes an Agree-based analysis for the phenomenon. Rodrigues de Paula investigates the grammatical status of the stative morpheme in Shimakonde, an under-documented Bantu language spoken in Mozambique. The author assumes that this morpheme expresses stativity on the verb when related to remote past tense. In contrast, when related to recent past tense, the stative morpheme is also compatible with a telic aspectual denotation.

In conclusion, this volume represents an attempt to understand different grammatical aspects of unrelated languages, and it is just for this reason that putting together these thirteen papers was a very gratifying task. It gave us the opportunity to learn from others about the inner mechanisms of grammar. We are now very pleased to pass this volume on to you, the reader, with the hope that it will please you as well.

We are greatly indebted to those who contributed during the editorial process. We are especially thankful to the distinguished team of reviewers for their invaluable work. We also give our thanks to the proofreaders listed below, and to the Portuguese Linguistics Graduate Program of the Federal University of Rio de Janeiro – UFRJ and the Editorial board of *Diadorim* for their support.

Reviewers (in alphabetical order)

Ana Maria Martins (Universidade de Lisboa)
 Ana Paula Quadros Gomes (Universidade Federal do Rio de Janeiro)
 Bruna Franchetto (Universidade Federal do Rio de Janeiro/Museu Nacional)
 Charlotte Galves (Universidade Estadual de Campinas)
 David Alberto Seth Langa (Universidade Eduardo Mondlane/Maputo)
 Denny Moore (Museu Goeldi)
 Erica Rodrigues (PUC-Rio)
 Evani Viotti (Universidade de São Paulo)
 Filomena Sândalo (Universidade Estadual de Campinas)
 Heloísa Maria Moreira Lima Salles (Universidade de Brasília)
 Izete Lemkhul Coelho (Universidade Federal de Santa Catarina)
 Jairo Moraes Nunes (Universidade de São Paulo)
 José Olímpio de Magalhães (Universidade Federal de Minas Gerais)
 Juanito Ornelas de Avelar (Universidade Estadual de Campinas)
 Malte Rosemeyer (Universität Freiburg)
 Marcelo Modesto (Universidade de São Paulo)
 Márcia Santos Duarte de Oliveira (Universidade de São Paulo)
 Marco Antonio Martins (Universidade Federal de Santa Catarina)
 Marcos Maia (Universidade Federal do Rio de Janeiro)
 Maria Aparecida Corrêa Ribeiro Torres Morais (Universidade de São Paulo)
 Maria Cristina Figueiredo e Silva (Universidade Federal do Paraná)
 Marília Facó Soares (Universidade Federal do Rio de Janeiro/Museu Nacional)
 Marina Rosa Ana Augusto (Universidade Estadual do Rio de Janeiro)
 Mary Aizawa Kato (Universidade Estadual de Campinas)
 Plínio Almeida Barbosa (Universidade Estadual de Campinas)
 Sílvia Cavalcante (Universidade Federal do Rio de Janeiro)

Proofreaders (in alphabetic order)

Isabela Soares-Bastos (PUC-Rio)
 Isabella Bouvier (PUC-Rio)
 Leslie Rose (PUC-Rio)
 Paula Terra (PUC-Rio)

Organizers (in alphabetic order)

Cilene Rodrigues (PUC-Rio)
 Fábio Bonfim Duarte (Universidade Federal de Minas Gerais)
 Maria Eugenia Lammoglia Duarte (Universidade Federal do Rio de Janeiro)



Truly Minimal Pronouns

Anders Holmberg^a
On-Usa Phimsawat^b

ABSTRACT

This paper examines the properties of inclusive generic constructions, focusing on languages where the inclusive generic pronoun is a null category. We investigate empirical data from a set of languages with and without agreement to test Phimsawat's (2011) hypothesis that the inclusive generic pronoun lacks all phi-features, and therefore has the least restricted reading, due to there being no restriction on the reference. We show that this hypothesis cannot hold true universally, as phi-features trigger agreement in inflecting languages. We show that there is a correlation between presence of agreement and restriction to human reference for null inclusive generic pronouns, based on comparison of a set of languages without agreement (Thai, Mandarin Chinese, Korean, Vietnamese, and Sinhala) with a set of languages with agreement (Finnish, Brazilian Portuguese, Hebrew, Basque, and Tamil). An explanation in terms of feature architecture is proposed for this correlation. A prediction for generic PRO is discussed and shown to be inconclusive or false.

KEYWORDS: generic pronoun; null inclusive generic pronoun; languages with agreement; languages without agreement

RESUMO

Este artigo examina as propriedades das construções genéricas inclusivas, enfocando línguas em que o pronome genérico inclusivo é uma categoria vazia. Analisamos dados empíricos de um conjunto de línguas com e sem concordância a fim de testar a hipótese de Phimsawat (2011), segundo a qual o pronome genérico inclusivo não tem traços-phi, e, portanto, tem leitura menos restrita devido à ausência de qualquer restrição na sua referência. Mostramos que esta hipótese não se confirma universalmente, uma vez que os traços-phi desencadeiam concordância em línguas com flexão. Mostramos ainda que há uma correlação entre presença de concordância e restrição à referência humana nos pronomes genéricos inclusivos, com base na comparação de um conjunto de línguas sem concordância (tailandês, mandarim chinês, coreano, vietnamita e sinhala) com um conjunto de línguas com condordância (finlandês,

^a Newcastle University/University of Cambridge

^b Burapha University, Thailand

português brasileiro, hebraico, basco e tâmil). Para tal correlação, uma explicação relacionada à arquitetura de traços é proposta. A discussão sobre uma predição em favor de PRO genérico mostra que esta é inconclusiva ou falsa.

PALAVRAS-CHAVE: pronome genérico; pronome genérico inclusivo nulo; língua com concordância; línguas sem concordância

1. Introduction

The following sentences exemplify the so called inclusive generic pronoun, overt in (1), covert in (2) and (3).

(1) One shouldn't be afraid of making mistakes. [English]

(2) Tämän koneen voi hoitaa yhdellä kädellä. [Finnish]

this machine can.3SG operate with one hand

'One can operate this machine with one hand.'

(3) díawníi ñaan hãa yâak mâak thâa mây còb trii. [Thai]

nowadays job seek difficult very if NEG finish B.A

'To seek a job is difficult nowadays if one hasn't finished a B.A.'

It is called inclusive because the generic reference includes the speaker, the addressee, and other people. It is, thereby, the most general of pronouns, semantically. The question we will address is how this property is encoded in the feature make-up of the pronoun. There are basically two hypotheses. One is that it is the most richly specified pronoun, specified for first, second, and third person. The other is that it is the least specified one, therefore the least restricted one, allowing reference to the speaker, the hearer, and other people. We will explore a version of the latter hypothesis, following Phimsawat (2011). We refer to these pronouns as 'truly minimal pronouns'.¹

Anders Holmberg's research for this paper was funded by the European Research Council Advanced Grant No. 269752 'Rethinking Comparative Syntax' (ReCoS). Thanks to the organisers and the audience at Encontro Intermediário do GT-TG at Universidade Federal de Minas Gerais, 2015, and *Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017*.

The question is, what features does this minimally specified pronoun still have? A restriction that the inclusive generic pronoun has, at least in some languages, is that it can only include humans in its reference. We will show that this is true of some, but not all languages. Focusing on languages where the inclusive generic pronoun is a null category, we will demonstrate that there is a correlation between having subject agreement and having the reference of the inclusive generic subject pronoun restricted to humans. The task undertaken is to explain this correlation.

2. Inclusive, quasi-inclusive and exclusive

The inclusive generic pronoun can be contrasted with the quasi-inclusive generic pronoun ‘we’, as in (4), and the exclusive generic pronoun ‘they’ as in (5).

- (4) We like smoked fish in Finland.
- (5) They died young in the Middle Ages.

Generic ‘we’ is called quasi-inclusive because it includes the speaker but not necessarily the addressee. (4) would typically be uttered by a Finn to a foreigner. It can be paraphrased as ‘people in general in Finland, of which I am one’. Generic ‘they’ is exclusive in that it excludes the speaker and the hearer. The pronoun in (5) can be paraphrased as ‘people in general in the Middle Ages’. The quasi-inclusive and exclusive generic pronouns both typically require the specification of a domain, either geographical or temporal, where the temporal domain typically denotes a historical period (see Holmberg and Phimsawat 2015). In Thai, a radical pro-drop language, the quasi-inclusive pronoun has to be overt, in an out of the blue situation, as shown by (6).

especially to Fábio Bonfim Duarte for inviting Anders Holmberg. We wish to thank the following people for having contributed with data and discussion of the data: Seiki Ayano, Pauli Brattico, Sonia Cyrino, Maia Duguine, Ricardo Etxepare, Yujia Han, Saara Huhmarniemi, Rebeen Kareem, Shin-Sook Kim, Tawee Kueakoolkiat, Kadri Kuram, Marcello Modesto, Makiko Mukai, Trang Phan, Michelle Sheehan, Ur Shlonsky, Halldor Sigurðsson, Salinee Somtopcharoenkul, Harold Thampoe, Hofa Meng Jung Wu. Thanks also to the ReCoS team: Ian Roberts, Theresa Biberauer, Jenneke van der Wal, Sam Wolfe, Georg Höhn.

¹ This distinguishes them from Angelica Kratzer’s ‘minimal pronouns (e.g. as in Kratzer 2009), which are pronouns used as bound variables.

(6) **raw** kin cee nay dʉan tùlaakhòm.

we have vegetarian food in month October

‘We have vegetarian food in October.’

With a null subject (6) would either be interpreted as inclusive generic (‘One has vegetarian food...’) or as having a referential 1st person subject (‘I have vegetarian food...’). The quasi-inclusive pronoun can be null if it is bound or controlled by an overt one (see Holmberg and Phimsawat 2015).

(7) **raw** kin cee nay dʉan tùlaakhòm lǎŋ Ø thamboonsàjbàat.

we have veg. food in month October after offer food to monk

‘We have vegetarian food in October after offering food to monks.’

The exclusive pronoun can be overt or covert (see Holmberg and Phimsawat 2015 for more details).

(8) bon kò ní sùanyài (**khǎw**) plùuk chaa khǎay.

on island DEM mostly they grow tea sell

‘On this island they grow and sell tea.’

In this, the exclusive and quasi-inclusive pronouns contrast with the inclusive pronoun, in Thai, as the inclusive pronoun can be null in out of the blue sentences, in fact must be, as there is no overt counterpart.

The present paper will focus on the inclusive generic pronoun. The quasi-inclusive and exclusive pronouns are mentioned here to show that they can be clearly distinguished empirically from the inclusive one.

3. The inclusive generic pronoun in Thai has no phi-features

What features does an inclusive generic pronoun have? The meaning is ‘people in general, including me and you’. It has, thereby, the most general reference of all pronouns. There are two hypotheses how to encode this property as phi-features: One is that it is the most richly specified pronoun, specified for first, second, and third person, however this is formally expressed (see Hoekstra 2010). The other is that it is the least specified one, therefore allowing reference to the speaker, the hearer, and everyone else. A version of the latter hypothesis is proposed in Nevins (2007), where impersonal pronouns have an underspecified person feature (see Fenger 2016 for discussion). We will assume another version of the latter hypothesis, according to which the inclusive generic pronoun has no phi-features in some languages, namely language without agreement, including Thai, while it has minimal phi-features in languages with agreement. Phimsawat (2011) argues, for Thai, that personal pronouns have the featural make-up (9) while the inclusive generic pronoun has (10)² (see Déchaine and Wiltschko 2002, Holmberg 2005, 2010a,b).

(9) [uD, [ϕ [N]]]

(10) [uD, [[N]]]

uD (‘unvalued D’) is a referential feature, which is valued either by a referential index, which may be assigned freely or under anaphoric binding, or else by quantificational binding. In generic pronouns, and generic expressions more generally, the feature is bound by a generic operator, an adverbial operator GEN_x (= ‘It is generally true of x’) in the C-domain (following Moltmann 2006). The phi-features include person, number, and in some languages, gender or class. We will discuss the properties of the feature/head N below. We will take this theory as a starting point. As we shall see, it cannot be the case universally that the generic inclusive pronoun is phi-featureless, because in some languages it triggers agreement.

As argued by Phimsawat, the absence of phi-feature specification explains why the inclusive generic pronoun is obligatorily null, in Thai: Having no phi-features

² In Phimsawat’s (2011) notation the D-feature is R, for ‘referential’.

means that there are no features to spell out, on the assumption that the uD feature and the categorial N-feature are, or at least can be, not associated with any phonological features.

This analysis of the inclusive generic pronoun is part of a theory, articulated in Phimsawat (2011), according to which arguments in Thai can be null if and only if (a) they have an antecedent which is sufficiently local, from which they can inherit a referential index, or (b) they have no phi-features but are bound by a generic operator.

An observation which can be explained immediately within this theory is that the quasi-inclusive pronoun cannot be null in Thai, in an out of the blue context. This follows since (a) the pronoun has the phi-feature value 1PL (excluding the addressee), and (b) being generic, it has no antecedent (see Holmberg and Phimsawat 2015). Since the value [1PL] cannot be deleted without irretrievable loss of information, it must be spelled out.

4. Inclusive generic pronouns and reference to humans

We have said, and illustrated with examples, the claim that the inclusive generic pronoun includes the speaker, the addressee, and other people in its reference. What about inanimate things and non-human animals? Can they be included as well? Is it an integral property of the inclusive generic pronoun, or possibly generic pronouns more generally, that they only include humans in their reference, or is it just a consequence of the choice of predicates, so far? Predicates like ‘be afraid of making mistakes’, ‘operate with one hand’ and ‘seek a job’ select a human subject. It is clearly not the case that generic reference in general is restricted to humans: *Tigers are dangerous*, *Cars are expensive* are examples of non-human generic subjects.

If it turns out that inclusive generic pronouns are restricted to human reference, this should be encoded by some feature or features, following the logic of Phimsawat (2011). We could then not maintain the explanation that the inclusive generic pronoun is null because it has no restricting features.

We will start by considering what the inclusive generic pronouns look like in some other languages.

(11)

English:	<i>one, you</i>
Tamil:	<i>oruvan</i> [also ‘one (person)’], \emptyset (with 3SG agreement)
Sinhala	<i>kenek</i> [also ‘one (person)’], \emptyset
Swedish:	<i>man</i> [also ‘man’], <i>du</i> ‘you’
Turkish:	<i>insan</i> [also ‘human’], \emptyset (with 3SG agreement)
Japanese:	<i>hito</i> [also ‘human’], \emptyset
Italian	<i>si</i> , ‘REFL’, <i>tu</i> ‘you’
Finnish:	\emptyset (with 3SG agreement), <i>sä</i> ‘you’
Brazilian Portuguese:	\emptyset (with 3SG agreement), <i>você</i> ‘you’
Basque	\emptyset (with detransitivized verb)
Thai:	\emptyset
Chinese	\emptyset , <i>ren</i> [also ‘person’]
Central Kurdish:	<i>hamu kas</i> ‘any person’
Vietnamese	<i>chung ta</i> [‘you+me+others’], \emptyset

English is a representative of languages where the pronoun is a cognate of the numeral ‘one’. Other languages in this category include Tamil, where the commonest form of the overt generic inclusive pronoun is *oruvan*, which is the masculine form of the numeral ‘one’, which can also refer to women but not to non-persons. In Sinhala, too, the inclusive generic pronoun is *kenek* ‘one (person)’. Swedish, Turkish, and Japanese represent languages where the overt form of the inclusive generic pronoun is a cognate of the noun ‘human’ or, as in Swedish, ‘man’. Italian represents languages (including most Romance and Slavic languages) where a reflexive clitic *si* (or a cognate thereof) is used to express inclusive genericity.

- (12) a. Si lavora sempre troppo. [Italian]
 SI work.3SG always too,much
 ‘One always works too much.’
- b. W tym domu umiera się spokojnie. [Polish: Krzek 2012]
 in this house die.3SG SIEꞑ peacefully
 ‘In this house one dies peacefully.’

It is debatable whether the reflexive pronoun itself is the generic pronoun, or whether it is a voice-related, detransitivizing category which serves to license a null generic pronoun (see Cinque 1988, d’Alessandro 2008, Krzek 2012, 2013). There are also languages where the passive is systematically used to express inclusive generic meaning. An example is Standard Arabic (see Fassi Fehri 2009). Basque, which is included in (11), also represents languages where the generic reading is marked by a special, impersonal verb form.

Finnish, Brazilian Portuguese, Basque, Chinese, and Thai represent languages where the inclusive generic pronoun may be (and in some languages including at least Thai and Standard Finnish, must be) null. Central Kurdish represents languages where there is no designated inclusive generic pronoun, but where a quantificational expression meaning something like ‘everyone’, ‘anyone’, or ‘whoever’ is used. Vietnamese represents a possibly less common form of the inclusive pronoun. *Ta* means ‘you+me’ and *chung* is a pronominal associative plural marker. This is, thus, quite explicitly an inclusive pronoun.

Many languages, but not all, have the 2SG pronoun as an alternative inclusive generic form, overt or null with 2SG agreement. Interesting though it is, we will put aside the 2SG generic pronoun in this paper (see Gruber 2013).³

In some languages the generic pronoun can be overt or null. This is the case in Japanese, for example. This is not a matter of optionality: in some contexts the pronoun

³ In Brazilian Portuguese, especially in colloquial varieties, the 2nd person pronoun *você* is considerably more common as an inclusive generic pronoun than the null 3rd person pronoun. The latter is restricted to sentences with imperfective aspect, while *você* can be used with any aspect (thanks to Fábio Bonfim Duarte for the information).

must be pronounced, in other contexts it can be null, even when not bound by another generic pronoun (Seiko Ayano, p.c.). It is at present unclear what determines the distribution of overt and covert inclusive generic pronouns. We leave this issue for future research.

The list in (11) indicates that humanness is common as a feature of the inclusive generic noun/pronoun, as several of the pronouns are etymologically derived from a noun meaning ‘human’ or ‘man’. In Tamil, the masculine inflection restricts reference to humans (Tamil has ‘semantic gender marking’, where masculine and feminine can only refer to male and female persons, respectively). In Vietnamese, the associative plural of *ta* ‘you+me’ can only refer to persons. It is not necessarily the case that a generic pronoun which is derived from a noun meaning ‘human’ would be restricted to human reference, though, since it may have been grammaticalized as an even more generic pronoun, including also non-human referents. Whether this has actually happened is an empirical issue. We will return to it briefly in section 7.

To test whether the human restriction is endemic to inclusive genericity we need to employ a predicate which can be applied to a human as well as a non-human subject. Since the inclusive generic pronoun always includes the speaker and the addressee (or it would not be inclusive), the predicate must be compatible with human reference. But for the purposes of this test, it must also be compatible with non-human reference.⁴ One such predicate is ‘grow’. Humans can grow, but so can animals and plants. It is conceivable that the word for growth in humans and plants might not be the same in all languages. However, in the languages we have looked at so far, the same verb can be applied to all living beings. The test sentence we will use is a version of (13):

(13) One grows well, if one gets good care and a lot of nutrition.

⁴ Hualde and Ortiz de Urbina (2003: 587-588) test whether the Basque generic construction with an impersonal verb form must have a subject with human reference. However, in their two test sentences they employ a verb meaning ‘bloom’ and a verb meaning ‘bark’. The result is ungrammatical, from which they conclude that the construction must have a human subject. But the sentences could also be ungrammatical because (a) the construction cannot exclude reference to humans, to be inclusive generic, and (b) these predicates cannot be applied to a human subject. See the text for evidence that the impersonal verb form in Basque is restricted to human reference, though.

The context would be a person proudly showing his garden to a visitor, offering the sentence as an explanation why the garden is so lush. The sentence is meant to be a generalisation over humans, animals, and plants. In English, (13) cannot be used in this way: the generic pronoun *one* can only refer to humans (which shows, incidentally, that the etymological link to a noun meaning ‘human’ is not a crucial factor).

In this paper we will, however, only consider inclusive generic constructions with a null subject. This is to test Phimsawat’s (2011) hypothesis that inclusive generic pronouns are null because they have no phi-features (see Fenger 2012 for discussion of the features of overt generic pronouns).

Consider the following list of examples. The extension, humans only or humans and plants, is indicated. The sentences are meant to be uttered ‘out of the blue’, i.e. the subject should not be anaphoric.

- (14) *thâa dâayráb khwaamrák khwaam?awcaysày kôo cá too rew.* [Thai]
 if get love care then FUT grow fast
 ‘If one gets love and care, one will grow up faster.’ [humans and plants]
- (15) *rúguǒ néng huòdé gèng duō de yíngyǎng, nàme huì zhǎng de*
 if can get more much DE nutrition then be.likely grow DE
gèng kuài. [Mandarin Chinese]
 more fast
 ‘If one gets more nutrition, one will grow faster.’ [humans and plants]
- (16) *yeongyangpwun -ul seopchwiha-myeon, ppali calaᅇ-ta.* [Korean]
 nutrition -ACC take -if quickly grow.PRES DECL
 ‘If one gets more nutrition, one will grow faster.’ [humans and plants]
- (17) *vadi poshana labuvuth honthata hadai.* [Sinhala]
 more nutrition get-PTCP-CON well grow-PRS
 ‘If one gets more nutrition, one will grow faster.’ [humans and plants]

- (18) Nếu hấp-thụ được nhiều chất dinh - dưỡng, thì sẽ [Vietnamese]
 if receive obtain many CLF nutrition COND FUT
 phát-triển nhanh.
 grow fast
 ‘If one gets much nutrition, one will grow fast.’ [humans and plants]
- (19) Sitä kasvaa nopeammin jos saa paljon ravintoa. [Finnish]
 EXPL grow.3SG quicker if gets much nutrition
 ‘One grows quicker if one gets much nutrition.’ [humans only]
- (20) im meqablim harbe ahava ve maym az gdelim maher. [Hebrew]
 if receive.3PL much love and water then grow.3PL faster
 ‘If one gets much love and water, one will grow faster.’ [humans only]
- (21) Com boa alimentação cresce mais rápido. [Brazilian Portuguese]
 with good nutrition grow.3SG more quick
 ‘One grows faster with good nutrition.’ [humans only]
- (22) Behar bezala zainduz gero, hemen ongi hazitzen da. [Basque]
 appropriately take.care.IMP after here well grow.HAB is
 ‘If one is treated appropriately, one grows well here.’ [humans only]

According to our informants, the Thai, Mandarin, Korean, Sinhala, and Vietnamese examples may well be said about plants as well as animals and (necessarily) humans. The Finnish and the Hebrew examples cannot include plants. The Brazilian Portuguese example is not acceptable for all speakers (some speakers want an overt pronoun here, which would be *você* ‘you’ to convey the inclusive reading), but for those

who accept it, it can only refer to humans.⁵ The Basque example also cannot include plants.

One salient property that distinguishes Mandarin, Korean, Sinhala, Vietnamese, and Thai from Finnish, Hebrew, Brazilian Portuguese and Basque is that the former set lacks subject-verb agreement.⁶

Tamil provides some interesting evidence that agreement is, or at least can be, crucial.

- (23) a. kooda satthu kidaithaal, nalla valarum.
 more nutrition get.PRTC.CON well grow.FUT.3N
 ‘If they get more nutrition they will grow well.’ [plants, not humans]
- b. kooda satthu kidaithaal, nalla valaruvan.
 more nutrition get-PTCP-CON well grow.FUT.3SG.M
 ‘If one gets more nutrition, one will grow well.’ [humans only]
- c. kooda satthu kidaithaal, nalla valaramudium.
 more nutrition get.PTCP.CON well grow.INF.can
 ‘If one gets more nutrition, one will grow well.’ [humans and plants]

The null subject in (23a) can only refer to plants and animals because the gender agreement on the verb is incompatible with human reference. The null subject in (23b) can only refer to humans, because the gender agreement on the verb is incompatible

⁵ Marcello Modesto (p.c.) has provided the following example from Brazilian Portuguese as a case where a null generic pronoun can refer to plants and animals as well as humans.

(i) Se está vivo, um dia morre.
 if is alive one day dies
 ‘Whoever/whatever is alive, will die one day.’

This means that Brazilian Portuguese and Finnish are not exactly alike in relevant respects, and suggests that the correlation between agreement and human reference is not universal. We will return to this case in section 7.

⁶ Three other languages which have a null inclusive generic pronoun and agreement, and are reported to allow reference to humans only are Bengali (Wim van der Wurff, p.c.), Assamese (Hemanga Dutta, p.c.), and Icelandic (Halldór Sigurðsson, p.c.). For various reasons we don’t have examples from these languages directly comparable with the nutrition examples in (15)-(20).

with non-human reference. In (23c), the head of the predicate is a modal auxiliary which does not show agreement. Now the null generic subject can refer to humans as well as animals and plants.

Why would agreement make a difference to generic reference in languages which do not show the kind of gender agreement on T that Tamil does, though?

The following is a possible hypothesis, which can, however, be rejected: In the languages without agreement the null subject in (14)-(18) is ambiguous between an inclusive generic pronoun referring to humans in general and an exclusive generic pronoun referring to plants (or non-humans) in general. This hypothesis can be rejected, at least in the case of Thai, on the grounds that there is no exclusive generic pronoun, null or overt, which would refer to plants/non-humans.

(24) thîi kò nîi yùudiikindii.

at island this live well

‘They live well on this island.’

This sentence cannot be taken to be an exclusive generic statement about plants or animals, only about people (see Holmberg and Phimsawat 2015). To refer to plants and/or animals, the subject would have to be overt.

The following is another possible hypothesis, which can also be rejected. The subject in (14)-(18) is not a generic pronoun at all, but a multiply ambiguous referential pronoun: ‘I’, ‘you’, ‘he’, ‘it’, ‘they’, etc., covering all people, animals, and plants. This can be rejected because referential pronouns other than first person and in some circumstances second person cannot be null in out of the blue sentences; they need a topic antecedent in the immediate discourse context (Phimsawat 2011, Holmberg and Phimsawat 2015). A first person, and in certain cases, a second person subject, can be null in out of the blue sentences because, in informal terms, the speaker and the addressee provide contextual antecedents for the null subjects. In more formal terms, the null subject can be bound by a ‘speaker feature’ or ‘addressee feature’, a syntactic

representation of the speaker and the hearer in the C-domain (Sigurðsson 2004, 2007; Holmberg and Phimsawat 2015).

5. Inclusive reference in languages with agreement

We assume a Chomskyan theory of agreement (Chomsky 2001). Subject-verb agreement is formally a set of unvalued phi-features of T, person, number, and in Hebrew also gender. These features need to be assigned a value in the course of the syntactic derivation. They are assigned a value by the subject DP, being the closest DP which is ‘active’, not having been assigned a Case by some independent means. The valued phi-features of T are spelled out as an inflection on the finite verb or auxiliary, in the languages under discussion here. If the unvalued phi-features are not assigned a value, the derivation will crash at PF, as they, and thereby the finite verb, cannot be spelled out.

This means that there must be a null generic subject in the structure, which has inherently valued phi-features. The agreement in the Hebrew example shows that it has 3PL.M. In Finnish and Brazilian Portuguese it has 3SG.

Could the 3SG in Finnish and/or Brazilian Portuguese generic sentences be default agreement, though? Default agreement is well known from many languages, employed when, for some reason, the phi-features of T (in the case of subject agreement) cannot be valued by the subject DP. This could be because the subject DP is assigned Case independently, and is thereby deactivated, or because there is no subject DP. Default agreement is typically 3SG. This can be seen in the Finnish sentence (25):

(25) Minun pitää ostaa uusi auto.

I.GEN should.3SG buy new.NOM car.NOM

‘I should buy a new car.’

Some predicates assign genitive case to the subject, in which case it cannot assign phi-feature values to T. In that case, the phi-features of T get the default value

3SG (Laitinen and Vilkuna 1993). This suggests that the 3SG agreement in construction with the inclusive null generic subject could be default agreement. The same could then be true of Brazilian Portuguese. However, as demonstrated in Holmberg (2010b), the default agreement analysis is not right for Finnish. The argument is based on the fact that default agreement and ‘true’ agreement, including 3SG agreement, have clearly different effects elsewhere in the clause: If the subject of a transitive verb does not trigger agreement the object will get nominative case, as in (22). If the subject does trigger agreement, which entails that the subject gets nominative case, the object will get accusative case, as in (26).

- (26) Minä voin ostaa uuden auton.
 I.NOM can.1SG buy new.ACC car.ACC
 ‘I can buy a new car.’

As shown in (27), sentences with a null inclusive generic subject show the same variation as sentences with an overt subject, which is to say that the null subject triggers agreement just like an overt subject. In (27a) the predicate assigns genitive case to the (null) subject, hence it does not trigger agreement, and the object has nominative case. The verb has the default 3SG form.

- (27) a. Nyt pitää ostaa uusi auto.
 now should.3SG buy new.NOM car.NOM
 ‘One should buy a new car now.’
- b. Nyt voi ostaa uuden auton.
 now can.3SG buy new.ACC car.ACC
 ‘One can buy a new car now.’

In (24b) the subject triggers agreement, which is 3SG because the generic subject is 3SG. In return, the subject gets nominative, and the object consequently gets accusative.

Under the present theory of agreement, the existence of subject agreement marking on the verb which can be shown not to be default agreement, is evidence that there is a subject, even though nothing is spelled out (in the case of Finnish there is no overt form of a 3SG inclusive generic subject), and shows what phi-features it has, while tests such as the nutrition sentence test, can be used to show what other restricting features it has. We take it that we have established that it has the feature [+Hum] (we will later provide a reason for taking it to be the value of a binary feature rather than a privative feature). There are other tests which can be employed to establish whether an understood, but covert subject is actually syntactically represented. Such tests have been applied to the Finnish inclusive generic pronoun, and have showed consistently that there is a syntactically represented subject (Hakulinen and Karttunen 1973, Vainikka 1989, Vainikka and Levy 1999, Laitinen 1995, 2006, Holmberg 2010b). This covert subject can bind anaphora, control a PRO subject in a purpose clause, and license agentive adverbials (see Holmberg 2010b for examples, with details). There is consensus among the linguists who have worked on the inclusive generic construction in Finnish that it has a syntactically represented subject.

We can explain why there has to be a subject with phi-features in the languages with subject agreement. We have not explained why that subject must be restricted to human reference.

6. Explaining the relation between inclusive reference, phi-features and humanness

First, what we call Human in grammar would be more appropriately termed something like Conscious Being, to also include talking animals and extraterrestrials and other such imaginary entities which have crucial human properties. With this proviso, we will continue to use the label Human or [\pm Hum].

There are various ways to integrate the feature Human in the structure of pronouns. One is that this feature is a component of N, the nominal ‘base’ of nominal

expressions, perhaps appropriately seen as the root of a pronoun, a minimal root. *He* and *she* would have the root feature Human, or [+Hum], non-human-referring pronouns like English *it* would have a [-Hum] root. We may want to make a distinction between pronouns that get their interpretation from an antecedent and pronouns that do not. In the former case the component N, the root component of the pronoun, may be taken to be a copy of the NP of the antecedent, deleted under identity with this antecedent (see Panagiotidis 2002, Elbourne 2008 for different versions of this idea). In the case of the generic pronoun, there is no antecedent.) Therefore it needs a root of its own. The [+Hum] feature would provide this. The fact that the inclusive generic reading includes, by definition, the speaker and the addressee in the extension of the pronoun means that in the case of this pronoun, the feature [-Hum] is not an option.

But what is the connection with agreement? What about all the languages where the generic pronoun is so inclusive that it can include plants along with humans and animals? In this case the pronominal root would seem to be unspecified for humanness, [\pm Hum], allowing reference to entities of any kind. The generalisation that we want to express, though, suggested by our data, is that a pronoun cannot have phi-features without specification of the feature [\pm Hum].

The following is an alternative. First, the minimal root of a pronoun is, universally, [ENTITY]. Second, there are two ways that a pronoun can refer to everything and/or everybody: one is not to have any phi-features, hence no restriction. The other is to have minimal phi-features, just enough to satisfy the requirements of agreement, yet allowing reference to the speaker, the addressee, and a maximally general set of ‘non-participants’. The feature [participant], widely assumed as part of pronominal systems, following Harley and Ritter (2002), distinguishes between speaker and addressee on the one hand, and everyone/everything else on the other hand. In Harley and Ritter (2002) all the features are privative. Third person is when the feature [participant] is absent, i.e. ‘third person is no person (see Nevins 2008 for discussion). Such a system does not allow for a pronoun with phi-features which allows reference to the speaker, the addressee, and everyone/everything else. The system must include a feature which can be underspecified for person: [\pm participant] (see Nevins 2008 for other arguments that this device is needed). On its own, this feature will not exclude reference to non-human entities, and therefore must be supplemented by at least one more feature.

Assume that the phi-feature set of a pronoun has to include at least one specified feature. The pronominal phi-features are person, number, and class (Harley and Ritter 2002). The inclusive generic pronoun, although formally singular is not semantically singular. Arguably this rules out the use of a pronoun specified for singular number as an inclusive generic pronoun. Assume that the first division among the class features is between human and non-human, as seen in the many pronominal systems which make a distinction between human and nonhuman third person pronouns. The inclusive generic pronoun cannot be specified [-Hum], as it must allow inclusion of the speaker and addressee. But it can be specified as [+Hum]. The minimal feature make-up of a pronoun with phi-features which will allow inclusive, generic reference will therefore be [\pm Participant, +Human].⁷

This presupposes that the unvalued phi-features of T are, or at least can be, formally valued by this minimal phi-feature set, where the spell-out of the so valued T is the third person singular suffix on the finite verb (in most but not all of the relevant languages; in Hebrew it is plural). That is to say, the third person singular form that the finite verb has in Finnish, discussed in section 5, would be a form of default agreement after all, in that the subject valuing the features of T would not be specified for person or number, but only for class (the [+Hum] value), which, however, has no morphological effect in Finnish.⁸

In languages without agreement, there is no reason why a generic pronoun would have to have any phi-features. All it needs is the root feature [ENTITY] and merged with it, the [uD]-feature. When the D-feature is bound by the generic operator this results in a reading which can be rendered as ‘entities in general including the speaker and the addressee’, the minimally specified DP giving the maximally inclusive reading.

⁷ Hebrew is a language with a null inclusive pronoun which triggers plural agreement, an option which would appear to be consistent with the semantics of inclusivity. The idea that one specified feature is enough would then seem to predict that the inclusive pronoun in Hebrew could remain unspecified for [Hum]. The data we have indicates that this is a false prediction.

⁸ According to the theory of null subjects in Holmberg (2010a,b), Roberts (2010b), based on the theory in Roberts (2010a), null subjects in languages with agreement are derived by copy deletion. The valued phi-features of T and the subject pronoun form a chain of two copies, where one, the subject, is deleted, provided its features are a subset of the phi-features of T. Since the subject, if it is third person, is valued for gender (i.e. class) in many languages, T must be valued for gender as well, for the subject to be deletable, even when this is not morphologically realised, as is the case in many languages. The notion that T has, or may have, an invisible class feature in languages with phi-features in T thus has independent motivation.

7. A prediction for generic PRO

The theory predicts that arbitrary/generic PRO as found in the subject position of non-finite clauses should not have its reference restricted to humans, in languages or constructions where non-finite clauses do not exhibit agreement, as is the case in English, for example. In the absence of agreement there is no compulsion for the subject to have any phi-features, and therefore no compulsion to be restricted at all, beyond ENTITY. If the non-finite clauses exhibit agreement, as they can do in some languages, we expect inclusive generic PRO in those clauses to be restricted to humans.

In order to test the prediction we need a predicate which can apply to humans and non-humans, for instance plants.

(28) On a day like this it's important [PRO to get enough water].

Can this be said as an explanation for watering the houseplants repeatedly, or for starting up the expensive sprinkler system on a wheat field? Native English-speaking informants that we have consulted agree that it can be, although some report a moment of hesitation before the judgment. All agree that (28) contrasts clearly with the overt generic construction in (29), employing *one*, which can only have human reference.

(29) On a day like this it's important [that one gets enough water]. (humans only)

(28) also contrasts with (30), although less clearly than in the case of (31).

(30) On a day like this it's important [that you get enough water]. (humans only?)

Some informants report a difference between (28) when PRO refers to houseplants (marginally OK) and when it refers to a wheat field (not OK). We return to this point below.

However, predicates that take complements with generic PRO typically have an implicit, if not explicit, experiencer argument controlling PRO: 'It is important/good/necessary/etc. for X_i [PRO_i to ...]'. The issue whether PRO is restricted

to human reference or not may then more accurately be the issue whether the implicit/null generic experiencer is restricted to human reference or not. Whether it does may be an interesting question but it has no immediately obvious consequences for the issue at hand, which is the relation between agreement, phi-features, and reference to humans.

A predicate taking infinitival complements in English which does not have an experiencer argument is *common*.

(31) It's common [PRO to deteriorate with age].

The predicate of the embedded clause is selected to allow a human or non-human subject, which can even be inanimate. The question now is, can (31) be said, for example, by someone inspecting a leaking roof, as a generic statement including roof tiles along with people and any other entities subject to aging? The prediction made by the theory sketched in section 4 is that it can be. Informants consulted (a limited number) agree that it can be, although sometimes after a moment's hesitation. All informants agree that there is a contrast between (31) and (32a,b), with an overt generic subject, such that the latter would be restricted to human subjects, hence not felicitous in the leaking roof context.

- (32) a. It's common that one deteriorates with age.
 b. It's common that you deteriorate with age.

Some informants also, again, report a difference between (32a and b), such that (32a) categorically excludes anything but human reference, while (32b) can be used, perhaps with an element of jocularly, for at least some non-humans, such as house plants.

For the class of languages which have subject agreement and a null inclusive generic subject pronoun in finite clauses, we predict a difference between inclusive

generic pro and PRO: The former should admit only human reference, the latter should be more permissive. We have tested this with Finnish. The context is a farmer saying either (33a) or (33b), as an explanation why he is starting up his expensive sprinkler system to water his wheat fields.

- (33) a. Näin kuumassa ilmassa on tärkeää saada tarpeeksi vettä. [Finnish]
 so hot.INE weather.INE is important get.INF enough water
 ‘In hot weather like this it’s important to get enough water.’
- b. Näin kuumassa ilmassa on tärkeää että saa tarpeeksi vettä.
 so hot.INE weather.INE is important that get.PRS.3SG enough water
 ‘In hot weather like this it’s important to get enough water.’

The prediction is that there would be a difference between (33a), with the infinitival complement, and (33b), with the finite complement, such that (33a) would be appropriate in this context but (33b) would not. As it happens there was disagreement among the informants whether there was a clear difference between (33a,b). Some informants confirmed the prediction, but other informants accepted them both.⁹

However, as in the case of the English examples, we cannot tell whether this is a matter of features of the implicit experiencer argument or the null subject. The following is a test using the predicate *on yleistä* ‘(it) is common’, which can be read without any implicit argument.

- (34) a. On yleistä rapistua vuosien myötä.
 is common deteriorate.INF years by
 ‘It’s common to deteriorate with age.’

⁹ Thanks to Saara Huhmarniemi for her help with these tests.

b. On yleistä että vuosien myötä rapistuu.

is common that years by deteriorate.3SG

‘It’s common that one deteriorates with age.’

As above, the question is whether (34a,b) can include roofs as well as people in the generic reference. Among the limited number of informants consulted there was variation. Some did not accept either of them as generic statements including roofs, while some accepted (34b) with that interpretation. This is not predicted by the theory articulated above. A more careful investigation will have to be left for future research.

In section 4, note 4, we mentioned a counterexample, provided by Marcello Modesto (p.c.) to the generalisation that a null inclusive 3SG generic pronoun in a language with agreement can only have human reference.

(35) Se está vivo, um dia morre. [Brazilian Portuguese]

if is alive one day dies

‘Whoever/whatever is alive, will die one day.’

This sentence can be said when talking about plants, animals, and or humans. The following is a similar Finnish example, provided by Pauli Brattico (p.c.):

(36) Sitä syntyy ja kuolee aikanaan. Mikään/kukaan. [Finnish]

EXPL is.born and dies some.day nothing/no-one

ei elä loputtomasti

NEG lives for.ever

‘One day you are born, one day you die. Nothing/no-one lives for ever.’

Brattico reports that he could use this sentence, for example, in conversation with a child when the family's cat or some other important living entity, even a house plant, is dying. The extension of the relevant feature of the generic pronoun is thus not humans but something like 'humans and our associates'. It is perhaps significant that the predicates in (35) and (36) are 'be born', 'be alive' and 'die', predicates denoting defining properties of animate beings, including plants just as much as humans.

As mentioned, we will leave a more detailed investigation of generic pronouns and the human feature in non-finite constructions for future research.

8. Conclusions

The starting point is the hypothesis, articulated in Phimsawat (2011), that the inclusive generic pronoun is the least specified nominal category, which therefore has the most general reference, including the speaker, the hearer, and everyone else. The observation is that there is cross-linguistic variation as to whether the pronoun is or is not restricted to humans. Focusing on languages which have a null inclusive generic pronoun in finite clauses, we have found that the null inclusive generic pronoun is restricted to human reference in some of them, but not all. The generalisation, based on data from primarily ten languages, five without agreement, four with subject-verb agreement, and one (Tamil) with or without agreement) is that the pronoun is restricted to human reference in the languages that have subject-verb agreement in finite clauses. The explanation proposed is (a) in languages with subject agreement, i.e. unvalued phi-features in T, the inclusive generic pronoun has to have at least one specified phi-feature, to value the phi-features of T; (b) if the pronoun is to be inclusive, it cannot be specified for number, which entails that it must be specified for class; (c) if the pronoun is to be inclusive, i.e. include the speaker and the addressee, it must be specified [+Hum].

References

- ACKEMA, P. and A. NEELEMAN. 2013. Person features and syncretism. *Natural Language and Linguistic Theory* 31, 901-950.

- BIBERAUER, T., HOLMBERG, A., ROBERTS, I., & SHEEHAN, M. 2010. *Parametric syntax: Null subjects in minimalist theory*. Cambridge: Cambridge University Press.
- BRODY, M. 2013. Silent People: The pseudo-impersonal. In J. BRANDTLER, V. MOLNEÁR, C. PLATZACK (eds.), *Approaches to Hungarian: Volume 13: Papers from the 2011 Lund conference*. John Benjamins Publishing Company.
- CINQUE, G. 1988. On *si* constructions and the theory of *ARB*. *Linguistic Inquiry* 19: 521-81.
- D’ALESSANDRO, R. 2007. *Impersonal si constructions*. Berlin/New York: Mouton de Gruyter.
- ELBOURNE, P. 2008. The interpretation of pronouns. *Language and Linguistics Compass*: 119–150.
- DÉCHAINÉ, R. M., and M. WILTSCHKO. 2002. Decomposing pronouns. *Linguistic Inquiry* 33 (3): 409–442.
- FASSI FEHRI, A. 2009. Arabic silent pronouns, person and voice. *Brill’s Journal of Afroasiatic Languages and Linguistics* 1: 1-38. Also in A. FASSI FEHRI. 2012. *Key features and parameters*. Amsterdam: John Benjamins, 257-290.
- FENGER, P. 2015. How impersonal does one get? A study of man-pronouns in Germanic. <http://ling.auf.net/lingbuzz/002802>
- GRUBER, B. 2013. *The spatiotemporal dimensions of person. A morphosyntactic account of indexical pronouns*. PhD dissertation. LOT, Utrecht University.
- HAKULINEN, A. and L. KARTTUNEN. 1973. Missing persons: on generic sentences in Finnish. *Papers from the ninth regional meeting of the Chicago Linguistic Society*: 157-71.
- HARLEY, H. & RITTER, E. 2002. Person and number in pronouns: a feature-geometric analysis. *Language* 78:482-526.
- HOEKSTRA, J. 2010. On the impersonal pronoun *men* in modern West Frisian. *Journal of Comparative Germanic Linguistics* 13: 31–59.

- HOLMBERG, A. 2010a . ‘Null subject parameter’. In T. Biberauer et al. *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge, UK: Cambridge University Press. 88–124.
- HOLMBERG, A. 2010b. ‘The null generic subject pronoun in Finnish: a case of incorporation in T’. In T. BIBERAUER et al. *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge, UK: Cambridge University Press. 200–230.
- HOLMBERG, A, and O. PHIMSAWAT. 2015. Generic pronouns and phi-features: Evidence from Thai. *Newcastle and Northumbria Working Papers in Linguistics*, vol.21. <http://www.ncl.ac.uk/linguistics/research/workingpapers/Volume21.1.htm>
- HUALDE, J. I. and J. ORTIZ DE URBINA. A grammar of *Basque*. Berlin: Mouton de Gruyter.
- IWASAKI, S. & INGKAPHIROM, P. 2009. *A Reference Grammar of Thai*. Cambridge: Cambridge University Press.
- KRATZER, A. 2009. Making a pronoun: fake indexicals as windows into the properties of pronouns". *Linguistic Inquiry* 40 (2): 187–237.
- KRZET, M. 2013a. Interpretation and voice in Polish SIEĆ and –NO/–TO constructions. In I. K. CHAHINE (ed.), *Contemporary Studies in Slavic Linguistics*. Amsterdam: John Benjamins 185-198.
- KRZEK, M. 2013b. Generic subjects and Voice in Polish impersonal constructions. In C. S. RHYS, P. IOSAD, and A. HENRY (eds.), *Microvariation, Minority Languages, Minimalism and Meaning: Proceedings of the Iris Network in Formal Linguistics*. Cambridge: Cambridge Scholars Publishing. 186-206.
- LAITINEN, L. 2006. Zero person in Finnish. *Grammar from the human perspective*. In M.-L. HELASVNUO and L. CAMPBELL (eds.). Amsterdam: Benjamins, 210-31.
- MOLTMANN, F. 2006. Generic one, arbitrary PRO, and the first person. *Natural Language Semantics* 14: 257–281.
- NEVINS, A. 2007. The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory* 25 (2), 273-313.

- PANAGIOTIDIS, E. P. 2002. *Pronouns, clitics and empty nouns*. Amsterdam: John Benjamins.
- PHIMSAWAT, O. 2011. *The syntax of pro-drop in Thai*. Ph.D. dissertation, Newcastle University.
- ROBERTS, I. 2010a. *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge, MA: MIT Press.
- ROBERTS, I. 2010b. A deletion analysis of null subjects. In T. BIBERAUER et al. *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge, UK: Cambridge University Press. 58–87.
- SIGURÐSSON, H. A. 2004. The syntax of Person, Tense, and speech features. *Italian Journal of Linguistics* 16:219–251.
- SIGURÐSSON, H. A. 2015. About pronouns. <http://ling.auf.net/lingbuzz/001593>.
- VAINIKKA, A. 1989. *Deriving syntactic representations in Finnish*. Ph.D. dissertation: University of Massachusetts, Amherst.
- VAINIKKA, A. and Y. LEVY. 1999. Empty subjects in Finnish and Hebrew. *Natural Language and Linguistic Theory* 17: 613-71.

PassiveP and the Distinction between Eventive, Resultative and Stative Passives

João Claudio de Lima Júnior^a
Marina Rosa Ana Augusto^b

ABSTRACT

This paper focuses the distinction between verbal and adjectival passive sentences. For that differentiation to be accounted for, a previous proposal considering a specific functional node for eventive passives - passiveP (LIMA JÚNIOR; AUGUSTO, 2015) is retrieved. It is arguably assumed that the approach based on passiveP, besides maintaining a uniform analysis to active and passive sentences (as VoiceP in Collins (2005)), deals well with intervention issues, and is prosperous in addressing parametric variation and language acquisition facts. As far as the tripartite distinction among the different types of passives is concerned (EMBICK, 2004; DUARTE; OLIVEIRA, 2010), it is proposed that a main bipartite distinction between eventive and adjectival passives may be retained, which is here attributed to the presence of passiveP. Concerning stative and resultative passives, an agreement operation between the auxiliary verbs and the participle (LUNGUINHO, 2011) is assumed to allow for different readings to be obtained. A fourth group of passive-like sentences, involving participles, which have lost their connection with their original verbs, is also syntactically distinguished and treated as actual copular constructions.

KEYWORDS: participles; stative passives; resultative passives; eventive passives

RESUMO

Este artigo enfoca a distinção entre passivas verbais e passivas adjetivais. Para tornar essa diferenciação possível, recupera-se uma proposta prévia que assume um nó funcional específico para passivas eventivas - passiveP (LIMA JÚNIOR; AUGUSTO, 2015). Argumenta-se que a abordagem baseada em passiveP, além de manter uma análise uniforme entre ativas e passivas (como VoiceP em Collins (2005)), lida satisfatoriamente com questões de intervenção, além de responder bem a fatos relativos à variação paramétrica e à aquisição de linguagem. Em relação à divisão ternária proposta para os diferentes tipos de passiva (EMBICK, 2004; DUARTE; OLIVEIRA, 2010), propõe-se que a divisão binária entre passivas verbais/eventivas e passivas adjetivais pode ser mantida, diferença essa que é atribuída aqui à presença de passiveP. No concernente à subdivisão entre passivas estativas e resultativas, uma operação de concordância (*agree*) entre verbos auxiliares e participípio (LUNGUINHO, 2011) permite que cada leitura seja adequadamente obtida. Um quarto grupo de construções aparentemente passivas, envolvendo

^a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)/ Laboratory of Psycholinguistics and Language Acquisition (LAPAL), limajr.lapal@gmail.com.

^b Universidade Estadual do Rio de Janeiro (UERJ)/ Laboratory of Psycholinguistics and Language Acquisition (LAPAL), marinaaug@uerj.br

The first author thanks CNPq (process number #152661/2016-6) for the financial support.

particípios que tenham perdido a conexão com seus verbos originais, é também sintaticamente distinguido e tratado como envolvendo construções copulares.

PALAVRAS-CHAVE: particípios; passivas estativas; passivas resultativas; passivas eventivas

Introduction

This paper focuses on the syntactic structure of verbal (1) and adjectival (2) passive sentences. Couched in the assumptions of the Minimalist Program (CHOMSKY, 1995; subsequent work), the present paper elaborates on a previous proposal on the topic (see LIMA JÚNIOR; AUGUSTO, 2015)¹⁰, in which the incorporation of recent formal mechanisms to the analyses of verbal passive sentences has been scrutinized (see BOECKX, 1998; COLLINS, 2005; GEHRKE; GRILLO, 2009). An alternative formal solution, which adopts a specific functional category – PassiveP – for verbal passives, is assumed. This possibility is taken into account for the extension of the analysis for adjectival passives, which is carried out in this paper.

(1) *Verbal Passive*

O gato foi machucado pelo cachorro.

The cat_Masc. BE_Past_Eventive hurt_Masc. by the dog

(2) *Adjectival Passive 1: Resultative*

O gato ficou machucado.

The cat_Masc. BECOME_Past hurt_Masc..

(3) *Adjectival Passive 2: Stative*

O gato está machucado.

The cat_Masc. BE_Present_Stative hurt_Masc..

PassiveP functions as an alternative to the vP category of the transitive actives. As such, it constitutes a syntactic phase (see LEGATE, 2003). Although it is not capable of assigning case (its distinctive feature in relation to vP), passiveP is endowed with an

¹⁰ It is important to anticipate that much of the conclusions reached here benefit, to a great extent, from the past discussions conducted in Lima Júnior; Augusto (2015). Nevertheless, this work has been organized in a way to keep its independence in relation to the past work, retrieving the most important discussions whenever it is necessary although some relevant details are omitted.

edge feature, projecting an extra specifier. Such an extra spec allows the internal argument (logical object) to be moved to the subject position cyclically. It is hypothesized here that not only does passiveP play a central role in the derivation of verbal passives, but also in the distinction between eventive passives and passives with an adjectival reading.

The main purpose of this paper is then to verify to which extent the proposed node PassiveP is relevant for an analysis of passives which distinguishes eventive (1), resultative (2) and stative (3) passives (EMBICK, 2004; see also DUARTE; OLIVEIRA, 2010 - for Portuguese). In order to do so, we firstly review the main advantages for assuming the passiveP-approach: the possibility (a) of giving a uniform analysis for passive and active sentences; (b) of dealing well with intervention issues; and (c) of adequately addressing facts concerning parametric variation and language acquisition. As far as adjectival passives are concerned, the analysis presented departs from Embick (2004) and Duarte; Oliveira (2010), and resorts to Lunguinho's analysis to distinguish stative and resultative passives by means of an agreement operation between the participial node and the auxiliary verb.

The paper is organized as follows: In the next section, the analysis of verbal passives is recaptured in the light of the passiveP-approach; the reasons for dispensing with the smuggling approach (COLLINS, 2005) are also briefly reviewed. In section 3, the main differences between eventive, resultative, and stative passives are discussed, taking into consideration the analyses proposed by Embick (2004) for English, and by Duarte; Oliveira (2010) for Portuguese. In section 4, a reformulated and uniform analysis of eventive, resultative, and stative passives, assuming PassiveP, is presented. Section 5 is dedicated to final remarks.

1. Verbal passives

Recent analyses for Portuguese passive structures (LUNGUINHO, 2011; DIAS; NAVES, 2014) have adopted the smuggling approach proposed by Collins (2005) for English. Basically, the smuggling movement, which is at the center of the smuggling approach, functions as a last resort operation for reasons of Case-assignment of the internal argument in verbal passives. A number of criticisms has been, nevertheless, evoked against this approach (GEHRKE; GRILLO, 2009; LIMA JÚNIOR; AUGUSTO,

2015). One major argument contrary to the smuggling movement is its lack of a clear teleological motivation.

Roughly, the smuggling approach appears as a solution to a 'problem' the active-like derivation of verbal passives (see BOECKX, 1998; COLLINS, 2005) has: the intervention caused by the external argument¹¹. In other words, the external argument intervenes between the probe (T) and the goal DP (internal argument). In this sense, the internal argument DP cannot move past the external argument because, concerning syntactic distance, the latter is closer to the probe than the former. As an active-like derivation of passive sentences seems to be desirable (see BOECKX, 1998; COLLINS, 2005; LIMA JÚNIOR; AUGUSTO, 2012; 2014; 2015), the movement of the internal argument DP to [spec, TP] becomes a formal complication, inasmuch as it should be blocked by the external argument DP, contrary to facts.

In order to handle 'the intervention problem' in verbal passives, Collins (2005) argues for the movement of a maximal projection XP containing the internal argument DP instead of the movement of the internal argument DP itself. As the whole 'XP containing the internal argument DP' and 'the external argument DP' have different bundles of features, the movement of the XP cannot be blocked by the external argument DP. It is not known for sure, though, what triggers such a movement and what would prevent this movement from being driven under different circumstances.

The XP-movement in Collins's analysis, furthermore, presupposes the existence of a specific functional projection, called VoiceP. VoiceP is a functional category composed by uninterpretable features only, headed by the preposition *by*. In simple words, the presence of VoiceP in the numeration licenses partP forcing its movement, as a maximal projection, to [spec, VoiceP]. It is what constitutes the basic notion behind the smuggling movement: the internal argument DP is smuggled during the movement of partP; that is, the internal argument DP is supposed to be taken along with partP to [spec, VoiceP]. At this position ([Spec, VoiceP]), no intervention is expected anymore. Thus, the internal argument DP is free to move to [spec, TP].

Although Collins' analysis is doubtless insightful to handle the intervention problem, it causes additional difficulties (GEHRKE; GRILLO, 2009; LIMA JÚNIOR; AUGUSTO, 2015): the need for look-ahead, the problematic constituency of the by-

¹¹ The problem of intervention stems from the Relativized Minimality principle (see RIZZI, 1990; 2004), which states that a phrase XP endowed with a feature [+F] cannot move past another phrase YP with the same feature [+F].

phrase; the fact that by-phrases are not exclusive of verbal passives; the fuzzy nature of the head of VoiceP in languages that admit both a passive morpheme and the by-phrase, as Kiswahili and Japanese¹².

VoiceP seems, then, to bring about too many problems to be taken as a desirable solution in order to maintain an active-like derivation for passives while circumventing the intervention issues.

Nevertheless, it has been hypothesized that the postulation of a specific functional category for verbal passives is at the locus of the solution to the problems the smuggling analysis failed to solve. This functional projection is passiveP, an alternative node to vP, which will be presented in details in the next section.

2. PassiveP: functional node for verbal passives

First of all, it is important to state that there seems to be a consensus that any alternative account for verbal passives must assume an active-like derivation (see BOECKX, 1998; COLLINS, 2005). This also implies that a solution to evade the intervention problem must be offered. Additionally, one should ask whether it is reasonable to propose a specific functional category for passives, be it VoiceP or, as proposed here, passiveP.

An advantage of having a specific node for passives is to account for what is observed in typological studies (SIEWIERSKA, 1984; *apud* JAEGGLI, 1986): many languages in the world do not exhibit passive sentences (see KEENAN; DRYER, 2007). Collins argues that VoiceP is the expression of a parameter 'voice'. Thus, in languages that do exhibit passives, a positive value for the parameter 'voice' must be set. In this respect, Collins' (2005) work seems to be in line with what is observed in terms of parametric variation (p. 117), and paves the way for relevant discussions concerning the acquisition of passives (see HYAMS; SNYDER, 2006; SNYDER; HYAMS, 2008; 2015; GEHRKE; GRILLO, 2009; LIMA JÚNIOR; AUGUSTO, 2012; 2014; LIMA JÚNIOR, 2016; CORRÊA, LIMA JÚNIOR, AUGUSTO, 2016).

Once a positive value for the parameter 'voice' must be set, it is possible to conclude that VoiceP must be acquired. The node proposed by Collins (2005), however, is of an exclusive syntactic character, apparently with no counterpart in other domains. The preposition "by", head of Voice, is dummy. 'By', thereby, functions as a Case marker, only pronounced if the external argument is overt. Thus, Voice brings about a

¹² For a detailed presentation of these problems, we refer the reader to Lima Júnior; Augusto (2015).

legibility issue, which is problematic for acquisition models of passive sentences (see LIMA JÚNIOR, 2016). In other words, to be acquired, the information pertaining to VoiceP should be legible at the interface levels. The fact that VoiceP is formed by uninterpretable features only is thus undesirable for the premises of a language acquisition theory couched in Minimalist assumptions (see CORRÊA, 2009, 2014; CORRÊA; LIMA JÚNIOR; AUGUSTO, 2016).

In languages like Portuguese and English, the child may then find samples of by-phrases at PF, so, at a first sight, VoiceP being headed by the preposition 'by' does not turn out to be a problem for acquisition models of passives. In some languages (Latvian, for instance), though, long passives are grammatically forbidden; that is, by-phrases are by no means generated; its presence would make the passive sentence ungrammatical¹³ (see KEENAN; DRYER, 2007). In this sense, according to the smuggling approach, it is really hard to predict how children may set a positive value for the parameter voice in Latvian.

Thus the characterization of VoiceP does not seem to be adequate to language acquisition theories and models of acquisition of passives. Moreover, it is argued that passiveP would not suffer from the drawbacks highlighted in Collins' analysis.

One natural association to be established is between the functional node passiveP and a particular morphology identifying passives, as many languages show (BOECKX, 1998). It has been postulated then by Lima Júnior; Augusto (2015) that this morphology has to be inserted into the head of passiveP¹⁴. In these languages, passiveP may be easily identified in the input children receive (Sesotho, for instance) (see DEMUTH, 1989; 1990). Nevertheless, what is to be said in relation to languages that exhibit participial passives, in which no particular morpheme is specifically dedicated to the generation of passives?

Concerning PF, the main characteristic of passives, in languages that display participial passives, like English and Portuguese, is the non-adjacent dependency between auxiliary + participle. In Portuguese, this complex has been characterized as

¹³ It is important to notice that Latvian passives are very similar to English passives, and even more similar to Portuguese passives; that is, Latvian passives do not have a specific passive morpheme; they are formed by an auxiliary verb and the past participle. The past participle agrees in gender and in number with the subject, as in Portuguese.

¹⁴ The idea that the preposition 'by' may be the head of the functional category of passives is, thus, automatically eliminated, insofar as the preposition 'by' is not specific of verbal passives (see BRUENING, 2014; MCINTYRE, 2012).

the morphophonological identity of passives (see LIMA JÚNIOR, 2012; 2016; LIMA JÚNIOR; CORRÊA, 2015; LIMA JÚNIOR; AUGUSTO, 2015).

Concerning LF, it is possible to notice some kind of semantic constraint, which may be attributed to the presence of the functional category, passiveP (see LIMA JÚNIOR; AUGUSTO, 2015). Without entering into details here, notice that taking the work of Cançado (2002; 2005) into account, Lima Júnior; Augusto (2015) claimed that the functional category passiveP is endowed with a semantic feature ‘trigger’ to be obligatorily associated to the external argument. Compare the examples (4) and (5).

- (4) John married Mary.
 (5) Mary was married by John.

By no means can sentence (4) be interpreted in the same way as sentence (5). In other words, *John* and *Mary* are in a legal and/or spiritual commitment with each other in the active sentence (4). As sentence (5) is a passive, the external argument associated to the by-phrase (*by John*), must not be understood as Mary's husband, but someone in authority to celebrate a ceremony which ends up making Mary a married woman. If the semantics of a verb disallows the feature 'trigger' to be read in the external argument, the derivation, in spite of its syntactic convergence, tends to crash at LF.

Take the example (7), which intends to be the passive sentence of (6).

- (6) John has a car.
 (7) */? A car is had by John.

The verb *have* does not allow the external argument (*John*) to be associated to the [trigger] feature. As *John* does not trigger the ‘event’ of *having a car*, regardless of the fact that there exists no syntactic problem in deriving (7), semantically, such a sentence is not allowed. This semantic constraint is extremely important for PassiveP to be semantically distinguished from vP and acquired in languages such as Portuguese and English (see LIMA JÚNIOR, 2016; CORRÊA; LIMA JÚNIOR; AUGUSTO, 2016).

In conclusion, differently from VoiceP, PassiveP is well-motivated at the interface levels. Apart from that, the intuition that a specific functional category is the expression

of a parameter may be flawlessly incorporated by the PassiveP-approach¹⁵. On the other hand, the fact that PassiveP is an alternative to vP (instead of a functional category above it, as assumed for VoiceP), makes the issue of intervention to be expected to reappear. In the next subsection, the syntactic nature of the PassiveP-approach is, thus, discussed and it is shown that the intervention issues are trivially circumvented.

3. The passiveP-approach

The argument structure of a verbal passive is very similar to the argument structure of the transitive actives. Compare (8) to (9) below:

- (8) [PRO_{ext.arg} passiveP. [passive-V-part [PartP [~~V-part~~ [VP [~~V~~ DP_{int.arg}.
 (9) [DP_{ext.arg} vP [v-V-part [PartP [~~V-part~~ [VP [~~V~~ DP_{int.arg}.

The only relevant differences up to the point of the argument structure sketched above are that passiveP in (8) 'replaces' vP in (9); besides that, the arbitrary PRO compulsorily occupies the spec of passiveP in (8), while in (9) a DP occupies the spec of vP.

For an active sentence (10), V and the internal argument DP are merged projecting VP, as illustrated in (11). The theta-role is immediately discharged through merge.

- (10) The farmer has harvested the flowers.
 (11) [VP [harvest the flowers_{{μCase}]{θ-role-theme}}]

Subsequently, VP is merged with PartP. V, then, moves to the head of partP. Notice, in (12), that the internal argument DP has its Case still unmarked. Only *v* may assign accusative Case, as in Chomsky (1995).

- (12) [PartP [harvested [~~harvest~~ the flowers_[μCase]]

¹⁵ As 'by' is rejected as specific information of verbal passives, no problem arises for languages in which there is a specific morpheme for verbal passives apart from the by-phrase, such as Kiswahili and Japanese.

As soon as ν is merged inserting (and theta-marking) the external argument DP into the derivation, the complex V-part raises to the head of ν P, and the Case of 'the flowers' is immediately assigned by the head ν .

(13) [The farmer_{{ μ Case}{ θ -role}} ν P [harvested- ν [PartP [~~harvested~~ [VP [harvest the flowers_{ μ Case}

Likewise, in a passive sentence (14), V and the internal argument DP are merged projecting VP, as illustrated in (15). The theta-role is also immediately discharged through merge.

(14) The flowers were harvested.

(15) [VP [harvest the flowers_{{ μ Case}{ θ -role-theme}}

Subsequently, VP is merged with PartP. V, then, moves to the head of PartP. Again, notice, in (16), that the internal argument DP also has its Case unmarked.

(16) [PartP [harvested [~~harvest~~ the flowers_{ μ Case}

When passiveP is merged with the derivation in course (see 17), two main differences may be set in comparison with the derivation of actives: (i) PassiveP, differently from ν P, cannot assign Case to the internal argument; (ii) passiveP compulsorily selects an arbitrary PRO (see COLLINS, 2005; LIMA JÚNIOR; AUGUSTO, 2015) as its external argument, associating 'trigger' to it. Moreover, as illustrated in (18), passiveP is taken as constituting a phase, and thus is endowed with an edge feature. So an extra Spec may be generated and the internal argument DP, an active element without case, moves to this outer extra Spec.

(17) [PRO_{ θ -role} passive' [harvested-passive [partP [~~harvested~~—[VP [harvest the flowers_{{ μ Case}{ θ -role}}

(18) [the flowers_{{ μ Case}{ θ -role}} passiveP_{extra spec} [PRO_{ θ -role} passive' [harvested-passive [partP [~~harvested~~—[VP [harvest the flowers_{{ μ Case}{ θ -role}}

In Lima Júnior; Augusto (2015), several empirical arguments were retrieved from the literature in regard to the fact that both passive and active argument structures constitute phases. For instance, both passiveP and vP exhibit PF isolability and full argument structure at LF, which meets the propositionality criterion for phasehood (see RICHARDS, 2004, EPSTEIN, 2006); both categories also provide reconstruction sites, which are typical of an intermediate phase edge (see FOX, 2002; LEGATE, 2003); they allow for the same freedom of reordering of verb and object as well as other transphasal movements (see RICHARDS, 2004; 2006). Therefore, being passiveP a phasal node, it may project an extra specifier that can momentarily accommodate the internal argument DP.

Having set apart passiveP and vP, it is possible to continue with the derivation of passive sentences. Firstly, the auxiliary is merged with the derivation in course. As proposed by Linguinho (2011), the auxiliary verb does not project an external argument. The complex V-part-passive adjoins to the head of the auxiliary, as illustrated in (19).

- (19) [VP [be harvested-passive [the flowers_{{μCase}{θ-role}} passiveP_{extra spec} [PRO_{θ-role} passive' [~~harvested-passive~~ [partP [~~harvested~~-[VP [~~harvest the flowers~~

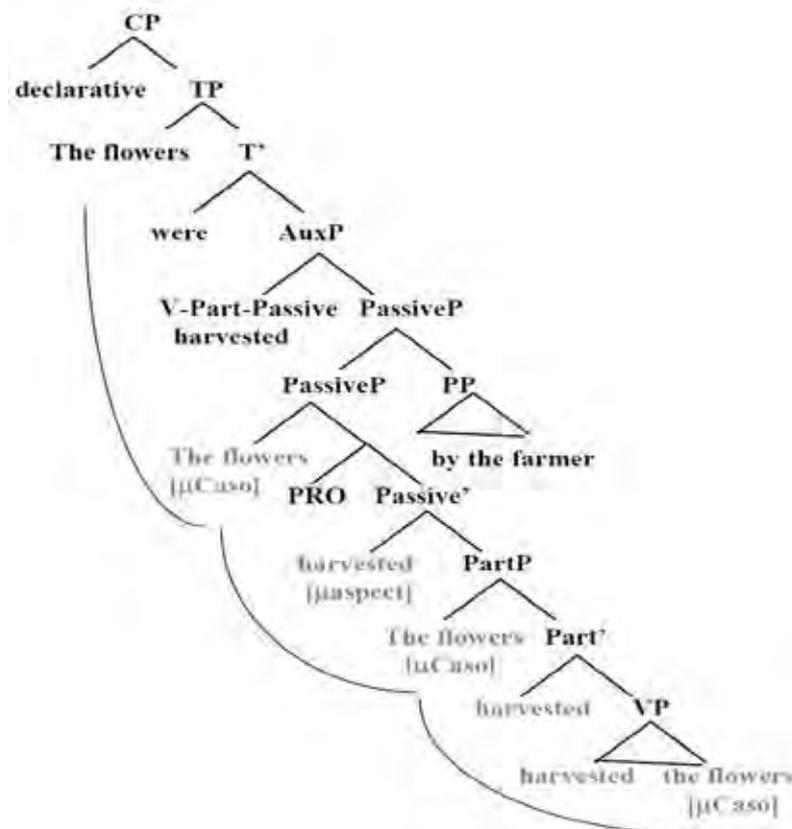
Due to the extended projection principle, as soon as T is merged, the search for a goal to occupy [spec, TP] starts. In the derivation sketched so far (see 19), the empty category, arbitrary PRO, is at [spec, passiveP]. The internal argument, which ultimately is going to be raised to [spec, TP], has found, however, an escape hatch, the extra specifier projected by passiveP. At this position, the internal argument DP is as distant of the probe (T) as PRO. This DP, thereby, may cyclically move to [spec, TP]. Once the internal argument DP, and the arbitrary PRO are equidistant from the probe, intervention is no longer obtained, as it is illustrated in (20).

- (20) [The flowers_{nominative} TP [T were [be harvested-passive [~~the flowers~~ passiveP [PRO passiveP [~~harvested-passive~~-[partP [~~harvested~~-[VP [~~harvest the flowers~~

It is important to highlight that the structure presented in (20) is the prototypical passive. In other words, if a grammar generates passives, it will generate short passives (see KYPARSKY, 2013). However, languages like Portuguese and English also generate long passives, which are passives that have by-phrases or, in other words, an apparent overt external argument. In this respect it is important to discuss where the by-phrase is to be positioned and which status it owns.

As stated by Lima Júnior; Augusto (2015), the status of the by-phrase is the following: while semantically it is an argument, syntactically it is an adjunct. In other words, by-phrases are adjuncts semantically oriented to the external argument; that is, as by-phrases are oriented to external arguments, they tend to be read at LF as if they were the argument of the verb itself. This claim is not controversial, provided that the adjunct occupies [spec, passiveP], as illustrated in (21) below:

(21)



Up to this point, the derivation of passives appears to be complete. No intervention or any complication of the sort is expected once passiveP provides the

necessary escape hatch for the internal argument DP (the flowers) to move to [spec, TP]. Yet, an aspect of this analysis is quite fuzzy. One should question what motivates the movement of the internal argument DP to the intermediate position [spec, passiveP], and then to [spec, TP]. Recall that one of the most serious criticisms that the smuggling approach has received was the lack of interpretive motivation for the XP-movement. In the next section, the interpretive motivation of the movement of the internal argument DP in verbal passives is discussed in details.

4. Motivating the movement of the internal argument DP in the derivation of verbal passives

It seems that different types of EPP are at work for allowing the internal argument DP to eventually reach [Spec, TP] in a passive sentence. There has been a long traditional view associating EPP to the sentence subject position, and more recently to the idea of an edge feature EPP, which is responsible for cyclic movement. In any case, the moved element must be active to the system, that is, it must portray an unchecked/unvalued feature. Particularly, the movement involving the definition of the subject of a sentence is standardly assumed to be motivated by Case requirements, which suggests the relevance of the interface with morphology. Many works on the issue, nonetheless, have been showing that Nominative-marking does not necessarily involve movement to [spec, TP] (see ZAENEN; MALING; THRÁINSSON, 1985; RICHARDS, 2004; CARDINALLETTI, 2004; RIZZI, 2006). Case may be assigned by the operation *agree*, which may be driven at long distance (see CHOMSKY, 1995). Given this fact, Case-marking is not, to say the least, the whole picture concerning the movement of the internal argument DP to the subject position.

In this paper, benefiting from the subject criterion proposal put forward by Rizzi (2006), we hypothesized that besides an unvalued Case feature, the internal argument DP bears another unvalued formal feature, which is read off at the subject position. Thus, concerning the intermediate movement of the argument DP to the edge of the phase (in the case of verbal passives), it occurs due to the fact that an extra Spec may always be generated to host any DP which bears some unvalued feature, and otherwise would be kept trapped in the lower phase. The unvalued features – Case and a formal D feature - allow this movement, although the position may not be able to check/value any

of them. This movement is very similar to cyclically wh-movement in complex sentences.

As far as the next movement to the subject position is concerned, it is worth following Rizzi's proposal, which aims at considering which kind of interface requirements the EPP_subject movement would be satisfying.

Elaborating on Chomsky's (2002) ideas, Rizzi (2006) states that the subject position is endowed with special discourse properties (quasi-topicality, and the like). Another external system requirement, again according to Rizzi (2006), has been stated by Rothstein (1983): events must be expressed in a subject-predicate format. These arguments are in consonance with various results of psycholinguistic studies concerning the informational relevance of the linguistic position at stake (see BOCK, 1986; KELLY; BOCK; KEIL, 1986; BOCK; WARREN, 1985; BATES; MCWHYNNY, 1982). In general, experimental results show that the subject position tends to be occupied by the element that is the most active in the working memory; and also the most prototypical and animate element in a proposition. In sum, TP has an undeniable semantic/pragmatic relevance. As stated by Foley and Van Valin (1985), it seems that elements syntactically defined as subjects have informational privileges, which are assumed here to correspond to the high structural position they occupy in the syntactic tree.

In order to talk about the nature of subjects, Rizzi (2006) resorts to the fact that subjects and topics have something in common concerning the relation between them and their predicates. Rizzi uses the terminology "aboutness" to name this relation. According to Rizzi, subjects and topics share "aboutness", but the two notions differ in that D(iscourse)-linking is a necessary component of topics, but not of subjects.

Thus, DPs occupying [Spec, TP] (or a Topic position) would be interpreted as "about which" some sentence or predicate is generated. The difference between an active and a passive sentence has to do, then, with to which argument the [aboutness] feature is associated.

Getting back to the case that brought up this issue, in the analysis being proposed for passive sentences, both arbitrary PRO and the internal argument DP are equidistant from the probe T while at [spec, passiveP] (see example 21). If no feature of DP is particularly associated with TP, the system may opt for arbitrary PRO instead of the

internal argument DP. Yet, as it has been proposed here, the aboutness feature may be a distinctive feature between both verb arguments.

In summary, the derivation of verbal passives proceeds with a successive-cyclic movement of the internal argument DP, which has been properly motivated. PassiveP plays a fundamental role since, as a phase, it provides the intermediate site for the movement of the internal argument, before it reaches [spec, TP].

In the next sections, the relevance of passiveP to the distinction between verbal passives and passives with an adjectival reading is examined.

5. Towards a minimal and uniform analysis of verbal passives and passives with an adjectival reading

Our purpose in this paper is to verify to what extent the proposal of passiveP is relevant for achieving a uniform and minimalist analysis to both eventive and adjectival passives. With that in mind, we recap Hornstein, Martins and Nunes's analysis (2008) for eventive passives, which is a very insightful and elegant solution for the intervention problem raised by the external argument in active-like derivations of verbal passives. In their proposal, the passive morpheme *-en* occupies [spec, *prtP*] (*prtP* being a light verb representative category for passives), as illustrated in (22) (see also LIMA; RUBIN, 2008).

- (22) a. The flowers were watered.
 b. [CP [C [TP the flowers [T were [VP ~~be~~ [prtP watered [prt [VP [~~water~~ the flowers

The passive morpheme *-en* at [spec, *prtP*] does not bear the feature [person]. Therefore, it does not pose problems concerning intervention effects (or minimality issues) (CHOMSKY, 1995; 2002; RIZZI, 1990), nor does it force any violation of the maximization principle (see CHOMSKY, 2000), differently from what is observed in Boeckx (1998), for example (see LIMA; RUBIN, 2008; LIMA JÚNIOR; AUGUSTO, 2015 for a discussion of Boeckx's analysis).

Being so, Hornstein et al.'s analysis seems to be simpler and more elegant than the one being proposed here and thus should be favored. It is worth highlighting,

though, that their analysis seems to presuppose that the passive participle morpheme is a kind of external argument, and thus would imply that a totally different proposal would have to be pursued to account for adjectival passives, for example. Actually, Hornstein, Martins and Nunes's analysis (2008) requires three specific participial categories to account for the different sentences that are derived from participles. In this sense, there should be a partP for active participles, such as in (23a.), a passive partP for sentences like (23b.), and an adjectival partP for sentences such as (23c.).

- (23) a. John has broken the window.
 b. A window was broken by John.
 c. The window remains broken.

Notice that an agentive entity may be syntactically represented in eventive passives as in (24), but not in stative adjectival passives as in (25). If it is assumed otherwise that the passive morpheme is somehow the external argument of passive sentences, or that the interpretation of an external argument is directly derived from the presence of this morpheme, every passive sentence with, or without, an adjectival reading should imply the interpretation of an external argument, which does not seem to be the case.

- (24) a. A porta foi queimada.
 b. The door was burned. (an event of the burning of a door by someone)
 (25) a. A porta estava aberta.
 b. The door was open(ed). (the state of the door)

[see LIMA JÚNIOR; AUGUSTO (2015), examples (1-2) in the original]

Moreover, if adjectival passives derived from unnaccusative verbs are taken into account (26a. and a'.), neither a semantic nor a syntactic external argument is allowed, despite the fact that the participle is present (26b. and b'.).

- (26) a. The tree fell down.
 b. The tree remains fallen.

- a'. A árvore caiu.
 b'. A árvore está/permanece caída.

At this point, it seems that, as Lima Júnior; Augusto (2015) argue, an external argument has to be allotted at a syntactic layer necessarily distinct from the participle itself, even though the proposition of specific categories to account for particular structures should be avoided to the limit (see CHOMSKY, 1981).

Following Collins (2005), Lima Júnior; Augusto (2015) give a uniform account to participles. Roughly, it was argued that there is no reason for active and passive participles to be distinguished, regardless of the fact that passive participles exhibit gender and number agreement in languages such as Portuguese. Participles are then inserted into the derivation, contributing to the aspectuality of the sentence in conjunction with the auxiliary verb (see LUNGUINHO, 2011). It seems to us that any difference concerning verbal passives and passives with an adjectival reading shall not be centered in the participle. In the next subsection, we entertain the possibility of attributing such a difference to the functional node passiveP, but before going into the details of this distinctive analysis between verbal and adjectival passives, this classification will be presented.

6. Verbal passives and passives with an adjectival reading: a tripartite division?

It was traditionally accepted that verbal and adjectival passives are derived in different modules. Verbal passives would be derived in the syntax, while adjectival passives in the lexicon (WASOW, 1977; among many others), possibly by some kind of verbal feature demotion operation.

According to a more recent view, this dual division was considered too simplistic (see EMBICK, 2004; DUARTE; OLIVEIRA, 2010). In his work, Embick (2004) argues for a ternary distinction between eventive passives and two types of adjectival participles: resultative and stative¹⁶. Embick offers syntactic, morphophonological, and interpretive reasons for that to be so. He is particularly interested, though, in the very syntactic nature of resultative participles in comparison with eventive participles and

¹⁶ This distinction seems fundamental from the point of view of an acquisition path (see ISRAEL; BROOKS; JOHNSON, 2000; CAPRIN; GUASTI, 2006; Lima Júnior, 2012), and it also reverberates in terms of computational cost predicted by processing models, which has already been attested experimentally (see LIMA JÚNIOR; CORRÊA, 2015).

stative participles. According to him, the different syntactic structures pertaining to the passive predicates give rise to the three distinct interpretable semantic relations at stake.

From a semantically interpretive position, passive predicates are formed by two features: [agentivity] and [eventivity]. In this sense, if a predicate is semantically marked for both features, an eventive predicate is obtained. When only eventivity is positively marked, then, a resultative predicate is obtained. In case of no semantic feature marking at all, a stative predicate is thus obtained, as illustrated in (27).

- (27) a. [+agentivity; +eventivity] → eventive passive
 b. [-agentivity; +eventivity] → resultative passive
 c. [-agentivity; -eventivity] → stative passive

From a morphophonological position, although Embick seems very convincing at elaborating on the ternary division in English, the discussion here could solely resort to Portuguese, a language in which the three patterns above are clearly morphophonologically different from each other. As it can be seen in (28-30), three different AUX+PART (auxiliary + participle) complexes are involved in the derivation of passives in this language.

- (28) As flores **foram** molhadas (ser+do) → eventive passive
 The flowers were watered. (BE_{eventive}+part)
- (29) As flores **ficaram** molhadas (ficar+do) → resultative passive
 The flowers become wet. (BECOME+part)
- (30) As flores **estão** molhadas (estar+do) → stative passive
 The flowers are wet. (BE_{stative}+part)

Although it seems clear from a morphophonological and from a semantic perspective that Portuguese exhibits the three passive predicates, Duarte; Oliveira (2010) offer the results of a series of diagnostic tests, which distinguishes the properties of the three passive predicates. Some of the tests they have run are: control of the purpose clause; acceptance of instrumental PPs; aspectual restrictions of the verb; formation of resultative constructions; negative prefixation *-in*; acceptance of by-

phrases; acceptance of adverbial phrases oriented to the agent. These tests were tentatively organized in the table 1 below in the same order they were mentioned here.

Type of test	a. Stative	b. Resultative	c. Eventive	Examples
Control of the purpose clause	X	X	√	a. *O teste está corrigido para irritar o João. b. *O teste ficou corrigido para irritar o João. c. O teste foi corrigido para irritar o João.
Instrumental PPs	X	X	√	a. *?O teste está corrigido com caneta. b. *O teste ficou corrigido com caneta. c. O teste foi corrigido com caneta.
Restrictions on the aspectuality of the verb	X	X	√	a. O teste está corrigido/*temido. b. O teste ficou corrigido/*temido. c. O teste foi corrigido/temido.
Resultative constructions	√	X	X	a. O professor construiu o teste correto. b. *O professor construiu o teste corrigido. c. *O teste foi construído corrigido.
Negative prefixation –in	√	X	X	a. O teste está incorreto. b. *O teste ficou incorrigido. c. *O teste foi incorrigido.
By-phrase/ Adverbial phrases oriented to the agent	X	X	√	a. *O copo está quebrado pelo menino/*de propósito. b. *O copo ficou quebrado pelo menino/*de propósito. c. O copo foi quebrado pelo menino/de propósito.

Table 1: Duarte; Oliveira's (2010) tests distinguishing passive predicates.

These tests function quite perfectly for both European and Brazilian Portuguese variants. They are assumed here as valid, except for the fact that some stative and resultative sentences, contrary to Duarte; Oliveira's (2010) judgment, seem to accept by-phrases and even instrumental PPs. In the example (31) below, for instance, a clearly

stative passive *está pintado* (is painted) is followed by a PP, instrumental phrase, *com pincel* (with a paintbrush). In (32), an evidently resultative passive *ficou fechada* (got/became closed) is followed by a PP, by-phrase, *pela polícia* (by the police). In (33), again a resultative passive *ficou penteado* (got/became combed) is followed by a PP, instrumental phrase, *com a bandana* (with a bandanna).

- (31) (...) com duas portas em que está pintado com pincel um carro de feno.¹⁷
 (...) with two doors in in which is painted with paintbrush a hay car
- (32) A avenida que dá acesso ao hospital ficou fechada pela polícia.¹⁸
 (...)The avenue that gives access to the hospital got closed by the police
- (33) (...) que meu cabelo ficou penteado com a bandana do show.¹⁹
 (...) that my hair got combed with the bandanna from the concert

Plenty of examples as such are provided by Portuguese speakers, regardless of their level of education. In Lima Júnior; Augusto (2015), the status of the by-phrase is discussed in details,²⁰ and it seems well incorporated to the derivation of verbal passives of the present paper. However, the fact that the presence of by-phrases and instrumental PPs is disputable in relation to the (un)grammaticality of stative and resultative passives does not seem to be sufficient reason to invalidate the ternary division proposed by Embick (2004). As far, such a division seems to survive quite well to the discussions made with respect to the semantic, morphophonological and syntactic behavior of passive predicates. What must be assessed now is the kind of analysis Embick (2004) and/or Duarte; Oliveira (2010) will offer for the three passive types

Before entering the field of the analyses, it is important to mention that both the work of Embick (2004), and of Duarte; Oliveira (2010) entertain the derivation of passives in terms of the Distributed Morphology paradigm (see HALLE; MARANTZ,

¹⁷ Example extracted from the master thesis by Isaac Vieira da Silva, *A obra de Jerônimo Bosch à luz de escritos dos séculos XV, XVI e XVII*, for the faculty of Philosophy in the University of São Paulo, 2014, page 120.

¹⁸ Example extracted from the website of O globo, accessed on the 25th of December, 2015 in <http://oglobo.globo.com/brasil/pm-da-guarda-do-governo-de-minas-gerais-atira-dentro-de-hospital-2840890#ixzz3vLp3Ojji>

¹⁹ Extracted from the blog 'Esclerose Múltipla e Eu', accessed on the 25th of December, 2015 in <http://esclerosemultiplaeu.blogspot.com.br/2011/03/foi-show.html>.

²⁰ It is possible to cogitate that by-phrases in adjectival passives own a default theta-role.

1993). Although some preoccupations are specific of a particular language, both analyses are offered in a very similar way.

In order to derive the three kinds of passives (stative, resultative and eventive), AspP, *v*P and VoiceP projections were assumed in Embick's (2004) analysis. AspP bearing a [Stative] feature would be responsible for the stative interpretation of the stative passive predicate. No *v* is present, thus no eventivity. AspP bearing a [Fient] feature along with *v* bring eventivity, in contrast with stative passives. VoiceP, which inserts the 'agent' into the derivation of passives, and AspP (also bearing a [Fient] feature]) are the categories by means of which eventive passives are derived.

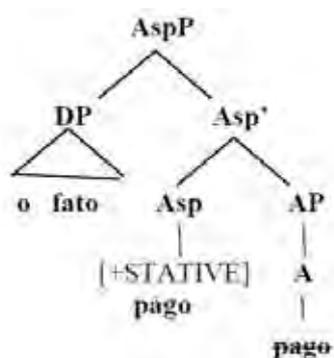
According to Embick, since the stative and the resultative participles exhibit structural differences other than the presence of Asp, it would not be necessary to assign all of the interpretive differences to the Asp. However, he refers to the aspectual head that appears in statives as Asp_S and the one that appears in resultatives as Asp_R. In sum, he argues that Asp_R defines a state out of an eventive subcomponent, while Asp_S defines a simple state. Duarte; Oliveira (2010) do not entertain that discussion, and simply call both aspectual phrases Asp, one introducing a stative feature [state] while the other introduces a fientive feature [Fient].

The stative participle is the simplest structure of the three being compared here. Take, for instance, the stative passive in (34). The derivation proposed for (34) starts as illustrated in (35) (from DUARTE; OLIVEIRA, 2010). Yet, in the resultatives, Asp introduces the feature [Fient], as it has been mentioned before. Fient is analogous to the BECOME-operator in the sense it denotes a *transition* event; that is, *that moves toward a state* (EMBICK, 2004, p. 366)²¹. The other difference is the presence of *v*. The derivation of a resultative passive (see 36) starts as illustrated in (37) (from DUARTE; OLIVEIRA, 2010).

(34) O fato está pago.

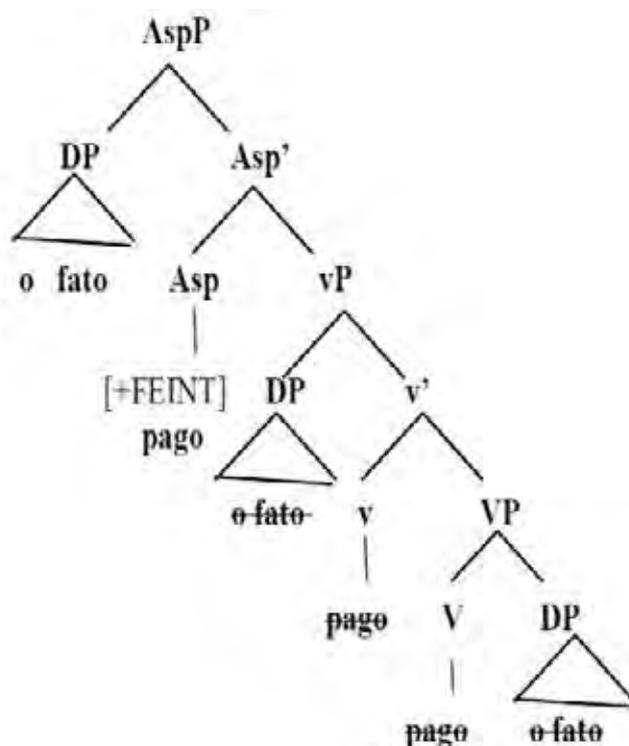
²¹ With using the nomenclature *Fient*, Embick wants avoid the misunderstanding raised by the fact that the operator BECOME is usually associated with 'achievements' and 'accomplishments' (see VENDLER, 1967; GEHRKE; GRILLO, 2009).

(35)



(36) O fato ficou pago.

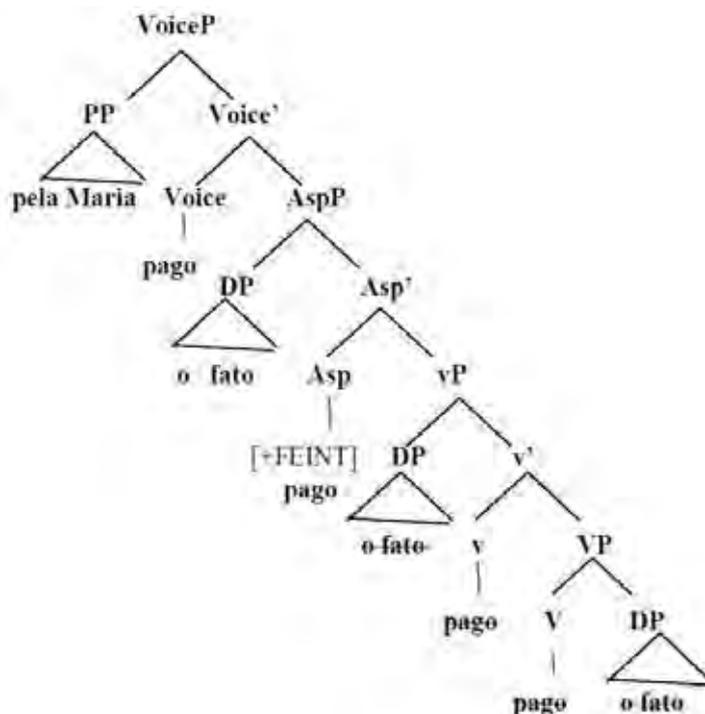
(37)



As it could not be different, the eventive passive is the most complex of the three. So as to account for the agentivity of eventive passives, Embick (2004) resorts to VoiceP. vP and AspP are also obligatorily involved. The derivation of an eventive passive (38) starts as illustrated in (39) (from DUARTE; OLIVEIRA, 2010).

(38) O fato foi pago.

(39)



Concerning the tripartite division for Portuguese in Duarte; Oliveira (2010), it could be concluded that much of what Embick (2004) proposed for English can be incorporated into an analysis taking Portuguese into account. The present paper, however, is not completely in line with this conclusion.

According to the proposal just reviewed in this section, an aspectual projection bearing a [Fient/BECOME] feature is present in both *eventive* and *resultative* passives. In other words, if the participial form of a verb allows for a resultative reading to be obtained, it should be the case that this verb allows for an eventive passive to be generated as well, contrary to facts.

Observe the passive predicates formed with the verb *preocupar* (worry) in (40). A similar frame is obtained in English (41).

- (40) a. O João está preocupado com o filho.
 b. O João ficou preocupado com o filho.
 c. *O João foi preocupado pelo filho.

- (41) a. John is_{state} worried with his son.
 b. John got worried with his son.
 c. *John was worried by his son.

Looking at the examples above, it can be stated that the verb *preocupar* (worry) allows for stative/resultative interpretation (40a. and b./41a. and b.) to be obtained. It implies that both an Asp_S, and an Asp_R may be projected, respectively resulting in a stative and in a resultative passive, as discussed before and as it was illustrated in (35) and (37). The verb *preocupar* cannot form eventive passives, though (see 40c. and 41c.) (see also BELLETTI; RIZZI, 1988; CANÇADO, 1995). In the terms stated in this section, VoiceP cannot select the projection of *preocupar* although it may project AspP [+Fient]. It is not inferable, from Duarte; Oliveira's (2010) (nor from Embick's (2004) work), what prevents the predicate *preocupar* from forming eventive passives provided that, semantically, it meets all the criteria established in Embick (2004) and/or Duarte; Oliveira's analyses; that is, it admits eventivity (*v*), and the Fient/Become-operator to be inserted into the derivation.

In the same line of reasoning, if every derivation of eventive passives must exhibit Asp bearing [Fient], it should be the case, then, that every eventive passive also forms resultative passives, once again, contrary to facts.

Take the perceptual verb *ver* (see) (*and other perceptual verbs*), for instance (42). It does not allow a stative/resultative reading (36a. and b.), but it does form eventive passives (39c.).

- (42) a. *O João está/esteve/estava/permaneceu visto na festa.
 b. *O João ficou visto na festa.
 c. O João foi visto na festa.

What seems to hold for the problems observed here is that there is more than a progressive distinction represented in the functional nodes present for each type of passive. It is what the analyses proposed by Embick (2004), and followed by Duarte; Oliveira (2010) for Portuguese, assume.

Although the differences between stative and resultative passives (from both a morphophonological and a semantic standpoint) are undeniable, the fact that some stative passives allow for by-phrases and instrumental PPs show that the boundaries between what is a stative and what is a resultative passive are not yet firmly drawn, and may depend much more on the context than we would like to accept. Take the anecdotal example in (43).

- (43) Ih! Choveu, né? Mas, - que bom! Não preciso mais regar as plantinhas que, agora, tão molhadas pela chuva.

Gee! It rained, han? That's good! I don't need to water the plants that are now wet by the rain.

For a less controversial example, let's examine (44).

- (44) a. A cama estava cuidadosamente arrumada.
The bed was carefully made.
b. *A cama estava cuidadosamente bonita.
The bed was carefully pretty.

Again, the presence of *estar* (stative be) in (44a.) makes clear that it turns out to be a stative predicate. The presence of 'cuidadosamente' (carefully) should be banned provided that this adverb is traditionally associated to an agent. However, there is no problem in 44a. -thousands of examples as such are normally produced in Portuguese. When the participle *arrumada* (made) is changed for the attribute *bonita* (pretty), the sentence becomes ungrammatical, though. The contrast between (44a.) and (44b.) shows that the verbal nature of the participle has not been completely lost as Duarte; Oliveira's (2010) work seems to imply. Notice that Embick (2004) explicitly says that *the structure for the stative should not involve v and its concomitant eventivity* (p. 363). Based on that, it can be speculated that what Embick (2004) has been calling stative participles are not participles anymore, and have migrated to the category of pure adjectives. Duarte; Oliveira (2010) mention that this category does exist in Portuguese, as it can be seen in (45). Some of them are still present in the language under both forms, and, sometimes, with different meanings.

- (45) a. cativado → cativo (captivated)
 b. tingido → tinto (dyed)
 c. corrigido → correto (corrected)
 d. nascido → nato (born)
 e. bendito → bento (blessed)

In sum, as far as Portuguese is concerned, it could be mentioned that there are two great groups of passive predicates, which refer us back to the traditional binary division: verbal passives and passives with an adjectival reading. This major division can be supported by the presence of a passiveP in eventive passives. The semantic restriction provided by that functional category prevents eventive passives of verbs as *preocupar* (worry) from being derived. The passiveP-approach also accounts for the fact that eventive passives of perceptual verbs as *ver* (see) may be derived even though it does not form stative, or resultative passives.

In the group of passives with an adjectival reading, in particular, it sounds quite evident that there are stative and resultative passives conforming two different groups - especially if the existence of the auxiliaries '*estar*' and '*ficar*' are taken into consideration. It seems that the difference between them, apart from this morphophonological difference, is solely aspectual.

A fourth group could still be identified. This fourth group holds for the participles that have completely become adjectives, and, hence, cannot be analyzed as participles anymore.

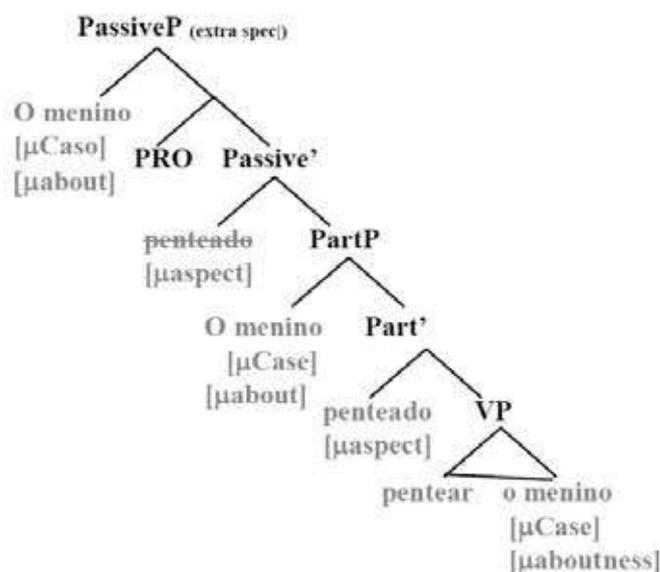
In the next section, a reformulated analysis couched in Minimalist assumptions is offered for eventive, resultative, and stative passives, as well as for the fourth group cited in the last paragraph.

7. The tripartite division reformulated

The analysis to be offered here recovers the proposal of the passiveP node. When it comes to verbal passives, the ones that claim for an eventive reading, passiveP is the maximal projection of the argument structure of transitive verbs, alternatively to vP of the actives, as illustrated in (46). As it can be seen in (46), PassiveP selects a participial

phrase, whose aspectuality is not yet defined [μ aspect] at this point. Lunguinho (2011) is thus assumed in the sense that an auxiliary verb has to assign aspect to the participle by means of agreement. The auxiliary-*ser* does that in Portuguese by selecting the passive phrase.

(46)

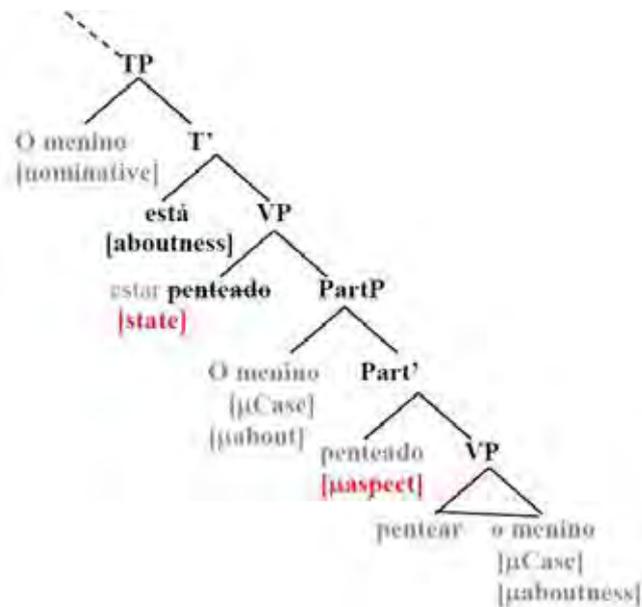


When it comes to passives with an adjectival reading; that is, resultative and stative passives, their derivations follow a very similar and unified step-by-step. It is argued that, in both, passiveP is not projected (neither is ν P), which accounts for the major binary division between verbal passives and passives with an adjectival reading.

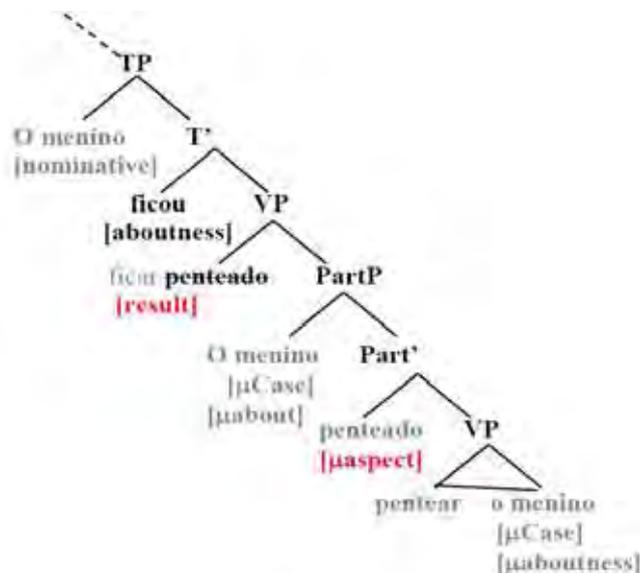
The locus of the difference between the two adjectival passives, which constitute the subgroup of resultative and stative passives, is aspectual. Particularly in Portuguese, this difference is morphologically evident. In English, it is not, unless an aspectual analysis of the sort is assumed for *get-passives*. This issue cannot be properly addressed in this paper, though, and will be left for further research (see also DIAS, 2012).

Therefore, *estar* and *ficar* are two auxiliary verbs in the sense that they must select verbal phrases. The verbal phrases selected contain PartP, which is [μ aspect]. The auxiliary verbs will, thereby, value aspect, as illustrated in (47-48), fostering either a stative or resultative interpretation of the predicate at LF.

- (47) Stative Passive: the boy is_{state} combed.

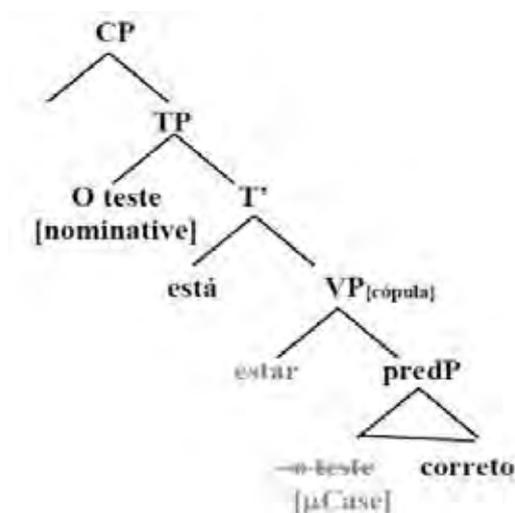


- (48) Resultative Passive: the boy became_{result} combed



When it comes to the group of participles that have lost their connection with the original verbs, such as *correto* (corrected), *tinto* (dyed), *bento* (blessed), *cativo* (captivated), among others, the analysis should be the same as the one attributed to real adjectives, as illustrated in (49).

- (49) Real Copular Construction: The test is_{copula} correct.



In this last specific case, the auxiliaries *estar* or *ficar* (*o teste ficou correto*) should not be considered auxiliaries in the actual sense of the word, but copulas, inasmuch as they select a predP. No verbal phrase is involved.

In conclusion, both VoiceP and passiveP allow for the distinction between eventive passives and passives with an adjectival reading. Remember that voiceP holds for the preposition *by*. As it has been pointed out, though, adjectival passives may also present a *by*-phrase. The head of passiveP, for its turn, contrary to voice, is not assumed to be the locus of this preposition. Therefore, passiveP seems to be more appropriate for the distinction proposed here. Additionally, contrary to Embick's and Duarte; Oliveira's works, our proposal does not necessarily imply that both eventive and resultative readings be expected for the same verbs. The distinction between stative and resultative readings among the adjectival passives is derived from the types of auxiliaries selected. Eventive readings follow from the presence of the semantic feature "trigger" associated to passiveP (LIMA JÚNIOR; AUGUSTO, 2015; for additional discussion on the relevance of features in the interface semantic-syntax see also Cançado (2005), Naves (2005), and Dias (2012)).

8. Final Remarks

The main objectives of this paper were basically twofold. Firstly, passiveP was carefully examined as the potential specific node for verbal passive sentences. Secondly, this very category was pointed out as the locus of the distinction between verbal passives and passives with an adjectival reading.

The legitimacy of passiveP has been attested from a theoretically internal point, but also in its teleological nature. The fact that passiveP is motivated at the interface levels with language faculty has positive consequences for typological studies and models of language acquisition.

PassiveP was shown to be a functional category alternative to vP, but, as the latter, passiveP is a phase node (also see LIMA JÚNIOR; AUGUSTO, 2015). Being a phase, passive projects an extra specifier that functions as an intermediate site for the cyclic movement of the internal argument. The ultimate movement to [Spec, TP], which was also discussed in the paper, is motivated by the presence of an aboutness feature in the sense of Rizzi (2006). This feature is present in the DP that is eventually moved to the highest A-position in the tree, [Spec, TP]. It is hypothesized here that such a claim may have important consequences for online models of linguistic computation (see CORRÊA; AUGUSTO, 2007; 2011; 2013; LIMA JÚNIOR; CORRÊA, 2015; LIMA JÚNIOR, 2016).

Concerning verbal and adjectival passives, the traditional division was advocated. PassiveP may be considered the locus of the distinction between these types of passives. The ternary division, however, was not exactly rejected, provided that it seems clear, especially if Portuguese is considered, that there is a difference between resultative, and stative passives. The paper, however, notices that this difference is fundamentally based on the aspectual nature of the two complexes, and it does not seem to amount to the syntactic structure, contrary to Embick (2004), at least with respect to passives. Moreover, a fourth group of apparent passives was identified in Portuguese, featuring participles that have lost their verbal connection with the original verb. These participles seem to have been lexicalized as actual adjectives and contrary to Duarte; Oliveira (2010), they have not been considered real stative passives, but copular constructions.

Finally, we would like to mention the compatibility of the analyses proposed for the acquisition path drawn by Israel, Johnson and Brooks (2000) (among others). The kind of analyses offered here seem to predict a gradual difficulty for copular

constructions, adjectival passives, and, finally, eventive passives. The acquisition of passives is assumed to rest largely on the distinction of the interpretation of participles in the language and the features associated with the AUX+PART complex (LIMA JÚNIOR, 2012; LIMA JÚNIOR; AUGUSTO, 2014; LIMA JÚNIOR; CORRÊA, 2015; CORRÊA, LIMA JÚNIOR; AUGUSTO, 2016). It is possible to speculate that children may start with the assumption that all participles are predP until more specific features are represented in the lexicon, allowing them to syntactically derive more complex passives (for more details, see LIMA JÚNIOR, 2016).

References

- BATES, Elizabeth; MACWHINNEY, Brian. "Functionalist approaches to grammar".
In.: E. Wanner & L. Gleitman (eds.) *Language Acquisition: The State of the Art*.
New York: CUP, 1982.
- BELLETTI, Adriana; RIZZI, Luigi. "Psych-verbs and θ -Theory." *Natural Language & Linguistic Theory*, 6(3), 291-352, 1988.
- BOCK, J. Kathryn. "Syntactic persistence in language production". *Cognitive Psychology* 18, 1986, p. 355-387.
- BOCK, J. Kathryn; WARREN, Richard K. "Conceptual accessibility and syntactic structure in sentence formulation". *Cognition* 21, 1985, p. 47-67.
- BOECKX, Cédric. A minimalist view on the passive. University of Connecticut, Papers in Linguistics, Massachusetts Institute of Technology, 1998.
- BRUENING, Benjamin. "Word formation is syntactic: adjectival passives in English". *Natural Language & Linguistic Theory*, May 2014, Volume 32, Issue 2, 363-422, 2014.
- CANÇADO, Márcia. "A Teoria da Proeminência de Grimshaw e os Psico-Verbos do Português Brasileiro". *DELTA*. 11.2: 279-299, 1995
- CANÇADO, Márcia. "Uma Aplicação da Teoria Generalizada dos Papéis Temáticos: Verbos Psicológicos". *Revista do GEL*. Número Especial: Em Memória de

Carlos Franchi.. (Eds.) ALTMAN C., M. HACKEROTT e E. VIOTTI. São Paulo: Humanitas/Contexto, 2002.

CANÇADO, Márcia. “Propriedades semânticas e posições argumentais”. *D.E.L.T.A*, v. 21, n. 1, 2005.

COLLINS, Chris. “A smuggling approach to the passive in English”. *Syntax* 8, 81-120, 2005.

CHOMSKY, Noam. *Lectures on government and binding*. Dordrecht: Foris, 1981.

CHOMSKY, Noam. *The Minimalist Program*. Cambridge, Mass.: MIT Press, 1995.

CHOMSKY, Noam. “Minimalist Inquiries: The Framework”. In MARTIN, Roger; MICHAELS, David; URIAGEREKA, Juan (eds.). *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, Ma.: MIT Press, 89-155 2000.

CHOMSKY, Noam. "Derivation by phase". In.: KENSTOWICZ, Michael *Ken Hale: A life in language*, 1-54. Cambridge, MA: MIT Press, 2002.

CHOMSKY, Noam. "Beyond Explanatory Adequacy". In.: BELLETTI, Adriana. (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, 104-131. Oxford: OUP, 2004.

CHOMSKY, Noam. “On Phases” In.: FREIDIN, R.; OTERO, C.P.; ZUBIZARRETA, M. (eds.), *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press, 2008.

CORRÊA, Letícia M.S. “Bootstrapping language acquisition from a minimalist standpoint: On the identification of phi-features in Brazilian Portuguese”. In: PIRES, A.; ROTHMAN, J.. (Org.). *Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies across Portuguese*. 1 ed. Berlin: Mouton de Gruyter, v. 1, p. 35-62, 2009

CORRÊA, Letícia M.S. (2014). “Interface Information and Computational Cost: An Integrated Procedural Approach to Language Acquisition with Some

Implications for SLI". In *New Directions in the Acquisition of Romance Languages*. In COSTA, João; FIÉIS, Alexandra; FREITAS, Maria João; LOBO, Maria; SANTOS, Ana Lúcia, Cambridge publishers, CA, 2014

CORRÊA, Letícia M.S; AUGUSTO, Marina R.A. Computação linguística no processamento on-line: soluções formais para a incorporação de uma derivação minimalista em modelos de processamento". *Cadernos de Estudos Linguísticos* 49:167–183, 2007.

CORRÊA, Letícia M.S; AUGUSTO, Marina R.A. "Possible loci of SLI from a both linguistic and psycholinguistic perspective". *Lingua* (Haarlem. Print), v. 121, p. 476-486, 2011

CORRÊA, Letícia M.S; AUGUSTO, Marina R.A. "Manifestações do DEL (Déficit/Distúrbio Específico da Linguagem) no domínio da sintaxe à luz de um modelo integrado da computação on-line." *Revista da Abralin*, vol. XII (2), Jul/Dez de 2013.

CORRÊA, Letícia M.S.; LIMA JÚNIOR, João C.; AUGUSTO, Marina R.A.. "A Aquisição da linguagem por meio do processamento de informação das interfaces: sobre o processo de aquisição de passivas". *Scripta*, vol. 20, número 38, 2016, p. 306-336.

DIAS, Bruno. *Aspectos das construções passivas com verbos psicológicos*. 2012. Dissertação (mestrado) - Universidade de Brasília, Instituto de Letras, Departamento de Linguística, Português e Línguas Clássicas, Programa de Pós-Graduação em Linguística, 2012.

DIAS, Bruno; Naves, Rozana R.. The auxiliary ser (be) and ficar (get) in passive construction with psychological predicates in Portuguese. Paper presented in the X Workshop on Formal Linguistics, UFRGS, Porto Alegre, RS, 2014.

DUARTE, Inês; OLIVEIRA, Fátima. Particípios Resultativos. *Textos Selecionados XXV Encontro Nacional da Associação Portuguesa de Linguística*. Porto, APL, 2010.

- EMBICK, David. "On the structure of resultative participles in English". *Linguistic Inquiry* 35, 355-92, 2004.
- EPSTEIN, Samuel. "On I(nternalist)-functional Explanation in Minimalism". To appear in *Linguistic Analysis* (ed. Kleanthes Grohmann).
- FOLEY, William A.; VAN VALIN, Robert D. "Information packaging in the clause". In.: T. Shopen (Ed.), *Language typology and syntactic description*. Vol. I: Clause structure. New York: Cambridge University Press, 1985.
- FOX, Dan. *Economy and Semantic Interpretation*. Cambridge, MA: MIT Press, 2002.
- GEHRKE, Berit; GRILLO, Nino. "How to become passive". In GROHMANN, K. *Explorations of Phase Theory: Features, Arguments, and Interpretation at the Interfaces*, Berlin e New York: de Gruyter, 231-268, 2009.
- HORNSTEIN, Norbert; MARTINS, Ana M.; NUNES, Jairo. "Perception and Causative Structures in English and European Portuguese: phi-Feature Agreement and the Distribution of Bare and Prepositional Infinitives". *Syntax* 11, 198-222, 2008.
- ISRAEL, Michael, JOHNSON, Christopher; BROOKS, Paricia J. "From states to events: The acquisition of English passive participles". *Cognitive Linguistics* 11, 2000.
- JAEGGLI, Oswald. A. "Passive". *Linguistic Inquiry*, Volume 17, Number 4, 587–622, 1986.
- KELLY, M. H. BOCK, Kathryn; KEIL, F. C. "Prototypicality in a linguistic context: Effects on sentence structure". *Journal of Memory and Language*, vol. 25, 59-74, 1986.
- LEGATE, Julie A. "Some interface properties of the phase". *Linguistic Inquiry*, 34, 506-515, 2003.
- LIMA, Ricardo J.; RUBIN, Maraci C.B.P. "Efeitos de intervenção no sistema de Agree: o caso das passivas." *ReVEL*, Vol. 6, n.10, Março de 2008.

- LIMA JÚNIOR, João C. *Revisitando a aquisição de sentenças passivas do PB: uma investigação experimental com foco na compreensão*. 2012. Dissertação (Mestrado em Estudos da Linguagem) – Departamento de Letras, PUC-Rio, 2012.
- LIMA JÚNIOR, João C. *Aquisição e processamento de sentenças passivas: uma investigação experimental com infantes, crianças e adultos*. Tese (Doutorado em Estudos da Linguagem) – Departamento de Letras, PUC-Rio, 2016.
- LIMA JÚNIOR, João C.; AUGUSTO, Marina R. A. "Aspectos estruturais de sentenças passivas relevantes para questões em aquisição da linguagem: uma análise do Português". (2012). Anais do VIII ENAL / II EIAL Encontro Inter/Nacional sobre Aquisição da Linguagem. Juiz de Fora: Programa de Pós-Graduação, 186-199, 2011, 2012.
- LIMA JÚNIOR, João C.; AUGUSTO, Marina R. A. "The relevance of aspectual and semantic features for the comprehension of short passive sentences involving agentive and non-agentive verbs." In: *New Directions in the Acquisition of Romance Languages*. COSTA, João; FIÉIS, Alexandra; FREITAS, Maria João; LOBO, Maria; SANTOS, Ana Lúcia. (Orgs.) Cambridge publishers, CA, 2014.
- LIMA JÚNIOR, João C; AUGUSTO, Marina R. A. "Is 'smuggling' really necessary? The most recent analyses of passive sentences reconsidered in terms of phasehood and cyclic movement". *ReVEL*, edição especial n. 9, 2015. [www.revel.inf.br].
- LIMA JÚNIOR, João C; CORRÊA, Letícia M.S. "A natureza do custo computacional na compreensão de passivas: um estudo experimental com adultos". *Letras de Hoje*, Porto Alegre, vol. 50, n.1, p. 91-101, jan-mar, 2015.
- LUNGUINHO, Marco V. *Verbos Auxiliares e a Sintaxe dos Domínios não-Finitos*. 2011. Tese de Doutorado. São Paulo: Universidade de São Paulo.
- KEENAN, Edward; DRYER, Matthew. "Passive in the world's languages". *Syntactic Typology and linguistic description, Volume 1: Clause structure*, 2nd ed., ed. by T. Shopen, 325-61. Cambridge: CUP, 2007.

- MCINTYRE, Andrew. “Adjectival passives and adjectival participles in English”. In ALEXIADOU, Artemis; SCHAEFER, Florian (eds.) *Non-canonical passives*. Amsterdam: Benjamins, 2012.
- NAVES, Rozana R. *Alternâncias Sintáticas: Questões e Perspectivas de Análise*. 2005. Tese (Doutorado em Linguística) - Universidade de Brasília, Brasília: UnB, 2005.
- RICHARDS, Marc. (2004) *Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax*. Ph.D. dissertation: University of Cambridge.
- RICHARDS, Marc. “Dynamic Linearization and the Shape of Phases”. To appear in *Linguistic Analysis* (ed. Kleanthes Grohmann).
- SNYDER, William; HYAMS, Nina. Minimality effect in children's passives. In Elisa Di Domenico, Cornelia Hamann and Simona Matteini. *Structures, Strategies and Beyond: Studies in honour of Adriana Belletti*, 343–368, 2015.
- RIZZI, Luigi. *Relativized minimality*. MIT Press, Cambridge, Mass, 1990.
- RIZZI, Luigi. "Locality and left periphery". In.: *Structures and beyond: The cartography of syntactic structures*, vol. 3, ed. by Adriana Belletti, 223–251. Oxford: Oxford University Press, 2004.
- RIZZI, Luigi. “On the Form of Chains: Criterial Positions and ECP Effects”, in L. Cheng, N. Corver, eds, *On Wh Movement*, MIT Press, Cambridge, Mass, 2006.
- ROTHSTEIN, Susan. (1983) *The Syntactic Forms of Predication*, PhD dissertation, MIT, Cambridge, Massachusetts.
- WASOW, Thomas. "Transformations and the lexicon". In.: CULICOVER, Peter, WASOW, Thomas; AKMAJIAN, Adrienne (eds.) *Formal Syntax*, , New York: Academic Press, 1977, p. 327-360.

NUMP AND POSSP IN DIALECTAL BRAZILIAN PORTUGUESE

Bruna Karla Pereira^a

ABSTRACT

In standard Brazilian Portuguese (BP), as well as in other Romance languages, possessives have uninterpretable number features, which are valued via nominal agreement. However, dialects of BP, especially the one spoken in Minas Gerais, have shown that 2nd person possessives, in postnominal position, do not have number agreement with the noun. In order to account for these facts, I will argue that, in this grammar, number features on 2nd person possessives are reanalyzed as being: (i) associated with the person (rather than the noun) and (ii) valued. From the first postulation, ‘seu’ is expected to be the possessive for 2nd person singular, and ‘seus’ for 2nd person plural. From the second postulation, no number concord is expected to be triggered on the possessive. In addition, based on Danon (2011) and Norris (2014), I will argue that cardinals divide BP DPs into two domains in that phrases located above NumP are marked with the plural morpheme, while phrases below it are unmarked. In this sense, because prenominal possessives precede cardinals (NumP), they must be marked with the plural morpheme for nominal agreement; whereas postnominal possessives, which follow NumP, must be unmarked. Free from the plural marking associated with nominal agreement, postnominal 2nd person possessives favor the reanalysis of the morpheme ‘-s’ as indicating the number associated with person features.

KEYWORDS: 2nd person possessives; ϕ -features; NumP; cardinals; plural morpheme

RESUMO

Em português padrão, assim como em outras línguas românicas, os pronomes possessivos carregam traços não interpretáveis de número, que são valorados via concordância nominal. No entanto, certos dialetos do português do Brasil (PB) mostram que o possessivo de 2^a pessoa, principalmente em posição posposta, não concorda em número com o nome. Por exemplo, no dialeto mineiro, um N no singular pode coocorrer com possessivo no plural, que se refere a 2^a pessoa do plural (‘de vocês’). Do mesmo modo, um N no plural pode coocorrer com possessivo no singular, que se refere a 2^a pessoa do singular. Para explicar esses fatos, argumentarei que, nessa gramática, os traços de número no possessivo de 2^a pessoa são (i) traços da pessoa e não do nome e são (ii) valorados. Com base na primeira formulação, prediz-se que ‘seu’ seja o possessivo de 2^a pessoa do singular, e ‘seus’ do plural. Com base na segunda formulação,

^a Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM), brunaufmg@yahoo.com.br

This article is an updated version of the one available at <http://www.revistas.usp.br/flp/article/view/117112>. Accessed on: December 15th, 2016.

This research was supported by CAPES Foundation (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Ministry of Education of Brazil, Brasília - DF, Zip Code 70.040-020), under the register 0751/2015-04 in the program Pesquisa Pós-Doutoral no Exterior, and carried out at MIT (Massachusetts Institute of Technology, Cambridge, Massachusetts, USA), with a visiting scholar affiliation.

não se desencadeia concordância em número no possessivo. Além disso, seguindo Danon (2011) e Norris (2014), argumentarei que os cardinais dividem DPs do PB em dois domínios, sendo que os sintagmas situados acima de NumP são marcados com o morfema de plural em concordância nominal, enquanto os situados abaixo de NumP são impedidos de terem esta marca. Assim, pelo fato de o possessivo pré-nominal estar antes do cardinal, ele é obrigatoriamente marcado com o morfema de plural, enquanto o possessivo pós-nominal não tem esta marca. Livre da marca morfológica de concordância nominal, o possessivo pós-nominal de 2ª pessoa favorece a reanálise do ‘-s’ como indicador do número da pessoa.

PALAVRAS-CHAVE: Possessivos de 2ª pessoa; traços- ϕ ; NumP; cardinais; morfema de plural

Introduction

In standard Brazilian Portuguese (BP) and in other Romance languages, the possessive pronoun agrees in number with the determiner, the noun, and other DP-internal phrases. However, dialects of BP, especially the one spoken in Minas Gerais, show a different pattern of number agreement in DPs with 2nd person possessives. In this pattern, there is no agreement in number between the possessive and the noun. As observed in (1), the determiner ‘a’ and the noun ‘gerência’ are singular, whereas the possessive ‘suas’ is marked with the plural morpheme ‘-s’. In (2), it is the other way round: the possessive ‘sua’ is singular, whereas the noun ‘fotos’ is marked with the plural morpheme. This paper presents a proposal to explain why and how this phenomenon occurs.

(1) A gerência suas só atende clientes grandes¹ (Belo Horizonte, 2016-06)

The-FEM-SG management-FEM-SG your-FEM-PL only serve-3SG customer-PL big-PL

‘Your department works only with large business customers’

¹ Data from real speech situations are referred throughout this paper with place (city) and date (month and year) of utterance. I have been collecting them over the past five years or so, in Minas Gerais (MG), as part of this research. They come mainly from the central area (Belo Horizonte and surroundings), but they were also collected in other regions of this state, as indicated by the cities names written besides each example. They have been registered by writing them down or typing them right after every single occasion when they were heard, such as: academic events, classes, business meetings, counter service utterances, spontaneous conversations, and so forth. These examples are aimed at illustrating, from real utterances, the structures under analysis, rather than building a database. Besides, with the term ‘dialectal BP’, this paper refers to a dialect spoken in a region (that may be geographically mapped as a state, in this case, Minas Gerais). This is a general use of the term, considering different dialects in Brazil as well as linguistic variation inside a state.

- (2) Ana, preciso de fotos² sua pra colocar no site do (Lavras, 2016-03)
 Ana need-1SG of picture-FEM-PL your-FEM-SG to put in-the website of-the
 meu casamento
 my wedding
 ‘Ana, I need some of your pictures so that I can post them in my wedding
 website’

In order to account for these structures, this paper is organized as follows: section 1 describes the possessive paradigms in BP (1.1) and then the pattern found in dialectal BP (1.2). Section 2 presents the theoretical background on valuation and interpretability of features (2.1) and on cardinals viewed as a boundary for the DP-internal distribution of the plural morpheme (2.2). Section 3 argues that number features, on the 2nd person possessive, are reanalyzed as interpretable and valued features associated with the person rather than the noun (3.1); and explains why this reanalysis is more frequent in postnominal position (3.2). Finally, section 4 presents hypotheses for further research.

1. The possessive system in BP

This section presents an overview of the possessive paradigm in Brazilian Portuguese (1.1) and then the specificities shown in the dialect spoken in Minas Gerais (1.2).

1.1 The possessive paradigm in BP

The possessive system in BP has pronominal and prepositional forms, as observed in Table 1.

² Example (2) was collected from a written message sent via What’s App. Other examples of this pattern were found in spoken utterances, such as (i).

- (i) Amanhã, ele verá dois serviços seu (Belo Horizonte, 2015-09)
 Tomorrow he see-FUT-3RDSG two task-PL your-SG
 ‘Tomorrow, he is coming to see your two works’

In (i), it is possible that the plural morpheme ‘-s’, in ‘serviços’, was not pronounced, as a result of a phonetic assimilation with the initial ‘s’ in ‘seu’. This possibility is also expected, under the view that, in plural DPs, phrases to the right of cardinals are not marked with the plural morpheme, in non-standard BP (section 2.1).

	Nominative pronouns	Possessive pronouns	Prepositional forms
1 SG	eu ‘I’	meu ‘my’	
2 SG ³	você ‘you’	seu ‘your’	de você ⁴ ‘of you-SG’ ‘your’
3 SG	ele ‘he’	seu ‘his’	dele ‘of-he’ ‘his’
1 PL	nós ‘we’	nosso ‘our’	da gente ‘of-the folks’ ‘our’
2 PL	vocês ‘you’	seu ‘your’	de vocês ‘of you-PL’ ‘your’
3 PL	eles ‘they’	seu ‘their’	deles ‘of-they’ ‘their’

Table 1: The possessive system in BP

The pronominal forms may be both prenominal and postnominal and show nominal agreement in gender and in number. In contrast, prepositional forms are strictly postnominal and do not have nominal agreement, but work as follows: 3rd person prepositional forms share the same gender and number features with their co-referent (whether it is overt in the sentence or indexed in the context); 2nd person prepositional forms do not inflect in gender, but share the same number features with their co-referent (whether it is overt – as in vocative position – or indexed in the context); and the 1st person plural prepositional form inflects neither in gender nor in number, as shown in Table 2.

³ Because this work is focused on ‘seu’, ‘teu’ (2nd person singular for the nominative ‘tu’) was not added to the tables. Besides, ‘teu’ is available in BP, but it can be uncommon in certain regions.

⁴ Some authors consider that ‘de você’ is ungrammatical, in a structure such as “*pai de você” (PERINI, 1985, p. 5), while others do not (KATO, 1985, p. 115; NEVES, 2000, p. 473). There are examples of possessive forms with ‘de você’ shown in the mentioned references (i) and elsewhere (ii).

(i) “sei os podres de todos, de você e de seus amigos” (NEVES, 2000, p. 473).

Know-1SG the-PL evil-PL of everyone, of you-SG and of your-PL friend-PL
‘I know the evils of you all, of yours and of your friends’

(ii) “Eu sempre vou ser de você” <<https://www.cifraclub.com.br/adalberto-adriano/culpados/>>. Accessed on: December 5th, 2016.

I always go-FUT be of you-SG
‘I will always be yours’

	Nominative pronouns	Possessive pronouns	Prepositional forms
1 SG	eu ‘I’	minha(s) ‘my’	
2 SG	você ‘you’	sua(s) ‘your’	de você ‘ <i>of you-SG</i> ’ ‘your’
3 SG	ela ‘she’	sua(s) ‘her’	dela ‘ <i>of-she</i> ’ ‘her’
1 PL	nós ‘we’	nossa(s) ‘our’	da gente ‘ <i>of-the folks</i> ’ ‘our’
2 PL	vocês ‘you’	sua(s) ‘your’	de vocês ‘ <i>of you-PL</i> ’ ‘your’
3 PL	elas ‘they’	sua(s) ‘their’	delas ‘ <i>of-they-FEM</i> ’ ‘their’

Table 2: The possessive system in BP with number and gender inflection

Tables 1 and 2 roughly summarize what is described by Perini (1985), Kato (1985), Cerqueira (1993), Silva (1996), Müller (1997) and Castro (2001), and do not intend to represent all the views of these authors.

Therefore, in standard BP, the possessive ‘seu’ agrees in number and gender with the noun and may refer to either 2nd person plural or 2nd person singular. This is observed in (3), which allows the two possible readings shown in (4a) and (4b).

- (3) Preciso de dois favores seus (‘seus’ = ‘de você’ or ‘de vocês’)
 Need-1SG of two favor-MASC-PL your-MASC-PL (your-PL = ‘of you-SG’ or ‘of you-PL’)
 ‘I need two favors from you’
- (4) a. Amigo₁, preciso de dois favores seus₁! (‘seus’ = ‘de você’)
 Friend need-1SG of two favor-MASC-PL your-MASC-PL (your-PL = ‘of you-SG’)
 ‘My friend, I need two favors from you’
- b. Amigos₁, preciso de dois favores seus₁! (‘seus’ = ‘de vocês’)
 Friends need-1SG of two favor-MASC-PL your-MASC-PL (your-PL = ‘of you-PL’)
 ‘My friends, I need two favors from you (guys)’

This reanalysis occurs, when the possessive is postnominal. In contrast, the prenominal position does not make it often available. Among several data that I have been collecting over the past years, there were few examples in which the mentioned reanalysis occurs with prenominal possessives, as the one in (6):

- (6) O seus carro não pode ficar estacionado aqui (Ouro Preto, 2015-06)

The-MASC-SG your-MASC-PL car-MASC-SG not may stay parked here

‘Your (guys’s) car is not allowed to be parked here’

Except by this restriction, the possessive with reanalyzed number features occurs inside a DP, which may contain definite (7) or indefinite articles (11), indefinite pronouns (8), nominal ellipsis (9), and nouns without determiners (2, 10).

- (7) “Para a sorte suas, eu não vou estar aqui na próxima votação” (Belo Horizonte, 2015-12)

For the-FEM-SG luck-FEM-SG your-FEM-PL I not go-fut be here in-the next voting

‘For your (guys’s) luck, I will not be here in the next election’

- (8) “Eu não quero nada seus” (Belo Horizonte, 2014-04)

I not want-1SG anything your-PL

‘I do not want anything from you (guys)’

- (9) “O meu olhar é diferente do seus” (Belo Horizonte, 2014-11)

The-MASC-SG my-MASC-SG view-MASC-SG is different of-the-MASC-SG [e] your-MASC-PL

‘My view is different from yours’

- (10) “É interesse seus aprovar a proposta” (Belo Horizonte, 2015-12)

Is interest-MASC-SG your-MASC-PL approve the proposal

‘It is in your interest to approve the proposal’

- (11) “Gostaria de uma informação suas” (Belo Horizonte, January 2016)

Like-1SG of a-fem-SG information-FEM-SG your-FEM-PL

‘I would like to get a piece of information from you (guys)’

To sum up, in this grammar, the postnominal possessive ‘seu’: refers to 2nd person plural, when it has the plural morpheme ‘-s’; and to 2nd person singular, when it does not have ‘-s’.

2. Theoretical background

This section presents the theoretical background on valuation and interpretability of features (2.1) as well as on the position of cardinals as a boundary for the DP-internal plural marking (2.2).

2.1. The valuation and interpretability of features

The term phi-feature is used to cover broadly the three main categories that involve agreement (person, gender, and number) and that are analyzed under the concepts of valuation and interpretability. In this paper, I will follow the feature sharing approach (PESETSKY; TORREGO, 2007) for the understanding of these operations.

Concerning valuation, according to Pesetsky and Torrego (2007, p. 263), “Certain features on lexical items appear to come from the lexicon unvalued, and receive their value from a valued instance of the same feature, present on another lexical item”. For instance, gender is a property of the noun (N), and comes valued with the noun from the lexicon. In contrast, D and A are lexically unvalued for gender, and they get “valued as a consequence of a syntactic process of agreement with the gender feature of N” (PESETSKY; TORREGO, 2007, p. 263). D and A are also lexically unvalued for number, and they get “valued as a result of agreement with N” (PESETSKY; TORREGO, 2007, p. 263).

Concerning interpretability, the distinction between interpretable and uninterpretable features is related to “whether or not a feature of a particular lexical item

makes a semantic contribution to the interpretation of that item” (PESETSKY; TORREGO, 2007, p. 264). For instance, the number feature of A does not make any contribution to its meaning, while number and person contribute to the interpretation on the DP.

Having made this brief summary on the concepts of valuation and interpretability, it is important to inquire how one can tell whether the noun is lexically valued for number and also how number agreement is triggered inside the DP.

As for the first question, an explanation, according to Pesetsky and Torrego (2007), is related to *pluralia tantum* nouns, such as ‘scissors’. These nouns are always plural, which indicates that English nouns come lexically valued for number in the derivation; whereas there is no *pluralia tantum* D or A, which means that they cannot be lexically valued for number. Nevertheless, not every language has *pluralia tantum* nouns. According to Pesetsky and Torrego (2007), in Spanish, for instance, genuine *pluralia tantum*⁷ nouns “seem not to exist”, and this entails a different understanding on the source of number features in this language. Based on previous researches, Pesetsky and Torrego (2007) suggest that Spanish number is in fact a feature of NumP. Similarly, Blühdorn et al. (2008) assume that BP does not have genuine *pluralia tantum*⁸, which means that BP nouns are not lexically valued for number. From these facts, I assume that the *locus* of number in BP is NumP, as in Spanish.

As for the second question, under the assignment view (CHOMSKY, 2001), agreement takes place when a probe with uninterpretable features seeks its goal with interpretable features, in order to become valued. Once uninterpretable features have been assigned a value, they must be deleted. As opposed to the assignment view just described, one of the consequences of adopting a feature sharing approach is that, after valuation takes place, the feature is not deleted, but is still available for another probe:

In this respect, the output of the feature sharing version of Agree [...] is the same as the output of the assignment version of Agree [...] H now contains valued F. Of course, F on H may

⁷ In BP, ‘óculos’ (‘glasses’) is not always understood as plural. In fact, against what is prescribed in traditional grammars, it is often used with modifiers in the singular. In addition, most speakers understand that “o óculos” (the-SG glasses) refers to only one object (a pair of glasses), while “os óculos” (the-PL glasses) refers to more than one object (pair).

⁸ Blühdorn et al. (2008, p. 15) explain that, in BP, the plural morpheme may be either added to a noun in the singular or taken from a noun in the plural. Among the examples given, nouns such as “pêsames” (‘condolences’), “afazeres” (‘doings’), and “núpcias” (‘nuptials’) may appear without ‘-s’, while nouns such as “gado” (‘cattle’), “gente” (‘folks’), and “fôlego” (‘breath’) may appear with ‘-s’.

now serve as the goal for some later operation of Agree triggered by an unvalued, higher instance of F serving as a new probe. The result will be a single feature F shared by three positions, as the process could iterate further (PESETSKY; TORREGO, 2007, p. 268).

Another consequence of adopting a feature sharing approach is that there is an independence between valuation and interpretability. Thus, the lexicon is expected to have four types of features, as follows.

uF val	uninterpretable, valued	iF val	interpretable, valued
uF []	uninterpretable, unvalued	iF []	interpretable, unvalued

(PESETSKY; TORREGO, 2007, p. 269)

Danon (2011) explains and exemplifies this reasoning:

Assume [...] that D enters the derivation with unvalued gender and number features, and with a valued person feature. D would then probe for the ϕ -features of lower projections: NP for gender, and either NP or NumP for number. Unlike the system of Chomsky 2000, 2001, these features on D would not delete at this stage; they will continue to exist as instances of shared ϕ -features [...]

Consider, for instance, the derivation of the following French sentence:

[...] La fille a parlé
 the.F.SG girl.F.SG has.3SG spoken
 ‘The girl spoke’

Putting aside Move operations, the two relevant steps in the derivation are:

1. The unvalued gender and number features of the D *la* probe for those of the N *fille*; following Agree, the valued gender and number features have two instances each, on D and on N.
2. The unvalued ϕ -features of the T *a* probe for those of the D *la* [...]

Following step 1, the ϕ -features of D are all valued—not as a result of being interpretable (as the gender and number features are not interpretable on D), but simply as a result of being instances of shared, valued features [...]

Finally, interface conditions, such as the lexical category of each head, determine where each of these features should be interpreted: person on D, number on N (or Num), and gender on N.

(DANON, 2011, p. 308-309)

In sum, D and A probe NumP (in some languages or NP in others) as the goal for number feature valuation. The next section (2.2) will show how this mechanism applies to BP.

2.2. Cardinals and the DP-internal distribution of the plural morpheme

According to Danon (2011) and Norris (2014), in several languages, such as Finnish and Estonian⁹, cardinals work as a boundary dividing the DP into two domains, such that phrases preceding the cardinal are marked with the plural morpheme, while phrases following it are unmarked. In Pereira (2017), I assume that this prediction applies to non-standard BP. For instance, in (12), phrases above NumP are marked with the plural morpheme, while phrases under its c-command domain are unmarked¹⁰, as represented in (13).

⁹ (i) Finnish:

Ne kaksi pien-tä auto-a seiso-ivat tiellä. (Brattico 2010) (DANON, 2011, p. 301)
 those.PL two.SG small-PART.SG car-PART.SG stand-PAST.3PL road.ADESS
 ‘Those two small cars stood at the road’

(ii) Estonian:

nee-d viis ilusa-t maja. (NORRIS, 2014, p. 144)
 this-PL.NOM 5.NOM beautiful-PAR house.PAR
 ‘these five beautiful houses’ (Erelt et al. 1993b:143)

¹⁰ Two grammars co-exist in BP: in non-standard BP, the cardinal is a boundary for the DP-internal distribution of the plural morpheme; in standard BP, every single DP-internal phrase is marked with the plural morpheme ‘-s’, regardless of the cardinal position (i, ii).

(i) Os únicos (dois) baldes vermelhos
 The-PL only-PL (two) bucket- PL red-PL
 ‘The only (two) red buckets’

(ii) Os (dois) únicos baldes vermelhos
 The-PL (two) only-PL bucket- PL red-PL
 ‘The only (two) red buckets’

(12)

- | | |
|---|---|
| <p>a. Os único balde vermelho
The-PL only-ϕ bucket-ϕ¹¹ red-ϕ
‘The only red buckets’</p> <p>b. Os únicos balde vermelho
The-PL only-PL bucket-ϕ red-ϕ
‘The only red buckets’</p> <p>c. *O únicos baldes vermelho
The-ϕ only-PL bucket-PL red-ϕ
‘The only red buckets’</p> | <p>a’. Os dois único balde vermelho
The-PL two only-ϕ bucket-ϕ red-ϕ
‘The only two red buckets’</p> <p>b’. Os únicos dois balde vermelho
The-PL only-PL two bucket-ϕ red-ϕ
‘The only two red buckets’</p> <p>c’. *O únicos dois baldes vermelho
The-ϕ only-PL two bucket-PL red-ϕ
‘The only two red buckets’</p> |
|---|---|

(13) For (12a/a’): [DP_DOs [NumP(**dois**) [AgrP [APúnico [AgrPbalde_i [APvermelho [NP_{t_i]]]]]]]]¹²}

Therefore, in (12a, a’), the determiner, which precedes the cardinal, is marked with the plural morpheme; and, in (12b, b’), both the determiner and its most adjacent adjective, which precede the cardinal, are marked. In contrast, (12c, c’) are ruled out both (i) because phrases located to the left of the cardinal are not marked with the plural morpheme, when they should be¹³, and (ii) because phrases located to the right of the cardinal are marked with the plural morpheme, when they should not be.

¹¹ This paper glosses unmarked plurals with the symbol ‘ ϕ ’.

¹² According to Cinque (2005), the universal order of the DP-internal functional projections is: [DP NumP AP NP]. Other possible linear word orders are explained by movement of the NP as an XP to the Spec of AgrP positions, which are merged with each functional projection in the DP-structure. For instance, in (13), the postnominal position of ‘vermelho’ is derived by moving the NP to Spec, AgrP, a position higher than the AP. In contrast, (12b’) has the following word order [DP AP NumP NP], with an AP preceding NumP. This order is not generated under NP-movement. Cinque (2005, p. 381) explains that “Neither head movement nor movement of a phrase not containing the (overt) NP is possible (except perhaps for focus-related movements of phrases to a DP-initial position)”. Considering Giusti (1996, p. 121), I argue, in Pereira (2017), that the word order in (12b’) is derived by moving the AP ‘únicos’ to Spec, TopP, a position higher than NumP.

¹³ Following the same reasoning, (i) is ruled out because ‘único’, preceding the overt cardinal, should be marked with the plural morpheme.

- (i) *os único **dois** balde vermelho
The-PL only- ϕ two balde- ϕ vermelho- ϕ
‘The only two red buckets’

This assumption reveals that the plural marking is explained by the syntactic hierarchy of the DP. As a result, this conclusion challenges current proposals, which, under the basis of an “autonomous morphological component [...] partly independent from syntax” (COSTA; FIGUEIREDO SILVA, 2006, p. 44), argue for a “singleton” plural morpheme in BP (14). As pointed out by Castro and Pratas (2006, p. 18), this description does not account for the fact that: “In most cases the plural marker seems to surface as a singleton, but in others the plural is marked in two different positions [...] these patterns must be subject to further investigation”.

(14) “Os primeiro livro da biblioteca” (COSTA; FIGUEIREDO SILVA, 2006, p. 28)

The-PL first- ϕ ¹⁴ book- ϕ of-the library

‘The first books of the library’

In addition, assuming the concepts presented in section (2.1), the valuation of number features, in an example such as (12a, a’), can be described in (15), which means that number features come lexically valued neither with D nor with N, but with Num.

- (15) a. D, A, and N are lexically uninterpretable and unvalued for number: *uF []*
 b. NumP has interpretable and valued number features: *iF val*
 c. D probes NumP, the closest in its c-command domain, and gets valued: *uF val*
 d. A and N are under the c-command domain of NumP and constitute a chain sharing the same unvalued number features
 e. A probes Num and gets its number features valued¹⁵: *uF val*

¹⁴ In Costa and Figueiredo Silva (2006, p. 28), the glosses given for ‘primeiro’ and ‘livro’ in (14) are with ‘-SG’ (“The-PL first-sg book-sg”). Under the view assumed here, glossing these phrases with ‘-SG’ is unjustified, because they are not inflected for singular in these data. They are just not marked morphologically for plural, but become valued with plural features.

¹⁵ According to Norris (2014: 104-105, and references therein), “the relationship between a c-commanding probe and a c-commanded goal encoded in Agree is only a preference. Under this view, adjectival heads search their c-command domains for suitable goals as normal. Upon finding nothing to

- f. As a consequence of being in a chain with A, N and the lowest A get their number features valued as well: *uF val*

Therefore, (12) complies with the assumption made by Danon (2011) and Norris (2014) that the cardinal divides the DP into two domains, such that phrases on its left are marked with the plural morpheme while phrases on its right are unmarked.

3. Analysis

In this section, I explain: firstly, how the mentioned reanalysis occurs, in terms of φ -feature interpretability and valuation (3.1); and secondly, why the postnominal position is more suitable for this, in terms of the DP-internal distribution of the plural morpheme (3.2).

3.1. Possessive number features in dialectal BP

Possessives “combine two independent features for number: the first one is related to person, the other one is related to the DP”¹⁶ (ZRIBI-HERTS, 1998, p. 151, my translation). For instance, 1st person possessive pronouns have two number layers (Table 3): the number associated with the person (‘meu’ versus ‘nosso’) is interpretable, while the number associated with the noun (‘meu’ versus ‘meus’ or ‘nosso’ versus ‘nossos’) is uninterpretable.

1 st person		Person number	Noun number
meu	<i>my</i> -MASC-SG ‘my’	SG	SG
meus	<i>my</i> -MASC-PL ‘my’	SG	PL
nosso	<i>our</i> -MASC-SG ‘our’	PL	SG
nossos	<i>our</i> -MASC-PL ‘our’	PL	PL

Table 3: Combination of number features on 1st person possessives

Agree with, the search continues upward, i.e., the probe search for a c-commanding goal rather than a c-commanded goal”.

¹⁶ “les possessifs (...) combinent deux traits de nombre indépendants, le premier, solidaire de la marque de personne, l’autre étant celui du DP” (ZRIBI-HERTS, 1998, p. 151)

These two layers are not so distinctively specified with 2nd person possessives (Table 4). In standard BP, the layer on person features is unspecified for number, which means that ‘seu(s)’ is ambiguous between 2nd person plural and 2nd person singular.

2 nd person		Person number	Noun number
seu	<i>your-MASC-SG</i> ‘your’	-	SG
seus	<i>your-MASC-PL</i> ‘your’	-	PL

Table 4: Number features on 2nd person possessives in standard BP

In contrast, dialectal BP inverts this pattern (Table 5), which means that ‘seu’ is specified for 2nd person singular, and ‘seus’ for 2nd person plural.

2 nd person		Person number	Noun number
seu	<i>your-masc-sg</i> ‘your’	SG	-
seus	<i>your-masc-pl</i> ‘your’	PL	-

Table 5: Number features on 2nd person possessives in dialectal BP

To sum up, dialectal BP reanalyzes number features on the 2nd person¹⁷ possessive as being the number of the person and as being interpretable and valued. Therefore, the plural morpheme ‘-s’, on postnominal 2nd person possessives, does not reflect nominal number agreement¹⁸.

¹⁷ In dialectal BP, when ‘seus’ is for 2nd person plural and ‘seu’ for 2nd person singular, the person distinction between singular and plural is produced by the presence or absence of the morpheme ‘-s’ in the same pronominal form, as a result of reanalysis. In EP, this distinction is made in two different lexical forms: ‘vosso’ and ‘teu’. In standard BP, no formal distinction is made on the possessive pronoun, as ‘seu’ is used for both 2nd person plural and 2nd person singular. The reason why this reanalysis is excluded on 1st person possessives is that person distinction between singular and plural is already made in two different lexical forms: ‘nosso’ for 1st person plural and ‘meu’ for 1st person singular.

¹⁸ These two different grammars, the dialectal (Table 5) and the standard one (Table 4), co-exist.

3.2. 2nd person possessives in the DP-hierarchy: prenominal versus postnominal positions

This section is focused on the following question: why does the postnominal position favor the reanalysis explained above, while the prenominal position does not?

In section 2.2, I assumed that cardinals function as a boundary that divides the DP into two domains, such that phrases above NumP are marked with the plural morpheme, whereas phrases below NumP are unmarked. This prediction applies straightforwardly to possessives in plural DPs of non-standard BP. As shown in (16a), the article ‘os’ and the possessive ‘seus’, which precede the cardinal numeral, are marked for plural; whereas the noun ‘carro’ and the adjective ‘novo’, which follow the cardinal, are unmarked.

- (16) a. Não vi [os seus (dois) carro novo]
 Not saw the-MASC-PL your-MASC-PL two car- ϕ new- ϕ
 ‘I did not see your two new cars’

In prenominal position, the possessive precedes cardinals (16a). For this reason, it is always marked with the plural morpheme, in plural DPs. In fact, the definite article is allowed to be unmarked (16b), but never is the possessive (16c).

- (16) b. Não vi [o seus carro novo]
 Not saw the-MASC- ϕ your-MASC-PL car- ϕ new- ϕ
 ‘I did not see your new cars’
- c. *Não vi [os seu carro novo]
 Not saw the-MASC-PL your-MASC- ϕ car- ϕ new- ϕ
 ‘I did not see your new cars’

Therefore, (16b) may apparently represent a problem for the analysis I am assuming, because, as the article is located to the left of NumP, it should be marked with the plural morpheme as well as the possessive. Concerning this, it has been observed that, when the definite article co-occurs with pronominal possessives, “the definite article [...] is not the marker of definiteness, and is just an expletive” (COSTA; FIGUEIREDO SILVA, 2006, p. 40). Under this view, being an expletive determiner, the definite article is allowed to be dropped or to appear without plural marking.

My analysis of these facts is that the definite article, when co-occurring with pronominal possessives, forms, with the pronominal possessive, a single phrase (DP) in which the article is the specifier (Spec,DP)¹⁹, while the possessive is the head (D). The strongest evidence for this is the adjacency²⁰ between the article and the pronominal possessive, which has already been observed in the literature (CASTRO, 2001, p. 611). For instance, in (16d), the cardinal can not intervene between the article and the possessive, nor can an adjective (16e).

¹⁹ I disagree with an anonymous reviewer who claims that Spec,DP is the position for quantifiers. In a classical paper by Giusti (1991, p. 438), it is assumed that “Quantified nominals are of category QP, namely that Q [...] is a functional category that selects a definite nominal (DP) or an indefinite one (NP)”.

²⁰ I disagree with an anonymous reviewer who claims, based on the examples (ib) and (iib), that there is no adjacency between definite articles and pronominal possessives. Possessives, in (ib) and (iib), are originally postnominal rather than pronominal, despite the appearances. Linearly, one may think that, in (ib) and (iib), an adverb (‘quase’ or ‘ainda’) intervenes between the definite article and a pronominal possessive, but this is not the case. This superficial ordering derives from movement of a postnominal possessive with its preceding adverb (‘quase seus’ or ‘ainda minhas’) by pied-pieping. The functional projection containing the adverb and dominating the projection containing the possessive (‘quase seus’ or ‘ainda minhas’) move to Spec of an intermediate position (TopP) situated between the DP (‘os’ or ‘as’) and the NP (‘namorados’ and ‘mulheres’). An evidence for that comes from the canonical order in (ia) and (iia) where the possessive, preceded by an adverb, is shown to be in fact postnominal. Therefore, the possessive in (ib) and (iib), is not a D-head, but an XP, situated originally low (Spec,PossP), in the DP-structure. Its superficial pronominal position is a result of movement (with an adverb) from a postnominal position to a topic position (GIUSTI, 1996) that precedes the noun.

- | | | | |
|----|--|-----|---|
| i. | a. os namorados [quase seus]
the-PL boyfriend-PL almost your-PL
‘boyfriends that are almost yours’ | ii. | a. as mulheres [ainda minhas]
the-PL woman-PL still my-PL
‘women that are still mine’ |
| | b. os [quase seus]; namorados [t] _i
the-PL almost your-PL boyfriend-PL
‘boyfriends that are almost yours’ | | b. as [ainda minhas] _i ; mulheres [t] _i
the-PL still my-PL woman-PL
‘women that are still mine’ |

- (16) d. *Não vi [os dois seu carro novo]
 Not saw the-MASC-PL two your-MASC- ϕ car- ϕ new- ϕ
 ‘I did not see your two new cars’
- e. *Não vi [os único seu carro novo]
 Not saw the-MASC-PL only- ϕ your-MASC- ϕ car- ϕ new- ϕ
 ‘I did not see your unique new cars’

As the definite article and the pronominal possessive form together a single phrase, the specifier (the article) is optionally marked with the plural morpheme, while the head (possessive) is mandatorily marked. In view of this, the prediction in which NumP divides the DP into two domains applies: the DP (containing the article and the possessive) is above NumP, as shown in (20a), which makes it get the plural marking. The fact that the article may be optionally marked does not cause any problem for this prediction, because it is inside a phrase whose head is already marked.

To sum up, in (16b), the 2nd person possessive: (i) is in a plural DP; (ii) is pronominal; (iii) is a D-head; (iv) has its number features valued by NumP; (v) must be marked with the plural morpheme, which represents nominal number features; and (vi) may refer to either 2nd person singular or 2nd person plural.

Having said that, I will compare (16b) with (6), both repeated below. In these data, the DP-internal distribution of the plural morpheme ‘-s’ looks the same (17).

- (16) b. Não vi [o seus carro] novo
 Not saw the-MASC- ϕ your-MASC-PL car- ϕ new- ϕ
 ‘I did not see your new cars’

(6) [O seus carro] não pode ficar estacionado aqui (Ouro Preto, 2015-06)

The-MASC-SG your-MASC-PL car-MASC-SG not may stay parked here

‘Your (guys’s) car is not allowed to be parked here’

(17) D[--] Poss[-s] Noun[--]

Nonetheless, as discussed in previous sections, in (6), although the possessive is prenominal, it: (i) is in a singular DP; (ii) has interpretable and valued number features associated with the person; (iii) does not have its number features valued by NumP; (iv) does not have agreement in nominal concord; and (v) refers to 2nd person plural only.

Therefore, the structure in (17), for “o seus carro”, may refer to two readings (18). Thus, the formula in (17) represents in fact two possible different structures. The one in (19a) refers to the reading in (18a) and corresponds to (16b); whereas the one in (19b) refers to the reading in (18b) and corresponds to (6).

(18) a. The DP is plural, and the ‘-s’ on the possessive indicates DP-internal agreement.

b. The DP is singular, and the ‘-s’ on the possessive indicates 2nd person plural.

(19) a. D[- ϕ] Poss[-s] Noun[- ϕ]

b. D[-SG] Poss[-s] Noun[-SG]

In this sense, sentences like (6), that presents reanalysis of number features in the possessive ‘seu’, are uncommon, because the possessive in prenominal position is mandatorily marked with the plural morpheme for nominal concord, which makes it resistant to the reanalysis described above. A sharp contrast is shown by the postnominal possessive. It is prevented from being marked with the plural morpheme for nominal concord, which makes it free for the reanalysis to occur.

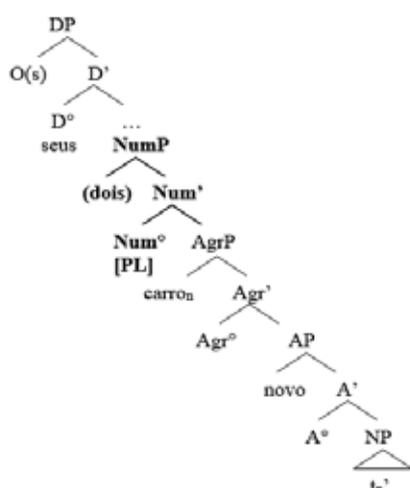
This contrast is also evident in the DP-hierarchy. In (16b), the possessive is a D-head (20a); whereas, in (1), repeated below, the postnominal possessive is the specifier of a functional projection (PossP²¹), which is merged lower in the DP (20b).

(1) A gerência suas só atende clientes grandes (Belo Horizonte, 2016-06)

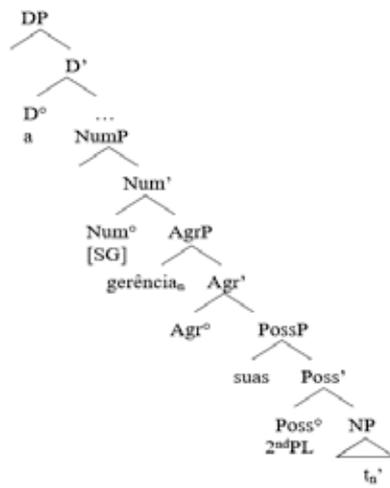
The-FEM-SG management-FEM-SG your-FEM-PL only serve-3SG customer-PL big-PL

‘Your department works only with large business customers’.

(20) a. For (16a, b):



(20) b. For (1):



In sum, this section began with the following question: why does the postnominal position favor the reanalysis (of the number features on the 2nd person possessive), while the prenominal position does not? The answer is the following: prenominal possessives are in a phrase (DP) located above NumP; as such, they must receive the plural marking associated with nominal agreement. In contrast, postnominal possessives are in a phrase (PossP) located below NumP; as such, they must not receive this mark. Being free from this mark, the latter is suitable for the mentioned reanalysis to take place.

²¹ PossP stands for Possessive Phrase, as seen in Coene and D’hulst (2003) and other references. Besides, Castro (2001), Costa and Castro (2001), and Brito (2007) recognize that prenominal possessives differ from postnominal possessives in that the latter is an XP – PossP in Brito (2007) – while the former is an X°. In (20), the XP is a PossP, and the X° is a D.

4. Hypotheses for a further discussion

An important aspect of the distribution of the possessive ‘seu’ that remains to be addressed is the relation between its position and its interpretation with regard to 2nd and 3rd person readings as well as the distinction between weak and strong forms.²²

As seen above, the 2nd person possessive, in dialectal BP, has its number features reanalyzed in postnominal position preferably. In addition, it is observed that the postnominal position, in BP, is more likely to license ‘seu’ as a 2nd person pronoun than ‘seu’ as a 3rd person pronoun. For instance, in (21)²³, when the possessive ‘seu’ is prenominal, ambiguity between 3rd and 2nd person readings is conveyed. However, when it is postnominal, 2nd person reading is the predominant interpretation²⁴.

²² Brito (2007) claims that European Portuguese (EP) has three grammars for possessives, which can be D (clitic - i), AgrP (weak pronoun - ii) or PossP (strong pronoun - iii). In the last two cases, possessive movement to AgrP (ii) and N movement to a higher position than the PossP (iii) are assumed.

- | | | |
|-------|--|----------------------|
| (i) | “os me livros”
the-PL my(reduced) book-PL | (Brito, 2007, p. 45) |
| (ii) | “a ainda minha mulher”
the still my wife | (Brito, 2007, p. 45) |
| (iii) | “uma casa minha”
A house my | (Brito, 2007, p. 46) |

The analysis given to (i), for prenominal possessives in EP as D, is the closest to the one I am assuming for prenominal possessives in BP, due to adjacency facts. However, in BP, no phonological reduction of the type given in (i) can be found, which means that prenominal possessives in BP cannot be analyzed as clitics. In fact, Castro and Costa (2001, p. 109) have already pointed out that pre-nominal possessives are weak (but not clitic) forms that exhibit typical behavior of X⁰: “Sendo formas fracas não-clíticas, esperase que exibam um comportamento típico de X⁰, mas não idêntico ao dos clíticos”. In addition, as opposed to movement of possessives or of N as a head, I am adopting Cinque’s (2005) proposal for movement of the NP (as an XP) only or conjoined with another phrase by pied-pipe, which does not allow for the possessive to move alone.

²³ The pairs in (21) were elaborated based on the idea that ‘seu’ in BP refers to 3rd person when its antecedent is a bound variable (MÜLLER, 1997).

²⁴ The following questions were asked to informants in a written test:

In (21a), whose problem is it?

- () the teacher’s
- () the listener’s
- () the teacher’s and the listener’s (both readings are possible).

In (21b), whose problem is it?

- () the teacher’s
- () the listener’s
- () the teacher’s and the listener’s (both readings are possible).

- (21) a. Cada professor₁ tratou de resolver o seu_{1/2} problema.
 Each teacher treat-PAST-PL of solve-INF the his/your problem
 ‘Each teacher sorted out his/your problem’
- b. Cada professor tratou de resolver o problema seu
 Each teacher treat-PAST-PL of solve-INF the problem his?/your
 ‘Every teacher sorted out his?/your problem’

Cardinaletti (1998) explains that the semantic interpretation of possessives is constrained in postnominal position: for instance, in Italian (22), “When occurring in postnominal position, possessives become restricted to human referents” (CARDINALETTI, 1998, p. 20):

	<i=John>	<i=frying pan>
(22) a. Il suo _i coperchio è molto pratico. The his/its lid is very practical	√	√
b. Il coperchio SUO _i è molto pratico.	√	*

Table 6: The interpretation of Italian postnominal ‘suo’

In (21), repeated below as (23), the interpretation of the possessive ‘seu’ seems to become restricted to 2nd person, when it occurs in postnominal position, as shown in Table 7:

Most informants have chosen the 3rd alternative for (21a) and the 2nd alternative for (21b). This was an initial test that should be verified with more speakers and other structures, such as the one in (24).

	<2 nd person>	<3 rd person>
(23) a. Cada professor tratou de resolver o seu problema	√ √	√ ?
b. Cada professor tratou de resolver o problema seu		

Table 7: The interpretation of BP postnominal ‘seu’ with [+animate] antecedent

Nonetheless, 3rd person reading is not excluded, in postnominal ‘seu’ (21b/23b), when the antecedent is [+animate]. If we then replace “cada professor” with an antecedent “cada numeral” that has [-animate] reference, as in (24b), 3rd person interpretation seems to be excluded in postnominal position.

	<2 nd person>	<3 rd person>
(24) a. Escreva ao lado de cada numeral o seu nome ²⁵ Write to-the side of each numeral the its/your name Write besides each numeral its/your name	√	√
b. Escreva ao lado de cada numeral o nome seu Write to-the side of each numeral the name your Write besides each numeral your name	√	*

Table 8: The interpretation of BP postnominal ‘seu’ with [-animate] antecedent

(24a) was taken from a school assignment. In this task, it is expected that students write in full the names of each numeral listed in the question, but it turns out that some students wrote their own names, repeated times, in the blank spaces. This shows that, in prenominal position, even when the antecedent has [-animate] reference, ‘seu’ allows

²⁵ Available in: <<http://www.bombounowa.com/imagens/agora-escreva-ao-lado-de-cada-numeral-o-seu-nome/>>. Accessed on: January, 1st 2017.

both 2nd and 3rd person readings. However, in postnominal position (24b), when the antecedent has [-animate] reference, only 2nd person reading is available.

In other words, 3rd person postnominal ‘seu’ searches for its antecedent, inside the IP, in a bound variable with [+animate] reference, while 2nd person postnominal ‘seu’ searches for its antecedent, outside the IP domain, in an operator located in the SaP [addressee].

In sum, according to Cardinaletti (1998, p. 21), in Italian, 3rd person pronoun ‘suo’, as a strong form, has its interpretation restricted to human referents in postnominal position. Likewise, in BP, the hypothesis to be verified is the following: 3rd person reading seems to be excluded in postnominal ‘seu’, when the antecedent is a bound variable with [-animate] reference. In this case, only 2nd person reading is available. In contrast, when the sentence has a bound variable with [+animate] reference, both 3rd and 2nd person readings are available in postnominal ‘seu’.

5. Conclusions

In standard BP, as well as in other Romance languages, possessives have uninterpretable number features, which are valued via nominal agreement. However, dialects of BP, especially the one spoken in Minas Gerais, show that postnominal 2nd person possessives do not have number agreement with the noun. In order to account for these facts, I analyzed the interpretability and valuation of number features in addition to the positions of the possessive in the DP-hierarchy.

With respect to the interpretability and valuation of features, I have claimed that, in this dialect, number features on 2nd person possessives are reanalyzed as being: (i) associated with the person and (ii) interpretable and valued. From the first postulation, ‘seu’ is expected to be the possessive for 2nd person singular, and ‘seus’ for 2nd person plural. From the second postulation, no nominal number agreement is triggered on the possessive, which means that there is neither “mismatch” of agreement with the noun, as one could presume, nor even agreement with something else, such as “possessor” or “addressee”.²⁶

²⁶ As such, the analysis carried out in this paper reformulates and prevails over other hypotheses proposed in previous stages of this research (PEREIRA, 2015, 2016a, 2016b). Concerning Pereira (2016b), it is *Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017*.

Furthermore, with respect to the DP-hierarchy, I have claimed that cardinals divide BP DPs into two domains, such that phrases preceding NumP are marked with the plural morpheme for nominal concord, whereas phrases following it are unmarked. The pronominal possessive precedes cardinals and must be marked, which makes it resistant to the reanalysis described above. In contrast, the postnominal possessive follows cardinals and must be unmarked, which makes it free for the reanalysis to occur.

To conclude, assuming that the 2nd person possessive has its number features reanalyzed explains why they are independent of the number for which the DP-internal phrases are inflected. Finally, assuming that the DP is divided into two domains, with regard to the DP-internal plural marking, explains why the postnominal position favors the mentioned reanalysis.

References

- BLÜHDORN, Hardarik. et al. Sintagmas nominais contáveis e não-contáveis no alemão e no português brasileiro. In: BATTAGLIA, Maria Helena; NOMURA, Masa (Org.). *Estudos linguísticos contrastivos em alemão e português*. São Paulo: Annablume, 2008. pp. 41-82.
- BRITO, Ana Maria. European Portuguese possessives and the structure of DP. *Cuadernos de Lingüística del I.U.I. Ortega Y Gasset*, v. 14, pp. 27-50, 2007.
- CARDINALETTI, Anna. On the deficient/strong opposition in possessive systems. In: ALEXIADOU, Artemis; WILDER, Chris (Org.). *Possessors, predicates and movement in the Determiner Phrase*. Amsterdam: Philadelphia: John Benjamins, 1998. pp. 17-53.
- CASTRO, Ana. Os possessivos em português europeu e português brasileiro: unidade e diversidade. In: XVI Encontro Nacional da Associação Portuguesa de Linguística, 2001, Lisboa. *Actas...* Lisboa: APL, 2001. pp. 599-613.
- CASTRO, Ana; Pratas, Fernanda. Capeverdean DP-internal number agreement: additional arguments for a distributed morphology approach. In: COSTA, João;

reasonable to assume, in terms of co-reference rather than agreement, that 2nd person possessives share the same number feature of their co-referent (which may be overt, as a vocative, in the SaP).

- FIGUEIREDO SILVA, Maria Cristina (Org.). *Studies on agreement*. Amsterdam; Philadelphia: John Benjamins, 2006. pp. 11-24.
- CERQUEIRA, Vicente. A forma genitiva ‘dele’ e a categoria de concordância (AGR) no português brasileiro. In: ROBERTS, Ian; KATO, Mary. *Português brasileiro: uma viagem diacrônica*. Campinas: UNICAMP, 1993. pp. 129-161.
- CHOMSKY, Noam. Derivation by phase. In: KENSTOWICZ, Michael (Ed.). *Ken Hale: a life in language*. Cambridge: The MIT Press, 2001. pp. 1-52.
- CINQUE, Guglielmo. Deriving Greenberg’s Universal 20 and its exceptions. *Linguistic Inquiry* 36 (3), pp. 315-332, 2005.
- COENE, Martine; D’HULST, Yves (Ed.). From NP to DP: The expression of possession in noun phrases. New York/ Amsterdam: John Benjamins, 2003.
- COSTA, João; FIGUEIREDO SILVA, Maria Cristina. Nominal and verbal agreement in Portuguese: an argument for distributed morphology. In: ____ (Org.). *Studies on agreement*. Amsterdam; Philadelphia: John Benjamins, 2006. pp. 25-46.
- DANON, Gabi. Agreement and DP-Internal Feature Distribution. *Syntax* 14 (4), pp. 297-317, 2011.
- GIUSTI, Giuliana. Is there a FocusP and TopP in the Noun Phrase structure? *University of Venice Working Papers in Linguistics*, v. 6, n. 2, p. 105-128, 1996.
- GIUSTI, Giuliana. The categorial status of quantified nominals. *Linguistische Berichte: Forschung, Information, Diskussion*. Opladen, p. 438-454, 1991.
- KATO, Mary. A complementaridade dos possessivos e das construções genitivas no português coloquial: réplica a Perini. *DELTA* 1 (1, 2), pp. 107-120, 1985.
- MAYNOR, Natalie. Battle of the pronouns: *Y’all* versus *you-guys*. *American Speech* 75(4), pp. 416-418, 2000.
- MÜLLER, Ana. *A gramática das formas possessivas no português do Brasil*. 1997. Dissertation (PhD in Linguistics) - UNICAMP, Campinas, 1997.
- NEVES, Maria. O pronome possessivo. In: _____. *Gramática de usos do português*. São Paulo: UNESP, 2000. pp. 471-489.

- NORRIS, Mark. 2014. *A theory of nominal concord*. PhD Dissertation, University of California, Santa Cruz.
- PEREIRA, Bruna Karla. O possessivo de 2ªPL no dialeto mineiro: DP e CP em análise. *Linguistic Studies 11: Contemporary approaches to Portuguese Linguistics*. Lisboa: Colibri/CLUNL, 2015 (In press).
- PEREIRA, Bruna Karla. A concordância em número com o possuidor: um estudo da sintaxe do DP. In: Martins, Marco et al. (Org.). *Estudos linguísticos: textos selecionados/Abralin-2013*. João Pessoa: Ideia, 2016a. pp. 146-163.
- PEREIRA, Bruna Karla. Second person possessives and allocutive agreement. In: Pilati, Eloisa (Org.). *Temas em teoria gerativa: homenagem a Lucia Lobato*. Curitiba: Blanche, 2016b. pp. 71-83.
- PEREIRA, Bruna Karla. The DP-internal distribution of the plural morpheme in Brazilian Portuguese. In: IOVTCHEVA, Snejana; STORME, Benjamin (Eds.). *Papers on Morphology. MIT Working Papers in Linguistics 81*, 2017 (In press).
- PERINI, Mário. O surgimento do sistema de possessivo coloquial: uma interpretação funcional. *DELTA 1* (1, 2), pp. 1-16, 1985.
- PESETSKY, David; TORREGO, Esther. The syntax of valuation and interpretability of features. In: KARIMI, Simin et al. (Org.). *Phrasal and clausal architecture*. Amsterdam: John Benjamins, 2007. pp. 262-294.
- SILVA, Giselle. Estertores da forma ‘seu’ na língua oral. In: SILVA, Giselle; SCHERRE, Marta (Org.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro, 1996. cap. 7, pp. 170 - 181.
- ZRIBI-HERTS, Anne. Les syntagmes nominaux possessifs en français moderne: syntaxe et morphologie. In: GUÉRON, Jacqueline; ZRIBI-HERTS, Anne. (Org.). *La grammaire de la possession*. Paris: Université Paris X – Nanterre, 1998. pp. 129 - 166.

On the Syntax of Subjects in Brazilian Portuguese: Using the 'Split' Pronominal System as the Basis for an Alternative Analysis

Eloisa Pilati^a

Rozana R. Naves^b

Heloisa M. Lima Salles^c

ABSTRACT

The goal of this article is to investigate the syntax of subjects in Brazilian Portuguese (BP), developing a unified analysis for referential and non-referential subjects, on the one hand, and for the so-called topic-subject constructions, VS word order clauses, third person subjects with a generic interpretation without the indefinite pronoun 'se', existential and meteorological predicates, on the other hand. By hypothesis, in these contexts, the subject is filled by a (null or overt) DP with (spatial/ temporal) locative interpretation. It is claimed that the relevant facts can be accounted for if the pronominal/inflectional system of BP is analysed as displaying a split in two subsystems: one, which is constituted by the first and second person, being inherently definite/ referential, and the other, which is constituted by the third person, being underspecified for the definite/referential feature. Our proposal is that the split in the pronominal/inflectional system is the key to account for the distribution of third person (spatial/ temporal) locative subjects, either lexical or null in BP, both in matrix and embedded clauses. In the discussion, we present previous analyses on the syntax of subjects of BP, pointing out that the absence of consensus is due to their theoretical implementation (which is often complementary), although their contribution converge with respect to the properties of the third person, as well as to the role of discourse orientation, allowing for the presently proposed unification.

KEYWORDS: null and overt subjects; Brazilian Portuguese; split pronominal/inflectional system

RESUMO

O objetivo deste artigo é investigar a sintaxe do Português Brasileiro (PB), desenvolvendo uma proposta unificada para os casos de orações com sujeitos referenciais e não-referenciais, por um lado, e de orações

^a Universidade de Brasília, eloisapilati@gmail.com

^b Universidade de Brasília, rozanarn@hotmail.com

^c Universidade de Brasília, heloisasalles@gmail.com

We would like to thank the audience of the GTTG 2015 Meeting, for the discussions, with special reference to Anders Holmberg, the invited speaker, for his fine contribution, as well as to Fabio Bonfim Duarte for his careful work leading the organizing committee. We are also grateful to Shigeru Miyagawa, David Pesetsky, C-T. Huang, Mary Kato, Marcello Modesto, Ian Roberts for their interest in our work and for their insightful observations. All errors are our own. This research is supported by CNPq, n°. 460416/2014-6.

do tipo tópico-sujeito, com ordem VS, com sujeito de terceira pessoa e interpretação genérica sem o pronome ‘se’, com verbos existenciais e meteorológicos, por outro lado. Por hipótese, nesses contextos, o sujeito é preenchido por um DP/ pronome (nulo ou manifesto) com interpretação locativa (espacial/temporal). Argumentamos que é possível entender os fatos relevantes se analisarmos o sistema pronominal/flexional do PB como sendo cindido em dois subsistemas: um, composto pela primeira e segunda pessoas, que são inerentemente definidas/referenciais; e o outro, constituído pela terceira pessoa, que é subespecificada para o traço definido/referencial. A proposta que defendemos é que a divisão desse sistema pronominal/flexional constitui a chave para explicar a distribuição de sujeitos locativos (espaciais/temporais) manifestos e de sujeitos nulos de terceira pessoa em PB, tanto em oração matriz como em oração encaixada. Na discussão, apresentamos uma retrospectiva das análises mais representativas sobre a sintaxe do sujeito (manifesto e nulo) no PB, ressaltando que a falta de consenso decorre dos recortes teóricos adotados (os quais são muitas vezes complementares), embora as contribuições sejam convergentes em relação às propriedades da terceira pessoa e ao papel da orientação para o discurso, o que respalda a unificação presentemente proposta.

PALAVRAS-CHAVE: sujeitos nulos e manifestos; Português Brasileiro; sistema pronominal/flexional cindido

Introduction

The syntax of (lexical and null) subjects in Brazilian Portuguese (henceforth BP) has been carefully investigated, back to Tarallo’s (1983) and Duarte’s (1993) pioneer studies. In this article, we provide a critical overview of the main analyses, pointing out that the (apparent) absence of consensus is due to their theoretical implementation (which is often complementary), with implications for the identification of the relevant data. Our proposal is that the contribution of these analyses converges with respect to the properties of the third person, as well as to the role of discourse orientation, allowing for a unified analysis of a number of constructions apparently distinct.

As widely shown, the following facts characterize the BP grammar (as opposed to European Portuguese (EP) and other canonic null subject languages (NSL)):

- (i) BP does not display third person null subjects in ‘out of the blue’ contexts, contrary to what has been recurrently reported for NSL, although third person null subjects

with generic interpretation are found in contexts without the pronoun 'se', as originally noted in Galves (1996, 2001), a deictic reference being required, as pointed out in Pilati and Naves (2013) and Naves, Pilati and Salles (2013), and Pilati, Naves and Salles (2015, 2017a/b);

- (ii) With referential subjects, BP displays a tendency for an overt realization of the subject position in the full paradigm, as a correlate of the loss/reduction of inflectional morphology on the verb, due to pronominal reanalysis (by which 'você', a honorific pronoun triggering third person agreement, is grammaticalized as a second person personal pronoun), as originally noted in Tarallo's (1983) and Duarte's (1993, 1995) seminal works; although overt subjects are more recurrently found with first and second person subjects, a tendency for subject drop was noted in [-human] third person contexts (cf. Cyrino, Duarte and Kato, 2000);
- (iii) BP displays the so-called topic-subject constructions, in which locative/possessive phrases in preverbal position trigger agreement on the verb, as originally noted in Pontes (1986), a phenomenon further including constructions with existential and meteorological predicates, as pointed out in Kato and Duarte (2008a, 2008b);
- (iv) BP displays VS word order as a type of locative inversion, in which a deictic anchoring is required, being provided by a (spatial/temporal) locative – whether overt or null, as proposed in Pilati (2006).

We will argue that the facts in (i), (ii), (iii) and (iv) are crucially linked to constructions with third person (overt/null) subjects, under a requirement on discourse orientation, which is satisfied by a (spatial/temporal) deictic locative in the structure of the predicate. This condition does not apply to first and second person subjects, as they are inherently deictic (and discourse oriented), their overt or null distribution being determined by independent properties, as can be inferred by the facts described in (ii).

Descriptively, the facts in (i) to (iv) demonstrate that the pronominal/inflectional properties of the subject in BP are not uniform, constituting empirical evidence for discarding the idea of a single property affecting the full paradigm. In particular, we will propose that a unified analysis for the above-mentioned facts is possible, if we take the pronominal/inflectional system of BP to be split into two subsystems: one constituted by first and second person, which are inherently marked as definite/referential; another constituted by the third person, which is underspecified for

the definite/referential feature, an idea that stems on Rabelo’s (2010) analysis of null subjects in BP.

Our proposal is that a split in the pronominal/inflectional system involving the above-mentioned features constitutes the key for accounting not only for the distribution of (spatial/temporal) locative third person subjects (whether overt or null), but also for the condition on the distribution of VS word order. As already mentioned, the proposal is based on two previous approaches: on the one hand, we assume Pilati’s (2006) analysis of VS word order in BP as a type of locative inversion, in which a requirement on discourse anchoring/ orientation is met through merge of a (spatial/temporal) locative (DP/*pro_{LOC}*) in subject position, while the logical subject remains *in situ*, in the domain of *v*P (or in a focus position in the middle field of the clause); on the other hand, we adopt Pilati, Naves and Salles’ (2015, 2017a/b) analyses, in which the properties of third person, as opposed to first and second person, subjects provide an account clustering different phenomena, namely: topic-subject constructions and meteorological and existential constructions with a preverbal locative DP triggering agreement on the verb, as well as third person null subjects constructions with a generic interpretation on the subject (without the indefinite pronoun ‘se’), in which a requirement on discourse anchoring/orientation is obtained either pragmatically or through insertion of a (spatial/temporal) locative DP in subject position (exactly as in VS constructions).

The contribution of the present analysis is then twofold: firstly, we will provide a formal account for the syntactic properties of the above-mentioned facts concerning BP subjects; secondly, we will provide a novel setting for the previous analyses, in which their fine insights and their theoretical complementarity will be acknowledged.

The article is organized as follows. The first section provides the relevant data, as well as their bibliographic source, followed by an overview of the literature on the syntactic and semantic properties of BP subjects, focusing on the main proposals, which are divided in three lines of research: the parametric change hypothesis, the BP as a topic-prominent language hypothesis, and the BP as a partial null-subject language hypothesis.¹ The next section discusses the main data, showing the importance of

¹ An anonymous reviewer notes that the hypotheses as presented in the structure of the paper describe parametric analyses, thus rendering inappropriate that the reference to parameter change in the title of the section be restricted to the first one. We are aware that all the proposals, being stated within the Principles and Parameters framework, provide an account for language change in terms of parametric change – including the one presently given. However, by using the label ‘parametric change hypothesis’ (only) in

viewing BP's pronominal/inflectional paradigm as a system split into two subsystems – one, containing the first and second person pronouns/inflections, and the other, containing the third person pronoun/inflection; viewing the paradigm in this fashion allows us to advance the understanding into the licensing of null subjects in this language. The following section presents a theoretical proposal for explaining the BP data, and is followed by the final considerations.

1. An overview of the studies examining overt and null subjects in BP

1.1. The relevant data

It is worth recognizing that, since the 1980's, a number of studies have investigated the syntactic and semantic properties of null subjects and overt subjects in BP. As already mentioned, these studies have demonstrated that there is a growing tendency in BP to use overt subjects, notwithstanding those contexts in which null subjects are the only admissible option (such as in existential constructions, for example). Despite the sheer number of studies dedicated to overt and null subjects in BP, there is still no single analysis which enjoys widespread acceptance with respect to the whole set of data involving the filling of the subject position – a conclusion we attribute to the fact that the previous works dealt with different range of empirical data. Somewhat surprisingly, though, there is relatively little disagreement over the status of the relevant facts themselves, which can be illustrated as follows:²

- A. Third person singular null subjects are disallowed in declarative clauses, when uttered in an 'out of the blue' context. Third person plural null subjects (with a

the first section we want to remark the seminal and pioneer contribution of this approach to the innovative properties of BP, as opposed to European Portuguese (EP). In the subsequent analyses, we adopt the label as provided by the authors in their analyses. This comment is also a reply to a suggestion by another anonymous reviewer according to which sections 1.2 to 1.4 should be collapsed as the three of them describe a parametric change undergone by BP taking into consideration the Null Subject Parameter. We would like to maintain the present structure of the presentation, because it is intended to provide the (chronological) evolution of the scientific thought.

² BP data are not glossed, but simply given a word-to-word translation. We believe that this is enough to establish the required understanding of the phenomena. In cases in which the data are taken from real speech, we will give a free translation to English.

generic/ indefinite interpretation), as well as first and second person null subjects, are allowed in this environment.³

- (1) a. $\emptyset_{1SG/PL}$ Comi_{1SG}/Comemos_{1PL} o chocolate./ Ate_{1SG/1PL} the chocolate.
 b. \emptyset_{3PL} Comeram_{3PL} o chocolate./ Ate_{3PL} the chocolate.
 c. * \emptyset_{3SG} Comeu_{3SG} o chocolate./ *Comeu_{3PL} the chocolate.

B. Independent of the person and number values borne by the null subject, null subjects are licensed in question and answer contexts, as well as under coordination.

- (2) a. A: Maria/ você/ o João/ as crianças comprou(aram) o bolo?
 ‘Did Maria/ you/ John/ the children buy the cake?’
 B: Comprei/ Comprou/ Compraram.
 ‘I/He/João/They did.’

³ An anonymous reviewer noted that “nenhuma língua licencia um sujeito nulo de terceira pessoa em contextos out of the blue” [our translation: no language licenses a third person null subject in ‘out of the blue’ contexts], pointing out the well-known example *Parla italiano*, which is recurrently used to illustrate canonic null subject languages (NSL), further adding that this sentence “só é gramatical se o sujeito de terceira pessoa tem necessariamente um antecedente identificado pelo contexto precedente ou pragmaticamente)” [our translation: is only grammatical if the third person subject necessarily has an antecedent which is identified in the context or pragmatically], an observation that also applies to the third person pronoun in *Lui parla italiano*, as the reviewer concludes. What we would like to say is that an antecedent/referent is indeed required for a referential interpretation, however utterances produced in ‘out of the blue’ do not imply non-referential subjects, rather what they essentially imply is wide focus. Hence, the example, as presently given, recalls the well-known example in most theoretical approaches to canonic NSL, back to Rizzi (1982), among many others. Interestingly a minimal pair, which is quoted in Pilati, Naves and Salles (2015), illustrates how speakers of BP e EP react to the same question, depending on how they interpret the reference of third person inflection on the verb (the data was tested with speakers of both dialects): while the EP reply indicates that the null subject is interpreted as referential, the BP one shows that the sentence is interpreted as existential. We will return to these facts.

- (i) A: Tem pão?/ Have_{3s} bread? (PILATI, NAVES and SALLES, 2015)
 PE: Tenho, mas não há ainda./ Have_{1s}, but there isn’t yet
 PB: Tem./ Have_{3s[=non-referential]} [= There is bread]

Besides, it is worth mentioning that the presence of an overt pronoun in canonic NSL such as Italian should give rise to a marked interpretation, thus disallowing the intended generalization as postulated by the reviewer.

- b. Você tem que sair (...) Tudo isso você tem que fazer, (cv) não pode parar assim. Tu não morreu, pô! (cv) Aposentou, mas tu ‘tá vivo, pô!
(DUARTE, 2003, p. 10)

‘You must leave (...) All this you have to do, (cv) [you] can’t stop like that. You haven’t died! (cv) [you] Retired, but you are alive!’

- C. Third person, referential null subjects are allowed in embedded clauses when the subject of the matrix clause controls the null subject – a context in which null subjects freely vary with overt subjects:

- (3) Maria_i disse que ela/ \emptyset _i come chocolate quando ela/ \emptyset _i está chateada./ *Mary said that she/ \emptyset _i eats when she/ \emptyset _i is annoyed.*

- D. There are also contexts in which a subject is bound by a quantificational antecedent. In such contexts, the third person null subject is the only option available:

- (4) a. Ninguém_i aqui acha que *ele(a)/ \emptyset _i vai perder./ *Nobody_i here believes that *(s)he/ \emptyset _i will loose.*
b. Quem_i acha que *ele(a)/ \emptyset _i vai perder?/ *Who believes that *(s)he/ \emptyset _i will loose.*

To the facts A to D above, we add the empirical observations in E, below, which are in the present work taken as additional syntactic contexts involving the filling of the (third person) subject position in BP – the unified account of the syntactic and semantic properties of the subject position in the relevant contexts being the strong point of the present analysis:

- E. Postverbal third person subjects are only possible in sentences denoting an event that has recently taken place or that coincides with the moment in which the sentence is uttered (cf. (5a-b)); by hypothesis, in such sentences, the preverbal position is filled by a (spatial/ temporal) locative, which is either null or overt, the

latter being filled either by a locative DP (triggering agreement) or by a pronoun/adverb (*aqui/ali/lá* ‘here’/ ‘there’ (proximal)/ ‘there’ (distal)) (also found in Italian, cf. PINTO, 1997). A type of alternation involving the subject position is found in topic-subject constructions (cf. (6)), in constructions with existential verbs (cf. (7)), with meteorological verbs (cf. (8)), and with subjects interpreted as arbitrary/indefinite, in which the pronoun ‘se’ is absent (cf. (9)).

- (5) a. Morreu Pavarotti. (acabou de acontecer)/ *Died Pavarotti. (it has just happened)*
 b. Ergue o braço o juiz. / *Raises his hand the referee.*
 c. Ali entrou Dante./ *There entered Dante.*
 (examples adapted from PILATI, 2006)
- (6) a. Cabe muita roupa nessa(s) mala(s)./ *Fits lots of clothes in these suitcases.*
 b. Essa(s) mala(s) cabe(m) muita roupa./ *This(These) suitcase(s) fit_{3SG(PL)} lots of clothes.*
 c. Furou o pneu do carro./ *Punched_{3SG(PL)} the car tire.*
 d. O(s) carro(s) furou/furaram o pneu./ *The car(s) tire(s) punched_{3SG(PL)}.*⁴
 (examples adapted from PONTES, 1986)
- (7) a. Tem monumentos lindos em Brasília/ lá/ *Has monuments beautiful in Brasília/there.*
 b. Brasília/ Lá tem monumentos lindos. / *Brasília/There has beautiful monuments.*
 (examples adapted from PILATI, NAVES and SALLES, 2015)
- (8) a. Chove pouco nessa(s) cidade(s)/ aqui./ *Rains little in this(these) town(s)/ here.*
 b. Essa(s) cidade(s) chove(m) pouco./ *This/ These town(s) rain_{3SG(PL)} a little bit.*
 c. Aqui chove pouco./ *Here rains a little bit.*

⁴ In a wide sense, we take the possessor phrase to denote a locative interpretation (cf. FREEZE 1992).
Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017.

(examples adapted from PILATI, NAVES and SALLES, 2015)

- (9) a. Vende CD nessa loja./ *Sell CD in this shop.*
 b. Essa(s) loja(s) vende(m) CD./ *Aqui vende CD./ This(These) shop(s) sell_{3SG(PL)} CD./ Here sells CD.*

(examples adapted from PILATI, NAVES and SALLES, 2015)

Summarizing: as is well known, there is a tendency to use overt subjects rather than null subjects in BP. Third person singular null subjects are ungrammatical in ‘out of the blue’ contexts, though they are licensed in specific contexts such as when they are in an embedded clause and controlled by the matrix subject. Third person (spatial/temporal) locative subjects are found in constructions such as (5) through (9); in these contexts, the verb exhibits plural agreement with non-pronominal plural locative subjects. Finally, the subject is obligatorily null in those contexts in which it functions as a variable bound by a quantificational antecedent.

Given the relative consensus over the status of the data presented above, coupled with the lack of consensus over how to best analyze these data, our intention in the present article is to argue that a split in the pronominal/inflectional system of BP is the key property characterizing the syntactic change undergone by BP, which in Minimalist terms amounts to parametric variation in the formal features of functional categories, as will be argued below.

1.2. The parametric change hypothesis

As is well known, the sociolinguistic research that was conducted originally by Fernando Tarallo and his collaborators (cf. TARALLO 1983 and subsequent works) has brought significant evidence for the hypothesis that the syntax of BP is undergoing a change in its pronominal system, which encompasses both pronominal subjects and pronominal objects – the latter, concerning the syntax of BP clitics. With regard to subjects, Tarallo’s research inaugurated a line of thinking based on an idea originally formulated in Chomsky (1981, 1986) and Rizzi (1982), according to which there is a direct relation between the tendency found in BP toward ceasing to be a null subject

language and the impoverishment in the verbal inflectional system as a result of a reanalysis in the pronominal system.

Duarte (1993, 1995), through an examination of dialogues taken from theatrical plays from the 19th and 20th centuries, and from synchronic data from speech, respectively, demonstrates that BP is indeed exhibiting the tendency argued for by Tarallo (1983). As the Table 1 shows, the relative frequency of overt referential subjects in BP has risen over the past two centuries:

(10)

Period	Percentage of null subjects
1845	74%
1882	67%
1937	50%
1955	20%
1975	23%
1992	25%

Table 1: Occurrence of null subjects in BP (adapted from DUARTE 1993: 112):

Duarte’s (1993) seminal work shows that a series of changes has taken place in the system of verbal agreement, giving rise to an impoverishment in the verbal’s inflectional morphology due to a reanalysis of the pronominal system – specifically, with the grammaticalization of *você* ‘you’ as a second person singular pronoun and of *a gente* ‘the people’ as a first person plural pronoun, both of which inducing third person singular agreement on the verb. The development of the inflectional paradigms is illustrated on Table 2:

(11)

Person	Number	Paradigm 1	Paradigm 2	Paradigm 3
1	Singular	Cant- o	Canto	Canto
2 direct	Singular	Canta- s	_____	_____
2 indirect	Singular	Canta- 0	Canta- 0	Canta- 0
3	Singular	Canta- 0	Canta- 0	Canta- 0
1	Plural	Canta- mos	Canta- mos	Canta- 0
2 direct	Plural	Canta- is	_____	_____
2 indirect	Plural	Canta- m	Canta- m	Canta- m
3	Plural	Canta- m	Canta- m	Canta- m

Table 2: The development of BP inflectional morphology (adapted from DUARTE, 1995)

Both Tarallo (1983) and Duarte (1993; 1995) contend that the facts above point to the conclusion that BP is undergoing a parametric change affecting the syntax of subjects. In particular, they argue that BP is ceasing to be a null subject language, starting to display the properties of a non null subject language.⁵ In this respect, Duarte (2003, p. 115) writes:⁶

⁵This approach followed in the footsteps of the theoretical framework inaugurated by Tarallo and Kato, which came to be known as Parametric Sociolinguistics because it made use of quantitative data drawn from instances of real speech, on the one hand, while offering analyses couched within the Principles and Parameters framework, as found in Chomsky (1981, 1986), on the other hand.

⁶The translation is ours. The original is: *Diferentes trabalhos com base em dados de língua oral (Lira 1982, Duarte 1995, 2000) e na escrita de cartas (Paredes Silva 1988) ou de peças de teatro popular (Duarte 1993) têm mostrado que o português do Brasil apresenta índices de preenchimento do sujeito pronominal bem superiores aos apresentados pelas chamadas línguas românicas de sujeito nulo, como o espanhol, o italiano e a variedade europeia do português. De modo geral, o fenômeno tem sido associado à simplificação ocorrida em nossos paradigmas flexionais verbais, que contam com a mesma forma para a segunda e terceira pessoas do singular e, com frequência cada vez maior, para a primeira do plural, graças ao crescente uso da forma “a gente” em detrimento de “nós” (Duarte, 2003, p.115).*

Different studies based on data from the spoken language (Lira 1982, Duarte 1995, 2000) and collected from letters (Paredes Silva 1988) and popular theatrical plays (Duarte 1995) have shown that Brazilian Portuguese exhibits rates of overt subject use much higher than what is exhibited by null subject Romance languages such as Spanish, Italian, and the European variety of Portuguese. Typically, this phenomenon has been associated with the simplification that has occurred in our verbal inflectional paradigms, which use the same form for second and third person in the singular and, with increasing frequency, for first person in the plural, thanks to the rise in the use of the form *a gente* 'the people' as a substitute of the form *nós* 'we'.

At this point, the main contribution of the sociolinguistic studies is that they revealed, through quantitative analysis, that BP has begun opting for the use of overt subjects over null subjects. Adopting the Principles and Parameters framework, and taking into consideration the effect of the pronominal system's reanalysis on the verbal inflectional paradigm, it was possible to analyze these quantitative results in terms of a parametric change, according to which BP develops towards becoming a non-null subject language. Seen from a minimalist perspective (CHOMSKY, 1995), in which parametric change involves the features of functional categories, the null subject parameter can be defined in terms of the operation AGREE and the formal features on T° . Since T° is involved, the connection between the null subject parameter and the richness of the inflectional morphology on the verb is straightforwardly captured.

This idea is given a formalization in Galves' (2001) analysis, in terms of the properties a Person head (implying that Agr is not present), which captures the fact that the third person morpheme in BP is incapable of assigning reference to a null subject. When only the third person inflection is used, it receives a generic or an indefinite interpretation (for example, as in (9) above). This behavior is different in EP, in which a null third person is a referential subject (just like in typical pro-drop languages) and the generic interpretation is obtained only when there is an overt pronoun 'se'.

The results obtained in previous works (particularly in DUARTE, 1993) are refined in Duarte, Mourão and Santos (2012), who show that the behavior of first, second, as opposed to third person subjects, is not uniform, as the former displays a higher tendency for being overt. "If, on one hand, it is true that the third person is affected by the change, on the other hand, it is also true that this change occurs in a

different way” (Duarte, Mourão and Santos, 2012, p. 25).⁷ According to these authors, third person overt subjects are inhibited in the presence of the [-human] feature on the antecedent. Conversely, the results also point to a correlation between the occurrence of (third person) overt subjects and the semantic features [+specific] e [+human] on the referent. These facts are then analysed in terms of a referential hierarchy, as proposed in Cyrino, Duarte and Kato (2000).

Within the theoretical framework presently adopted, the referential hierarchy does not have a proper formulation, as it cannot be translated in terms of formal features, in spite of being indeed revealing. Instead, we will propose an alternative analysis in which the formal features of the functional categories involved in subject licensing are taken into consideration. Interestingly, Duarte, Mourão and Santos’ (2012) observation that the change does not affect the person paradigm uniformly corroborates our hypothesis that the pronominal/inflectional system of BP manifests a split. In our approach, the absence of a referential feature on the relevant head is a necessary condition for the occurrence of third person non-argumental/locative subjects, an option that is not available for first and second person, as they are inherently referential.⁸

1.3. BP as a topic-prominent/discourse-oriented language hypothesis

Building on Negrão’s earlier work (NEGRÃO, 1999), Negrão and Viotti (2000) argue that the impoverishment in BP’s inflectional morphology is not the cause of the change in the licensing of null subjects in this language, being it quite the opposite. The impoverishment is a consequence of the fact that BP is becoming a discourse-oriented language – a typological pattern that characterizes languages such as Chinese, in which the topic position, situated in the periphery of the clause, determines the organization of the clause at the expense of the subject (and object) position.

For Negrão and Viotti (2000), it is the specialization of forms, rather than the loss of null pronouns, which explains the asymmetries in the distribution of null and overt

⁷ Our translation of Duarte, Mourão e Santos (2012, p. 25): *Se, por um lado, é inegável que a 3ª pessoa é afetada pela mudança, por outro é incontestável que ela se dá de forma diferente.*

⁸ We do not ignore the fact that generic/indefinite interpretation is available with first, second and third person overt pronouns in BP. We take this fact to be independent of the properties affecting the third person, as opposed to first and second person pronominal/inflectional system, as presently postulated. In particular, we believe that, in the latter case, the relevant property displays a contrastive distribution on the relevant functional head, contrary to the former case, in which the pronominal items are uniformly affected, allowing for both a definite/referential and a generic/indefinite interpretation.

subjects in BP. That is, it is not the case that BP is becoming a non-null subject language; rather, overt and null subjects in BP have simply undergone a specialization, whereby null subjects in this language can be interpreted as bound variables, and overt subjects cannot.

The interpretation of sentences such as those found in (12) constitutes one of the arguments put forward by the authors in support of their hypothesis:

- (12) a. Só o Maluf₁ acha que *ec*₁ vai ganhar as eleições. (os outros candidatos não acham)/ *Only Maluf believes that ec₁ will win the elections. (the other candidates do not believe so)*
- b. Só o Maluf₁ acha que ele₁ vai ganhar as eleições. (ninguém mais acha que ele irá ganhar)/ *Only Maluf believes that he will win the elections. (nobody else believes that he will win)*

The interpretation of (12a) is: “The only *x* such that *x* thinks that *x* will win the elections is Maluf”, or, in other words: Maluf is the only candidate who thinks that he, himself, will win the elections; none of the other candidates think they will win the elections. The interpretation of (12b), on the other hand, is that Maluf is the only person who thinks that *Maluf* will win the elections; none of the other candidates think that Maluf will win the elections. Thus, the empty category obligatorily functions as a bound variable, whereas the overt pronoun must be construed as referential.

Following Negrão (1999), Negrão and Viotti (2000) apply Huang’s (1989) proposal for Chinese to BP, arguing that BP is a type of discourse-oriented language; as such, it exhibits the properties of languages of this type. In particular, in discourse-oriented languages, the subject-predicate relation does not form the basic relation around which the sentence is structured. Accordingly, the predication relation around which the sentence is formed does not necessarily involve the nominal expression functioning as the subject, and occupying SpecTP, and the verbal expression functioning as the predicate (as in a predicate relation formally marked through agreement); rather, the predication relation around which the sentence is formed may

involve a constituent situated in the specifier of a projection in the left periphery of the clause and the rest of the sentence.

Modesto (2000, and subsequent works) develops Negrão's (1999) hypothesis that BP is a topic-prominent language, like Chinese. Based on an examination of null subjects in embedded clauses in BP, Finnish, and Chinese, Modesto (2008) argues that verbal agreement plays no role in the licensing and identification of null embedded subjects. According to Modesto, in these languages, the subject of matrix clauses is located in a non-marked topic position, which enables the null subject of embedded clauses to be licensed and identified. This derivational process of licensing and identifying null subjects is what, according to the author, defines these languages as discourse-oriented. Adapted examples from the author are given below (*op. cit.*, p. 382):

- (13) [O amigo do Feco₂]₁ disse que e_{1/*2/*3} ganhou a competição. [BP]
 [Feco₂'s friend]₁ said that e_{1/*2/*3} won the competition.
- (14) [Veljeni₂ vaimo]₁ oli niin iloinen, ettei e_{1/*2/*3} voinut nukkua. [Finnish]
 brother.gen spouse was so happy that not could sleep
 'My brother's wife was so happy that she could not sleep.'
- (15) [Zhangsan₂ fangwen de ren]₁ xiwang e_{1/*2/*3} neg ying. [Chinese]
 Zhangsan visit DE person hope can win
 'The person that Zhangsan visited expected that she could win.'

The author shows that embedded subjects always refer to the matrix subject in sentences such as these and contends that the embedded subjects cannot be interpreted as coreferential with another expression in the sentence including the object and that it cannot be interpreted as deictic. This leads Modesto to propose that matrix subjects in these languages are situated in an A'-position, and that the null subject is therefore a variable.

Two observations regarding the hypothesis put forward by Negrão (1999) and Negrão and Viotti (2000) are in order. First, these studies have the merit of having observed that the third person overt pronoun and the third person null pronoun have

undergone a specialization whereby the former must function as a referential pronoun, and the latter, as a bound variable – a novel observation. Second, the authors’ proposal refers to third person pronouns without making additional reference to number. This proves problematic, given that third person plural null pronouns do not exhibit the bound variable reading, contrary to what the authors’ proposal would lead us to expect (cf. (16b)):

- (16) a. Os meus carros₁, e_{1,2} bateram e eu nem vi./ O meu carro, e_{1,*2} bateu e eu nem vi.
*My cars, e_{1,2} crashed and I did not even see it/ My car e_{1,*2} crashed and I did not even see it*
- b. Os meus carros₁, eles_{1,2} bateram e eu nem vi./ O meu carro₁, ele₁ bateu e eu nem vi.
*My cars, they_{1,2} crashed and I did not even see it/ My car they_{1,*2} crashed and I did not even see it*

Another relevant point is that EP, like BP, is considered a topic-prominent language, as defended in Raposo (1987), and Kato and Raposo (2005). Therefore, the fact that BP is a topic-prominent language does not, by itself, explain the syntactic change undergone by BP (as opposed to EP).

In our view, Modesto (2000), and subsequent studies, contributed to the understanding of null subjects in BP in demonstrating that the licensing of third person null subjects in this language may be compared to that of Chinese (following NEGRÃO, 1999). Nonetheless, even if the author is correct with respect to the syntactic properties of third person pronouns in BP, the posited behavior of this pronoun does not explain the properties of the entire pronominal system in this language. By invoking discourse orientation as a crucial condition in overt subject realization, it is inferred that all persons display identical properties, being uniformly affected (see also (RABELO, 2010). As already mentioned, in present terms, the third person is distinct from first and second person in the expression of the referential properties in the pronominal/inflectional system.

1.4. BP as a Partial Null-Subject Language hypothesis

The hypothesis of analyzing BP as a partial null-Subject language (PNSL) is originally investigated in Rodrigues' (2004) study, in which it is shown that third person, as opposed to first and second person, null subjects in BP can be compared to Finish (following the lead of Holmberg and Nikkane (2002)). According to Rodrigues, in both languages, agreement with possessives and third person subjects is morphologically weak allowing non-emphatic overt pronouns to occur in subject position. In particular, differently from NSL, in which *pro* is assumed to be the inflectional morpheme on the verb, "in BP and Finnish 3rdP [3rd person] verbal agreement morphemes (Agr) underwent a degradation (...) and, as a consequence, were reanalyzed as part of the verb", the EPP feature being checked by the D feature of the reanalyzed Agr morpheme under V adjunction to T (RODRIGUES 2004, p. 353-354).

Coupled with the hypothesis that BP and Finnish are not null subject languages, the author concludes that in languages of this type, the EPP feature of the relevant head licensing the subject is checked by a strategy other than the one involving the specifier (DP) subject/ head (Agr) relation. As a consequence of the above-mentioned morphological change: "1stP [1st person] null subjects are formed via movement to a topic position and deleted later, in accordance to a topic-deletion analysis, [while] embedded 3rdP null subjects are formed via A-movement of a DP inserted as an argument of an embedded finite clause to a non-expected A-movement outside a Case domain, before Case is checked. Being the residue of A-movement, third person null subjects in these languages are 'salient copies of their antecedents'" (RODRIGUES 2004, p. 1) – and not null pronouns.

Again the contrast between third person, as opposed to first and second person subjects, is given a formal approach in the account of the syntax of subjects in BP, as proposed in Rodrigues (2004). Although we will not assume this formal proposal, we will retain the distinctive properties associated to PNSL (as opposed to NSL).

More recently, Duarte and Kato (2008a, 2008b) also propose that BP is a partial null-subject language. However, in their analysis, the label PNSL refers to a system that displays overt referential subjects, as well as non referential null subjects, which are properties that do not coincide with those described in Rodrigues (2004) and

Holmberg’s (2005) previous work.⁹ In particular, Duarte and Kato’s (2008a, 2008b) analysis is based on Cardinaletti’s (2004) proposal of distinguishing two projections for subjects: the specifier of SubjP, responsible for checking the feature “subject of predication”, and the specifier of TP, responsible for checking both the Case feature and phi-features. Accordingly, the null subject parameter distinguishes languages which possess a pronominal Agr(eement) (these being the canonic NSL, in which personal pronouns are grammaticalized as verbal inflection) and languages which possess weak pronouns and subject clitics (these being non-null subject languages). The relevant feature of BP is that third person subjects can be realized either as a pronominal Agr or as a weak pronoun, generic reference being obtained only through the use of pronominal Agr. In turn SubjP is projected for strong pronouns and locative subjects (cf. (17a)), but it is not projected for weak pronouns (as in ‘*cê* in (17b)), which occurs in SpecTP:

- (17) a. Londres tem prédios lindos./ *London has beautiful buildings.*
 b. ‘*Cê* tem prédios lindos em Londres./ *You have beautiful buildings in London.*

In Holmberg’s (2010) study, which develops his previous analyses on PNSL, further assuming BP alignment with Finnish, the properties of PNSL are directly related to the fact that null subjects in these languages are permitted with first and second person subjects, whether singular or plural, while third person null subjects are permitted only in embedded clauses. The relevant contexts involving null subjects are the following, with examples from BP: (i) the subject is non-thematic (cf. (18)); (ii) the subject is a generic pronoun corresponding to English *one* (cf. (19)); (iii) the subject is controlled by an argument in the clause that immediately dominates the one containing the (null) subject (cf. (20)):

(18) Está chovendo/ *(It) is raining*

(19) É assim que faz o doce/ *(It) is this way that (you) make the dessert*

⁹ We are grateful to an anonymous reviewer for requiring a clarification with respect to the distinct concepts of Partial Null Subject Languages, which proves to be crucial for the present discussion.

- (20) O João_i disse que (ele_i) tinha comprado uma casa/ *John Said that (he) had bought a house*

In Holmberg's (2010) proposal, the difference between full null-subject languages (NSL) and partial null-subject languages (PNSL) is determined in the properties of T: in NSL, T bears an uninterpretable D(efinite)-feature; in PNSL, it lacks an uninterpretable D-feature. Thus, T is defective in PNSL.¹⁰ According to the author, definite null subjects in PNSL are in SpecTP and check the EPP feature, while generic null subjects are incorporated into T and do not check the EPP feature. As such, for a sentence such as (21a) below, a definite reading will arise provided that the embedded null subject is licensed by the DP in the matrix subject position – the null subject, being realized in SpecTP, where it checks the EPP. For a sentence such as (21b), however, in which the null subject is assigned a generic interpretation, it is assumed that the null subject (*pro*) incorporates into T, the EPP being checked by the PP, *na praia* 'in the beach':

- (21) a. João me contou que *pro* vende cachorro quente na praia./ *John told me that pro sells hot-dogs in the beach.*
- b. João me contou que na praia vende cachorro quente./ *John told me that in the beach [one] sells hot-dogs.*

The analysis proposed by Duarte and Kato (2008a, 2008b) is challenged by Costa (2010). Comparing BP and EP, Costa argues that many of the constructions that are quoted as characterizing the syntax of subjects in BP are also found in EP, except for the ones with a left-dislocated subject doubled by an overt pronoun, and the ones bearing plural agreement on the verb with locative/genitive DPs in preverbal position.

¹⁰An analysis distinguishing definite and generic null subjects in BP based on feature defectivity is found in Galves (2001), who proposes that the Agr features in BP have become impoverished with respect to the category Person: "At least in the 3rd person singular, AGR seems to be too referentially impoverished to identify a null subject as a null, specific pronoun" ["Pelo menos na 3ª pessoa do singular, AGR, parece ser referencialmente pobre demais para identificar um sujeito nulo como um pronome nulo específico"] (our translation of GALVES, 2001, p. 110). However, the claim that clauses containing null subjects receive an indefinite interpretation does not suffice to explain the ungrammaticality of clauses in which the null subject is interpreted as generic in 'out of the blue' contexts (as in **Conserta sapatos* 'They_{generic} repair shoes', in contrast with *Aqui conserta sapato* 'Here, they_{generic} repair shoes'). We will return to these contrastive facts.

The author concludes that BP has undergone a microparametric change with respect to the licensing of null subjects, and not a parametric change in the more general sense, given that the contrast between the two varieties is crucially restricted to verbal agreement in a limited number of contexts.

Kato and Duarte (2014, 2017) acknowledge Costa’s (2010) observations regarding the constructions shared by BP and EP. However, they provide empirical arguments in support of their productivity in BP, as opposed to EP, not only for their wide spread among speakers, but also for their expressive manifestation with a wide amount of lexical items in each verb class, which seems not to be the case in EP. Accordingly, they suggest that the clausal structure in BP is essentially ‘personal’, in the sense of requiring an EPP licensing through verb-subject agreement, with SpecTP being filled either via (hyper)raising or via insertion of a full phrase or a weak pronoun. Kato and Duarte (2017) also reject the analysis that identifies BP and Finnish as PNSL, taking into consideration crosslinguistic data concerning the availability of subject raising and the presence/absence of lexical expletives.

Considering the results in Duarte and Kato (2008a, 2008b), Costa (2010), and Kato and Duarte (2014, 2017), our conclusion is that the set of data that we presently take into consideration display peculiar properties with respect to the pronominal/inflectional system in BP, pointing to the relevance of agreement with the subject, further corroborating the hypothesis that it is SpecTP that is at stake. In other words, not only the increasing tendency for subject filling with overt referential elements, but also the rise of (spatial/temporal) locative phrases in subject position, as a strategy of subject filling, triggering third person agreement on the verb, are the evidence that the pronominal/inflectional system of BP deals with the features of the third person in a different way, as opposed to first and second person.

1.5. Partial Summary

The studies discussed in the previous sections characterize null subject languages in terms of the properties of the functional head licensing the subject, or the status of the null category (whether a variable or not), thus affecting the manifestation of subjects in all persons. Such a characterization is unable to explain the fact that in BP first and second person null subjects behave differently from third person null subjects.

Before continuing, a summary of the properties discussed thus far is in order:

- a) there is a tendency in BP to use overt subjects, rather than null subjects, which suggests that a syntactic change has taken place, which may be explained in terms of a parametric change or in terms of a reorganization of sentence structure (where the language is now topic-prominent and no longer subject-prominent);
- b) the system of verbal inflection has undergone an impoverishment;
- c) there has been a specialization in the syntax of null subjects in BP;
- d) the occurrence of overt non-argumental phrases in subject position (in topic-subject clauses; in meteorological predicates) consistently trigger third person agreement on the verb.

We are aware that other analyses have been put forward considering particularly the data with locative and a generic reading on the subject. In Galves' (2001) analysis the third person morpheme in BP is unable of assigning reference to a null subject. Consequently, a generic/ indefinite interpretation arises if the subject is null (differently from EP, in which the SE morpheme is obligatory). Avelar and Cyrino (2008) look at the data involving a prepositional locative in preverbal position, positing that they are the counterpart of Bantu language's locative inversion constructions. We shall not go into the details of this and other analyses because they take each type of construction separately. Accordingly, what we tried to point out with the present discussion is that (most of) the previous analyses about the syntax of subjects in BP look at different sides of the pronominal system and make different predictions. In our point of view they are partially correct and we will explain why, further showing that these analyses can be unified.

A. The parametric change hypothesis

According to the parametric change hypothesis BP is changing from a pro-drop language to a non pro-drop language.

However, no explanation is provided for why certain embedded contexts permit only null subjects, as in embedded complement clause within a wh-question and a relative clause, as illustrated in (22) and (23), respectively:¹¹

- (21) a. Quem_i acha que corre rápido e_{i/*j} levanta a mão./ *Whoever thinks that (he) ran fast e_{i/*j} raise (their) hand.*
 b. *Quem_i acha que corre rápido ele_{i/j} levanta a mão./ *Whoever thinks (he) ran fast he_{i/j} raise (their) hand.*
- (22) Ninguém_i achou o livro que e_{i/*j} perdeu./ *No one has found the book that e_{i/*j} lost.*

B. BP as a topic-prominent/discourse-oriented language hypothesis

The analysis arguing that BP is a discourse-oriented language does not explain why locative adverbs/pronouns are intervening elements for control of the null embedded subject by the matrix subject (24a) – in the absence of the locative adverb/pronoun the possibility of embedded subject control by the matrix subject holds (24b).

- (23) a. O João_i disse que aqui vende_{arb/*i} fruta./ *The John said_{3SG} that here sell_{3SG} fruits.*
 b. O João_i disse que vende_i fruta./ *The John said_{3SG} that sell_{3SG} fruits*

¹¹ Original examples from Costa and Pratas (2013), corresponding to those in (22):

- (i) Kenhai ki ta atxa ma Ø_{i/*j} kore faxi labanta mo.
 who COMP TMA think COMP run fast raise hand
 ‘Whoever thinks he ran fast raise (their) hand.’
 Kenhai ki ta atxa ma e_{i/*j} kore faxi labanta mo.
 who COMP TMA think COMP PRON run fast raise hand
 ‘Whoever thinks he ran fast raise (their) hand.’

C. BP as a partial null-subject language

Our criticism to the hypothesis that BP is a Partial Null Subject Language (PNSL), in spite of agreeing with the description that it brings, is that it does not consider the importance/obligatoriness of locatives in the licensing of these sentences. This analysis has nothing to say about the role of locatives or expletives in the syntax of generic or indefinite sentences in this type of languages.

In order to show an analysis where these problems are accounted for we will argue that BP is a language with a split pronominal system, the role of locatives being explicitly connected to the properties of third person, as opposed to first and second person inflection.

2. BP as a language with a split pronominal system

Firstly it is important to notice that Pilati and Naves' (2013) analyses originally claimed that a uniform analysis should be provided for five phenomena related to the licensing of null subjects in BP, developing Rabelo's (2010) original proposal that there is a split in the inflectional system of BP with respect to the ability to assign reference to a subject. These constructions, which are similar to one another in that they predominantly occur with verbs in the third person, lead the authors to the following descriptive generalizations: (i) in BP, matrix null subjects receive a deictic interpretation, predominantly locative; such sentences are not licensed when the deixis is not rendered possible (a licensing requirement which can be modeled in syntactic or extra-syntactic terms); (ii) deictic expressions occupy the subject position, even when such expressions are not arguments of the predicate.

Specifically, the five relevant phenomena are: (i) clauses containing Verb-Subject order, which are analyzed as having a deictic reading or a reading implying simultaneity with the time of speech (cf. (25a), in contrast to (25b)) and as allowing a deictic, preverbal element to control the embedded null subject (cf. (26c), in contrast to (26a-b)); (ii) topic-subject constructions, in which non-canonical elements (possessives, locatives, and benefactives) occupy the pre-verbal position triggering agreement on the verb (cf. (27)); (iii) clauses containing a null subject with arbitrary reference, which is licensed by either a (null) pronoun or an adverb in preverbal position (cf. (28a)) or in a discursive context in which the null subject's reference can be determined through

deixis (as in (28b), for example, this sentence being found on a sign in a marketplace); (iv) clauses with quasi-arguments, in which locative phrases occur in subject position (as in (29)).

- (25) a. Entrou Dante. [Dante has just entered here, in this place]
 b. Dante entrou. [Dante entered in some place at some time]
- (26) a. As crianças_i dormem ali porque e_i querem./ *The children sleep there because e_i want.*
 b. *Ali dormem as crianças_i porque e_i querem./ *There sleep the children because e_i want.*
 c. Ali_i dormem as crianças porque e_i é mais quente./ *There sleep the children because e_i is warmer.*
- (27) a. O carro furou o pneu. / Os carros furaram o pneu./ *The car punctured_{3SG} the tire/ The cars pierced_{3PL} the tire*
 b. Essa mala cabe muita roupa. / Essas malas cabem muita roupa./ *This suitcase fits lots of clothes/ These suitcases fit lots of clothes*
- (28) a. Vende frutas./ *[Here] sells fruits.*
 b. Aqui faz conserto de roupas./ *Here [one] fixes clothes.*
- (29) a. Brasília não chove há mais de 90 dias./ *Brasília does not rain for 90 days.*
 b. Londres tem prédios lindos./ *London has beautiful edifications.*

The authors, thus, base their analysis on the typological classification found in Bhat (2004), under which the class of pronouns is not uniform but heterogeneous: on the one hand, there are pronouns which refer to the participants in the speech act; on the other hand, there are pronouns which do not refer to speech act participants. Using Bhat's classification as the basis of their analysis, the authors propose that BP treats the first and second persons as personal pronouns *stricto sensu* while treating the third person as having distinct properties, which are, by hypothesis, similar to those of other pronominal forms – indefinites and demonstratives, for example.

This analysis thus emphasizes the hypothesis that there is a split in BP's pronominal system. Although previous studies have singled out the third person in BP as having special properties (as we have shown in section 1), in their respective analyses of how this language's pronominal system works, they have nonetheless treated the system as uniform. As already mentioned, the hypothesis that BP exhibits a split pronominal system was originally investigated in Rabelo (2010), who argues, within Manzini and Savoia's (2005) approach to subject licensing, that BP's third person morphology lacks the denotational property necessary for identifying null subjects.

Pilati and Naves (2013), and Pilati, Naves and Salles (2015, 2017a/b) defend two hypotheses concerning the syntax of BP. The first hypothesis is that there is a split in the pronominal system whereby the first and second persons exhibit different syntactic and semantic properties as compared to those exhibited by the third person. The second one is that BP licenses locative elements in subject position.

The authors analyze the data in (30) in the following terms: with regard to the third person in BP, the authors contend that (i) it is underspecified for the relevant feature, receiving a referential interpretation when the relevant expression is either a full DP or an overt pronoun (cf. (30a-b) versus (30c)); (ii) it receives a generic/ indefinite reading when the expression is a null pronoun or a (null or overt) (spatial/ temporal) locative phrase in pre-verbal position (cf. (30d-f)).

- (30) a. Maria conserta roupas./ *Maria fixes clothes* [lexical DP (third person)]
 b. Ela conserta roupas./ *She fixes clothes* [overt third person pronoun]
 c. *Conserta roupas./ *Fixes clothes* [referential null third person pronoun – *out of the blue*]
 d. Conserta roupas./ *Fixes clothes* [non-referential null third person pronoun – *a sign hanging in a store*]
 e. Aqui conserta roupas./ *Here fixes clothes* [locative pronoun/adverb – generic subject]
 f. Antigamente consertava sapato, hoje joga fora e compra outro./ *In the past [one] used to fix shoes, nowadays [one] throws it away and buys another one* [temporal, preverbal pronoun/adverb – generic subject]

The authors thus argue that (i) null third person subjects in matrix clauses cannot be construed as definite; they are construed as generic or arbitrary when uttered in discourse anchored contexts and not in ‘out of the blue’ contexts (cf. (30c) *versus* (30d)); (ii) BP has developed (third person) null generic subjects without the ‘se’ pronoun (cf. (30e-f)); (iii) locative DPs may occupy the subject position under specific structural conditions (specifically, when the predicate is unnaccusative – whether monoargumental (cf. (27) and (29)) or biargumental (cf. (27b)) – and when the predicate is transitive, with the external argument interpreted as generic (cf. (30e)).

This analysis receives support from Finnish data, which, as was mentioned above, has been described as a partial null subject language. As Holmberg (2010) demonstrates, in null subject sentences containing a preverbal adverb, the sentence receives a generic/arbitrary reading (cf. (31a)). Moreover, these adverbs are in complementary distribution with expletives (cf. (31b)):

- (31) a. Tässä istuu mukavasti./ *Here sits comfortably* ‘One can sit comfortably here.’
 b. Sitä istuu mukavasti tässä./ *EXP sits comfortably here* ‘One can sit comfortably here.’

Considering the analysis presented thus far, we may add two syntactic and semantic characteristics of null and overt subjects in BP to those listed in (a) through (d) in section 1.4:

- e) there is a split in BP’s pronominal/inflectional system whereby the first and the second person behave differently from the third person singular;
 f) both BP and Finnish allow locatives to occupy the subject position when the external argument receives an indefinite interpretation; according to Pilati, Naves and Salles (2015, 2017a/b), locative subjects tend to be grammaticalized as an expletive; in BP the locative can be null (an overt expletive being unavailable).¹²

¹² For other analyses which suggest that locative pronouns/adverbs in BP behave similarly to expletives, see Buthers (2009) and Pereira (2011). Differently from these analyses, the locative pronoun/adverbs in *Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017*.

3. Towards an analysis for BP and Finnish data

In Holmberg's (2005) analysis, it is argued that there is a D(efiniteness)-feature on the head of the inflectional domain (T) and that this feature is relevant for characterizing a language as a full Null Subject Language (NSL). In non-null subject languages, the D-feature is absent while in full NSL, T bears a D-feature, which is licensed under agreement with the referential subject or under incorporation of a related pronominal category. In Partial Null Subject Languages (PNSL), in turn, T is not specified for the D-feature and, consequently, null subjects are restricted to indefinite environments. Since *pro* does not bear a D-feature, it may have a c-commanding antecedent. In sum, the main difference between Null Subject Languages, Partial Null Subject Languages and Non-Null Subject Languages is whether T bears a D feature or not.

In explaining why PNSL subjects are restricted to indefinite environments, Holmberg (2010) observes that Finnish does not have an overt Generic pronoun (G-pronoun) corresponding to English *one*, French *on*, German *man* or Italian *si*. In Finnish a generic pronominal subject does not have an overt expression in subject position. Thus according to Holmberg (2010), a possible analysis of (32a-c) is that they have a null generic pronoun subject. However, the fact that the adverbial phrase in pre-verbal position is obligatory is not taken into consideration.

- (32) a. Tässä istuu mukavasti./ *Here sits comfortably* 'One can sit comfortably here.'
- b. Kesällä herä ä aikaisin./ *summer-ADE wakes.up early* 'You wake up early in the summer.'
- c. Sitä ei kannat ain valittaa/ *EXP not should always complain* 'It's no use always complaining.'

Holmberg (2010) further assumes that the interpretation of *pro* is determined by its internal structure, which is illustrated in (33a), as opposed to (33b), which corresponds to lexical pronouns:

our analysis is taken to behave as quasi-argument expletives in predicates describing natural phenomena (cf. Rizzi, 1982), due to their distribution involving the existential status of the predicate.

- (33) a. $[\phi_P \phi]$
 b. $[_{DP} D [\phi_P \phi [_{NP} N]]]$

While a D-feature (associated with the D-level in the pronominal structure) renders a lexical pronoun definite, a null pronoun, being deprived of the D-feature, needs to acquire definiteness through other means. Full and Partial Null Subject Languages (NSL/ PNSL) differ as to whether T displays a D-feature in its make-up: in full NSL (such as Italian and European Portuguese), T is assigned a (uninterpretable) D-feature (which is bound by an (Aboutness-shift) topic in the CP articulated domain), thus licensing ϕ_P ; in PNSL (such as BP and Finnish) in turn T does not have a D-feature, hence the null pronoun is not assigned a definite interpretation. As Holmberg (2010, p. 95) says: “When T probes a ϕ_P subject, and has its unvalued ϕ -features valued by the subject, the resulting union of the ϕ -features of T and the subject yields a definite pronoun.”

Given these assumptions, in constructions such as (32), a G-pronoun is merged like a regular pronoun in vP in a transitive sentence, in VP in unaccusative sentences, receiving the (subject) phi/theta-role exactly like other pronouns or referential expressions. It consists of nominal features, meaning that it has number (singular), (third) person and an unvalued Case-feature (the value of which needs to be assigned). It is, however, deficient in that it lacks a D-feature (a property it shares with overt generic pronouns such as English *one*, German generic *man* etc.). As a pronoun, it also lacks a root: it is a bare phi-feature complex. This means that when T probes this pronoun, and has its f-features valued by it, while at the same time valuing the Case feature of the pronoun, T and the pronoun end up having the same feature values. Following Roberts (2010), it is further assumed that the grammar takes them to be copies, forming an argument chain, effectively as if the subject pronoun had moved by head-movement to T, except that there has only been Agree (mutual feature valuing), with no movement. The lower copy must be deleted according to standard rules of chain reduction.

The G-pronoun is a copy of T, and it thus forms a chain, in fact, an argumental chain (A-chain). Consequently, T is spelled out as an affix on the finite verb or auxiliary, while its copy, the subject G-pronoun in Spec vP , is not spelled out (or is

spelled out as a null subject). According to the author, this means that the reason why the G-pronoun is null in Finnish is because it is a D-less pronoun probed by finite T, the only interpretation available to a pronoun being an arbitrary or a generic one.

As argued in Holmberg (2010), the locative pronoun/adverb and the expletive occupy the same subject position, checking the EPP feature. Accordingly, null subjects are not found in these sentences, because Finnish has overt expletives. Thus, for Holmberg, when there is a null pronoun in subject position it is possible to have locative or expletives in the subject position, checking the EPP feature.

A problem of Holmberg's (2010) analysis is that null subjects are not always (null) indefinites in PNSL. This is illustrated in (34), below, as opposed to (32a-c).

- (34) [Veljeni₂ vaimo]₁ oli niin iloinen, ettei e_{1/*2/*3} voinut nukkua./ brother-gen wife
was so happy that.not could sleep/ 'My brother's wife was so happy that she
could not sleep.' (MODESTO, 2007, ex. 2, p. 02)

Holmberg (2010) further argues that the above-mentioned properties of Finnish are shared by BP (as well as by Marathi and Assamese, from the Indo-Aryan stock). Taking into consideration the data from BP we will argue that the properties of PNSL are not encoded in terms of the defective make-up of T features, as it does not capture the empirical facts of BP regarding the pronominal reanalysis and its implications to the agreement system.

The Proposal – Part I: There is a split in the BP inflectional system

Looking at the same descriptive properties, Pilati and Naves (2013) and Pilati, Naves and Salles (2015, 2017a/b) assume, (partially) following Bhat (2004) and Rabelo (2010), that there is a split in the BP inflectional system which renders the third person inflection unable to license referential/ definite null subjects, as opposed to first and second person. They argue that the emergence of this innovative property is directly related to the fact that the third person inflection is unable to license the EPP, which further relates to the loss of indefinite SE. The EPP requirement is then satisfied under insertion of a locative subject, which is realized either by a locative DP or by a (null) locative adverb/pronoun – the null variant being recovered in the discourse.

The third person inflection on the verb is valued as referential by either a full DP (35a) or an overt pronoun in subject position (35b); in the absence of a full referential DP/ pronoun, the third person inflection is valued by an overt/ null locative pronoun and the (third person) external argument is interpreted as generic/ arbitrary (35d) – which is further determined by the absence of the indefinite pronoun ‘se’ in the pronominal system:

- (35) a. Maria conserta roupas. [DP + third person = referential meaning]
Mary repairs clothes
- b. Ela conserta roupas. [Pronoun + third person = referential meaning]
She repairs clothes
- c. Aqui conserta roupas. [Locative or expletive pronoun = arbitrary reading]
here repairs clothes

A piece of evidence that the locative is in subject position is that it can control a null pronoun in embedded clauses, exactly as a lexical DP can in subject position (as originally observed in Pilati’s (2006) analysis of VS in BP as an instance of locative inversion) – cf. (26), repeated below as (36):

- (36) a. As crianças_i dormem ali porque e_i querem./ *The children_i sleep there because e_i want_{3SG}.*
- b. *Ali dormem as crianças_i porque e_i querem./ *There_i sleep the children because e_i want_{3SG}*
- c. Ali_i dormem as crianças porque e_i é mais quente./ *There_i sleep the children because e_i is warmer.*

Accordingly in PNSL indefinite subjects are third person null pronouns (possibly a ϕ P category, as described in Holmberg (2010)), while the EPP can be licensed by either a DP or a locative (null or overt) pronoun/adverb in SpecTP (or an expletive, as in Finnish).

Assuming the split in the pronominal/inflectional system, first and second person may be found as null subjects in NSL, being always referential, despite the fact that BP shows more overt subjects. In turn, as already mentioned, third person subjects do not behave as null subjects in PNSL, the null form never being referential in main clauses. It should be noticed that in Holmberg's (2010) terms, the referential properties of third person (null) subjects are provided by the D-feature on T to the subject. Accordingly, the latter property is parameterized (an idea that we shall not pursue, as our claim is that parametric variation is a property of the inflectional/ agreement system, in accordance with the pronominal system).

A fact that should be pointed out is that both BP and Finnish show some kind of loss in the verbal agreement system. As showed in the first section, BP lost the richness of its verbal agreement morphology under pronominal reanalysis (cf. Duarte, 1995), and Finnish has only one morpheme corresponding to third person plural or singular (cf. Holmberg and Nikkane, 2002).

These changes and the behavior of third person lead us to think that, regarding this person, the inflectional/agreement system of BP and Finnish may be compared to that of Chinese. Chinese is well-known for its lack of verbal agreement while still allowing empty subjects for all persons. Huang (1984) proposes that null subjects in Chinese can refer either to the matrix subject or to a salient discourse topic. So they can be sometimes *pro*, sometimes a *variable*. According to Huang (1984), the null subject in (37a) can refer to either the matrix subject (behaving as a *pro*) or to a salient discourse topic (behaving as a *variable*), while the null object in (37b) can only refer to a discourse topic.

- (37) a. Zhangsan_i shuo [e_{i/j} bu renshi Lisi]./ *Zhangsan say not know Lisi*
 'Zhangsan said he didn't know Lisi.'
- b. Zhangsan_i shuo [Lisi bu renshi e_j]./ *Zhangsan say Lisi not know*
 'Zhangsan said Lisi didn't know him.'

Huang (1984) argues that the empty subject in (37a) is a *pro*, if the nominal phrase *Zhangsan* is its antecedent. In addition, nothing prevents this null subject to be analysed as a *variable*, if it refers to someone else whose reference is identifiable in the

discourse. As for the null subject in (37a), it is a variable, as it has its reference identified in the discourse, by a topic, which is an A'-position.

So in (37a) the subject can be either viewed as *pro*, since it not only is free in its binding domain, but can also be co-indexed with the closest nominal phrase, which is the matrix subject; or interpreted as a variable, since it can be bound by some salient topic generated in the discourse. But when it is uttered in out of the blue contexts, the null subject has to refer to the matrix subject. Huang further claims that it is not straightforward to find a context in which the null subject refers to a 'salient discourse topic': in this case, a person who is relevant in the conversational context, but who is not mentioned as the missing argument. Therefore, the null subject in Chinese has a dual status. When A'-bound, it is a variable; when A-bound, it is a *pro*.¹³

Following Chomsky (1982), who argues that the identity of an empty category can in general be predicted on the basis of its position of occurrence or that of its antecedent, Huang (1989) observes that it is plausible to assume that the different empty categories are in fact 'allocategories' of a single *empty category* (analogous to the allomorphs of a single morpheme), whose identity can be determined locally, on the basis of the following 'functional definition':¹⁴

- a) An [NP e] is pronominal (=PRO or *pro*) if, and only if, it is free or locally θ -bound (i.e., bound by an element with an independent thematic role), and non-pronominal (=trace) if, and only if, it is locally non- θ -bound.
- b) A pronominal EC is PRO if, and only if, it is ungoverned, and *pro* only if it is governed.

¹³ Huang (1989) claims that "what we have is a general condition of identification or recoverability which requires an empty pronominal (PRO or *pro*) to be identified under appropriate circumstances, where identification can take different forms – if not by agreement then by an antecedent". In a clause with agreement markings, the agreement element (Agr) is the closest potential 'controller' (an Agr essentially amounts to an overt pronoun, with its markings for person, gender and number), so it must be the controller of its subject *pro*. A subject *pro* is licensed if the Agr is rich enough to be its 'controller', as in Italian-type languages. In a language like English, the mere appearance of Agr makes it the obligatory controller of its subject *pro*, but its degenerate nature makes it incapable of carrying out the task of control. Hence a *pro* is excluded in English. On the other hand, a *pro* in Chinese, does not have Agr occurring as its minimal potential 'controller', so it can look outside the clause for its controller, and we have cases of control in the standard sense.

¹⁴ We are aware that in minimalism terms the theoretical notion of government does not hold, crucially affecting the theoretical distinction between *pro* and PRO (cf. CHOMSKY, 1995). We shall not go into the details of the debate regarding the status of PRO. We will thus provisionally refer to both categories as *pro*, distinguishing them with respect to their syntax in each case.

- c) A non-pronominal EC is an anaphor (NP-trace) if, and only if, it is locally A-bound, and a variable (wh-trace) if, and only if, it is locally A'-bound.

Going back to the present analysis, it is important to observe that our claim is that there are two kinds of null pronouns or two different agreement systems. In languages with rich agreement systems the null *pro* or the verbal agreement morphemes will behave like overt personal pronouns, they can have independent reference even if they are null. However, in languages without an overt agreement system, as Chinese, or in PNSL, as BP, the third person agreement inflectional morpheme will have special properties, because they do not have an independent reference, as argued by Huang (1989) for Chinese.¹⁵

As shown by previous works (RODRIGUES, 2004; MODESTO, 2000, 2004, 2008; PILATI, NAVES and SALLES 2015, 2017a/b; among others), in PNSL the null subject obtains its reference when coindexed with the matrix subject (it can be analyzed as an A-bound *pro*):

- (38) João_i disse [CP que [TP e_{i/*j} virá]. ‘John_i said [he_{i/*j}] will come.’

Considering this, we can argue that in PNSL the third person agreement morphology behaves partially as the Chinese agreement system. Accordingly, these empty categories can behave in two different ways. This is stated in Part II of our proposal.

¹⁵ Since the Government and Binding (GB) theory, the notion of the Extended Projection Principle (EPP, Chomsky 1981, 1982) requires that every sentence should have a subject. Empty subjects are generally referred to as null pronominals, or *pro*, which take Case, theta roles, phi-features (person, gender, and number) just as regular pronouns do. The interpretation of *pro* is derived by certain principles of identification if the rest of the sentences can provide enough information to recover the content of the missing subjects. In agreement languages such as Italian and Greek, the empty subjects are identified via rich morphological marking on the verbal agreement system. In “agreementless” languages like Chinese, for all persons, or in PNSL as Brazilian Portuguese and Finnish, for third person, such a mechanism cannot apply. Due to the lack of agreement, these null subjects have to resort to the salient antecedents in the discourse and are regulated by either a certain pronominal binding interpretation or the operator-variable schema (cf. Huang 1984, 1989).

The Proposal – Part II: The third person null subject in BP is analyzed:

- a) as a null *pro* in *out of the blue contexts*, when the null morpheme seeks its reference in a c-commanding position locally θ -bound (i.e., bound by an element with an independent thematic role):

- (39) a. O João_i disse que vem_i para a festa./ *John_i told that [he]_i will come to the party*
 b. O João_i disse [CP que [TP pro_{[\phi P]i} [T' vem_{[\phi P]i} [vP pro_{[\phi P]i} [v' ~~vem~~ [VP ~~vem~~ [para a festa]]]]]]]

- b) as a variable in contexts in which it is locally A'-bound, as in question-answer sentences or in sentences with topics:

- (40) a. A Maria_i, o João_j disse que vem_i para a festa./ *Mary_i John_j told that [she]_i will come to the party*
 b. [CP A Maria_i [TP O João_j disse [CP que [TP vem [vP pro_{[\phi P]i} [v' ~~vem~~ [VP ~~vem~~ [para a festa]]]]]]]]]

Contexts in which null subjects show an arbitrary reading, as in (41) below, can also be explained in the lines of (39a) – (as a null *pro* in *out of the blue context*, when the null morpheme will seek its reference in a c-commanding position locally θ -bound). We assume that the arbitrary reading will arise because in these contexts the (interpretable) person feature of the locative DP/pronoun/adverb in subject position enters an agreement relation with the (uninterpretable) (third) person feature on T (provided by the agreement system), which incorporates the phi-features of *pro* (which is realized as a ϕ P category in the specifier of vP, in Holmberg' (2010) terms).¹⁶

¹⁶ It is worth observing that our proposal seems to have some counterexamples in the case of constructions with *dicendi* verbs with a deontic meaning (*mandar* ‘to order’, *pedir* ‘to ask’ and *dizer* ‘to say’), such as in: *O guarda_i mandou e_{*i,j} sair./ O guarda disse para e_{*i,j} sair.* ‘The policeman ordered e_{*i,j} [to] leave./ The policeman said to e_{*i,j} leave.’ Indeed, in these constructions, the null subject of the infinitive embedded clause is not controlled by the subject in the matrix SpecTP (*the policeman*); on the contrary, it receives an arbitrary/generic reading. We consider these facts to be related to the lexical properties of *dicendi* verbs, the (null/ implicit) internal argument of these verbs being the appropriate antecedent for the null embedded subject (The policeman_i ordered e_j [e_{*i,j} to leave]). The evidence for this analysis is that control of the embedded subject by the subject in the matrix clause is expected if the verb of the matrix clause is changed, as in: *O ladrão_i quis e_j sair./ ‘The thief_i wanted e_j [to] leave’.*

- (41) a. Aqui senta confortavelmente./ *Here sits comfortably*
 b. [CP [TP Aqui_[φP] [T senta [VP pro_[φP] [_V senta [VP senta] [confortavelmente]]]]]]

The rise of this operation is determined by the following conditions: (i) the loss of the indefinite pronoun ‘se’ (which is in complementary distribution with the locative pronoun/adverb in SpecTP); (ii) the formal identification of the external argument with a (spatial/ temporal) locative which is either an argument of the predicate (as in (41a), and an internal argument, as in topic-subject constructions (*Essa casas bate sol/ This house hits the sun [=The sun hits the house]*), or the (spatial/temporal) locative, which encompasses the (external) argument that is selected by the predicate (as in *Antigamente usava bengala/ In the old time used a stick*), giving rise, in some cases, to a sort of metonymic reading (as in *Essa loja/Aqui vende fruta; This shop/Here sell fruits*). This analysis is supported by the fact that in the presence of the number feature, agreement arises, as illustrated in (42):

- (42) a. Essas lojas venderam muitos CDs./ *These stores have sold many CDs*
 b. Essas cidades chovem muito./ *These cities rains a lot.*

This analysis further accounts for the contrastive facts in (43a) and (43b), in which the null subject is a *pro* licensed under the agree relation above described with the *locative* pronoun/adverb in the subject position and under binding by the local c-commanding antecedent, respectively.

- (43) a. O João_i disse que aqui $e_{j/*i}$ vende fruta. *John said that one sells fruits here.*
 b. O João_i disse que e_i vende fruta. *John said that he sells fruits.*

4. Final considerations

In this article, we have provided a review of well-known studies of the syntax of BP subjects, in which a tendency to displaying overt subjects is found in contexts in which the subject is consistently null in EP – as a result of the reanalysis of subject pronouns, which, in turn, gave rise to a drastic reduction in the language’s verbal inflectional paradigm.

Taking into consideration the changes which have taken place in the pronominal system, as well as in the inflectional system of BP, and assuming that BP exhibits (some) properties of Discourse-Oriented Languages, as stated in previous studies (see section 1), we proposed that the pattern concerning the distribution of definite and generic/arbitrary null subjects crucially affects third person pronouns – in terms of the hypothesis that BP is a language with a split pronominal system. This proposal enables us to explain the conditions that determine the distribution and interpretation of the null subject in this language, and its relation to the occurrence of (spatial/ temporal) locatives in subject position. Accordingly, it allows us to build up a unified analysis of a variety of phenomena consistently found in BP, such as: the obligatory presence of a locative/ temporal adverb (whether overt or null) in sentences with VS word order; the fact that the subject position is obligatorily filled by (pronominal/lexical) locatives DP in the topic-subject construction and in sentences containing a quasi-argument; and the properties of the subjects in constructions in which the agent receives a generic/arbitrary interpretation.

In sum, we have argued that null subjects in PNSL have two different behaviors: while first and second person verbal morphology can be analyzed as the verbal morphology in NSL, the third person verbal morphology (singular in BP and singular and plural in Finnish) displays the same properties as the verbal morphology of Chinese (a language without overt agreement). This null third person morpheme occurs: a) as a Chinese *pro* in main clauses, being grammatical if it can seek its reference through a DP in the subject position of a matrix sentence, further establishing a formal identification with a locative DP/adverb/pronoun (whether lexical or null) under merger; b) if CP is filled by either a WH word, a DP in topic position, or a quantifier.

References

AVELAR, J.; S. CYRINO. “Locativos preposicionados em posição de sujeito: uma contribuição das línguas Bantu à sintaxe do português brasileiro”. *Linguística 3 – Revista de Estudos Linguísticos da Universidade do Porto*, 2008, pp. 218 – 249.

BHAT, D. *Pronouns*. Oxford: Oxford University Press, 2004.

- BUTHERS, C. *Emergência da ordem [XP V (DP)] no PB contemporâneo e o Parâmetro do Sujeito Nulo: uma abordagem minimalista*. Master's Thesis. UFMG, 2009.
- CARDINALETTI, A. "Toward a cartography of subject positions". In: RIZZI, L. (ed.) *The Structure of CP and IP: the cartography of syntactic structures, vol 2*, New York: Oxford University Press, 2004, pp. 115-165.
- CHOMSKY, N. *The Minimalist Program*. Cambridge (MA): The MIT Press, 1995.
- CHOMSKY, N. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger, 1986.
- CHOMSKY, N. *Some concepts and consequences of the theory of government and binding*. Cambridge (MA): The MIT Press, 1982.
- CHOMSKY, N. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.
- COSTA, J. "PB e PE: orientação para o discurso importa?". *Revista Estudos da Língua(gem)*. Vitória da Conquista, v. 8, n. 1. 2010, pp. 123-143.
- COSTA, J.; PRATAS, F. "Embedded null subjects in Capeverdean". *Journal of Linguistics* 49, 2013, pp. 33-53.
- CYRINO, S; DUARTE, M. E.; KATO, M. "Visible subjects and invisible clitics in Brazilian Portuguese". In: KATO, M.; NEGRÃO, E. (eds.), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt am Main: Vervuert-Iberoamericana, 2000, pp. 55-104.
- DUARTE, M. E. "A evolução na representação do sujeito pronominal em dois tempos". In: PAIVA, M. C.; DUARTE, M. E. L. (orgs.), *Mudança Linguística em Tempo Real*. Rio de Janeiro: ContraCapa. 2003, p. 115-128.
- DUARTE, M. E. *Perda do Princípio 'Evite Pronome' no Português do Brasil*. Doctoral Thesis. Campinas: Unicamp, 1995.
- DUARTE, M. E. "Do pronome nulo ao pronome pleno: a trajetória do sujeito nulo no português do Brasil". In: ROBERTS, I.; KATO, M. (orgs.). *Português Brasileiro: Uma Viagem Diacrônica*. Campinas: Unicamp, 1993, pp. 107-128.

- DUARTE, M. E.; MOURÃO, G. C.; SANTOS, H. M. “Os sujeitos de 3ª pessoa: revisitando Duarte 1993”. In: DUARTE, M. E. (org.). *O Sujeito em Peças de Teatro (1983-1992): estudos diacrônicos*. São Paulo: Parábola, 2012, pp. 21-44.
- FREEZE, R. “Existentials and other locatives”. *Language* 68 (3), 1992, pp. 553-595.
- GALVES, C. “Clitic-placement and parametric changes in Portuguese”. In: SALTARELLI, M.; ZUBIZARRETA, M.-L. (Orgs.) *Selected Papers from the 24th Linguistic Symposium on Romance Languages*. Georgetown University Press, 1996, pp. 227-240.
- GALVES, G. *Ensaio sobre as Gramáticas do Português*. Campinas: Editora Unicamp, 2001.
- HOLMBERG, A. “The null generic subject pronoun in Finnish”. In: BIBERAUER, T. *et al. Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press, 2010, pp. 200-230.
- HOLMBERG, A. “Is there a little pro? Evidence from Finnish”. *Linguistic Inquiry* 36 (4), 2005, pp. 533-564.
- HOLMBERG, A.; NIKANNE, U. “Expletives, Subjects and Topics in Finnish”. In: SVENONIUS, P. (ed.) *Subjects, Expletives, and the EPP*, Oxford: OUP, 2002, pp. 71-105.
- HUANG, C.-T. J. “Pro-drop in Chinese: a generalized control theory”. In: JAEGGLI, O.; SAFIR, K. (eds.) *The Null Subject Parameter*. Dordrecht, Boston & London: Kluwer, 1989, pp. 215-238.
- HUANG, C.-T. J. “On the Distribution and Reference of Empty Pronouns”. *Linguistic Inquiry* 15, 1984, pp. 531-574.
- KATO, M.; DUARTE, M. E. “O sujeito no português brasileiro e sua tipologia”. In: PILATI, E.; SALLES, H. L.; NAVES, R. (orgs.). *Novos Olhares para a Gramática do Português Brasileiro*. Campinas (SP): Pontes, 2017, pp. 13-42.
- KATO, M.; DUARTE, M. E. “A variação entre construções finitas pessoais e impessoais no português brasileiro”, *WEB-Revista Socióleto*, v. 4, n. 12, pp. 153-177.

- KATO, M.; DUARTE, M. E. “Indefinite subjects in Brazilian Portuguese, a Topic and Subject-prominent language”. Paper presented at the VII Workshop on Formal Linguistics, Curitiba-PR, 2008a.
- KATO, M.; DUARTE, M. E. “Mudança paramétrica e orientação para o discurso”. Paper presented at the XXIV Encontro Nacional da Associação Portuguesa de Linguística, Braga, 2008b.
- KATO, M.; RAPOSO, E. “Objeto(s) e artigos nulos: similaridades e diferenças entre o português europeu e o português brasileiro”. In: MOURA, D.; FARIAS, J. (org.) *Reflexões sobre a Sintaxe do Português*. Maceió: Edufal, 2005, pp. 73-96.
- KISS, K. “The EPP in a topic-prominent language”. In: SVENONIUS, P. (ed.) *Subjects, Expletives and the EPP*. New York: Oxford University Press, 2002, pp. 107-124.
- MANZINI, M. R.; SAVOIA, L. M. *I Dialetti Italiani e Romanci* (3 vols.). Alessandria: Dell’Orso, 2005.
- MODESTO, M. “Topic prominence and null subjects”. In: BIBERAUER, T. (ed.) *The Limits of Syntactic Variation*. Amsterdam: John Benjamins Publishing Company, 2008, pp. 375-409.
- MODESTO, M. “Null subjects in Brazilian Portuguese and Finnish: they are not derived by movement”. In: DAVIES, W.; DUBINSKY, S. (org.). *New Horizons in the Analysis of Control and Raising*. Dordrecht: Springer, 2007 (available at <http://linguistica.fflch.usp.br/sites/linguistica.fflch.usp.br/files/Modesto%202007.pdf>)
- MODESTO, M. “Sujeitos nulos em línguas de tópico proeminente”. *Revista da ABRALIN*, vol. 3, 2004, pp. 119-145.
- MODESTO, M. *Null Subjects without ‘Rich’ Agreement*. Doctoral Thesis, University of South California, 2000.
- NAVES, R.; PILATI, E.; SALLES, H. “As cidades da Amazônia chovem muito”: Non-thematic subjects and the properties of agreement in Brazilian Portuguese (BP). Talk presented at Portuguese Linguistics in the United States, University of Georgia, Athens, November 15th, 2013.

- NEGRÃO, E. *O Português Brasileiro: Uma Língua Voltada para o Discurso*. Habilitation Thesis, São Paulo: Universidade de São Paulo, 1999.
- NEGRÃO, E.; VIOTTI, E. “Brazilian Portuguese as a discourse-oriented language”. In: KATO, M.; NEGRÃO, E. (eds.) *Brazilian Portuguese and the Null Subject Parameter*. Frankfurt, Editorial Vervuert/Iberoamericana, 2000, pp. 105-125.
- PEREIRA, B. *A Sintaxe Cartográfica de ‘lá’ no Português Brasileiro: um estudo da periferia esquerda*. Doctoral Thesis, UFMG, 2011.
- PILATI, E. *Aspectos sintáticos e semânticos da ordem verbo-sujeito no português*. Tese de Doutorado. Brasília: UnB, 2006.
- PILATI, E.; NAVES, R. R. “Desenvolvendo a hipótese da cisão da categoria pronominal no português brasileiro”. In: MOURA, D.; SIBALDO, M. *Estudos e Pesquisas em Teoria da Gramática. (Studies and Research in Grammatical Theory)* Maceió: Edufal, 2013, p. 233-253.
- PILATI, E.; NAVES, R.; SALLES, H. “Locative DPs and deictic adverbial pronouns in subject position in Brazilian Portuguese”. In: FERNÁNDEZ-SORIANO, O.; CASTROVIEJO MIRÓ, E.; PÉREZ-JIMÉNEZ, I. (eds.) *Boundaries, Phases and Interfaces: Case studies in honor of Violeta Demonte*. Amsterdam: John Benjamins, 2017a, pp. 64-84.
- PILATI, E.; NAVES, R.; SALLES, H. “A cisão no sistema pronominal do português brasileiro: evidências na diacronia”. In: PILATI, E.; SALLES, H. L.; NAVES, R. (orgs.). *Novos Olhares para a Gramática do Português Brasileiro*. Campinas (SP): Pontes, 2017b, pp. 151-169.
- PILATI, E.; NAVES, R.; SALLES, H. “Tem pão?/ – proLOC Tem (BP)/ – pro Tenho (EP)”: adverbial (locative) pronouns as subjects in Brazilian Portuguese. Paper presented at Encontro Intermediário do Grupo de Trabalho Teoria da Gramática, Belo Horizonte, UFMG, 2015.
- RABELO, P. *Argumentos (EPP) Nulos no Português do Brasil em Contextos Oracionais Finitos e Infinitivos*. Doctoral dissertation, Brasília: Universidade de Brasília, 2010.
- RAPOSO, E. “Case theory and Infl-to-Comp”. *Linguistic Inquiry* 18, 1987, p. 85-109.

- ROBERTS, I. 2010. “A deletion analysis of null subjects”. In: BIBERAUER, T. *et al.* *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press, 2010, pp. 58-87.
- RODRIGUES, C. *Impoverished morphology and A-movement out of case domains*. Doctoral Thesis. Universidade de Maryland, College Park, 2004.
- RIZZI, L. *Issues in Italian Syntax*. Dordrecht: Foris, 1982.
- TARALLO, F. *Relativization Strategies in Brazilian Portuguese*. Doctoral Thesis. University of Pennsylvania, 1983.

Revisiting Duarte (1995): for a Gradient Analysis of Indeterminate Subject in Brazilian Portuguese

Juliana Marins^a

Humberto Soares da Silva^b

Maria Eugenia Lammoglia Duarte^c

ABSTRACT

Recent studies on the representation of pronominal subjects in Brazilian Portuguese (BP) show the preference for overt indeterminate referential subjects. A “side effect” of the change in the Null Subject Parameter in BP is the progressive decline in the use of the standard strategies - structures with verb in the third person plural with a null subject and sentences with the pronominal generic clitic *se* - and the appearance of alternative strategies – an increasing frequency of sentences with overt nominative pronouns, especially *você* (you) and *a gente* (the people =we). Our purpose is to refine Duarte’s (1995) data of indeterminate subjects, collected from a sample of the speech of graduates from Rio de Janeiro. Our theoretical framework associates the Theory of Language Variation and Change (WLH, 2006 [1968]) with the Principles and Parameters Theory (CHOMSKY, 1981, 1995), which guides our analysis, from the hypotheses raised to the interpretation of the empirical data (DUARTE, 2016). Our results show that the different strategies are not in direct competition: they can be distributed in three different groups, according to a set of features they share, with respect to arbitrary and generic reference. At one extreme we find [+3rd person/+plural] category, **which excludes the speaker**, represented by the dying arbitrary clitic *se* and the pronoun *eles* (they), preferably overt. At the other extreme, we have a [+3rd person/+singular] category, **which may or may not include the speaker and the addressee**, represented by the generic clitic *se*, the zero strategy (with a 3rd person singular verb form) and *você* (you), which is preferably overt. Finally, we have a [+1st person/+plural] category, **which does include the speaker**, represented by *nós* (we) and *a gente* (the people, the folks=we), with considerable advantage with respect to the former. The variation in each category disposed along our scale is not a stable phenomenon: each point has a strong competitor to represent each degree of indeterminate reference as the change progresses.

KEYWORDS: indeterminate subjects; arbitrary reference; generic reference; Null Subject Parameter

RESUMO

Estudos recentes sobre a representação do sujeito pronominal no português brasileiro (PB) mostram a preferência por sujeitos de referência indeterminada foneticamente representados. Um “efeito colateral”

^a Universidade Federal do Rio de Janeiro – juespmarins@hotmail.com

^b Universidade Federal do Rio de Janeiro – humba6@gmail.com

^c Universidade Federal do Rio de Janeiro – eugenia@brazilmail.com

This research has the support of CNPq, proc. 307158/2014-4.

das mudanças envolvendo a remarcação do Parâmetro do Sujeito Nulo no PB é o declínio progressivo no uso das estratégias tidas como padrão – estruturas com o verbo na terceira pessoa do plural com sujeito nulo e sentenças com o clítico pronominal *se* – e o surgimento de estratégias alternativas, com a frequência crescente do uso do pronome nominativo expresso, especialmente *você* e *a gente*. Nosso objetivo é refinar a análise dos dados de sujeitos indeterminados levantados em Duarte (1995), com base em amostra da fala culta do Rio de Janeiro. Nosso quadro teórico associa a Teoria da Variação e Mudança (WLH, 2006 [1968]) com a Teoria de Princípios e Parâmetros (CHOMSKY, 1981), que guia nossa análise, desde levantamento de hipóteses até a interpretação dos dados empíricos (DUARTE, 2016). Nossos resultados revelam que as diferentes estratégias não estão em competição: elas se distribuem em três grupos, onde ocorre a competição entre as estratégias que os compõem, de acordo com o conjunto de traços que compartilham, o que está relacionado com o caráter arbitrário ou genérico de cada estratégia. Em um extremo dessa escala, encontramos a categoria [+3a pessoa/+plural], **que exclui o falante**, representada pelo agonizante clítico *se_{arb}* e pelo pronome *eles*, preferencialmente pleno. No outro extremo, temos a categoria [+3a pessoa/+singular], representada pelo clítico *se_{gen}*, pela estratégia zero (com o verbo na 3ª pessoa do singular) e *você*, que é preferencialmente pleno, **podendo incluir ou não o falante e o ouvinte**. Finalmente, temos a categoria [+1a pessoa/+plural], **que inclui necessariamente o falante**, representada por *nós* e *a gente*, com considerável vantagem de *a gente* sobre *nós*. A variação em cada categoria disposta ao longo da escala não é um fenômeno estável: em cada ponto há um forte competidor para representar cada grau de referência indeterminada, à medida que a mudança avança. PALAVRAS-CHAVE: sujeitos indeterminados; referência arbitrária; referência genérica; Parâmetro do Sujeito Nulo; variação morfossintática

Introduction

In the past three decades, several papers on the representation of definite reference pronominal subjects in Brazilian Portuguese (BP) – cf. Duarte (1993, 1995, 2012), among others – have shown that BP prefers the structural position of the subject filled by a pronoun, as shown in the examples in (1) for first and second person and in (2) for third person with [+/-human] referents.

- (1) a. Mesmo que **eu** não fizesse o pré-vestibular, **eu** acho
 even that I not take-Past.Subj the preparatory-course I think
 que **eu** passaria por causa da base que **eu** tinha.
 that I would-pass because of-the base that I had
 ‘Even if I had not taken the preparatory course, I think I would have
 passed because of the base I had.’

- b. **você** me disse que **você** está morando em copacabana.
 you me-cl told that you are living in copacabana
 ‘you told me that you are living in copacabana.’
- (2) a. [**Essa minha tia**]_i que mora aqui, **ela**_i é solteirona e eu acho que **ela**_i é
 This my aunt who lives here, she is spinster and I think that she is
 super-feliz, sabe? (...) **Ela**_i é uma pessoa que ajuda os outros pra caramba
 super-happy, know-2SG? She is a person that helps the others a lot.
Ela_i ficou solteira porque **ela**_i quis.
 She stayed single because she wanted
 ‘This aunt of mine who lives here, (she) is a spinster and I think she is
 very happy. She is a person who helps everyone. She remained single
 because she wanted to.’
- b. [**a casa**]_i virou um filme quando **ela**_i teve de ir abaixo.
 the house became a movie when she had of go down
 ‘The house became a movie when it had to be demolished.’

Cases of null referential subjects are restricted to certain structural contexts, such as control structures, with the antecedent as a subject in the previous clause, as in (3), or in an adjacent sentence, as in (4) – cf. Modesto (2004; 2008) –, but they are always in variation with overt subjects, which already outnumber null subjects (cf. DUARTE, 2015):

- (3) a. mas **ele**_i sentiu [que \emptyset _i era o único novo ali, recém-casado]
 but he felt that was the only young there, newly-married
 ‘But he felt he was the only young guy there, newlymarried.’
- b. [**Esse filme**]_i emocionou muita gente [quando \emptyset _i ficou pronto]
 That film touched many people when was ready
 ‘That film touched many people when it was shown.’
- (4) a. **Geisel**_i era realmente um poderoso presidente da república.
 Geisel_i was really a powerful president of-the republic.
 \emptyset _i Demitiu ministros militares...
 fired ministers militaries
 ‘Geisel was in fact a powerful president. He even fired military
 ministers...’

- b. [O Rio de Janeiro]_i é uma beleza! Realmente Ø_i é uma cidade linda
 The Rio de Janeiro is a beauty! Really is a city beautiful
 ‘Rio de Janeiro is a beauty. Really! It’s a beautiful city.’

Such facts have been associated to a change in the setting of the Null Subject Parameter (NSP) in BP, as a consequence of changes in the system of nominative pronouns, which triggered a significant reduction in the verbal inflexional paradigm (see DUARTE, 1993).

Following this path, researches on the representation of indeterminate reference subjects (DUARTE, 1995, CAVALCANTE, 1999, 2007; COUTO 2004, VARGAS 2012), including what we refer today as arbitrary and generic reference (EGERLAND, 2003; HOLMBERG, 2005 among others), propose that, as a “side effect” of the changes in the NSP in BP, the system would have come up with alternative strategies to the ones described by the grammatical tradition, which are restricted to structures with the verb in the third person plural with a null indeterminate subject and sentences with the pronominal “indefinite” clitic *se*, the same strategies characterizing European Portuguese (EP). In fact, these studies have demonstrated a progressive decline in the use of standard strategies and an increasing frequency of sentences with an overt nominative pronoun, especially *você* (you)¹ and *a gente* (the people = we), expanding their use for second and first person plural reference, respectively, to express generic reference, as shown in (5):

- (5) a. Quando **você** é menor, **você** não dá muito valor a essas coisas.
 When you are little you not give much value to these things
 ‘When you are a child you do not value such things.’
- b. mas na época **a gente** não podia acreditar... **a gente** não acreditava
 but at-the time the people not could believe the people not believed
 nisso porque **a gente** era novo.
 in-that because the people was young
 ‘but then we/one could not believe that... we/one did not believe that
 because we were young.’

¹ Regions where *tu* is preferred over *você* to refer to 2nd person also prefer *tu* for generic reference. Even in Rio de Janeiro, where both pronouns are in variation, generic *tu* is usual, being attested in popular varieties. (DUARTE, 2003).

On the other hand, a new sentence pattern related to indeterminate reference, first noticed by Galves (1987) and later confirmed by other studies (DUARTE, 1995; CAVALCANTE 2007; LUNGUINHO and MEDEIROS JUNIOR, 2009, among others)² would exhibit the use of a null subject with the verb in the third person singular, illustrated in (6):

- (6) a. \emptyset_{gen} não pode entrar de sapato; \emptyset_{gen} fica de meia.
 not can.3sg.pres enter of shoe; stay of sock
 ‘One cannot enter (the museum) with shoes on; one keeps only socks.’
- b. \emptyset_{gen} Não usa mais máquina de escrever.
 not use.3SG.PRES more machine of to-write
 ‘No one uses typewriters anymore.’
- c. \emptyset_{arb} Construiu um Bob’s na Tijuca.
 built.3SG.PAST a Bob’s in-the Tijuca
 ‘They built another Bob’s in Tijuca.’ ‘Another Bob’s has been built in Tijuca.’
- d. \emptyset_{arb} Vai reformar o Centro todinho por causa da
 is-going.3ps.pres to-reform the center whole because of-the
Olimpíada.
 Olympic-Games
 They are going to rebuild all downtown because of the Olympic Games.’

Lunguinho and Medeiros Jr. (2009), in particular, associate the implementation of such strategy in BP to three facts: (a) changes in the pronominal system, with the consequent weakening of subject/verb overt agreement, as already mentioned; (b) the progressive loss of pronominal third person clitics, especially the reflexives (D’ALBUQUERQUE, 1984)³; and (c) the rigidity of the order SVO. VS order is

² This null generic subject, here referred as “zero strategy”, has been considered a property of partial null subject languages being referred to as “null generic subject” (HOLMBERG, NAYUDU and SHEEHAN, 2009; HOLMBERG (2010). Notice, in addition, that examples (6c) and (6d), pointed out by Lunguinho and Medeiros (2009), in fact, convey arbitrary reference, contrary to the null generic subjects in (6a) and (6b). They resemble a simplified version of a passive structure.

³ Lunguinho and Medeiros Jr (2009) mention the loss of the system of reflexive clitics; however, this is a process of change observed in some regions of the country, which prefer a null reflexive, including

restricted to sentences with unaccusative structures, including unaccusative verbs and lexical passives; in both cases, however, the internal argument bearing nominative case is, in fact, interpreted as a complement, which is evidenced by the lack of verbal agreement.

Taking into account all these changes, the aim of this paper is (a) to show the relation between the changes in the NSP in BP towards an overt subject grammar and the rise of new strategies of subject “indetermination”, considering all the changes related to this new configuration of the system (change in the pronominal system, reduction in the number-person distinctive morphemes in the inflectional paradigm, and the weakening of subject/verb agreement); (b) to evaluate the productivity of the third person singular strategy (which will be referred to as zero strategy as in Duarte (1995) in a system that clearly prefers overt subjects, bearing in mind that its emergence seems to be more recent, and to explore possible motivations for its origin⁴⁵; and (c) to try to dispose the strategies attested along a *continuum*, which has arbitrary reference at one end, and different degrees of generic reference till the other end of the *continuum*. In this way, we will show that they are not in variation as a single set; rather, they have specialized to convey different degrees of indetermination. And, even in each point of this *continuum*, the strategies in competition are morphosyntactic doublets, in the sense of Kroch (1994), following Aronoff (1976), which means that one form is bound to disappear.

Our theoretical framework associates the Theory of Language Variation and Change (WEINREICH; LABOV; HERZOG, 1968) and the Principles and Parameters Theory (CHOMSKY, 1981, 1995) – this necessary grammatical component guides our analysis from the hypothesis raised to the interpretation of the empirical data. This, in turn, will allow us to find answers for the “embedding” problem, crucial to the model of language change used here.

As Duarte’s (1995) empirical research is probably the first to relate overt indeterminate subjects with overt referential subjects (see DUARTE’s chapter 3), our analysis will revisit and review her data in the light of recent discussions concerning the

argumental ones; third person accusative and dative clitics as well as arbitrary/generic clitic *se*, on the other hand, are in process of extinction in spontaneous speech in all regions investigated – see D’Abuquerque (1984) and Duarte and Ramos (2015).

⁴ This strategy has been referred to as a “null generic pronoun” (HOLMBERG, 2005, and subsequent work). As we will show, it can convey generic or arbitrary meaning in BP.

⁵ See Galves’s (1987) hypothesis, according to which it was exactly the impossibility to identify a definite third person null subject that allowed a generic interpretation for the null subject.

NSP. Her analysis was based on a sample comprising 12 graduated speakers, recorded in 1992, stratified for age and gender, as part of the Norma Urbana Culta – Rio de Janeiro (NURC-RJ) *corpus*. The statistical analysis used the Goldvarb X Program (TAGLIAMONTE, 2006). Also, a qualitative analysis was also necessary to identify the discursive contexts, which could be favoring the specialization of the attested indetermination strategies. Therefore, besides extending the data in this “revisit” to Duarte’s sample, discursive factors have also been considered to refine the analysis.

1. The change in the representation of indetermination: relevant facts

1.1. The restructuring of the pronominal system in BP and the changes in the inflectional verbal paradigm

BP experienced a significant restructuring of its pronominal system of nominative forms particularly triggered by grammaticalization processes undergone by the originally nominal address form *Vossa Mercê* – Your Grace – Your Mercy, and the collective DP *a gente* (literally “the people”). Both entered the system as fully grammaticalized pronouns for second (*você* = you) and first person plural reference (*a gente* = we), giving rise to a competition with the former pronouns *tu* (2ps.) and *nós* (1pp.), as shown in Chart 1, adapted from Duarte (1993):

	Pronouns	Paradigm1 19th Century	Paradigm 2 20th Century/1	Paradigm 3 20th Century/2
1PS	eu	estudo	estudo	estudo
1PP	<u>nós</u>	estudamos	estudamos	<u>estudamos</u>
	a gente	-	estuda	estuda
2PS	tu	estudas	estudas	estuda(s)
	você	-	estuda	estuda
2PP	vós	estudais	-	-
	vocês	-	estudam	estuda(m)
3PS	ele, ela	estuda	estuda	estuda
3PP	eles, elas	estudam	estudam	estuda(m)

Table 1. Nominative Pronouns and inflectional paradigms in Brazilian Portuguese (verb “to study” – *estudar*)

It is clear that BP, in earlier stages, exhibited, according to written language (popular plays and letters) six different verbal morphemes. From the 20th century on, with the complete replacement of *vós* (2pp) by *você*s, the competition between *tu* and *você* (2ps) and between *nós* and *a gente* (1pp), the inflectional paradigm was gradually reduced to four, three and sometimes two oppositions (see DUARTE and SOARES da SILVA, 2016, for a comparison of Romance inflectional paradigms and their role in the expression of definite reference subjects). In the first place, former nominal expressions combine with third person, unmarked verb form; in addition, phonological processes have affected the second person singular morpheme <-s> associated with the verb form combined with *tu* (you) and second and third plural <-m>, the latter being subject to variation according to years of school attendance and monitoring of speech (SCHERRE and DUARTE, 2016).

According to Duarte (1993), until the first quarter of the 20th century, *você* still preserved its original value of an address form and was rarely used in theater plays. Such a situation is confirmed by Souza (2012), who analyzes a sample of letters written in Rio de Janeiro between 1870 and 1970: the pronoun *você* was initially less productive than the pronoun *tu*. Around the 1930s the two pronouns reach similar rates of use, but, more important, they were not in variation, since each form had a correspondent distinct use: *tu* was used in familiar contexts and *você* maintained its original mark of courtesy. From the 1930s on, the use of *você* outnumbers the use of *tu*, which starts to decrease, at least in the samples analyzed⁶⁷. In addition, the complementary distribution is lost: *você* starts to be used in the same contexts as *tu*, an evidence of the fully grammaticalization of *você* as a real second person pronoun.

Considering Chart 1 and Souza's results, we can say that the history of PB exhibits three subsystems for second person reference: one with the pronoun *tu*; another in which *tu* and *você* coexist; and a third one, in which *você* is the predominant form (see LOPES and CAVALCANTE (2011). More recently, Scherre et al. (2015) identify 5 subsystems, refining the type of coexistence of *você* and *tu* in different areas of the country.

With respect to the variation between the innovative pronoun *a gente* and the conservative pronoun *nós* for first person plural reference, we can find a less advanced

⁶ Other address forms have been introduced in the system, such as *o(a) senhor(a)* (the sir/the madam), *Vossa Excelência* (Your excellency), whose use is each time more restricted.

⁷ In fact, the pronoun *tu* has not been abandoned in the metropolitan area of Rio de Janeiro. Presently, its use is expanding to more central areas of the city (see PAREDES SILVA, 2003).

process than the one shown for the representation of second person. But the propagation of the change proceeds steadily across age groups, particularly in the function of subject, as pointed out by Lopes (2007, p. 68), “the replacement of *nós* by *a gente* is progressively being carried out in the last 30 years in Rio de Janeiro, not only among graduated (cultured) speakers, but also among non-cultured speakers”⁸. As shown by Callou and Lopes (2004), *a gente* is already implemented in the system since the 1970s⁹, when it outnumbers the overall rates of *nós* in the speech of graduates in Rio de Janeiro, even though older groups still show a competition between the two pronouns. Duarte’s 1992 sample for graduates from Rio de Janeiro reveals the progress of the change: the older group (over 56 years of age) still prefers *nós* (62%), the intermediate group (36-55) reaches 50%, and the younger group (25-35), only 2%. There is no doubt that *a gente* is the new 1st person plural pronoun in spoken BP today – a mapping of this phenomenon in the country can be found in Vianna and Lopes (2015).

1.2. Resetting the Null Subject Parameter

Duarte’s (1993) diachronic research examining theater plays comprising the 19th and the 20th centuries, distributed in seven periods, presents evidence of a parametric change in process. PB is clearly moving from null to overt referential subjects.

Figure 1, adapted from Duarte (1993), shows the progressive decrease of null subjects in the corpus analyzed:

⁸ Our translation from the original: “a substituição de *nós* por *a gente* se está efetivando progressivamente nos últimos 30 anos, seja entre os falantes cultos, seja entre os não-cultos no Rio de Janeiro” (LOPES, 2007, p.68).

⁹ This study was based on a sample of popular speech and college-graduated speakers which belong to PEUL and NURC-RJ projects, respectively; speakers were recorded in two different moments: 1970s and 1990s (NURC), available at: www.lettras.ufrj.br/nurc-rj, and 1980s and 2000s (PEUL), available at: www.lettras.ufrj.br/peul.

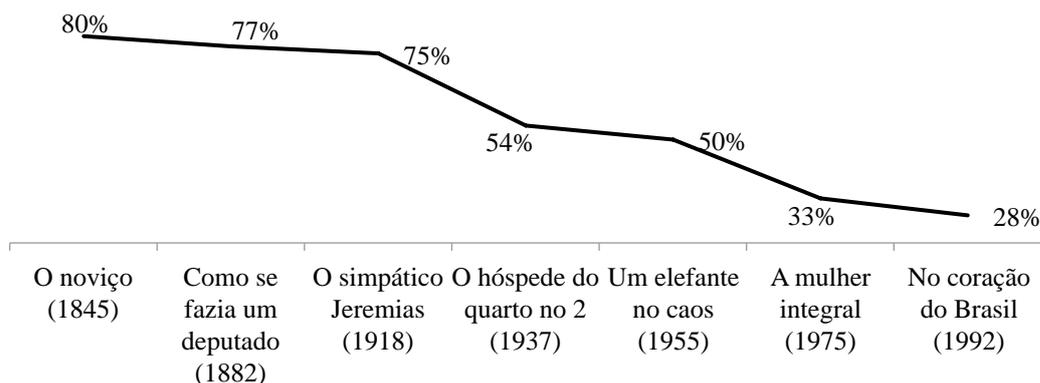


Fig. 1: Null subjects in theater plays along seven synchronies (19th and 20th centuries)¹⁰

Comparing the results in Figure 1 with the three paradigms illustrated in Chart 1, we can visualize three stages in the process of implementation of the change towards overt pronominal subjects: during the first three synchronies, which correspond to Paradigm 1, covering the 19th century and the first quarter of the 20th century – to which the author refers to as a “formally rich” paradigm, using Jaeggli and Safir’s (1989) proposal –, we see a consistent null subject grammar. The next two synchronies coincide with the first significant reduction of the verbal morphology, a consequence of the grammaticalization of *você*, which affects the identification of null subjects. However, null subjects are still identified, in a system with a “functionally rich” paradigm, with a reasonable number of ending distinctions and syncretisms (ROBERTS, 1992). In the last quarter of the century, the insertion of pronoun *a gente*, fully grammaticalized and combining with the unmarked third person verb morphology (zero), would bring about the complete loss of the “Avoid Pronoun” principle, which characterizes Romance Null Subject Languages (CHOMSKY, 1981; RIZZI, 1982).

It is important to add that the change affected first and second person faster than third person; therefore, licensing of null subjects is no longer related to a distinctive ending. Subsequent investigations would show the role played by the semantic feature [+human] in this process (CYRINO, DUARTE and KATO, 2000); it explained the quick implementation of overt first and second person subjects, inherently [+human], and also third person subjects bearing this feature – Duarte, Mourão and Santos (2012), re-examining an enlarged sample of theater plays, confirm that. Subjects bearing [-

¹⁰ Each period is represented by one play (indicated, in the chart, by its title and year of publication or first performance), written in Rio de Janeiro, by very popular playwrights born in the city.

human] feature undergo the change more slowly, but overt non-human subjects already compete with overt ones, a feature that does not fit a null subject Romance language. Besides the semantic feature, a null third person subject is more constrained by co-reference with the subject of the main clause or the presence of an identical antecedent in an adjacent clause bearing the same function – such contexts are illustrated in (3) and (4) in the Introduction to this article. This would be taken as an evidence of topic prominence in allowing null subjects in BP (MODESTO, 2008; HOLMBERG, NAYUDU and SHEEHAN, 2007; HOLMBERG and SHEEHAN, 2010). This is another structural factor to delay the fully implementation of the change in External-Language data. Synchronic samples, however, confirm that overt subjects already outnumber null subjects in every structural context (DUARTE, MOURÃO and SANTOS, 2012). Since indeterminate subjects are inherently [+human], one would expect a side effect of the change in progress in favor of overt indeterminate pronominal subjects. This is what is shown by Duarte's (1995; 2003) synchronic analyses and Vargas' (2012) diachronic study, which motivate the present article, as will be shown in the next sections.

2. Analysis

2.1. The gradation between arbitrary subjects and generic subjects

Up to this point, we have been dealing with subject indetermination strategies without taking into account possible discursive differences, just as it has been made by the researchers who studied subject indetermination using the Generative Theory (DUARTE, 1995; CAVALCANTE, 1999, 2007; COUTO 2004; VARGAS 2012). The referred papers were interested in the preference for overt pronominal subjects to null subjects rather than to establish any kind of distinction among the several possibilities. However, Duarte (1995) already mentions the zero strategy (sentences with the verb in the third person singular and null subject; see note 5) as constrained by the expression of aspect, modality and procedure discourse, which is confirmed by Cavalcante (2007). A refinement of such constraint would be offered by Lunguinho and Medeiros Jr. (2009), who add another possible reading to sentences with the verb in the third person singular and null subject (the zero strategy) to the ones already pointed out. The already mentioned meanings would receive the label “generic” (7a) and “episodic” (7b), which is not an instance of generic subject but could be included in the arbitrary group –

already instantiated in (6c-d) – and would be associated to the tense of the verb¹¹.

- (7) a. Ø Não tá contratando gente ainda não.
 not is hiring people still not
 ‘They are not hiring workers yet.’ ‘Workers are not been hired yet.’
- b. Ø Montou o armário lá de casa semana passada.
 set-up.3SG.PAST the wardrobe there of home week passed
 ‘They set up our wardrobe last week.’ ‘Our wardrobe was set up last week.’

Our concern in this article is not only to observe the zero strategy but the set of indetermination forms so as to find out their discursive motivations and their possible distribution along a gradient between two poles: arbitrary reference and generic reference.¹² By *arbitrary reference*, we mean subjects that cover a finite set of entities or an individual, whose identification cannot be individualized in the discourse, as we can see in (8) below:

- (8) a. Uma vez Ø_{arb} me **definiram** Austrália como a Inglaterra de
 one time me.CL.1SG defined.3PL Australia as an England of
 bermudas (...) bermudas
 ‘Once they defined Australia to me as England in bermuda shorts.’
- b. Então, (..) Ø_{arb} **criou-se** em 1920 ou 21, a Universidade do Brasil
 then created.SE.CL in 1920 or 21, the University-of-the Brazil
 ‘Then, in 1920, 1921, the University of Brasil was created.’

¹¹ The authors consider that the distinction is not associated to verb tense only, but with aspect as well, which is not new. In this paper, we will show that the aspect of the verb is really a factor that contributes to characterize the different nuances of subject indetermination.

¹² As far as we can recall, the use of such distinction (which has its origin in several semantic studies in the second half of the 20th Century) to refer to arbitrary and generic indetermination has become common since Holmberg (2005).

- c. Quando você termina o ciclo básico você faz outro vestibular.
 when you finish the cycle basic you tke another examination
 Mas não é um vestibular, né? **Ø_{arb} vai** pelo CR,
 but not is an entrance-examination, right? go.PRES.3PS by-the GPA
Ø_{arb} classifica pelo CR.
 classify.PRES.3PS by-the GPA
 ‘When you finish the basic course you go through another sort of
 entrance examination, see? But it’s not a real examination, they rank the
 student by the GPA.’

As shown by the examples in (8), although the subjects highlighted in the sentences cannot be determined by any element in the text/discourse, it is possible to establish a limit for their scope. In (8a), the speaker does not make it clear who defined Australia as “England in bermuda shorts”, but it is possible to infer that there was someone who said that¹³. In (8b), even if they do not tell who created the University of Brazil, there was a person – or an institution – who did it, also limiting the scope of the non-defined subject character. Finally, in (8c), it is interesting to notice the established contrast between the reference of the pronoun *você*, which appears as the subject of the clauses in the previous sentence, and the null subjects of the third person singular verbs in bold that follow. In this case, the speaker refers to the way the institution selects the students who should go from the basic cycle to the professional cycle, without identifying the agent. This makes the reference of the null subject of *ir* (to go) and *classificar* (to classify) more limited, although not precise.

At the opposite pole, we identify the **generic reference** subjects, which correspond to those whose reference cannot be determined by the text/discourse, being unlimited, in the sense that it can refer to an infinite set of entities. The examples in (9) show that:

¹³ It is also possible that this consideration might have been made by more than one person, but even so, there is a finite group of entities.

- (9) a. Mas agora \emptyset_{gen} não se tem mais inverno. Você vê que hoje
 but now not SE.CL has more winter. You see that today
 já estamos no inverno e tá um calor....
 already are.1PL in-the winter and is a heat
 ‘Nowadays one does not have winter any more. You see that we are
 already in the winter and it is so hot.’
- b. E tinha sorveteira que batia em casa. Tinha uma
 and there-was ice-creamer that beat.3SG.PAST at home. Had a
 hélice assim (...). \emptyset_{gen} Fazia sorvete em casa.
 propeller like-this (...) used-to-make.3SG icecream at home
 ‘And there was an appliance to make ice cream at home. There was a
 propeller (...) One used to make ice-cream at home.’
- c. Lá na América também eu cozinhava porque... lá você
 there in-the America also I cooked because...there you
 encontrava tudo pronto ou semi-pronto. Você punha as coisas
 found everything ready or semi-ready. You put.PAST the things
 no forno aquilo ia aparecendo a comida.
 in-the oven that would-go appearing the food
 ‘In the USA I also cooked because... there you already found everything
 ready or semi-ready. You would put the things in the oven and the food
 began do appear.’

In the three cases illustrated in (9), the subjects/agents cannot be identified and show an unlimited reference: in (9a), the winter is no longer noticed by *any carioca* or even by *anyone* who stays sometime or lives in Rio de Janeiro; in (9b), any person could make ice cream at home; and in (9c), the pronoun *você* (you), which does not allude to the interlocutor in this case, refers to anyone who was in America at that time and, therefore, could find ready-made food and put it in the oven.

It is interesting to notice that sentences with the indeterminate pronominal clitic *se* and with verbs in the third person singular and a null subject can convey both arbitrary and generic meaning. What seems to determine one or the other possibility is the interaction of the feature [+3rd person] with the aspect of the verb, as suggested by Lunguinho and Medeiros Jr. (2009). If the verb is perfective, the idea obtained will be

arbitrary, as illustrated by the zero strategy in (8a, b). On the other hand, the interaction with the imperfective aspect, illustrated in (9a, b), provides a generic reading of the subject, similar to the one obtained with *você* in (9c) and in the two first sentences in (8c) above.

In intermediate points of the gradient proposed here, there are subjects whose reference has a certain degree of specificity, which cannot be determined by the text/discourse, although it applies to an infinite set of entities, as in (10):

- (10) a. O que eles vão fazer depois é outra história, né? Brizola tá fazendo coisa pra caramba, a gente tá vendo que ele tá fazendo. doing thing too much the people (we) are seeing that he is doing ‘What they are going to do later is another story, but Brizola is doing a lot of things, we can see it.’
- b. (...) o Rio de Janeiro é uma cidade violenta, todo mundo diz. Nós ficamos nos enganando, dizendo: Ah, Nova York tem também violência also violence ‘... Rio de Janeiro is a violent city, everyone says so. We keep deceiving ourselves, saying: Well, New York also has violence’
- c. (...) a situação mais formal é quando você tem que tratar, com pessoas que hierarquicamente estão, acima, né? Então, se eu vou falar com o Senador, que se eu vou conversar com o Governador, eu tenho que tratar de uma forma, Vossa Excelência e tal (...) to address of a way Your Excellency and so ‘The most formal situation is when you have to address people who are hierarchically above you.... Then, if I’m talking to the Senator, to the Governor, I have to use a different address form, such as Your Excellency.’

Analyzing the examples in (10), we see that the reference of the subjects highlighted applies to an infinite set of entities, as shown in the examples in (9). The difference resides in the fact that in (10a) and (10b), the pronouns *a gente* (the people) and *nós* (we) include the speaker in the group of possible entities, which becomes the reference more specific. The same can be said about the example (10c), in which the pronoun *eu* (I), which associates the features [+1st person/+singular], used by the speaker to raise a hypothesis about the way any person should address another hierarchically superior using himself as an example, makes the reference even more specific.

It is understood, thereby, that the label “indeterminate subject” comprises a wider range of references, considered here as a scale determined by two opposite poles: the arbitrary reference and the generic reference. Between them, there would be at least two points in which the references that include necessarily the speaker would be located. It is also possible to think that the pronominal strategies to indicate indeterminacy of the subject in BP could be defined by a set of features – combining person and number – producing different forms to cover them. That is the idea defended in this paper. With respect to the interaction of these features and the possible forms of indetermination, we can, therefore, assume that the arbitrary interpretation is conveyed by the presence of the features [+3rd person/+plural] and it seems to be blocked by the features [+1st person/+2nd person]. When the strategy presents the feature [+3rd person], the interaction with the aspect of the verb seems to determine the arbitrary or the generic reference.

2.2. Data analysis related to reference: positioning the strategies on the scale

2.2.1. The standard strategies

Grammatical tradition includes both the third person plural with a null subject and the use of the clitic *se* as the subject indetermination mechanisms (the influence of European Portuguese is obvious). Lunquinho and Medeiros Jr. (2009) suggest, however, that these two strategies would not be interchangeable anymore having become specialized in the system: the third person plural would be associated to the episodic reference and the clitic *se*, to a more generic reference, in the authors' words. We assume that, in some period of BP history, the two strategies could have both types of reference (which is in fact shown in some empirical studies, as in Cavalcante's 1999

- b. Nessa rua que eu moro atualmente, **eles** construíram uma série de in-this street that I live now they built a series of prédios novos, com essas varandinhas que tão na moda. buildings new with these little-balconies that are in-the fashion ‘In the street I live now, they have built a number of new buildings with these small balconies so in fashion.’

A qualitative analysis of the structures with the clitic *se* showed an interesting result, shown in Table 2:

	Arbitrary	Generic	Total
Age group 3	1 occurrence – 4%	24 occurrences – 96%	25 – 100%
Age group 2	5 occurrences – 42%	7 occurrences – 58%	12 – 100%
Age group 1	–	2 occurrences – 100%	2 – 100%

Table 2: Distribution of clitic *se* according to reference in “apparent time”

Table 2 attests the unproductivity of clitic *se* to express arbitrary reference (hereafter referred as *se_{arb}*): there were no occurrences of this type in the speech of the younger group; the six cases attested in the sample occur in the speech of the intermediate age group – 5 occurrences - and only one in the speech of the older group. Such instances of *se_{arb}* can alternate with arbitrary third person plural strategy with the perfective past of the indicative tense (12a) or the perfective past reading of the indicative present form (12b).

- (12) a. Os produtos não vinham por causa da guerra. \emptyset_{arb} **Montou-se** the produces not came because of-the war. Created.3PS-SE.CL às pressas algumas indústrias de alimento, de roupas, de calçados pra quickly some industries of food, of clothes, of shoes to substituir esse produto que a classe brasileira importava. replace this product that the class Brazilian imported ‘The commodities were scarce because of the war. Industries were quickly created to supply the market with food, clothes and shoes imported by Brazil.’
- b. Então é todo um momento histórico muito importante no qual then is all a moment historical very important in-the which

\emptyset_{arb} **se cria** o partido comunista brasileiro (...)

SE.CL create.3SG the party communist Brazilian

‘Then it is a very important historical moment, in which the Brazilian communist party is created.’

On the other hand, the 33 occurrences exhibiting se_{gen} data convey the imperfective aspect, as we can see in (13):

- (13) a. A greve pra mim [...] \emptyset_{gen} só **deve se deflagrar** quando não
the strike for me [...] only should SE.CL deflagrate when not
houver mais recurso nenhum. Mas é um recurso válido que \emptyset_{gen} **se**
there-is more resort none. But is a resort valid that SE.CL
usa.

use.3SG.PRES.

‘In my opinion, the strike should be used as a last resort. One may only spark a strike when there is no appeal anymore. But it’s a valid resort one can use.’

- b. Quando eu era criança, \emptyset_{gen} **punha-se** a mesa pra tomar lanche.
When I was child set.3SG.PAST-SE.CL the table to have snack
‘When I was a child one would set the table to have an afternoon snack.’

As we have shown, the data does confirm the specialization of the two subject indetermination strategies regarded as standard: the strategy that bears the features [+3rd person/+plural] has specialized to represent arbitrary reference whereas the clitic *se*, bearing the features [+3rd person/+singular], clearly specializes for the generic reference; the rare uses of se_{arb} confirm that they are residual. This indicates that both strategies are no longer in variation, possibly constituting a case of complementary distribution in BP. However, we must not forget that even se_{gen} is disappearing in the speech of younger generations, as shown in Figure 2 and in Table 1.

2.2.2. The non-standard strategies

Considering that third person plural represents arbitrary reference and that clitic se_{gen} decreases significantly, we should expect the rise of an alternative form to represent this

type of reference. And also considering that third person clitic paradigm in BP (including accusative, dative and “indefinite” clitics) is clearly disappearing, one would expect overt nominative pronouns to represent indeterminate subjects. However, taking into account our approach, the new strategies would fit in different categories located in one point of our scale: (a) [+3rd person/+plural] category represented by *pro* + 3rd person plural verb or overt pronoun *eles* (they); (b) [+3rd person/+singular] category, represented by *se*, *pro* + 3rd person singular verb and *você* (you); and finally (c) [+1st person] and [+singular] or [+plural] category, represented by *eu* (I), *nós* (we) and *a gente* (the people = we). This makes it clear that we do not consider the different strategies as a single set. The idea is that indetermination, generally speaking, can be represented by the three categories characterized by a combination of person and number features mentioned in (a)-(c) and that the different strategies belonging to each category are in fact in variation. Considering that each group encodes different information from the ones that do not exhibit the same set of features, we start analyzing sentences with *pro* and verb in the third person singular – hereby referred to as *zero* strategy – and the pronoun *você*, in the face of the use of the conservative clitic *se_{gen.}*

- **The [+3rd person/+singular] group**

Similarly to what happens to the clitic *se* (as shown in Table 1), the zero strategy can express both arbitrary and generic reference. From the 40 occurrences of zero, 32 conveyed generic reference and only 8, arbitrary reference (see (8c) above). On the contrary, the use of *você* – in 387 occurrences – is always associated with generic reference, which may be explained by the presence of the [+2nd person] feature present in the original address form that underwent grammaticalization as mentioned in 1.1. Figure 3 presents the distribution of the three strategies with generic reference – *se*, zero and *você* –, through the three age groups.

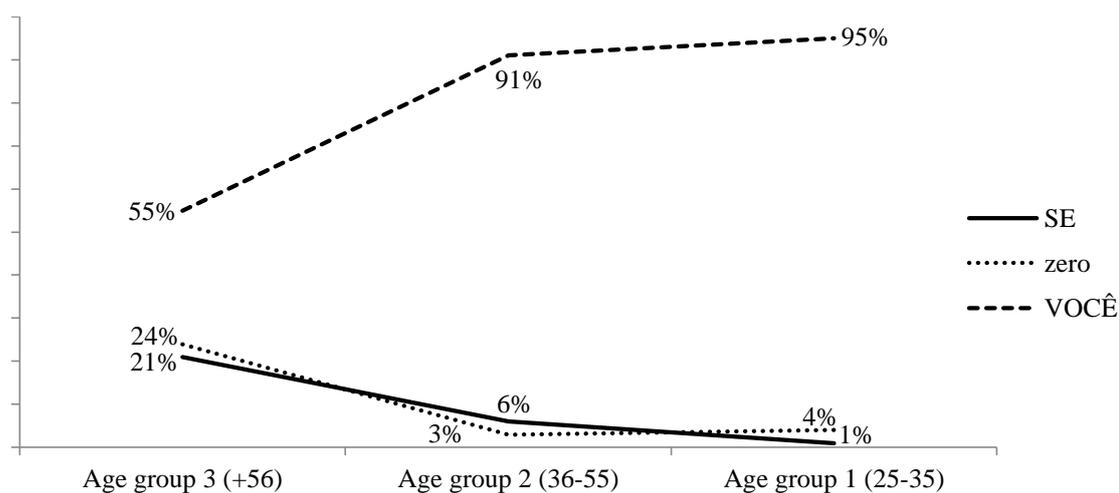


Fig. 3: Distribution of generic *se*, zero and *você* in “apparent time”

The distribution clearly reveals that the preferred strategy to replace *se_{gen}* is *você* and not the zero strategy, as suggested by recent analyses. Both *se* and zero show a significant decline, ranging from 21% and 24%, respectively, in the speech of the older group to an average of 4,5% and 2,5% among the speakers in the two younger groups. *Você*, on the other hand, with the highest rates among all speakers in the sample, shows an expressive increase across the three generations, reaching 95% in the speech of age group 1.

To explain such scenario, we should turn our attention to the possible origins of the zero strategy for subject indetermination. It is possible that, at some point in the evolution of the ongoing change, the identity between “synthetic” passive voice sentences and active voice sentences, both with the clitic *se*, associated with the reduction of the clitic paradigm, have favored the emergence of the zero strategy¹⁴. On the other hand, this situation would not fit a system that progressively presents clear preference for overt subjects. Therefore, it is possible that the expansion of the use of *você* is the solution the system has found to assure the generic interpretation and to avoid the empty subject position. Considering that *você* presents the second person feature, the arbitrary interpretation would be blocked, being restricted to the zero strategy, taking into account the interaction with the verbal aspect.

With respect to the context in which the form *se* still resists in the speech of older

¹⁴ We take Cavalcante’s (2007: 72) Table 3, which opposes only the zero and the *se* strategies (disregarding *você*), as an empirical evidence for this. Such table shows that the zero strategy rates increase in apparent time, replacing *se*, which decreases from older to younger speakers, an expected result. This does not mean that it outnumbers generic *você*

generations, the transitivity of the verb seems relevant. All occurrences of *se* (39 tokens) – either generic or arbitrary – exhibit a transitive verb, selecting an accusative or an oblique complement, as we can see in (14):

- (14) a. \emptyset_{gen} **Jogava-se** futebol na Rua Visconde Silva
 used-to-play.3SG-SE.CL soccer in-the Street Visconde Silva
 ‘One used to play soccer on Visconde Silva St.’
- b. Você colocava as suas jóias pra ir trabalhar, pra ir passear, pra ir
 you put-on the your jewelry to go to-work, to go for-a-walk, to go
 à rua, normalmente, sem problema nenhum. Nem \emptyset_{gen} **se**
 to-the street normally, without problem any. Not SE.CL
pensava em roubo
 used-to-think in theft
 ‘You put on your jewelry to go to work, to take a walk, to go out, without
 any kind of fear. One would not worry about theft.’

Taking into account the perspective we are adopting, based on Lunguinho and Medeiros Jr. (2009), the expectation would be that the tokens with the zero form would also occur with “direct” transitive verbs, as described by the authors. Therefore, if the zero strategy occurs only with “direct” transitive verbs, *você* would cover the other kinds of verbs. And it is what we find: considering the 40 tokens of the zero strategy with both types of reference, 36 appear with transitive verbs, most of them “direct” transitive, but there were still some occurrences of the zero form with transitive verbs which select an oblique internal argument, as in (15):

- (15) a. Tinha orquestra, \emptyset_{gen} não **precisava** daquele pra tocar.
 had.3SG orchestra, not needed.3SG of-that to play
 ‘There was an orchestra, one did not need that to play.’
- b. E as casas eram muito generosas de espaço, dava pra você
 and the houses used-to-be very generous of space, gave for you
 fazer festas, quando terminava tudo, \emptyset_{gen} **chegava** na cozinha,
 to-make parties when finished all, arrived.3SG in-the kitchen,
 ainda tinha feito uma areazinha pra lavar roupa
 still had sort-of a little-area to wash clothes

‘And the houses used to be more spacious, you could make parties (...); after the kitchen, there was still a little space serving as a laundry’

Contrary to the expectations, however, we found four instances with generic zero and non-transitive verbs, as exemplified with the copula in (16).

- (16) Depende, entendeu? Porque eu tô namorando há quatro anos e meio.
 depends, see? Because I am dating for four years and a-half
 E aí \emptyset_{gen} fica meio dependente de namorado.
 and so gets.3SG sort-of dependent of boyfriend
 ‘It depends. Because I’ve been dating for four years and a half. Therefore, you get sort of dependent on your boyfriend.’

The situation described seems to reinforce our claim: the strategy with *se* is progressively replaced by zero and, especially, by *você*; the few occurrences of zero show a transitive verb. On the other hand, *você* spreads in the system, appearing with all types of verbs.

- **[+1st person] [+/-plural]**

As we said above, the feature [+1st person] seems to block the arbitrary reading. All the 134 tokens with indeterminate subjects bearing this feature, receive generic interpretation, as we can see in (17):

- (17) a. Quando **a gente** levanta a coisa é que **a gente** vê tudo que
 when the people investigate the thing is that the people see all that
 aconteceu
 happened
 ‘When investigate it all, we understand all that really happened,’
 b. **Nós** ficamos nos enganando, dizendo: Ah, Nova York
 we keep us.CL.REFLEX.1PL deceiving saying: Well, New York
 tem também violência.
 has also violence

- ‘We keep deceiving ourselves, saying: Well, New York also has violence.’
- c. É muito caro vir para o Rio de Janeiro. (...) Agora o que
is very expensive to-come to the Rio de Janeiro Now what
vamos fazer?
go.1PL to-do?
Ø_{gen} **Vou** ver a Amazônia, tem que pegar um avião,
go.1SG to-see the Amazon, have.3SG that to-take a plane, quarto
horas e tal
four hours and so
‘It’s very expensive to visit Rio de Janeiro. (...) What are we going to do
then? I want to go to the Amazon, you must take a plane. It’s about a four-
hour flight away.’

We can notice that *a gente* and *nós* share the same set of features ([+1st person], [+plural]), belonging to the same group or category. As expected, this competition, showed in Figure 4, is bound to disappear, with the victory of the innovative strategy, following exactly the same tendency observed for definite 1st person plural (LOPES, 1993, 2003, among others).

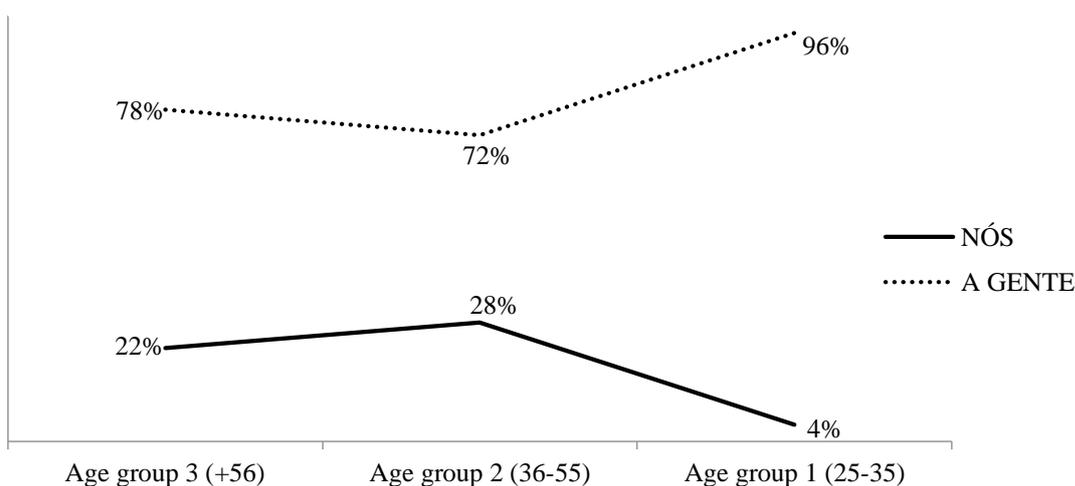


Fig. 4. Evolution of first person plural strategies in “apparent time”

As we can observe, the use of *a gente*, although completely implemented in the speech of all three age groups, already reaches 96% in the speech of the younger group.

This result confirms what was already attested by Duarte (1995), both for definite and indeterminate reference subjects, and by Vargas (2012), for indeterminate subjects along the 19th and the 20th centuries. This suggests that pronoun *nós* is in process of extinction in spontaneous speech.¹⁵

As for the use of [+1st person], [+singular], materialized by the pronoun *eu*, it cannot be explored here because we found only 4 tokens, all of them in the older age group, shown in (10c) and (17c). Therefore, it is not possible to make any sort of generalization about the use of this strategy, although the data show its generic character, with a higher degree of specificity than the first person strategies that combine with [+plural] feature.

Even though based on four tokens, 1st person singular *eu* seems to be used in situations in which the speaker offers an example to the proposed situation in the discourse, making the experience even more real, showing it could be lived by any person, including himself/herself, therefore limiting the generic reference.

Summarizing what has been said up to this point, we have proposed that BP has moved from a stage in which there were four possibilities to represent indeterminate pronominal subjects to a stage with seven possibilities. The strategies usually presented to express indeterminacy do not constitute a group nor can they be explained by the distinction between arbitrary *versus* generic reference only. We have a first stage with different categories disposed along a *continuum* as shown in Diagram 1, when BP exhibited four strategies (phase 1). As innovative strategies entered the system – *a gente*, *você* and zero –, what we find is variation in some points of the scale (phase 2). Our data suggest that in each point one of the forms will replace its competitors.

¹⁵ Writing recovers the clitic *se* and the 1st person plural *nós* (usually null) for indetermination (COUTO, 1994; DUARTE, 2007).

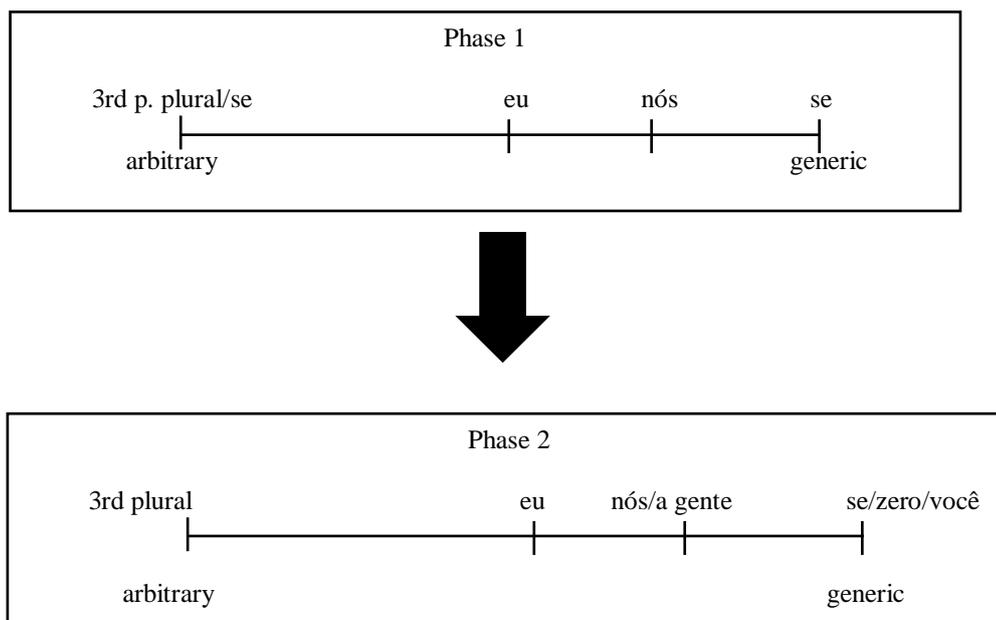


Diagram 1. Change in the representation of subject indetermination in two stages

Therefore, if our predictions are correct, it is possible that, in the future, BP will return to a reduced set of strategies for indetermination (phase 3), as illustrated in Diagram 2:

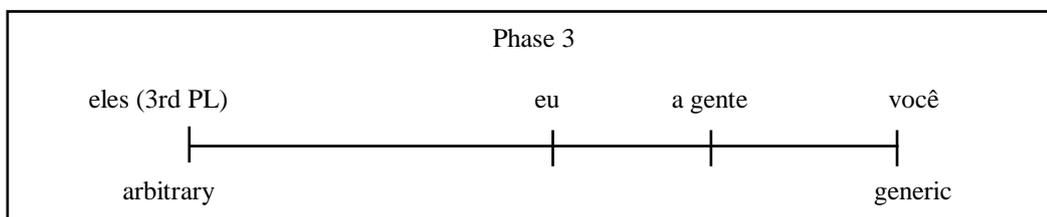


Diagram 2. Possible set of forms for subject indetermination in the future

2.3. Overt non-defined reference subjects: effects of the changes in the NSP

Another relevant factor concerning the observation of the data is the representation of the non-defined reference subjects. Duarte (1995) had already pointed out that, similarly to what happens to definite reference subjects, the “indeterminates” also presented higher rates of overt subjects (65% against 35% of null pronouns), especially among speakers in the younger group. In our revision of Duarte’s sample for the present analysis, we attested 84% of overt pronouns, an even more expressive percentage than that shown by Duarte

(1995).¹⁶ Table 2 exhibits the distribution of occurrences according to the strategy, including 3rd person plural (they):

Age Groups	G3 (+56)	G2 (36 -55)	G1 (25-35)
<i>Eles</i> (they)	13/58(22%)	19/26 (73%)	23/37 (62%)
<i>Você</i> (you)	59/65 (91%)	73/81 (90%)	223/241 (93%)
<i>Nós</i> (we)	9/12 (75%)	3/5 (60%)	2/2 (100%)
<i>A gente</i> (the people=we)	33/44 (75%)	17/17 (100%)	47/50 (94%)

Table 2. Overt pronouns (vs. total of occurrences) according to strategy and age group

Figure 5 shows the distribution of overt versus null pronominal strategies (*você*, *nós* and *a gente*), across the three generations:

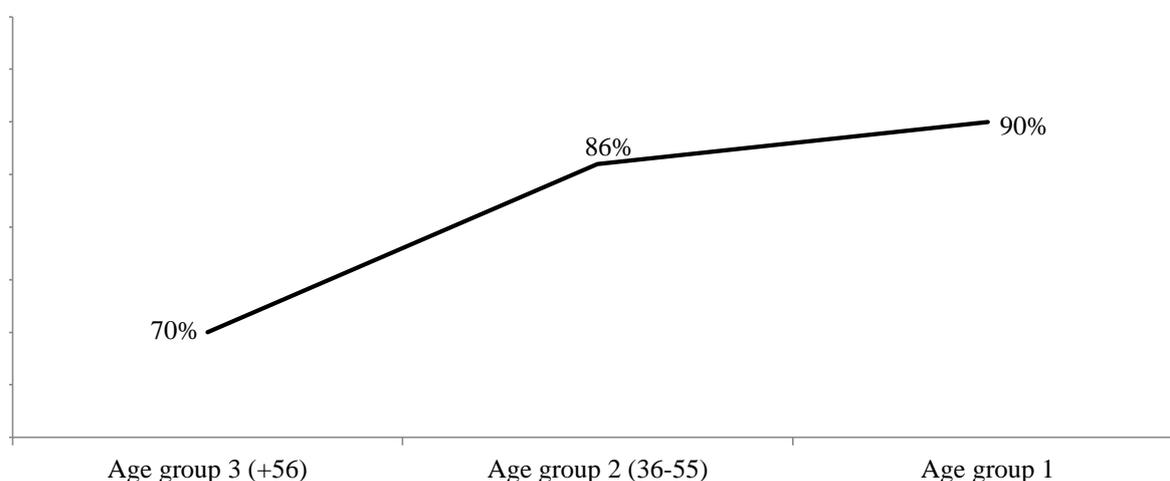


Fig. 5. Overt pronominal strategies (vs null) in apparent time

The percentages show that, in the early 1990s, not only overt subjects were already preferred to null pronouns, but also that the process of change in the speech of the younger group was close to completion. While for group 3 the mean percentage of overt subjects is 70% – 117 occurrences in a total of 167 –, the younger group reaches 90% – with 295 occurrences out of 330.

Figure 6 represents the results in Table 2 for overt pronouns. With respect to the category combining the features [+3rd person], [+plural], we can see the most significant change: overt 3rd person subjects, with 22% in the speech of speakers in the older group, rises to 73% and 62% among the speakers of the intermediate and younger groups.

¹⁶ Duarte (1995) did not include the clitic *se* as an overt strategy as we do in this analysis.

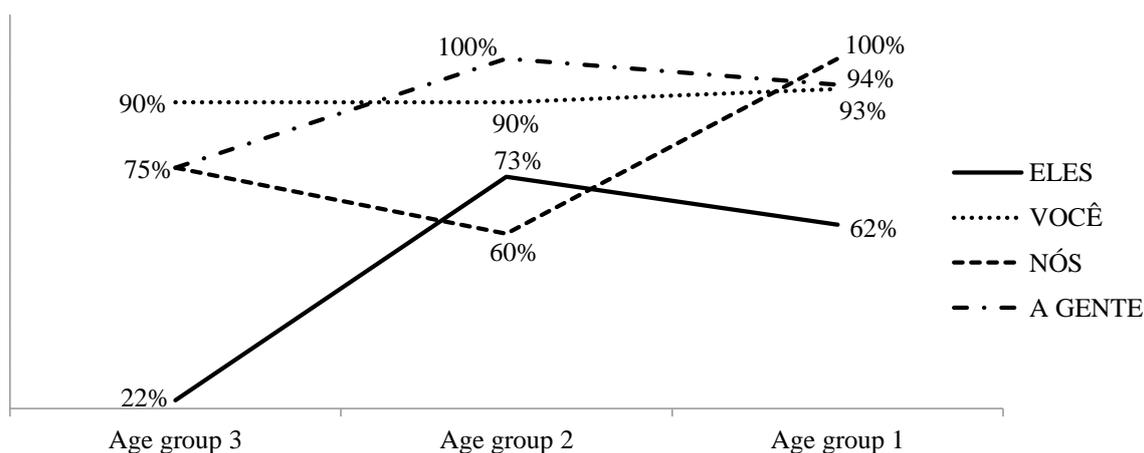


Fig. 6. Distribution of overt subjects in “apparent time”

As for the combination [+3rd person/+singular], we can see that overt *você* is already predominant in all age groups, with a percentage around 90%. Not only does this confirm Duarte’s (1995) results, but it also constitutes an important argument in favor of the victory of *você* as the strategy for [+3rd person/+singular] group over *se* and zero strategies, as previously mentioned. Therefore, it seems natural that *você* stands as the principal strategy in the extreme of the proposed gradient.

Finally, the combination of features [+1st person/+plural], with the competition between *nós* and *a gente*, presents an interesting picture. The pronoun *a gente*, which, as seen before, is replacing *nós*, for definite and indefinite reference, presents an increase in the rates of overt subjects – from 75% among the older group to 100% and 94% in the speech of intermediate and younger groups, respectively. Considering the use of *nós*, one can say that, in spite of its overt expression being preferred, reaching 100% in Group 1, our sample attested only 19 occurrences of such strategy (shown in Table 2) and the younger group shows only two instances of this pronoun, attesting its almost complete extinction in spontaneous speech.

3. Final considerations

This article aimed to contribute to studies concerning the representation of indeterminate subjects in BP, trying to relate the change between the reset of the NSP in this variety of Portuguese to the appearance of new strategies – preferably overt – as alternatives to the conservative forms described in traditional grammars and consistent with European Portuguese, which has been taken as a model to still current descriptions (see CARDOSO, 2017, for the diachrony of indetermination strategies in European Portuguese). Our

analysis made it possible to show that the several strategies under the label “subject indetermination” reveal different types of reference, which led us to dispose them along a scale or a *continuum*: in one pole, we find arbitrary reference, which corresponds to an indeterminate subject in the space of the discourse, but, in a way, limited to a set of entities, and, in the opposite pole, we have the generic reference, which also corresponds to an indeterminate subject in the space of the discourse, but whose scope is unlimited. In intermediate points in this scale, we find indeterminate subjects, including necessarily the speaker, whose scope can be wider or narrower, being closer to one pole or the other.

Through the revision of the corpus previously analyzed by Duarte (1995), consisting of the speech of graduates born in Rio de Janeiro, distributed in three age groups, in a pioneer empirical research bringing evidence of the resetting of the NSP in BP, we have seen that the different strategies attested are not in direct competition: it was possible to propose three different groups, with three strategies would be in competition in each one, defined by a set of features. One of them, characterized by [+3rd person/+plural], is represented by the pronoun *eles* (they), null or overt, and the clitic “se”. The [+3rd person/+singular] cluster of features contains the strategies with the clitic *se*, the zero strategy (3rd person singular) and *você* (you), which, in the analyzed sample, is almost categorically overt. Still concerning this group of features, it was interesting to notice that the obsolescence of the clitic *se* seems to have opened the way to the forms zero and *você*, but, differently from recent researches suggest, the zero strategy neither appears as an emerging resource, nor represents a strong candidate to replace *você*. The reason for that lies exactly the fact that the zero strategy decreases among the younger groups and this fact is consistent with a system moving towards overt pronominal subjects. It would be surprising to find it moving to an opposite direction. This zero strategy shows a more restricted use, being constrained with respect to aspect, modality and transitive verbs, which can explain its marginality in the system. Finally, the [+1st person/+plural] group can be represented by the strategies *nós* and *a gente*, with considerable advantage for the latter, which presents rates over 70% in the three age groups. The pronoun *nós* rarely appears, but when it does, it is predominantly overt.

In sum, the variation existing in each group or category disposed along our scale is not a stable phenomenon. In fact, each point has a strong competitor to represent each degree of indeterminate reference as the change in progress moves towards completion.

References

- ARONOFF, Mark. *Word formation in generative grammar*. MIT Press, 1976.
- CALLOU, Dinah; LOPES, Célia. “Contribuições da sociolinguística para o ensino e a pesquisa: a questão da variação e mudança linguística.” *Revista do GELNE*, 5 (1-2), p.63-74, 2004.
- CARDOSO, Marianna. *Estratégias de indeterminação em peças brasileiras e portuguesas: uma análise contrastiva*. Master Thesis – FL, UFRJ, Rio de Janeiro, 2017.
- CAVALCANTE, Silvia. *A indeterminação do sujeito na escrita padrão: a imprensa carioca dos séculos XIX e XX*. 1999. Master Thesis – FL, UFRJ, Rio de Janeiro, 1999.
- CAVALCANTE, Silvia. “O sujeito nulo de referência indeterminada na fala culta carioca”. *Diadorim: Revista de Estudos Linguísticos e Literários*. Rio de Janeiro, v.2, p. 63-82, 2007.
- CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.
- CHOMSKY, Noam. *The Minimalist Program*. Cambridge, Mass: The MIT Press, 1995.
- COUTO, Marco Aurélio de Souza. *Estratégias pronominais de indeterminação do agente*. Master Thesis – FL, UFRJ, Rio de Janeiro, 2004.
- CYRINO, Sônia; DUARTE, Maria Eugênia L.; KATO, Mary. A. “Visible subjects and invisible clitics in Brazilian Portuguese”. In: KATO, M. A.; NEGRÃO, E. (eds.). *Brazilian Portuguese and the Null Subject Parameter*. Madri: Iberoamericana/Frankfurt: Vervuert, 2000, p. 55-73.
- D’ALBUQUERQUE, Alair da Cruz Reis Cavalcanti. “A perda dos clíticos num dialeto mineiro.” *Sociolinguística e ensino do vernáculo*, 78/79, Rio de Janeiro, Tempo Brasileiro, p. 97-120, 1984.
- DUARTE, Maria Eugênia L. “Do pronome nulo ao pronome pleno: a trajetória do sujeito nulo no português do Brasil”. In: ROBERTS, I.; KATO, M. A. (eds.). *Português brasileiro: uma viagem diacrônica*. Campinas: UNICAMP, p. 107-128, 1993.
- DUARTE, Maria Eugênia L. *A perda do princípio Evite Pronome no português brasileiro*. 1995. PhD Dissertation – IEL, UNICAMP, Campinas, 1995.
- DUARTE, Maria Eugênia L. “A evolução na representação do sujeito pronominal em dois tempos”. In: PAIVA, M. C.; DUARTE, M. E. L. (eds.). *Mudança linguística em tempo real*. Rio de Janeiro: ContraCapa/FAPERJ, p. 115-128, 2003.
- DUARTE, Maria Eugênia L. “Sujeitos de referência definida e abstrata: aspectos *Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017*.”

- conservadores e inovadores na escrita padrão.” *Revista Linguística*, UFRJ, v. 3, n. 1, 88-115, 2007.
- DUARTE, Maria Eugênia L. (ed.). *O sujeito em peças de teatro (1833-1992): estudos diacrônicos*. São Paulo: Parábola, 2012.
- DUARTE, Maria Eugênia L. “Avanços no estudo da mudança sintática associando a Teoria da Variação e Mudança e a Teoria de Princípios e Parâmetros”. *Cadernos de Estudos Linguísticos*, Unicamp, v. 57, n. 1, 85-111, 2015.
- DUARTE, Maria Eugênia L. “Sociolinguística ‘Paramétrica’”. In MOLLICA, M. C.; FERRAREZI JR. C. (eds.) *Sociolinguística, sociolinguísticas – uma introdução*. São Paulo: Contexto, 2016, p. 33-44.
- DUARTE, M. Eugênia L.; MOURÃO, Gabriela; SANTOS, Heitor. “Os sujeitos de terceira pessoa: revisitando Duarte 1993”. In: DUARTE, M. E. L. (ed.) *O sujeito nulo em peças de teatro (1843-1992): estudos diacrônicos*. São Paulo: Parábola Editorial. 2012, pp. 21-44.
- DUARTE, M. Eugênia L.; SOARES da SILVA, Humberto. “Microparametric variation in Spanish and Portuguese: The Null Subject Parameter and the role of the verb inflectional paradigm.” In KATO, M. A.; ORDÓÑEZ, F. (eds.) *The Morphosyntax of Portuguese and Spanish in Latin America*, p. 1-26, 2016.
- EGERLAND, Verner. “Impersonal pronouns in Scandinavian and Romance.” *Working Papers in Scandinavian Syntax*, 71: 75-102, 2003.
- GALVES, C. “A sintaxe do português brasileiro”. *Ensaio de Linguística*, 13, 31-49, 1987.
- HOLMBERG, Anders. “Is there a little *pro*? Evidence from Finnish.” *Linguistic Inquiry*, 36, 533-564, 2005.
- HOLMBERG, Anders. “The null generic pronoun in Finnish: a case of incorporation.” In: BIBERAUER, T. et al. (eds.) *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: CUP. p: 200-230, 2010.
- HOLMBERG, Anders; NAYUDU, Aarti; SHEEHAN, Michelle. “Three partial null subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi.” *Studia Linguistica*, 63, 59-97, 2007.
- HOLMBERG, Anders; SCHEEHAN, M. Control into finite clauses in partial null subject languages. In T. BIBERAUER et al. (eds) *Parametric Variation: null subjects in Minimalist theory*. Cambridge: Cambridge University Press. 125-152, 2010.

- JAEGGLI, Osvaldo; SAFIR, K. “The Null Subject Parameter and Parametric Theory”.
In: JAEGGLI; SAFIR (eds.). *The Null Subject Parameter*. Dordrecht: Kluwer,
1989, p. 1-44.
- KROCH, Anthony. “Morphosyntactic Variation”. In BEALS, K. et al. (eds.) *Papers from
the 30th Regional Meeting of the Chicago Linguistic Society v. 2: The Parasession
on Variation in Linguistic Theory*. 1994, 1-23.
- LABOV, William. *Principles of Linguistic Change: Internal Factors*. Cambridge MA:
Blackwell, 1994.
- LOPES, Célia. “Nós por a gente no português falado culto.” In: *Anais do II Congresso
da ASSEL/Rio*. Rio de Janeiro: Faculdade de Letras, 1993, p. 75-81.
- LOPES, Célia. *A inserção de a gente no quadro pronominal do português: percurso
histórico*. 1999. PhD Dissertation, FL, UFRJ, Rio de Janeiro, 1999.
- LOPES, Célia. *A inserção de ‘a gente’ no quadro pronominal do português*.
Frankfurt/Madri. Vervuert/Iberoamericana, vol. 18, 2003.
- LOPES, Célia. “A gramaticalização de ‘a gente’ em português em tempo real de longa e
de curta duração: retenção e mudança na especificação dos traços intrínsecos”.
Fórum Lingüístico UFSC, v. 4, p. 47-80, 2007.
- LOPES, Célia; CAVALCANTE, Sílvia. “A cronologia do voçamento no português
brasileiro: expansão de você-sujeito e clítico-te”. *Linguística*, Madrid, v. 25, p.30-
65, 2011.
- LUNGUINHO, M. Vinicius; MEDEIROS JUNIOR, Paulo. “Inventou um novo tipo de
sujeito: características sintáticas e semânticas de uma estratégia de indeterminação
do sujeito no português brasileiro”. *Revista de Estudos em Língua e Literatura*, v.
23, 147-191, 2009.
- MODESTO, Marcello. “Sujeitos nulos em línguas de tópico proeminente”. *Revista da
ABRALIN*, vol. 3, 121-148, 2004.
- MODESTO, Marcello. “Topic Prominence and Null Subjects”. In: BIBERAUER, T.
(Ed.) *The limits of syntactic variation*. Amsterdam/Philadelphia: John Benjamin
Publishers, 2008, p. 375-410.
- PAREDES SILVA, Vera L. “O retorno do pronome *tu* à fala carioca.” In RONCARATI,
C.; ABRAÇADO, J. *Português Brasileiro – contato linguístico, heterogeneidade e
história*, p.160-169, 2003.
- RIZZI, Luigi. *Issues in Italian syntax*. Dordrecht: Foris, 1982.
- ROBERTS, Ian. *Verbs and Diachronic Syntax*. Dordrecht: Kluwer, 1992.

- SCHERRE, Marta et al. “Variação dos pronomes *tu* e *você*.” In: MARTINS, M. A.; ABRAÇADO, J. (eds.). *Mapeamento sociolinguístico do português brasileiro*. São Paulo: Contexto, p. 133-173, 2015.
- SCHERRE, Marta; DUARTE, Maria Eugênia L. “Main Current Processes of Morphosyntactic Variation.” In WETZELS, L; MENUZZI, S; COSTA, J. (eds) *The Handbook of Portuguese*. Malden, MA: John Wiley & Sons Inc. p. 526-544, 2016.
- SOUZA, Janaína P. F. de. *Mapeando a entrada do você no quadro pronominal: análise de cartas familiares dos séculos XIX-XX*. Master Thesis. Rio de Janeiro: UFRJ, 2012.
- TAGLIAMONTE, Sali. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press, 2006.
- VARGAS, Amanda de S. C. “A evolução na representação das estratégias pronominais de indeterminação”. In: DUARTE, M. E. L. (ed.). *O sujeito nulo em peças de teatro (1843-1992): estudos diacrônicos*. São Paulo: Parábola Editorial. p. 45-68, 2012.
- VIANNA, Juliana Barbosa S.; LOPES, Célia. “Variação dos pronomes *nós* e *a gente*.” In: MARTINS, M. A.; ABRAÇADO, J. (eds.). *Mapeamento sociolinguístico do português brasileiro*. São Paulo: Contexto, p. 109-131, 2015.
- WEINREICH, Uriel; LABOV, William.; HERZOG, Marvin. “Empirical foundations for a theory of language change”. In: LEHMAN, W.; MALKIEL, Y. (eds.). *Directions for Historical Linguistics*. Austin: University of Texas, p. 97-195, 1968.

A Comparative Diachronic Analysis of *Wh*-Questions in Brazilian and European Portuguese

Mayara Nicolau de Paula^a

ABSTRACT

Based on the Principles and Parameters framework (CHOMSKY, 1995), this article presents a diachronic analysis of European Portuguese (EP) *Wh*-questions and a comparison with the same structures in Brazilian Portuguese (BP). I also present a brief analysis of *Wh*-question patterns in sociolinguistics interviews recorded in two periods (years 1970/80 and 2010). The initial hypothesis, inferred from recent theoretical descriptions, was that WhVS order would be more frequent in EP and that WhSV, whenever attested, would be constrained by the presence of the cleft-structure. Taking into account the fact that EP is considered a consistent Null Subject Language (NSL), 1st and 2nd person subjects as well as anaphoric subjects would be preferably null. Therefore, EP would exhibit a different behavior from BP, which has become a WhSV system; the rare cases of SV are attested with unaccusative verbs and a lexical DP subject; and this change, as shown by the analyses reported here, runs parallel to the re-setting of the value of the Null Subject Parameter in BP. The sample analyzed for the diachronic study, comprising theater plays written across the 19th and the 20th centuries, is comparable to the Brazilian sample. For the interviews, I used the samples *NURC* and *Concordância* for PB and *Cordial-Sin* and *Concordância* for EP. The methodology to codify and submit the data to statistical treatment follows the variationist approach (TAGLIAMONTE, 2006; GUY AND ZILLES, 2007). The results show that EP prefers the WhV pattern, confirming its status of a NSL; as for overt subjects, WhVS is the preferred pattern; however, we can observe a slow decrease of VS in the last quarter of the 20th century, suggesting the implementation of a competition with WhSV, triggered by the introduction of clefting in the second half of the 19th century. Once introduced in the system, the cleft structure expands to all Wh patterns, which is confirmed in contemporary speech data.

KEYWORDS: *wh*-questions; VS/SV order; cleft structure; null subject parameter; European Portuguese-Brazilian Portuguese

RESUMO

Com base na Teoria de Princípios e Parâmetros (CHOMSKY, 1995), apresento uma análise diacrônica das interrogativas-Q do Português Europeu (PE) seguida de uma comparação com as mesmas estruturas no Português Brasileiro (PB). Paralelamente, faço uma breve análise dos padrões de interrogativas-Q em entrevistas sociolinguísticas gravadas em dois momentos (anos 1970/80 e 2010). A hipótese inicial, a partir de descrições recentes de base teórica, era a de que a ordem QVS no PE seria a mais frequente, enquanto a ordem QSV estaria sempre condicionada à presença da clivagem. Sujeitos de 1^a. e 2^a. pessoas bem como os anafóricos seriam preferencialmente nulos no PE, um sistema descrito como de sujeito nulo

^a Universidade Federal de Minas Gerais. E-mail: maynicolau@gmail.com

consistente. Nesse aspecto, o PE teria um comportamento diferente do PB, que perdeu a ordem QVS, hoje atestada apenas em estruturas com verbos inacusativos, desde que o sujeito seja um DP lexical, uma mudança paralela à remarcação do valor do Parâmetro do Sujeito Nulo no PB. A amostra analisada para o estudo diacrônico é constituída de peças portuguesas escritas ao longo dos séculos 19 e 20, comparável à amostra brasileira que nos serve de ponto de comparação. No caso das entrevistas sociolinguísticas, foram utilizadas as amostras *NURC* e *Concordância* para o PB, e *Cordial-Sin* e *Concordância* para o PE. A metodologia para o tratamento dos dados segue o modelo variacionista (TAGLIAMONTE, 2006; GUY E ZILLES, 2007). Os resultados mostram que o PE prefere o padrão QV, com sujeitos nulos; quanto aos sujeitos expressos, a ordem QVS é o padrão preferido; observamos, no entanto, uma curva descendente no último quartel do século 20, sugerindo o início de uma competição com QSV, desencadeada pela entrada da clivagem, a partir da segunda metade do século 19. Uma vez introduzida no sistema, a clivagem se expande para os três padrões de interrogativas-Q, o que é confirmado pelos dados da fala contemporânea.

PALAVRAS CHAVE: interrogativas Q; ordem QVS/ASV; clivagem; parâmetro do sujeito nulo;
Português Europeu and Português Brasileiro

Introduction

This paper describes the Subject (S) Verb (V) order in European Portuguese (EP) *Wh*-questions in matrix clauses, based on a sample of popular plays comprising the 19th and 20th centuries, and compares the results attested for EP (NICOLAU DE PAULA, 2016) with Duarte's (1992) findings for Brazilian Portuguese (BP), later refined by Pinheiro and Marins (2012), who included *Wh*-questions with null subjects using the same sample of plays used by Duarte. This would allow an examination of three possible patterns: WhVS, WhSV and WhV

The general hypothesis guiding this study is based on Duarte and Kato (2002), who suggested that the change in BP attested by Duarte (1992) – from WhVS to WhSV – was closely related to the rise in the use of overt referential pronominal subjects in all kinds of sentences (declaratives, yes/no questions and *Wh*-questions) in the same sample (cf. DUARTE, 1993). A comparison of both analyses allowed the hypothesis that rates of overt pronominal subjects in declaratives, which preceded the change towards SV in *Wh*-questions, would have functioned as a trigger for this change. The high rates of overt pronouns would lead a child to interpret a null subject in a WhV sentence as preposed to the verb Wh (S) V, while older generations's null subjects would be in a structure like WhV (S). Therefore, overt subjects and WhSV order would be closely related. Since EP is a consistent null subject system (see BARBOSA, 1995; DUARTE,

1995; HOLMBERG, 2010; LOBO, 2013, among others), one would expect a different behavior, with subjects preferably in WhV (S) pattern. WhSV pattern, whenever attested, would be constrained by the presence of the cleft structure, according to descriptions of standard EP (AMBAR, 1992; BRITO, DUARTE AND MATOS, 2003).

The paper is organized as follows: the next section presents Duarte's (1992) diachronic analysis of the change from WhVS to WhSV in BP, followed by Pinheiro and Marins's (2012) work, which confirms the co-relation suggested by Kato and Duarte (2002), investigating *Wh*-questions with null subjects in the same sample of Brazilian plays analyzed by Duarte (1992). I then review descriptions related to VS order in EP *Wh*-questions and present the diachronic sample collected for the present analysis. In the next section, both the diachronic and synchronic results for EP are presented and compared with BP. Finally, I discuss the differences between BP and EP *Wh*-questions with respect to the loss of null subjects in the former and the predominance of null subjects in the latter, the importance of the introduction of the cleft structure to trigger WhSV order in both varieties and its propagation to WhV and even WhVS questions in EP.

1. *Wh*-questions and VS–SV order in Brazilian Portuguese

Duarte's (1992) analysis of *Wh*-questions in BP popular plays suggested that the introduction of the cleft structure might have been the trigger of the loss of VS. Fig. 1 shows the rise of SV order across seven synchronies (the year of the play analyzed represents each synchrony).

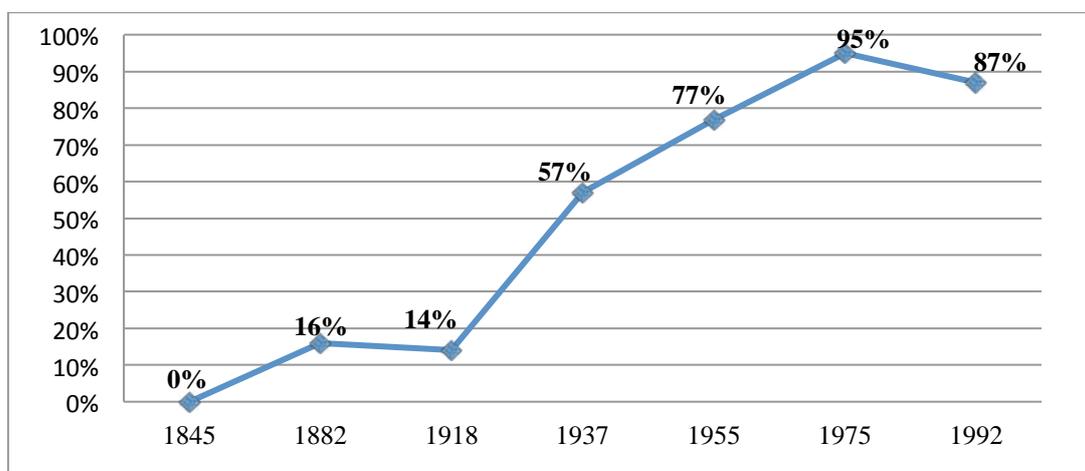


Fig. 1. WhSV (vs WhVS) across two centuries (adapt. from Duarte, 1992:41)

In the first half of the 19th Century (1845), the pattern attested is VS, as in (1):

- (1) O que **pensa tua filha** do nosso projeto? (1845 - BP)
 what think.3SG your daughter of-the our plan
 ‘What does your daughter think of our plans?’

Only in the 2nd half of the century are clefts introduced in the system, occurring in 4 instances showing WhSV order (2a) and one instance of WhVS (2b); in the same synchrony, two WhSV questions with no cleft (both with the interrogative “why”), instantiated in (2c), are attested:

- (2) a. O que **é que tu tens** nesta barriga? (1882 - BP)
 what is that you have.2PS in-this belly
 ‘What do you have in this belly?’
- b. E onde **é que foi a mulher do Seabra?** (1882 - BP)
 and where is that went the wife of-the Seabra
 ‘And where did Seabra’s wife go?’
- c. E para que **você quer ser** ministro, seu Chico? (1882 - BP)
 and for what you want to-be minister, Mr. Chico
 ‘And why do you want to be a Minister, Mr. Chico?’

The low frequency of WhSV continues in the play representing 1918, with 6 instances (out of 29 questions), 3 with cleft and 3 without it. In either case, the interrogative is adverbial. Only in 1937 does WhSV with the present of cleft outnumber WhVS, and both patterns are in complementary distribution. The examples in (3), both attested in the same play, instantiate that:

- (3) a. Mas que **veio você fazer** aqui novamente? (1937 - BP)
 but what came you to-to here again
 ‘But what have you come to do here again?’
- b. Que **é que você veio fazer** aqui novamente? (1937 - BP)
 what is that you came to-do here again
 ‘What are you doing here again?’

From 1955 onwards, this complementarity disappears; we still find the old VS pattern in non-cleft questions (4a,b), but WhSV is again licensed with or without cleaving material (4c,d):

- (4) a. Que **fez seu filho** com os documentos que lhe dei? (1955 - BP)
 what did your son with the documents that to-him.CL gave.1SG
 ‘What have your son done with the documents I gave him?’
- b. De onde **surgiu você**? (1955 - BP)
 from where appeared you
 ‘Where have you come from?’
- c. Onde **é que você andou** até agora? (1955 - BP)
 where is that you have-been until now
 ‘Where have you been?’
- d. Onde **você andou**? (1955 - BP)
 where you have-been
 ‘Where have you been?’

In the last quarter of the 20th century (1975 and 1992), the change towards SV is almost complete: WhVS order is restricted to monoargumental verbs, particularly unaccusatives and the copula, since the subject is a full DP (5a,b).¹ Besides, the use of clefting becomes optional (5c):

- (5) a. Onde **andar** **a Neiva**? (1992 - BP)
 where will-be the Neiva
 ‘Where will Neiva be?’
- b. Como **é que vai ser a reprodução da espécie**? (1975 - BP)
 how is that is-going to-be the reproduction of-the species
 ‘How will the reproduction of the species be?’

¹ If we take into account that the subject DP with unaccusatives is an internal argument and that BP still allows VS order with such verbs, with the same constraints, we can say the change has reached completion.

- c. E o que **você quer?** (1975 - BP)
and what you want
'And what do you want?'

In the same period, two new structures enter the system: reduced cleft-structures (without the copula) (6a) and Wh-*in-situ* sentences (6b):²

- (6) a. Se eu estiver mesmo grávida, o que **que eu vou fazer?** (1992 - BP)
if I am really pregnant, what that I go to-do
'If I am really pregnant, what am I going to do?'
- b. **Você vai botar** o quê hoje? (1975 - BP)
you go to-wear what today
'What are you going to wear today?'

Pinheiro and Marins (2012) analyze the diachrony of WhV interrogatives,. Their results suggest a parallel change involving the decrease of WhVS order and the decrease of null subjects (see Fig. 2).

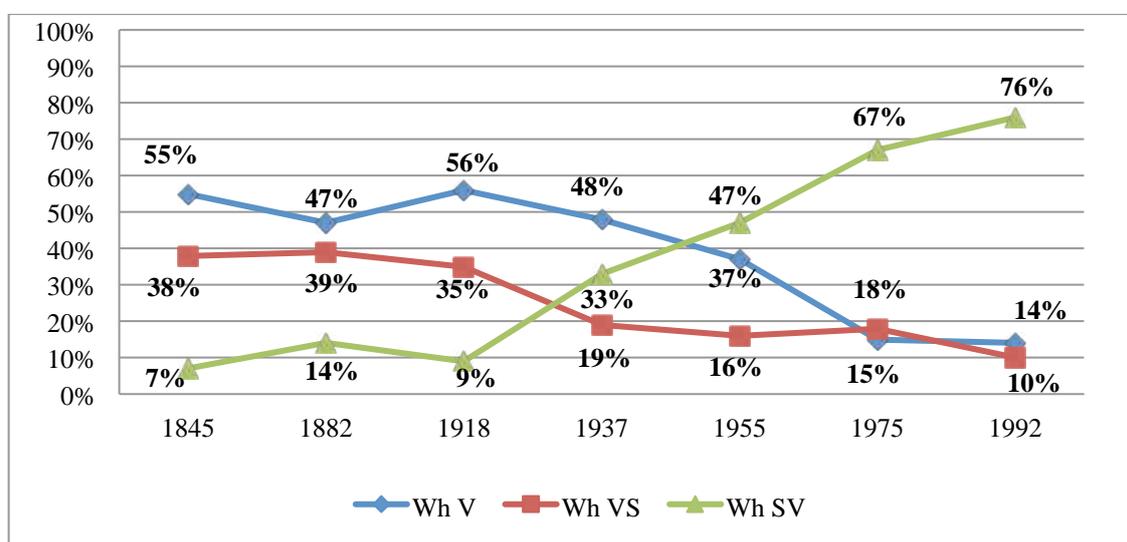


Fig. 2. WhV, WhVS and WhSV sentences in BP (Adapted from Pinheiro and Marins, 2012: 172)

² The author of this article carried out the research for both patterns in the sample analyzed by Duarte (1992).

In Fig. 2, WhSV word order (with and without cleft) starts with a relative frequency of 7% and rises steadily to reach 76% in the final synchrony. Notice that at the same time, the usage frequencies of null subject (WhV) and WhVS constructions descend, reaching almost equal percentage rates in the final period analyzed. Null 1st, 2nd and 3rd person subjects in *Wh*-interrogatives (7) are predominant in the first four periods of time, ranging from 55% to 48%):

- (7) a. Com quem **tenho** o prazer de falar? (1845 - BP)
with whom have.1SG the pleasure to speak
'With whom do I have the pleasure to speak?'
- b. Para que **estudaste** tanto, rapaz? (1882 - BP)
for what studied.2SG so-hard, boy
'Why have you studied so hard, boy?'
- c. Onde se **esconderia** ? (1845 - BP)
where SE.Cl.Refl would-hide.3SG
'Where would she hide?'

To sum up: null subjects are predominant throughout the 19th century and the first half of the 20th century. When an overt subject is used, VS is preferred to SV order in the three first synchronies; from then on, the usage frequency of SV order begins to rise and propagates quickly, particularly from the 1950s on. As shown above, the remaining cases of WhVS attested in the last quarter of the 20th century can no longer be considered real "inversions" since they are restricted to unaccusative verbs with a lexical DP subject, an internal argument in fact still allowed in declaratives with such verbs. This is also the moment when two innovations are attested in the analyzed sample: the reduced cleft and interrogatives with *Wh-in-situ*.

2. Theoretical questions and methodology

2.1. *Wh*-questions in European Portuguese

Formal descriptions about the order in EP *Wh*-questions usually start with the contrast exhibited in (8). According to Ambar (1987; 1992), a sentence like (8a) is ungrammatical whereas (8b) is grammatical. This implies that movement of V to C is obligatory in EP, unless the cleft structure is present (8c):

- (8) a. *O que **a Maria comprou**?
 what the Maria bought
 b. O que **comprou a Maria**? (EP)
 what bought the Maria
 c. O que *é que* **a Maria comprou**? (EP)
 what bought the Maria
 ‘What did Maria buy?’

The author’s hypothesis for the obligatory movement is that there is an empty NP category in the *Wh*-structure; therefore, the raising of the verb would fill this empty slot. With complex *Wh*-structures (*Wh* + NP), the order is optional:

- (9) a. Que vinho **bebe o João** habitualmente? (EP)
 what wine drink.3SG the João usually
 b. Que vinho **o João bebe** habitualmente? (EP)
 what wine the João drink.3SG usually
 ‘What wine does João usually drink?’

A more recent study by Brito, Duarte and Matos (2003: 472) confirms Ambar’s analysis, and adds that not only SV but VS can occur with cleft (10b):³

- (10) a. Onde *é que* **a Maria trabalha**? (EP)
 where is that the Maria work
 b. Onde *é que* **trabalha a Maria**?
 where is that work the Maria
 ‘Where does Maria work?’

Contrary to what we have seen for BP, however, Brito, Duarte and Matos (2003) show that the reduced cleft structure is not grammatical in EP:

³ Mioto and Lobo (2016) maintain the ungrammaticality of WhSV without the cleft structure.

(11) **Quem que* chegou?

Who that arrive?

‘Who arrived?’

As for the occurrence of *Wh-in-situ*, according to the authors, they can appear in ordinary *Wh-* questions or in echo questions, just like in BP, and the interpretation will depend on the intonation.

(12) a. **A Maria sai** quando? / A Maria sai QUANDO? (EP)

the Maria leaves when

‘When does Maria leave?’

b. **Ela demorou** tanto por quê? / Ela demorou tanto POR QUÊ?

she took so-long why?

‘Why did she take so long?’

The descriptions by Ambar and Brito, Duarte, Matos above suggest VS is the usual pattern and WhSV with cleft is a less frequent and more constrained order. None of the descriptions reviewed here are concerned with frequency of use or with spoken language nor should they be. This analysis will investigate not only the preferred patterns but their distribution over time EP popular theater plays, which, as shown in a number of studies (cf. Duarte, 2012), allow to conclude that they approach speech.

In addition to the descriptions offered for EP, the theoretical support for this research comes from the Principles and Parameters framework. The central hypothesis is based on the results found for BP, shown in section 2. Since EP is a consistent null subject language (HOLMBERG, 2010; ROBERTS and HOLMBERG, 2010) and does not seem to be undergoing any changes involving the setting of the Null Subject Parameter, I expect null subjects to be very productive, particularly when deictic (1st and 2nd persons) or anaphoric; overt subjects will appear preferably in WhVS pattern and the occurrence of WhSV will be, according to descriptions, constrained by clefting. This situation would reflect the results depicted by Pinheiro and Marins (2012) for BP in the 19th century and the first half of the 20th century, in Figure 2. In other words, I expect the use of *Wh*-questions in contemporary EP to pattern older stages of BP.

2.2. Methodology

I collected data from Portuguese plays produced across the 19th and the 20th centuries, distributed in seven periods of time, following the periodization proposed by Duarte (1992). The comparison is based on Duarte's (1992) and Pinheiro and Marins's (2012) results for BP.

The sample comprises 40 plays, some in recognized publications and others collected in the library of the University of Lisbon. The distribution over time is not regular, since it follows the periodization of Duarte's sample of Brazilian plays. The 19th Century comprises two periods and the 20th Century, five. The results for several investigations show considerable regularity in the 19th century, and in the first quarter of the 20th century; only from the 1930s on, can we attest changes in BP pronominal system and consequent by-products or evidence of the embedding of such major change. More than one author had to be included to represent each period of time in order to obtain a balanced amount of data per synchrony.⁴ The plays are listed in Appendix 1.

The data selection and processing was based on variationist methodology (TAGLIAMONTE, 2006; GUY AND ZILLES, 2007). Besides the parameter of word order in overt subject interrogatives (WhVS/WhSV), *Wh*-questions with null subjects were included in order to test my predictions. In addition, I controlled for other structural factors, such as the type of Wh constituent, the presence of cleft-structure, the grammatical person of the subject, the use of a full DP or a pronoun for 3rd person over subjects, and the transitivity of the verb.

3. Results

3.1. *Wh*-questions in European Portuguese

A total of 681 data were collected, with the following distribution: 53% WhV, 38% and 9% WhSV (cases of Wh *in situ* were treated separately). The overall distribution confirms the general hypothesis that WhV would be the most frequent pattern,

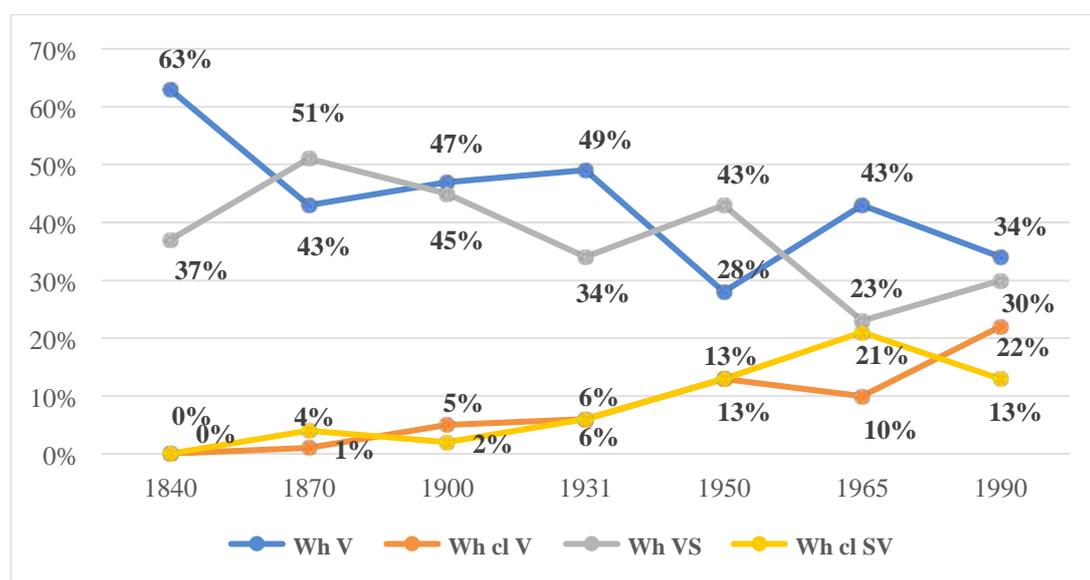
⁴ A research in course investigates individual playwrights who produced in two, sometimes, three periods considered in our sample, so as to observe their behavior across time.

exemplified in (13 a, b), without and with cleft, while (13 c, d) present WhVS and WhSV with cleft, respectively:

- (13) a. Então por que a **agrediu?** (1905 - EP)
 then why her.CL.3SG hurt.3SG
 ‘Then why did he hurt her?’
- b. O que *é que* **quereis dizer** no tal programa? (1871 - EP)
 what is that want.2PL to-say in-the such program
 ‘What do you want to say in that program?’
- c. Então que **respondeu Sofia** à minha carta? (1843- EP)
 then what answered Sophia to-the my letter
 ‘Then, what was Shophie`s answer to my letter?’
- d. O que *foi que* **tu disseste?** (1954 - EP)
 what was that you said?
 What did you say?

The distribution of the patterns attested along the periods, in Table 1 and Figure 3, provides some interesting results:

Patterns	1840-1858		1870-1898		1900-1928		1931-1944		1950-1963		1965-1988		1990-1998	
	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%	N°	%
Wh V	57	63	39	43	44	47	49	49	31	28	41	42	31	34
Wh cleft V	-		1	1	5	5	6	6	15	13	10	10	20	22
Wh V S	34	37	48	51	42	45	34	34	48	43	23	23	27	30
Wh cleft V S	-		-		-		-		-		1	1	-	
Wh S V	-		1	1	-		4	4	-		-		-	
Wh cleft S V	-		4	4	2	2	6	6	14	13	21	21	12	13
Wh-in-situ	-		1	1	1	1	1	1	4	3	3	3	1	1
Total	91	100	94	100	94	100	100	100	112	100	99	100	91	100

Table 1. Distribution of *Wh*-question patterns across seven synchroniesFig.3. Patterns of *Wh*-questions in EP plays across seven synchronies

The four main patterns in Table 1, represented in Figure 3, show that the null subject (*WhV* pattern) and the overt subject in *WhVS* pattern are predominant, even though the *WhV* pattern declines over time, starting with 63% in the first synchrony to reach 34% in the years 1990, the final synchrony in our sample. The *WhVS* pattern is

more stable across time. The introduction of cleft structures in the second synchrony, just as in BP, triggers not only Wh-cl-SV order, but appears in Wh-cl-V sentences (with a null subject) as well. Both patterns start with only 4% and 1% respectively and maintain low usage frequencies until the 1930s. From then on while VS and null subjects keep more regular frequencies. A closer examination of the three last time frames, with the presence of cleft in two patterns, announces what the results for speech recorded in two time frames for EP and BP will show. Its increasing use, either with and over subject in patten Wh SV and with a null subject may reinforce the role of clefting as a multi-functional strategy to codify focus in Portuguese, as pointed out by Kato and Martins (2016).

Only one sentence was attested with VS order and the presence of cleft in the sample analyzed:

- (14) Onde *é que* **está** **isso** na comédia? (1957 - EP)
 where is that is this in-the comedy
 ‘Where is all that in the comedy?’

Another structure in Table 1, Wh SV, without cleft, not represented in Figure 3, shows only five occurrences, one attested in the second synchrony (ex. 13a) and four in second quarter of the 20th century (fourth synchrony), the latter produced by the same author, one of them with a complex Wh-structure (ex. (13,b,c):

- (15) a. Quem **nós temos** por aí apresentável? (1874 – EP)
 Who we have around here presentable
 ‘Who do we have around here presentable enough?’
- b. Porque **ela** te **horroriza**? (1944 - EP)
 why she you.CL.2SG frighten.3SG
 ‘Why does she frighten you?’
- c. Que noção exata **você tem** do que seja liberdade? (1944 - EP)
 what notion exact you have of-the what is freedom
 ‘What do you know about the meaning of freedom?’

Finally, the last line in Table 1 illustrates *Wh-in-situ* questions, which also appear in the sample from the second period on, keeping very low rates along the time frames (only eleven occurrences attested):

- (16) a. Acusas-me, então **de quê?** (1973 - EP)
 Accuse.2SG-me.CL.1SG then of what
 ‘Of what do you accuse me then?’
- b. E isso é pra **quando?** (1998 - EP)
 and this is for when
 ‘And when is this for?’

3.2. Comparing BP and EP in the writing of theater plays

Figures 2 and 3 in the previous section make it clear that EP and BP show very distinct behavior with respect to the patterns attested for *Wh*-questions. EP is very consistent in the preference for null subjects and VS order, even though we have attested some change in the second half of 20th century. BP, on the other hand, has become a *WhSV* system. Both grammars have in common the introduction of cleft structures, but their roles seem to be different in each system. In EP, as we have seen, it is on the way to become a marker of *Wh*-questions, either with null or overt subjects. Figure 4 shows the course of clefts in EP.

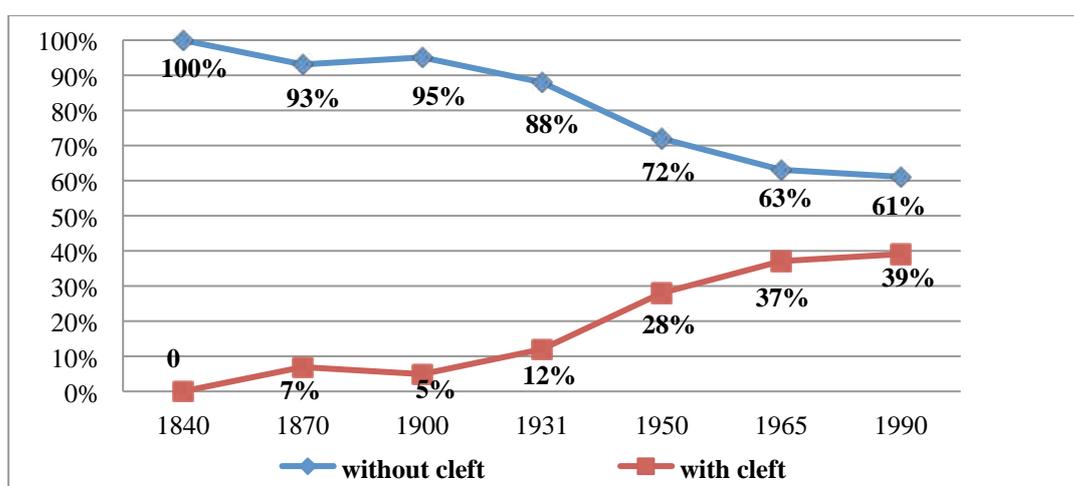


Fig. 4. The evolution of clefts in EP

Questions without clefts are still more frequent in the sample analyzed for EP, which is in part related to the significant occurrence of WhVS order, but cleft questions do show a steady rising trajectory. As for BP, cleft structures have had an important role in the implementation of SV order, and, since null subjects and VS order have almost disappeared, they can be considered the canonical type of *Wh*-questions, even though WhSV can still occur without cleft. Figure 5 shows this evolution in BP, with results from Duarte's (1992) analysis, described in the second section of this article.

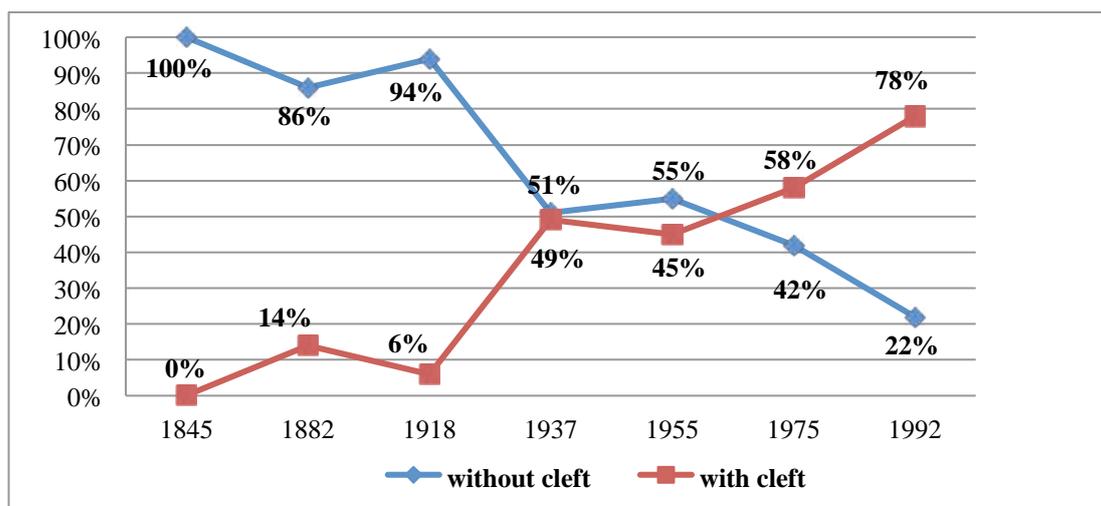


Fig. 5. The evolution of clefts in BP

The propagation of clefting in BP reaches higher rates than in EP, which would suggest that it is a component in focalization in both systems and this feature is not related to the Null Subject Parameter. In fact, based on Kato and Ribeiro's (2009) analysis, Kato (2014) raises the hypothesis according to which *Wh*-questions and focus-structures would have the same checking head (FocusP). Therefore, they should undergo parallel changes across time. This is what the author observes with respect to focalization with VS and with cleft-structures. The canonical cleft structure (copula and complementizer) appears in the 18th century in *Wh*- questions which allows Kato's (2014) suggestion that this innovation will be the preferred focalization pattern from the 19th century on. The results presented here support her hypothesis.

3.3. *Wh*-questions in Brazilian and European speech

In order to compare written to spoken language use, I additionally carried out an investigation of *Wh*-questions in speech. Since in sociolinguistic interviews the

interviewed participants rarely ask questions, I analyzed the interviewers questions. Two samples for European and Brazilian Portuguese recorded in the 1970s and 1980s and in 2010 were used. EP is represented by the interviews belonging to Cordial-Sin and Concordância Projects; BP is represented by Nurc-RJ and Concordância Project⁵. Since the interviewers were teachers or students involved in the research, we can say that the *Wh*-questions collected have been produced by graduates. The results can be seen in Table 2:

PATTERNS	EP		BP	
	Cordial-Sin 1980s	Concord. 2010s	NURC-RJ 1970S	Concor. 2010s
Wh cleft V	42%	53%	2%	0%
Wh cleft SV	20%	16%	73%	44%
Wh cleft VS	19%	9%	4%	0%
Wh-in-situ	17%	5%	4%	10%
Wh V	0,5%	6%	0%	3%
Wh VS	1,5%	10%	1%	1%
Wh SV	0%	1%	16%	42%
TOTAL (%)	100	100	100	100
TOKENS	157	120	83	77

Table 2. *Wh*-question patterns in spoken EP and BP

Regarding EP, I first observe that in both synchronies, the interviewers used much more cleft questions than other question types. Regarding subject expression, null subjects are by far the most frequent pattern followed by SV and VS, still significant particularly in the data of the 1980s. *Wh-in-situ* follows, but its use decreases in the second synchrony. The other patterns without cleft are less significant, with the exception of WhV and WhVS in the 2010s. These results confirm the observation about the propagation of clefts in the course of change suggested by Figure 4. I give some examples from spoken EP in (17):

⁵ The Cordial-Sin (*Corpus dialetal para o estudo da sintaxe*) investigates the dialectal syntactic variation in European Portuguese. Available at: <http://www.clul.ul.pt/pt/recursos>. The NURC-RJ Project (*Projeto da Norma Urbana Oral Culta do Rio de Janeiro*) is part of a national project that recorded a large sample of college educated speakers in the five largest capitals in Brazil, including Rio de Janeiro. The interviews were collected in the 1970s and the 1990s. For our analysis the 1970s interviews have been used. They are available at: <http://www.letas.ufrj.br/nurc-rj>. Finally, the data to represent the second synchrony – the 2010s – come from a bi-national project called Concordância, whose purpose was to compare the speech of Rio de Janeiro and Lisbon. The samples, stratified for age, gender and level of education, are available at: <http://www.concordancia.letas.ufrj.br>.

- (17) a. O que *é que* **começaste** por fazer? (Concordância - EP)
 what is that started.2SG for to-do
 ‘What have you started doing?’
- b. O que *é que* **ele ensinava**? (Concordância - EP)
 what is that he taught
 ‘What did he teach?’
- c. E como *é que* se **chama essa operação de ir lá tirar o mel**? (Cordial-Sin EP)
 and how is that se.CL.call this procedure of to-go there to-take the honey
 ‘And how is this procedure of collecting honey called?’
- d. Como **surgiu o gosto pela advocacia**? (Concordância - EP)
 how appeared the vocation for-the advocate
 ‘How did your vocation for law started?’ (When did you decide to be a lawyer?)
- e. E **a peneira é** para quê? (Cordial-Sin - EP)
 and the sieve is for what
 ‘And what is the sieve for?’

As for BP, the distribution confirms Duarte’s (1992) and Pinheiro and Marins’s (2012) findings regarding *Wh*-questions in BP. Null subjects are practically absent and WhSV is the normal word order both with (73% and 44% in each synchrony) and without clefts (16% and 42%). Other patterns are irrelevant, except for 10% of *Wh-in-situ* attested in the second synchrony. I give some examples in (18):

- (18) a. Como *é que* **você escolheu** uma loja de artesanato? (Concordância - BP)
 how is that you chose a store of artcraft
 ‘Why have you chosen to work with artcraft?’
- b. Quando **você vem**? (NURC-RJ - BP)
 when you come
 ‘When are you coming?’
- c. **Vocês viveram** aonde? (Concordância - BP)
 you lived where
 ‘Where did you live?’

Another feature characterizing Brazilian Wh-questions is the existence of reduced clefts (i. e. without the copula), absent in EP:

- (19) a. E *que que você acha* da política atual? (Concordância - BP)
and what that you think of-the politics nowadays
'And what do you think about our politics today?'
- b. Por *que que você acha* que tem esses problemas lá? (Concordância - BP)
for what that you think that have.3SG these problems there
'Why do you think such problems exist there?' (have = there are)

Final remarks

The analyses and comparisons with previous research presented in this paper allows the following conclusions:

(a) EP *Wh*-questions show three main patterns: WhV is the most frequent pattern, obviously with 1st and 2nd person and anaphoric subjects, and expected behavior from a consistent Null Subject Language; the introduction of cleft-structures is decisive in the implementation of WhSV order, but it expands to all patterns, including WhV and WhVS. Therefore, SV order in *Wh*-questions cannot be associated to the Null Subject Parameter in EP.

(b) WhVS order in BP, on the other hand, is concomitant to the loss of null subjects (Duarte, 1992; 1993; Pinheiro and Marins, 2012). The introduction of the cleft-structure, triggered the change towards WhSV, as shown previously; although still frequent, and usually "reduced" in speech (i. e. without the copula), clefting is no longer a constraint to SV order in BP. As WhSV order increases, *Wh-in-situ* also becomes an important strategy, particularly in the last quarter of the 20th century.

(c) The comparison of the use of *Wh*-questions in recent popular plays to their use in the spoken language of interviewers from Brazil and Portugal has revealed that the use of *Wh*-questions in plays reflects is comparable to their use in spoken language. In addition, the speech data reveals even more clearly the propagation of cleft usage to all patterns of *Wh*-questions in EP, which seems to be an important element to mark focalization (cf. Kato and Ribeiro (2009) and Kato (2014)).

(d) Cleft is also an important structure to focalize *Wh* constituents in BP, but speech shows a regular distribution between its presence and absence. The most important difference between EP and BP is the robust occurrence of null subjects and VS,

followed by SV – preferably with cleft - in the former, and the absence of WhV, the predominance of SV in the latter, with or without cleft; VS order in BP resists with monoargumental (unaccusatives and copula) verbs and lexical subjects in variation with SV.

Appendix 1. Portuguese plays used in the analysis

Period I (1841-1857)	Authors (year of birth)
Os logros numa hospedaria (1841)	Paulo Midosi (1790)
Uma cena de nossos dias (1843)	Paulo Midosi (1790)
Casar ou meter freira (1848)	Antonio Pedro L. de Mendonça (1826)
Nem tudo que reluz é ouro (1849)	João de Andrade Corvo (1824)
O misantropo (1852)	Paulo Midosi (1790)
A domadora das feras (1857)	Luís Augusto Palmeirim (1825)
Period II (1870-1897)	Authors
A liberdade eleitoral (1870)	Teixeira de Vasconcelos (1816)
Clero, Nobreza e Povo (1871)	César de Lacerda (1829)
Quem desdenha... (1874)	Pinheiro Chagas (1842)
O festim de Baltazar (1892)	Gervásio Lobato (1850)
A Senhora Ministra (1897)	Eduardo Schwalbach (1860)
Period III (1900-1923)	Authors
A festa da atriz (1903)	Jorge Santos
Terra Mater (1904)	Augusto de Lacerda (1864)
Os que furam (1905)	Emídio Garcia (1838)
Os Degenerados (1905)	Mário Gollen
O álcool (1912)	Bento Mântua (1878)
Cavalheiro respeitável (1914)	André Brun (1881)
O doido e a morte (1923)	Raul Brandão (1867)
Period IV (1931-1944)	Authors
Continuação de comédia (1931)	João Pedro de Andrade (1902)
Três gerações (1931)	Ramada Curto (1886)
A prima Tança (1934)	Alice Ogando (1900)
A invenção do guarda chuva (1944)	Luís Francisco Rebello (1924)
O ausente (1944)	Joaquim Paços d'Arcos (1908)
Period V (1954-1957)	Authors
Alguém terá que morrer (1954)	Luís Francisco Rebello (1924)
É urgente o amor (1957)	Luís Francisco Rebello (1924)
Mário o eu próprio – o outro (1957)	José Régio (1901)
O meu caso (1957)	José Régio (1901)
Period VI	Authors
A guerra santa (1965)	Luís de Sttau Monteiro (1926)
A menina feia (1970)	Manuel Frederico Pressler (1907)
A sogra (1973)	Alice Ogando (1900)
Prólogo Alentejano (1975)	Luís Francisco Rebello (1924)
A lei é a lei (1977)	Luís Francisco Rebello (1924)
O grande mágico (1979)	Luís Francisco Rebello (1924)

Grito no outono (1980)	Romeu Correia (1917)
Period VII (1996-1998)	Authors
Um filho (1996)	Luísa Costa Gomes (1954)
Quinze minutos de glória (1998)	Jaime Rocha (1949)
O céu de Sacadura (1998)	Luísa Costa Gomes (1954)

Portuguese Plays

MIDOSI, Paulo. Uma cena de nossos dias. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1843.

MIDOSI, Paulo. Os logros numa hospedaria. Biblioteca do Museu Nacional do Teatro. Escrita em 1841.

MIDOSI, Paulo. O misantropo. Biblioteca do Museu Nacional do Teatro. Escrita em 1852.

MENDONÇA, Antônio Pedro L. Casar ou meter freira. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1848.

CORVO, João de Andrade. Nem tudo que reluz é ouro. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1849.

PALMEIRIM, Luis Augusto. A domadora de feras. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1857.

VASCONCELOS, Teixeira de. A liberdade eleitoral. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1870.

LACERDA, César de. Clero, Nobreza e povo. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1871.

CHAGAS, Pinheiro. Quem desdenha... In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1874.

- LOBATO, Gervásio. O festim de Baltazar. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume I 1800-1899*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1892.
- SCHWALBACH, Eduardo. A senhora ministra. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1897.
- SANTOS, Jorge. A festa da atriz. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1903.
- LACERDA, Augusto de. Terra Mater. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II: 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1904.
- GARCIA, Emídio. Os que furam. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1905.
- GOLLEN, Mário. Os degenerados. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1905.
- MÂNTUA, Bento. O álcool. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1912.
- BRUN, André. Cavalheiro respeitável. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1914.
- BRANDÃO, Raul. O doido e a morte. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1923.
- ANDRADE, João Pedro de. Continuação de comédia. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1931.
- CURTO, Ramada. Três gerações. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1931.

- OGANDO, Alice. A prima Tança. In: REBELLO, Luis Francisco. *Teatro português em um acto: volume II 1900-1945*. Editora: Imprensa Nacional da Casa da Moeda, 2003. Escrita em 1934.
- D'ARCOS, Joaquim Passos. O Ausente. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1944.
- REBELLO, Luis Francisco. A invenção do guarda-chuva. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1944.
- REBELLO, Luis Francisco. Alguém terá que morrer. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1954.
- REBELLO, Luis Francisco. É urgente o amor. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1957.
- REBELLO, Luis Francisco. Prólogo Alentejano. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1975.
- REBELLO, Luis Francisco. A lei é a lei. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1977.
- REBELLO, Luis Francisco. O grande mágico. In: REBELLO, Luis Francisco. *Todo o Teatro*. Editora: Imprensa Nacional da Casa da Moeda. 1999. Escrita em 1979.
- RÉGIO, José. Mário, o eu próprio – o outro. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1957.
- RÉGIO, José. O meu caso. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1957
- MONTEIRO, Luís de Sttau. A guerra santa. Faculdade de Letras da Universidade de Lisboa. Escrita em 1965.
- OGANDO, Alice. A sogra. Biblioteca do Museu Nacional do Teatro. Escrita em 1973
- CORREIA, Romeu. Grito no outono. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1980.
- GOMES, Luísa Costa. Um filho. In: GOMES, L. *Duas comédias*. Editora: Relógio D'água. 1996.
- GOMES, Luísa Costa. *O céu de Sacadura*. Editora: Cotovia. 1998.
- ROCHA, Jaime. Quinze minutos de glória. Biblioteca da Faculdade de Letras da Universidade de Lisboa. Escrita em 1998.

References

- AMBAR, Maria Manuela. “Gouvernement et Inversion dans les Interrogatives QU- en Portugais”. *Recherches Linguistiques*, 16, pp. 5-15, 1985
- AMBAR, Maria Manuela. *Para uma sintaxe da inversão verbo-sujeito em Português*. Lisboa: Edições Colibri, 1992.
- BARBOSA, Pilar. *Null Subjects*. PhD Dissertation, Cambridge, Mass: MIT, 1995.
- BRITO, Ana Maria; DUARTE, Inês; MATOS, Gabriela. “Estrutura da frase simples e tipos de frases”. In: MATEUS, M.H.M *et alii*. (Eds.) *Gramática da língua portuguesa*. 5ª ed. rev. e ampl. Coimbra: Caminho, 2003.
- CHOMSKY, Noam. *The Minimalist Program*. Cambridge, Mass: The MIT Press, 1995.
- DUARTE, M. Eugênia L. “A perda da ordem V(erbo) S(ujeito) em interrogativas qu- no português do Brasil”. *D.E.L.T.A. 8 (Especial)*, p. 37-52, 1992.
- DUARTE, M. Eugênia L. “Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil.” In: ROBERTS, I; KATO, M. A. (Eds.) *Português Brasileiro: uma viagem diacrônica*. Campinas: Ed. da UNICAMP, 1993, p. 107-128
- DUARTE, M. Eugênia L. *A perda do princípio “Evite Pronome” no português brasileiro*. PhD Dissertation, UNICAMP, Campinas, 1995.
- DUARTE, M. Eugênia L. (Ed.) *O sujeito em peças de teatro (1833-1992): estudos diacrônicos*. São Paulo: Parábola, 2012.
- DUARTE, M. Eugênia L. and KATO, Mary A. “A diachronic analysis of Brazilian Portuguese wh-questions”. *Santa Barbara Portuguese Studies*, v. VI, p. 326-339, 2002.
- GUY, Gregory R.; ZILLES, Ana M. S. *Sociolinguística Quantitativa*. 1. ed. São Paulo: Parábola Editorial, v. 1, 2007.
- HOLMBERG, Anders. “Null Subject Parameters”. In: BIBERAUER, T. *et alli* (Eds). *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge/ New York: Cambridge University Press, 2010.
- KROCH, Anthony. “Reflexes of grammar in patterns of language change”. *Language Variation and change*, 1, 1989.

- KATO, Mary. “Focus and Wh-questions in Brazilian Portuguese”. In: DION, N. *et alli.* (Eds.). *Linguistic Variation: confronting fact and theory*. Routledge, 2014.
- KATO, Mary and RIBEIRO, Ilza. “Cleft sentences from old to modern Brazilian Portuguese”. In: DUFTER, A. e JACOB, D (Eds.). *Focus and background in Romance languages*. Amsterdam: John Benjamins, 2009, p.123-154.
- KATO, Mary and MARTINS, Ana Maria. “European and Brazilian Portuguese: An Overview on Word Order.” In: WETZELS, L. *et alli.* (Eds.) *The Handbook of Portuguese Linguistics*, Malden, MA: John Wiley & Sons, Inc. 2016, p. 14-40.
- LOBO, Maria. “Sintaxe e Semântica: fenômenos de omissão e elipse”. In: RAPOSO, E.P *et alli.* (Eds.) *Gramática do Português*, v. II. Coimbra: Fundação C. Gulbenkian, 2013, 2309-2335.
- LOPES ROSSI, Maria Aparecida Garcia. ‘Estudo diacrônico sobre as interrogativas do português do Brasil’. In: ROBERTS, I; KATO, M.A. (Eds.) *Português Brasileiro: uma viagem diacrônica*. Campinas: Ed. da UNICAMP, 1993
- LOPES ROSSI, Maria Aparecida Garcia. *A sintaxe diacrônica das interrogativas Q do português*. PhD Dissertation, Unicamp, São Paulo, 1996.
- NICOLAU DE PAULA, Mayara. *A ordem VS/SV e as interrogativas Q no PB e no PE: uma análise diacrônica*. PhD Dissertation, UFRJ, 2016.
- PINHEIRO, Diogo; MARINS, Juliana. “A trajetória das interrogativas QU- clivadas e não clivadas no Português Brasileiro”. In: DUARTE, M. E. L (Ed.) *O sujeito em peças de teatro (1833-1992): estudos diacrônicos*. São Paulo: Parábola, 2012, p.161-179.
- ROBERTS, Ian; HOLMBERG, Anders. “Introduction: parameters in Minimalist theory”. In BIBERAUER, T, *et alli.* (Eds.) *Parametric Variation: null subjects in Minimalist theory*. Cambridge: Cambridge University Press. 2010, 1-57.
- TAGLIAMONTE, Sali. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press. 2006.

The Acceptability of Clitic and Tonic Accusative 3rd Person Pronouns in Written Brazilian Portuguese

Cândido Samuel Fonseca de Oliveira^a

Ricardo Machado-Rocha^b

ABSTRACT

In the present study, we analyze the acceptability of 3rd person accusative clitic pronouns (3P-ACC) in written modality as compared to 3P tonic pronouns in the same context. Contemporary Brazilian Portuguese has lost 3P-ACC clitics in oral modality. For this reason, some linguists have preferred not to use 3P-ACC clitics in experimental items, even in reading experiments. Against this trend, we conducted an acceptability judgment task to investigate whether the low frequency of 3P-ACC clitics and the high frequency of 3P tonic pronouns in oral modality drive the evaluation of these pronouns in written modality. Our results suggest that, for students ranging from technical high-school to post graduation, both pronouns are highly accepted. Thus this leads us to maintain that both 3P-ACC clitics and 3P tonic pronouns are adequate for experimental items of reading tasks.

KEYWORDS: Brazilian Portuguese; clitic pronouns; tonic pronouns; acceptability judgment

RESUMO

No presente estudo, analisamos a aceitabilidade de pronomes clíticos acusativos de terceira pessoa (3P-ACC) na modalidade escrita em comparação com pronomes tônicos de 3P no mesmo contexto. O Português Brasileiro (PB) contemporâneo perdeu os clíticos de 3P na modalidade oral. Por essa razão, alguns linguistas têm preferido não usar os clíticos de 3P em itens experimentais, mesmo em tarefas de leitura. Na contramão dessa tendência, conduzimos uma tarefa de julgamento de aceitabilidade, com o objetivo de investigar se a baixa frequência dos clíticos de 3P e a alta frequência dos pronomes tônicos na modalidade oral influenciam a avaliação dos mesmos na modalidade escrita. Nossos resultados sugerem que, para alunos com nível escolar variando do ensino médio técnico até a pós-graduação, ambos os pronomes são altamente aceitos. Isso nos leva a sustentar que tanto os clíticos quanto os pronomes tônicos de 3P são adequados para itens experimentais de tarefas de leitura.

PALAVRAS-CHAVE: Português Brasileiro; pronomes clíticos; pronomes tônicos; julgamento de aceitabilidade

^a Professor de língua inglesa do Centro Federal de Educação Tecnológica de Minas Gerais - Unidade Contagem. Doutor em Linguística pelo Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais. coliveira@cefetmg.br.

^b Professor de língua portuguesa do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - Campus Sabará. Doutor em Linguística pelo Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais. ricardo.machado@ifmg.edu.br.

Introduction

Brazilian Portuguese (BP) and European Portuguese (EP) have evolved differently throughout the years with respect to many aspects. One of the main differences can be noticed in the usage of 3rd person accusative pronouns. Whereas EP has preserved the clitic system, BP has almost lost it and has applied different strategies to replace it. This new system includes the use of both null and 3P tonic pronouns in direct object function position. It is the contention of Oliveira (2007) that for question (1), for instance, EP would only license the answers in (2) and BP would license both (2) and (3). In other words, EP and BP allow the use of 3P clitic pronouns (2b) and repeated NPs (2c) (in BP (2b) is likely to be the least preferred option at least for oral genres), but only BP allows the use of null pronouns (3a) and 3P tonic pronouns (3b).

- (1) Você conhece o João?
 You know the João
 “Do you know João?”
- (2) a. Sim
 Yes
 “Yes”
 b. (Sim), conheço-o.
 Yes know him
 “(Yes), I know him”
 c. Sim, conheço o João
 Yes know the João
 “Yes, I know João”
- (3) a. Conheço [Ø]
 Know
 “I know him”
 b. Sim, (eu) conheço ele
 Yes I know he
 “Yes, I know him”
 (OLIVEIRA, 2007, p. 2)

The main analyses for this diachronic change relate the loss of 3rd person clitics to a modification in the directionality of phonological cliticization (NUNES, 1993, 2011) and to rearrangements in BP agreement system (GALVES, 1989; CYRINO, 1993, 1997; KATO, 1993). The presence of 3P-ACC clitics in current standard written and oral BP would be due to sociolinguistic factors, mainly associated with education (CORRÊA, 1991).

Due to these theoretical findings, it has been a common practice, as maintained by Maia and Lima (2014), not to use 3P clitic pronouns in linguistic and psycholinguistic experimental items. At first glance, this option seems natural since most experiments focus on speaker's natural behavior and, hence, the most natural pronouns should be preferred. However, these changes in BP pronominal system seem to have played a major role in the oral modality (CORRÊA, 1991; MAIA; LIMA, 2014) and, therefore, the exclusion of 3P clitic pronouns in reading experiments linguistics is still open to debate. The aim of this study is to shed some light on this topic.

Throughout this text we will refer to written and oral modalities in a more general manner, without references to specific genres. Thus in our experiment we used sentences out of context in order not to evoke rules related to specific genres. We conducted an acceptability judgment task with Brazilian students (ranging from technical high school to post graduation) in order to compare the preference for these pronouns. In the next section (section 1) we discuss BP pronominal system. We then describe the experiment we conducted to explore whether the changes occurred in the core grammar also influence the written modality (section 2). We conclude (section 3) with considerations about the implications of our findings.

1. 3P-ACC objects in contemporary BP registers

It is a consensus that the 3P accusative clitics *o*, *a* and their allomorphs are not part of contemporary vernacular BP (TARALLO, 1983; DUARTE, 1989; GALVES, 1989; CORRÊA, 1991; PAGOTTO, 1993; CYRINO 1993; NUNES, 1993, 2011; KATO 1993). As most of these studies claim, the usage of these forms is a result of schooling and limited to formal styles and written language (NUNES, 2011, p. 14).

Corrêa (1991) demonstrates that 3P-ACC clitics have a very low frequency in the speech of college students (11.2%) and do not appear at all in the speech of illiterate speakers. Machado-Rocha (2013) shows that they are completely absent in spontaneous speech of high-school students. As for written language, Nunes (2011), based on Corrêa

(1991), highlights that the frequency of clitics reaches a peak of 85.7% in the writings of college students, but these elements start to appear only in the writings of 3rd and 4th grade students and, by this time, children are already 9-10 years old. Based on these facts, Nunes assumes that these clitics are not part of the core grammar of BP. Instead, they are acquired like a second language structure attained through schooling, as opposed to the other clitics, which are products of a regular process of language acquisition.

Nunes (1993, 2011) proposes that the loss of 3P-ACC clitics in BP was the result of the change in the directionality of phonological cliticization in the language. As a consequence of this change, BP now allows the weak pronoun *ele* in object position and a greater range of null objects, (i.e. null pronouns, *pro*) (see for instance Galves (1989), Kato (1993) and Cyrino (1993, 1997)).

- (4) a. Eu entreguei **ele** pra Maria
 I delivered it to.the Maria
 “I delivered it to Maria”
- b. Eu entreguei (pro) pra Maria
 I delivered to.the Maria
 “I delivered it to Maria”
- (NUNES 2011, p. 14)

The core idea of Nunes (1993) is that the onset of the syllable of 3P-ACC clitics must be licensed, which in turn is achieved via phonological cliticization to the preceding word. In EP, it is a regular process, since this language has an enclitic pattern. This is to say that cliticization in EP is a process of leftward phonological affixation. BP, on the other hand, exhibits a rightwards phonological cliticization pattern (see e.g. Pagotto (1993)). Within this new system with rightward phonological cliticization, the question arises as to how the onset of the syllable of 3P-ACC clitics could be licensed. Nunes (1993) proposes that once a generation of children acquired rightward phonological cliticization, there could be no way for the onset of the syllable of 3P-ACC clitics to be licensed, and children then simply failed to acquire these forms.

More recently, Machado-Rocha (2016) and Machado-Rocha and Ramos (2016) correlate the absence of 3P-ACC clitics to the synchronic status of the object clitic

projection in BP, in accounting for the phenomenon of clitic doubling, as illustrated in (5):

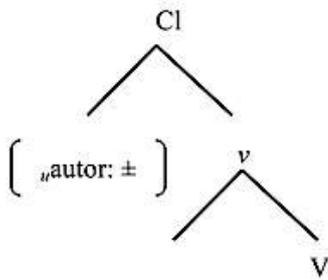
- (5) a. Ele me ajuda eu
 He 1PCL helps I
 “He helps me”
- b. Eu te ajudo você
 I 2PCL help you
 “I help you”
- c. *Eu o ajudo ele
 I 3PCL help he
 “I help him”

(MACHADO-ROCHA, 2016, p.13, 19.)

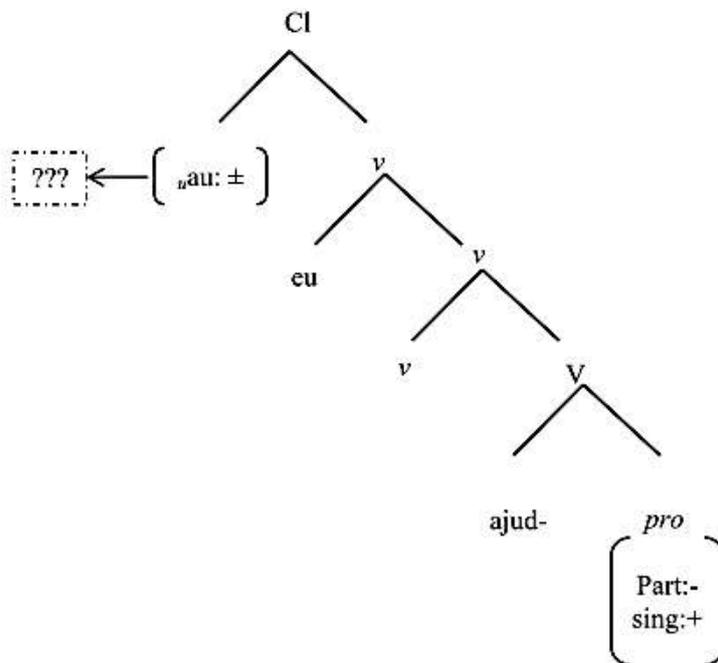
As (5) shows, BP clitic doubling occurs only for 1st and 2nd person pronouns, but not for 3rd person pronouns. The same is valid for single cliticization structures, when only oral registers are being considered, as we have discussed. Instead of assuming the obvious answer that (5-c) would be ungrammatical due to the absence of 3P-ACC clitic pronouns in BP lexicon, Machado-Rocha (2016) takes the opposite direction and argues that 3P-ACC clitics fail to obtain because the featural configuration of the syntactic projection of clitics in BP is incompatible with those items.

Following Sportiche (1996), we consider clitics to have their own functional projection and that all clitic structures are doubled constructions, either with an overt argument DP/pronoun or with a null element. Since clitic doubling in BP is optional, the clitic head is taken to be an optional agreement projection (according to McCloskey (1996)) in the domain of *v*. And because it generates only the 1st and 2nd person marks *me* and *te*, it is assumed that this projection hosts only one uninterpretable, unvalued person features, namely [_iautor:±], based on the ideas of Adger (2006).

(6)



Since 1P and 2P objects have a valued counterpart for the unvalued feature of ClP ($[\text{autor:}+]$ and $[\text{autor:-}]$, respectively), these elements can license a clitic (when the object is 1P or 2P *pro*) and can appear in clitic doubling constructions (when the object is 1P or 2P overt pronoun or a DP with 1P or 2P possessive determiner). However, when the object is a 3P overt element or *pro*, there will be no potential goal for the valuation of ClP's unvalued formal features, and the derivation crashes at this point:

(7) ^{*/?}Eu o ajudo (*pro*)

Because of this, when we leave aside the 3P null object, the only possible construction for 3P-ACC arguments is with the full pronoun (see Machado-Rocha

(2016) for details on how CIP can be present or absent in the derivation, depending on the presence of a valued counterpart of the feature [_uautor:±]).

When we consider the data that will be discussed in this paper, this analysis predicts that the featural configurations of the clitic projection in BP oral registers and in BP standard writing are not the same. As we have seen, Nunes (1993, 2011) assumes that the acquisition of 3P-ACC clitic forms in BP standard writing is the result of a process similar to second language acquisition. In these terms, we have to consider that in the grammar of standard BP writing, a different set of features is present in CIP, so that 3P-ACC clitics are possible.

This line of reasoning is coherent with Kato (2005)¹, which proposes that I-language is composed of a core grammar and a “marked periphery”, as assumed by Chomsky (1981). This core-periphery structure would account for a kind of “bilingualism” in what regards L1 vs. literate knowledge. For Kato, L1’s marked periphery is where parametric values may be present in a different fashion when compared to core grammar: these values would be marked, they would have a recessive character, and they might assume a competitive behavior during schooling in relation to those values defined in core grammar. The marked periphery would then account for phenomena of linguistic borrowing, change residues, creations, such that individuals of the same community may or may not present these phenomena marginally. The marked periphery is also responsible for different levels of literacy, and account, for instance, for the presence of structures that cannot be attributed to the core grammatical knowledge. (KATO, 2005, p. 132-134).

Due to the great range of differences between spoken and written Brazilian Portuguese, Kato argues that, for a Brazilian child, learning how to write is like learning a second language (KATO, 2015, p. 139). The marked periphery would be responsible for the creation of individual “variation islands”, providing the speaker with expressive nuances. Thereby, a literate adult would be a code-switching bilingual that alternates between the core grammar G1 and the marked periphery G2. Kato highlights that G2 does not have the same nature of G1, the former being comprised of superficial fragments of the initial parametric fixation. According to her, G2 is not actually the result of parametric selection, but represents “stylistic rules”, arbitrarily selected from

¹ We thank an anonymous referee for pointing out this correlation to us.

past grammars or borrowed from European Portuguese grammar. (KATO, 2005, p. 142).

In what follows we will show that the acceptability of 3P-ACC clitics and 3P tonic pronouns in written modality seems to indicate that both G1 and G2 are being considered by the subjects in their evaluations.

2. Investigating Clitic and Tonic Pronouns in Written Modality

Materials and Methods

Apparatus

We conducted an online Acceptability Judgment Task built at the website Easy Test Maker and conducted through Sona Systems. The independent variable was the direct object pronoun in the target sentences (3P-ACC clitics or 3P tonic pronoun). The dependent variable was the acceptability ratings provided by the participants.

The manner a sentence, a construction or any other linguistic unit sound to a speaker can offer important clues about his/her linguistic knowledge (SOUZA; OLIVEIRA, 2014; SOUZA et al., 2015). Consequently, the acceptability judgment task has been used by linguists who seek to investigate the grammar of different classes of speakers. This experimental paradigm consists in observing how speakers evaluate sentences in relation to their well-formedness (KELLER, 1998). The task is relatively simple: a representative sample of a speaking community is presented to a group of sentences (target sentences, control sentences and distractor sentences) and evaluate, usually with numbers, how acceptable each sentence sounds. After that, the obtained data is statistically analyzed, allowing us to observe potentially significant contrasts among the independent/manipulated variables.

It is important to notice that what is evaluated is the acceptability of the sentences and, not necessarily, their grammaticality. Sometimes these terms are used interchangeably, but they refer to different things in fact. Whereas grammaticality is a theoretical construct based on linguistic theories rules, acceptability is a perceptual construct. Thus, the grammaticality of a sentence is a logical consequence of linguistic theory assumptions and its acceptability is the conscious sensation speakers have when reading/listening to it (MYERS, 2009).

Participants

In this study, we carried out an acceptability judgment task with a between subject design. This strategy of not exposing participants to both pronouns aimed at reducing the possibility that they identified the target structure of the study. We understand that participants would behave more naturally if they processed and analyzed only one type of pronoun during the task. The activation of two types of pronouns during the task might encourage them to make explicit comparisons that do not reflect their natural behavior. 82 participants responded to List A, in which the target sentences had a 3P tonic pronoun in direct object position, and 81 participants responded to List B, in which the target sentences had a 3P-ACC clitic also in direct object position. Thus, in total, 163 participants took part in this study voluntarily.

The participants were all residents of the Belo Horizonte metropolitan area, a region in which Brazilian Portuguese is the prevalent language. They were 24 years old in average (SD=6 years). 109 were women (66,9%) and 54 were men (33,1%).

The participants had different levels of education. 33 (20,7%) people were technical high school students, 80 (50,3%) were undergraduates, 23 (14,5%) had a college degree and 23 (14,5%) had both a college degree and a complete or incomplete post graduation attendance. It is important to highlight that there are studies suggesting that the use of this kind of convenience sample does not interfere with the results. Souza et al. (2015), for example, demonstrated that students of linguistics did not behave differently from students of other areas in an acceptability judgment task.

Besides analyzing the general behavior of these participants, it was also possible to consider the impact of formal education in the acceptability judgment of the 3P-ACC clitic and 3P tonic pronouns as direct objects. Such analysis is non-trivial to the understanding of the focused pronouns in BP since there are studies indicating that the speakers' preferences concerning these two pronouns vary, in both oral and written modality, according to their level of education (CORRÊA, 1991).

Stimuli

In total, 64 sentences composed the experimental corpus of the acceptability judgment task. There were 8 target items, 56 distractors (Appendix 1). The target sentences were 8 instances of the depictive construction, as illustrated in (8). All the target sentences were formed by two clauses: the first one had an NP subject, a VP with a transitive verb in the past and an NP direct object; and the second one started with the conjunction

“e/and” followed by a VP formed by another transitive verb in the past with a pronoun as direct object – 3P tonic pronoun in List A and 3P-ACC clitic in List B – that referred to the direct object of the previous clause, and an AP that described the direct object. To avoid ambiguity, the two NPs in the first clause differed from each other in terms of gender and the pronoun in the second clause agreed only with the direct object. As described, these sentences were built based on the needs of the experiment and as such they are introspective data. Because of this, we will not distinguish among the various written genres in what regards to their relation to orality and literacy.²

- (8) Linda fatiou o salmão e comeu **ele** cru
 Linda sliced the salmon and ate **he/it** raw
 “Linda sliced the salmon and ate it raw”

- (9) Linda fatiou o salmão e **o** comeu cru
 Linda sliced the salmon and **him/it** ate raw
 “Linda sliced the salmon and ate it raw”

The distractor items included grammatical³ (24), ungrammatical⁴ (8) and non-standard sentences⁵ (24). We chose these sentences because they can all be easily interpreted regardless of their grammatical or standard status. All the items were pseudo-randomized in a manner that the target sentences would not be likely to be presented in sequence. Therefore, we also diminished possible biases related to the order of exposition and the repetition in sequence of the target structure.

Before starting the task, the participants were also presented to 4 sentences with comments about how the experimenter would judge their acceptability. The aim of these examples was to illustrate that the acceptability judgment should be more focused on

² A referee rightly drew our attention to a very important point about written genres and their relation to orality and literacy. Written genres could be characterized along a continuum orality-literacy, depending on how they approach or depart from spoken language (BORTONI-RICARDO, 2004). Based on this, Freire (2011) demonstrated that the presence of 3P-ACC clitics in BP texts consistently increases along this continuum, such that comics and strips show a low frequency of clitics while articles and editorials have the greatest number of these elements. In a future study, it would be very interesting to replicate our work using real data organized according to Bortoni-Ricardo’s *continuum*. In the present text however, as we use introspective data, this approach is not possible.

³ Ex: Juca odeia matemática e história; Ronaldo encaminhou os formulários ao cliente.

⁴ Ex: *Lúcia jogaram vídeo game no final de semana.

⁵ Ex: Os jogador treinou ontem no estádio; Rafaella contou o governo a verdade.

the sentences syntactic-semantic structure and not on the sentences content itself. Moreover, the participants underwent a training session in which they evaluated 4 sentences before starting the task.

Procedure

The procedure to perform the acceptability judgment task was relatively easy. First, the participants were presented to the instructions that explained how they should evaluate the sentences to be presented. Second, they read the commented sentences⁶, which suggested that the participants should focus on the structures used in each sentence and not on their pragmatic content. After that, the participants underwent the training session and started the task. The sentences were presented continuously and the participants rated their acceptability from 1 to 5 (i.e. we use a 5-point Likert Scale). 1 represented the lowest acceptability level, 5 represented the highest acceptability level and the other numbers represented intermediate acceptability.

Analysis and Discussion

The data obtained with the acceptability judgment task was registered and generated with Sona Systems, after that, it was tabulated in a Microsoft Excel spreadsheet. To perform the statistical analysis, we used the IBM SPSS statistical package version 21. Valid subjects' means across all critical items and items' means across all valid subjects were compiled from the raw data. The data is described in TAB. 1 and the means are illustrated in GRAPH 1.

⁶ These are some of the commented sentences participants read in BP followed by their translation:

Frase 1 – *Eu me morri amanhã.* *Eu daria nota 1 para a frase 1. A estrutura da frase soa extremamente estranha para mim. Não consigo sequer entender o sentido do que está sendo dito.*

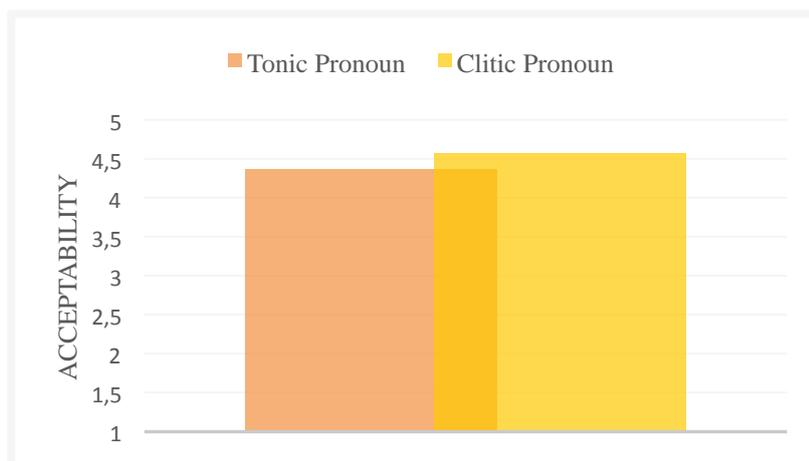
Sentence 1 – I died myself tomorrow. I would rate sentence 1 a 1. The structure of the sentence sounds extremely strange to me. I cannot even understand the meaning being conveyed.

Frase 2 – *O unicórnio foi baleado pela formiga.* *Eu daria nota 5 para a frase 2. Apesar de o evento descrito não ser comum, a estrutura da frase me parece totalmente aceitável.*

Sentence 2 – The unicorn was shot by the ant. I would rate sentence 2 a 5. In spite of the fact that the event being described is not common, the structure of the sentences seems to be totally acceptable.

Pronouns	Mean	Median	Minimum Score	Maximum Score
3P Tonic	4,38	4,63	2,25	5,00
3P-ACC Clitic	4,57	4,75	3,25	5,00

Table 1: Means, median, minimum and maximum scores for 3P-ACC clitics and 3P tonic pronouns.



Graph 1. Acceptability of tonic pronouns (LIST A) and 3P-ACC clitics (LIST B)

The compiled means were tested for normality with the Shapiro-Wilk test. The subjects' means across critical items differed significantly from the normal distribution for both List A ($W=.783$, $p<.001$) and List B ($W=.838$, $p<.001$). The same occurred to items' means across valid subjects for List A ($W=.814$, $p<.05$) and List B ($W=.695$, $p<.01$). Due to the non-normality of the data and the nature of the data (Likert scale), we proceeded with our statistical analysis using non-parametric tests.

In order to reach the goal of our investigation, which was to compare the acceptability of the 3P-ACC clitics with the acceptability of the tonic pronoun, the Mann-Whitney test was performed. The general results, which included the means of all valid subjects, did not yield a significant difference in the judgment of 3P-ACC clitics and tonic pronouns by subjects ($U=2958$, $W= 6279$, $Z=-1216$, $p=.22$), but it did by items ($U=11$, $W=47$, $Z=-2207$, $p<.03$). Both the tonic pronouns and the 3P-ACC clitics were highly accepted by the participants of our study.

We understand that these results indicate that the oral modality influences partially the written modality. On the one hand, the tonic pronouns, which are frequent

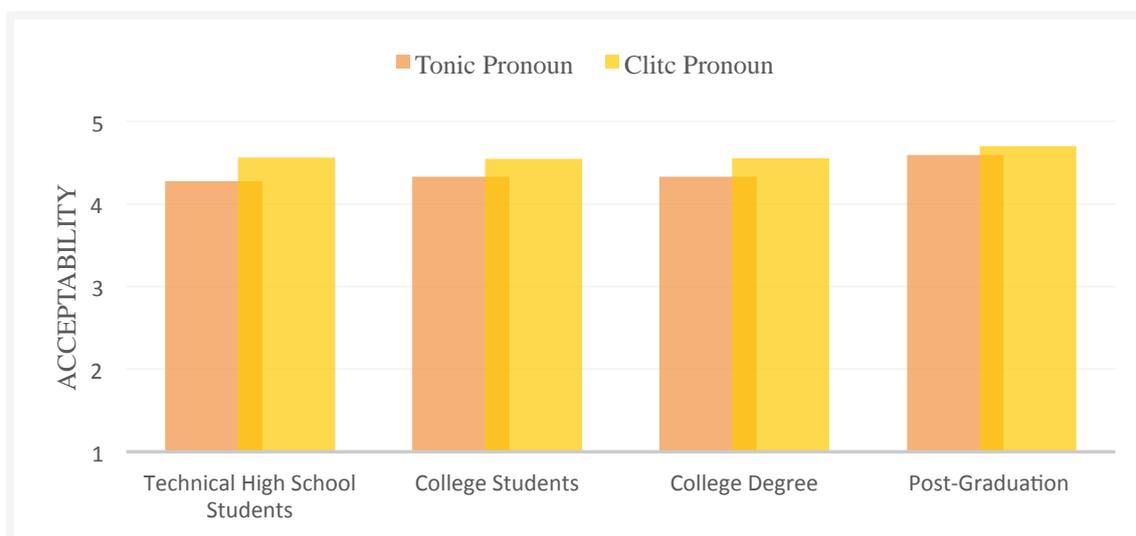
Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017.

in oral modality (even though far less frequent than the null object, as shown by the referred researches), but not in written modality (at least for texts that follow the grammar of standard BP), had high acceptability. This suggests that the high frequency of these tonic pronouns in oral modality may be capable of enhancing their acceptability in written modality. On the other hand, the 3P-ACC clitics, which are frequent in the written modality (at least for texts that follow the grammar of standard BP), but not in the oral modality, also presented high acceptability. This can be interpreted as evidence that the absence of the clitic pronoun in the oral modality does not interfere in its acceptability in the written modality for people with at least technical high school level of education. Therefore, the results suggest that both pronouns may be suitable for experimental items in readings tasks for this population.

We now proceed with the analysis considering the participants' level of formal education. The results were considerably similar when we compared the different groups. The data from technical high school students did not yield a significant difference by subjects ($U=139$, $W=239$, $Z=-.231$, $p=.25$) and it yielded only a marginally significant difference by items ($U=14$, $W=50$, $Z=-1899$, $p<.06$). The data from college students did not yield a significant difference by subjects ($U=50$, $W=86$, $Z=-.652$, $p=.52$), but it did by items ($U=12$, $W=48$, $Z=-2100$, $p<.04$). The data from people with college degree did not yield a significant difference by subjects ($U=809.5$, $W=1,799.5$, $Z=-.638$, $p=.52$), neither did it by items ($U=17$, $W=53$, $Z=-1583$, $p=.11$). The data from the participants with a complete or incomplete post-graduation degree did not yield a significant difference by subject ($U=65$, $W=56$, $Z=.00$, $p=1.0$), neither did it by items ($U=16$, $W=52$, $Z=-1694$; $p=.09$). Thus, in general the groups did not show a significant difference in acceptability for the sentences with the 3P-ACC clitics and tonic pronouns that we tested, especially in the analysis of subjects as a random factor. The few differences found in the analysis considering the items as a random factor showed that the 3P-ACC clitics had the acceptability a little higher than the tonic pronouns. The results are described in TAB. 2 and the means are illustrated in GRAPH 2.

Groups	Tonic Means	Tonic Medians	Clitic Means	Clitic Medians
Technical High School Students	4.27	4,56	4.56	4,75
College Students	4.33	4,56	4.54	4,63
College Degree	4.33	4,63	4.56	4,75
Post-Graduation	4.60	4,63	4.70	4.81

Table 2: Means and Standard Deviation for 3P-ACC clitics and tonic pronouns according to the level of schooling.



Graph 2. Acceptability judgment means for 3P-ACC clitics and tonic pronouns by participants with different levels of education.

We then conducted the Kruskal-Wallis test to observe if the acceptability of the pronouns varied significantly across the four groups tested. The results for the tonic pronoun did not yield a significant difference by subjects ($\chi^2=.468$, $p<.93$) and it yielded only a marginally significant difference by items ($\chi^2=8324$, $p<.06$). The results for the 3P-ACC clitics did not yield a significant difference by subjects ($\chi^2=.801$, $p=.85$), neither did it by items ($\chi^2=2070$, $p<.56$). Thus, our results suggest that both 3P-ACC clitics and tonic pronouns are similarly well accepted by all the education level groups observed.

The fact that 3P-ACC are not frequent in the oral modality does not seem to influence its acceptability in the written modality to a point that it would decrease their ratings in our acceptability judgment task. Their presence in the written modality seems to be enough to guarantee its proper perception by native speakers of BP with at least on going technical high school level of education. The high productivity of tonic pronouns in the oral modality, dissimilarly, seem to extend to the written modality and so these speakers perceive them as well accepted, despite the fact that they are considered to be errors in prescriptive grammars. These results are thus very coherent with the analyses of Kato (2005) and Freire (2011), which claim that schooling can rescue lost forms in orality into literate writing. And we can assume also that both core grammar knowledge G1 and peripheral stylistic knowledge G2 are being consistently applied by the subjects in their evaluations. Therefore, it seems that both 3P-ACC clitics and tonic pronouns can be part of experimental items of reading tasks designed for participants with a level of education ranging from technical high school to post graduation, especially in experiments that do not involve a specific genre, which was the case of our acceptability judgment task.

In face of the empirical reality and most of the analyses about the loss of 3P-ACC in BP, it would not be coherent to assume that the subjects involved in this experiment are employing a single grammatical knowledge that recognizes both clitic and tonic pronouns. As we have discussed in section 2, members of a same community, which represent a single speech community from a sociolinguistic point of view, were divided into two groups, in order to avoid comparison biases. In terms of linguistic competence, our outcomes indicate that the participants are considering two different kinds of knowledge, so that both forms are highly accepted: their native core grammar G1, that generates the tonic form, and the second language-like knowledge of the standard register G2, which allows for the acceptance of clitics. This assumption is compatible with the line of reasoning of Nunes (1993, 2011), and with approaches for individual variation (cf. DORIAN, 1994; HENRY, 2005, and the references cited there), which predicts individual bilingualism, when an individual is proficient in more than one dialect of a given language.

3. Conclusion

The comparability of results obtained through different protocols for data gathering is non-trivial for any long-term scientific enterprise, since they contribute to the

cumulative advancement of knowledge. Our study sought to add to the understanding of the pronominal system in BP and, more specifically, it addressed the discussion about whether 3P-ACC clitics are suitable in linguistics and psycholinguistic experiments.

The fact that 3P-ACC clitics have been seen as unproductive in BP has made some researchers consider these items inadequate for experimental purposes and argue that only 3P tonic pronouns should be used, as discussed in Maia and Lima (2014). However, our results showed that 3P-ACC clitics could be suitable for experimental items of reading tasks, which is in accordance with some previous findings. Corrêa (1991) analyzed several oral and written spontaneous narratives produced by people with different levels of formal education. Her results showed that in the oral narratives, the 3P tonic pronouns are prevalent for the primary and middle school students, but the 3P tonic pronouns and the 3P-ACC clitics are used in the same proportion by the university students. In the written narratives, dissimilarly, the 3P-ACC clitics are prevalent for all the students above middle school and 3P tonic pronouns are almost inexistent among the university students. Furthermore, Maia and Lima (2014) showed that 3P-ACC clitic pronouns are more easily processed than the 3P tonic pronouns by people with a college degree. These results suggest that 3P-ACC clitics are acquired during the process of formal education, whereas the 3P tonic pronouns are acquired in a more natural manner. If Kato (2005) is correct, this can be accounted for as a case of bilingual knowledge, with tonic pronouns being part of the core grammar and ACC clitics as a structure belonging to a marked periphery.

The results yielded by the acceptability judgment task conducted in the present study indicate that both 3P-ACC clitics and 3P tonic pronouns have high acceptability for students with technical high school or higher levels of education. The few significant differences encountered between the two pronouns in analysis of items as random factor favored 3P-ACC clitics. Thus, it is possible to assume that 3P-ACC clitics can be used in the composition of experimental items of reading tasks to be performed by students with higher levels of education. Our results, therefore, also help to validate acceptability judgment tasks whose critical items had 3P-ACC clitics (OLIVEIRA, 2013; 2015; 2016; OLIVEIRA et al., 2017; SOUZA; OLIVEIRA, 2017).

Notwithstanding the relevance of these observations concerning the acceptability of the 3P-ACC clitics and 3P tonic pronouns as direct objects, important questions remain to be answered. If our expectations are correct, the results presented here reflect individual competence for multiple dialects/registers. This in turn implies that the

subjects employed two different types of grammatical knowledge in evaluating structures with clitics and tonic pronouns, structures that belong to different dialectal grammars. Thus, more detailed explorations of whether or not such knowledge is accessed automatically must be conducted. During an acceptability judgment task, participants can rely on their explicit knowledge and, hence, these results do not seem to be reliable evidence of the speakers' implicit knowledge. An experiment focused on the online processing of the 3P-ACC clitics may reveal important information about the presence of these pronouns in the speakers' implicit knowledge. Also, this study does not reveal when 3P-ACC clitics start to be noticed. Therefore, it is important to gather data about the perception of speakers with lower levels of education in order to understand when the 3P-ACC start to be well-accepted.

References

- ADGER, D. Combinatorial Variability. *Journal of Linguistics* v. 42, n. 3, p. 503-530, 2006.
- BORTONI-RICARDO, S. M. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola Editorial, 2004.
- CORRÊA, V. *O objeto direto nulo no português do Brasil*. 1991. 90p. Thesis (MA in linguistics) – Linguistic Department, State University of Campinas. Campinas.
- CHOMSKY, N. *Lectures on government and binding*. Dordrecht: Foris, 1981.
- CYRINO, S. Observações sobre a mudança diacrônica no português do Brasil: objeto nulo e clíticos. In: ROBERTS, I; KATO, M. (eds). *Português brasileiro: uma viagem diacrônica*. Editora da UNICAMP: Campinas, 1993, p. 163-184.
- CYRINO, S. *O objeto nulo no português do Brasil: um estudo sintático-diacrônico*. Editora UEL: Londrina, 1997.
- DORIAN, N. Varieties of variation in a very small place: social homogeneity, prestige norms and linguistic variation. *Language*, v. 70, n. 4, p. 631–696, 1994.
- DUARTE, M. Clítico acusativo, pronome lexical e categoria vazia no português do Brasil. In: TARALLO, F. (org.) *Fotografias Sociolinguísticas*. Campinas: Pontes, 1989. p. 19-34.

- FREIRE, G. Acusativo e dativo anafórico de 3.^a pessoa na escrita brasileira e lusitana. *Revista da ABRALIN*, v. 10, n. 1, p. 11-32, 2011.
- GALVES, C. O objeto nulo no português brasileiro: percurso de uma pesquisa. *Cadernos de Estudos Linguísticos*, v. 17, p. 65-90, 1989.
- HENRY, A. Idiolectal variation and syntactic theory. In: CORNIPS, L; CORRIGAN, K. *Syntax and Variation - Reconciling the biological and the social*. John Benjamins: Amsterdam/Philadelphia, 2005. p. 109-122.
- KATO, M. The distribution of pronouns and null elements in object position in Brazilian Portuguese. In: ASHBY, W; PERISSINOTTO, M; RAPOSO, E (eds). *Linguistic Perspectives on the Romance Languages*. John Benjamins: Amsterdam / Philadelphia. 1993. p. 225-235.
- KATO, M. A gramática do letrado: questões para a teoria gramatical. In: MARQUES, M. A; KOLLER, E; TEIXEIRA, J; LEMOS, A. S. (orgs). *Ciências da Linguagem: trinta anos de investigação e ensino*. CEHUM (U. do Minho): Braga, 2005. p. 131-145.
- KELLER, F. *Grammaticality judgments and linguistic methodology*. Centre for Cognitive Science University of Edinburgh. Research Paper EUCCS-RP. 1998. p. 1-15.
- MACHADO-ROCHA, R. *Spontaneous speech – high-school students from Belo Horizonte*. Unpublished corpus, 2013.
- MACHADO-ROCHA, R. *O redobro de clítico no português brasileiro dialetal*. Dissertation (Ph.D. in Linguistics) 2016. 130p. College of Letters, Federal University of Minas Gerais, Belo Horizonte.
- MACHADO-ROCHA, R. and RAMOS, J. Clitic doubling and pure agreement person features. *Revista de Estudos da Linguagem*, v. 24, n. 2, p. 378-416, 2016
- MAIA, J. and LIMA, M. Referenciação e técnicas experimentais: aspectos metodológicos na investigação do processamento correferencial em português. *Revista de Estudos da Linguagem*, v. 22, n. 1, p. 67-93, 2014

- MCCLOSKEY, J. Subjects and Subject Positions in Irish. In: BORSLEY, R; ROBERTS, I (eds.) *The syntax of the Celtic languages: a comparative perspective*. Cambridge University Press: New York, 1996, p. 241-283.
- MYERS, J. Syntactic judgment experiments. *Language and Linguistics Compass*, v. 3, n. 1, p. 406-423, 2009.
- NUNES, J. Direção de cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro. In: ROBERTS, I; KATO, M (eds). *Português brasileiro: uma viagem diacrônica*. Editora da UNICAMP: Campinas. 1993. p. 207-222.
- NUNES, J. On the diachronic reanalysis of null subjects and null objects in Brazilian Portuguese: triggers and consequences. In: RINKE, E; KUPISCH, T (eds.). *The development of grammar: language acquisition and diachronic change - In honor of Jürgen M. Meisel*. John Benjamins: Amsterdam/ Philadelphia. 2011. p. 331-354.
- OLIVEIRA, C. *A Construção Resultativa e sua Representação por Bilíngues do Par Linguístico Português do Brasil e Inglês*. 2013. 118p. Thesis (MA in Linguistics). College of Letters, Federal University of Minas Gerais, Belo Horizonte.
- OLIVEIRA, C. Influência translinguística na representação da construção resultativa preposicionada na L1 (português brasileiro) e na L2 (inglês). *Prolíngua*, v. 10, n. 1, p. 121-133, 2015.
- OLIVEIRA, C. *Processing, representation and learnability of the resultative construction by Brazilian Portuguese-English bilinguals*. 2016. 202p Dissertation (Ph.D. in Linguistics). College of Letters, Federal University of Minas Gerais, Belo Horizonte.
- OLIVEIRA, C., SOUZA, R. and OLIVEIRA, F. Bilingualism effects on L1 representation and processing of argument structure. *Journal of the European Second Language Association*, v. 1, n. 1, 2017.
- OLIVEIRA, S. Objeto direto nulo, pronome tônico de 3ª pessoa, SN anafórico e clítico acusativo no português brasileiro: uma análise de textos escolares. *Revista Virtual de Estudos da Linguagem – ReVEL*. v. 5, n. 9, 2007.

- PAGOTTO, E. Clíticos, mudança e seleção natural. In: ROBERTS, I; KATO, M (eds). *Português brasileiro: uma viagem diacrônica*. Editora da UNICAMP: Campinas, 1993, p. 185-203.
- SOUZA, R. and OLIVEIRA, C. The learnability of the resultative construction in English L2: a comparative study of two forms of the acceptability judgment task. *Revista da Abralín*, v. 8, n. 2, p. 375-410, 2014
- SOUZA, R. and OLIVEIRA, C. Are bilingualism effects on the L1 byproducts of implicit knowledge? Evidence from two experimental tasks. *Revista de Estudos da Linguagem*, v. 25, n. 3, p. 1685-1716, 2017.
- SOUZA, R; OLIVEIRA, C; SOARES-SILVA, J; PENZIN, A; SANTOS, A. Estudo sobre um parâmetro de tarefa e um parâmetro amostral para experimentos com julgamentos de aceitabilidade temporalizados. *Revista de Estudos da Linguagem*, v. 23, p. 211-244, 2015.
- SPORTICHE, D. Clitic Constructions. In: ROORYCK, J; ZARING, L. (eds.). *Phrase structure and the lexicon*. Kluwer: Dordrecht, 1996, p. 213-276.
- TARALLO, F. *Relativization strategies in Brazilian Portuguese*. Dissertation (PhD in Linguistics). 1983. 270p. University of Pennsylvania, Philadelphia.

Appendix 1

- Linda fatiou o salmão e comeu ele cru.
- Nilton deu o colégio vários livros.
- Mariana deu sua amiga um livro de contos.
- Marina deu um brinquedo ao garoto.
- Jurema deu um computador ao colégio
- Os menino chegou cedo à escola hoje.
- Jair gosta de tocar guitarra.
- Sérgio falaram Alemão muito bem.
- Erica perdeu o cão e encontrou ele morto.
- Jamil encaminhou o colegiado um e-mail.
- Ricardo encaminhou seu advogado os documentos.
- Ronaldo encaminhou os formulários ao cliente.
- Ronilson encaminhou uma solicitação à prefeitura.
- Os jogador treinou ontem no estádio.
- Juca odeia matemática e história.
- Lúcia jogaram vídeo game no final de semana
- Júlia perdeu o telefone e encontrou ele quebrado.
- Rafaella contou o governo a verdade.
- Daniel contou sua esposa a fofoca.
- Daniela contou um segredo ao professor.
- José contou a história à polícia.
- As amiga comeu pizza no restaurante.
- Úrsula estuda Italiano todos os dias.
- Antônio usaram o computador a noite
- A gata pegou o rato e comeu ele vivo.
- A prefeitura concedeu o clube um prêmio.
- O diretor concedeu as secretárias um aumento.
- Janaína concedeu uma entrevista aos repórteres.
- O presidente concedeu um benefício à universidade.
- As criança brincou na rua pela manhã.
- Adilson fala alemão muito bem.
- Ronaldo gostaram de tocar guitarra.
- Renato comprou a pizza e comeu ela fria.
- Tatiana mostrou o hospital o problema.
- Luzia mostrou a menina as bonecas.
- Lourdes mostrou o projeto ao seu chefe.
- Moacir mostrou uma proposta à empresa.
- Os policial prendeu o criminoso à noite.
- Márcio ama seu irmão e sua irmã.
- Adélia beberam café pela manhã

- Natália gostou do computador e comprou ele novo.
- Tarcísio vendeu o açougue uma geladeira.
- Carlos vendeu seu irmão uma moto.
- Celso vendeu o carro ao irmão.
- Cássio vendeu o carro à concessionária.
- Os turista visitou a lagoa da cidade.
- Tati joga vídeo game aos sábados.
- Tânia odiaram matemática e história
- Eu ensopei a camisa e usei ela molhada.
- Luís enviou o Brasil uma carta.
- Marilda enviou a coordenadora seu projeto.
- O professor enviou os arquivos aos alunos.
- O cliente enviou uma reclamação à fábrica.
- As filha assustou o pai com a notícia.
- Timóteo usa o computador à noite.
- Michel estudaram Italiano todos os dias
- Iara preparou o café e bebeu ele quente.
- O bilionário ofereceu os EUA uma nova fábrica.
- Helena ofereceu os convidados caipirinhas.
- A garçonete ofereceu a sobremesa aos convidados.
- O professor ofereceu um software ao laboratório.
- As vizinha viu o meu cachorro fugir.
- Raimundo toma café pela manhã.
- Linda amaram o irmão e a irmã dela.

The prosody of Elliptical Constructions in Brazilian Portuguese: an Experimental study

Lílian Teixeira de Sousa^a

ABSTRACT

Ellipsis phenomena are studied according to two conflicting hypotheses: the semantic approach, which claims that the identification of the ellipsis-antecedent is semantic; and the syntactic approach, which argues in favor of a structural identification between ellipsis and its antecedent. Alternative analyses have argued that there are different types of ellipsis, some of them licensed by information structure, especially topic and focus. In this paper, in order to test the hypotheses presented above, I realized a prosodic study of elliptical sentences produced by 12 female native speakers of BP between 20 and 35 years old. The test consisted of five sentences to be completed by the participants plus 28 images containing two scenes; in each scene there was at least one different item (action, complement or subject). The results showed an interesting pattern. In the case of ellipsis, the given information was generally deaccentuated, while the contrasted information (action, complement or subject) were marked intonationally by High (H) or Low-High (LH) contour. If we take into account that contrasted items are moved to the left periphery, leaving a trace in their original position, then we can deduce that in all these cases there are deaccentuated syntactic structures.

KEYWORDS: ellipsis; intonation; information structure

RESUMO

Os fenômenos de elipse são estudados a partir de duas hipóteses conflitantes: A hipótese semântica, que defende que a identificação elipse-antecedente é de natureza semântica, não sendo necessário propor a existência de estrutura sintática; e a hipótese sintática, que argumenta a favor da existência de identificação estrutural entre a elipse e o antecedente. Teorias alternativas têm relacionado condições de licenciamento de elipse a questões de estrutura informacional, especialmente considerando os conceitos de foco e tópico. Neste artigo, como forma de testar as hipóteses acima, realizou-se um estudo prosódico de sentenças elípticas produzidas por 12 falantes do PB do sexo feminino e com idade entre 20 e 35 anos. O teste consistiu de cinco sentenças para serem completadas pelos participantes e 28 imagens contendo duas cenas; em cada cena havia pelo menos um item diferente (ação, objeto ou sujeito). Os resultados mostraram um padrão interessante. No caso de elipse, a informação dada foi geralmente desacentuada, enquanto a informação contrastada (ação, objeto ou sujeito) foi entoacionalmente marcada por contorno alto (H) ou baixo-alto (LH). Se considerarmos que itens contrastivos são movidos para a periferia à esquerda, deixando um traço na posição original, então, pode-se deduzir que nesses casos há estrutura sintática desacentuada.

PALAVRAS-CHAVE: elipse; entonação; estrutura informacional

^a Universidade Federal da Bahia, liliantsousa@gmail.com

Introduction

Ellipsis phenomena, or the omission of some element of a sentence recoverable in context, are usually understood according to one of two conflicting hypotheses: the semantic hypothesis, which argues that ellipsis-antecedent identification is semantic in nature (HARDT, 1993, HARDT & ROMERO, 2004, LOBECK, 1995); and the syntactic hypothesis, which argues for the existence of structural identification between the ellipsis and the antecedent (TANCREDI, 1992, CHOMSKY; LASNIK, 1993). Both approaches present consistent arguments, but which do not always cover all the cases analyzed, which has led to the emergence of alternative theories relating ellipsis licensing conditions to questions of informational structure, especially considering the concepts of focus and topic (WINKLER, 2005, KEHLER, 2000). One way or another, both approaches recognize issues of an informational nature as fundamental to the licensing of some types of ellipsis. Brazilian Portuguese (BP), on the other hand, presents interesting specificities with regards to the prosodic and structural creation of focus, which can be an important object of analysis of elliptical structures, leading in turn to a greater understanding of the syntactic or semantic nature of the ellipsis. Thus, the study described in this article has the objective of verifying the existence of a syntactic structure or pro-forms in ellipsis-containing constructions in BP. To this end, I carried out experiments involving intonational aspects of these structures in recordings made with 12 speakers, female¹, aged between 18 and 35 years, all from the metropolitan region of Belo Horizonte in the state of Minas Gerais.

This article is organized as follows: in section 1, Ellipsis and Grammar Theory, I present the main theoretical studies of the phenomenon of the ellipsis. In section 2, I describe the research methodology. In section 3, the results obtained are described. Finally, the final remarks are presented in section 4.

1. Ellipsis and Grammar Theory

The study of interfaces between the subcomponents of grammar has recently occupied an important place among generative studies, whether in the syntax-semantics, syntax-phonology or phonology-language processing interfaces, or in intonation-discourse structure, among others. The main question guiding such work is in understanding the grammatical mechanisms involved in the derivation of phenomena that relate to these

¹ I recorded only female speakers because their F0 pattern are more regular than male speakers.

subcomponents and in the way these subcomponents relate to one another. Accordingly, ellipsis phenomena appear as a broad field of analysis, since the licensing of the ellipsis may result from various subcomponents of grammar, depending on the theoretical model adopted.

The phenomenon of the ellipsis involves restrictions and licensing conditions, such as the need for identifying the ellipsis with its antecedent. The nature of this identification, however, has been marked by disagreement over which component is directly involved. Semantic analyses argue that ellipsis-antecedent identification is of a strictly semantic nature, whereas the more syntactically-oriented analyses argue in favor of a structural identification condition for the licensing of the ellipsis.

The lack of consensus in this matter stems mainly from the lack of uniformity in the restrictions on the occurrence of ellipses. The approach of structural identification, for example, is based on the impossibility of the ellipsis when no formal identification is made between the preceding frame and the elided structure:

- (1) *The incident was reported by the driver, and the pedestrian did too [report the incident] Ex. (6) Kertz (2010)

To Sag (1976), the fact that the ellipsis is precluded in contexts where the voice of the antecedent sentence does not match that of the sentence with the ellipsis is evidence for the need for structural identification, as shown in example (1) above. However, even when there is no symmetry in voice between the antecedent frame and the frame of the ellipsis, in some cases, the ellipsis is possible:

- (2) This problem was to have been looked into, but obviously nobody did
Ex. (24) Kehler (2000)

Examples like the one shown above are often used by supporters of the semantic hypothesis against the need for structural identification in ellipsis licensing. In an attempt to solve this impasse, semanticists such as Kehler (2000) attribute the restriction of the ellipsis to discourse coherence. For Kehler, inferential processes of coherence are sensitive to syntactic structure. Syntacticists, on the other hand, often come to a gradual acceptance of certain contexts of ellipsis. Either way, the impasse remains.

The main consequence of the choice of either approach is the understanding of how these structures are generated and which grammar components are relevant; while the syntactic approach raises the Phonological Reduction Hypothesis (TANCREDI 1992, CHOMSKY; LASNIK 1993, LASNIK 1999), the semantic approach argues for the existence of pro-forms (HARDT 1993, HARDT; ROMERO 2004, LOBECK 1995). For those who advocates the first proposal, elliptical sentences are formed by PF (Phonological Form) component rules, which phonologically delete the redundant information. As for the second proposal, it assumes that there are pro-forms to be derived essentially as empty pronouns generated at the base, there being, therefore, no internal syntactic structure. Both approaches have important consequences for grammar theory, because if there is a grammatical structure in the context of an ellipsis, one must assume that there are unuttered phrases and cores; otherwise, the syntax comes down to what is heard.

A central argument in the distinction between the two approaches is the possibility of extraction to outside the field of the ellipsis. If extraction is possible, it is believed that there is sufficient syntactic structure to host the feature; on the other hand, if extraction is impossible, the hypothesis is that it lacks syntactic structure, that is, that a pro-form is present.

Among the studies taking the syntactic approach, I highlight the work of Tancredi (1992) and Merchant (2001, 2008). For Tancredi, the ellipsis comprises an interface phenomenon involving PF and LF (Logical Form); the claim is that the elided element is not really deleted, but rather not pronounced, that is, it is suppressed in PF. Merchant (2001), analyzing cases of sluicing (IP ellipsis), suggests that the licensing of the ellipsis is given by checking an E constituent in a head to head relationship, causing the deletion in PF. This E constituent, in the author's claim, is related to the Focus Condition, in which an IP can be deleted if it is given (*e-given*). In the author's definition:

- (3) An expression X is *e-given* iff X has a salient antecedent A and, modulo existential type-shifting,
- a. A entails E-clo(X), and
 - b. X entails E-clo(A)

Thus, in Merchant's hypothesis, IP (Inflectional Phrase) or VP (Verbal Phrase) ellipsis ensures that an IP or VP can only be deleted if it is given. In later works, Merchant (2008) deals with issues of structural identification and argues that no parallelism between the antecedent and the target is allowed to depend on the level at which the deletion occurs; if the deletion targets a node below VoiceP, VP is not marked for VoiceP and the parallelism is not allowed.

Recent studies have associated the non-uniformity of the identification data between the antecedent and the elided structure with restrictions of information structure such as, for example, a requirement of topic/comment parallelism.

One way or another, both the syntactic and the semantic approaches associate ellipsis with informational structure. One group assumes that anaphoric rules take place in the semantic/pragmatic component and interacts directly with PF without reference to syntax. The other group assumes that the rules for movement of topic and focus, as well as anaphoric interpretation, anaphoric deaccentuation and particularly conditions of accessibility and identity over the ellipsis, occur in the informational structure-syntax interface, with the immediate effect of silence in PF. However, a third hypothesis, supported by Winkler (2005), integrates these two positions. According to the author, the ellipsis is an interface phenomenon that results from complex interactions between core grammatical components and the component of information structure. The idea is that the result of interpretation is marked in the form of traces on the respective phase, which is sent to the PF. PF derives phonological structures, phase by phase. Thus, in the case of ellipses, PF derives silence instead of phonological structures.

Kertz (2010) also associates ellipses to information structure. For the author, both syntactic and semantic models of the ellipsis reveal that it is characterized by different focus structures. In data involving non-parallelism in terms of voice, for example, she notes that the target subject is in focus and is interpreted contrastively with the passive agent of the preceding sentence.

(4) The driver reported the incident, and AN ONLOOKER_{loc} did too

(5) #The incident was reported by the driver, and AN ONLOOKER_{loc} did too

Exs. (57) e (58) Kertz (2010)

According to the author, the stress on *an onlooker* suggests the occurrence of focus and leads to the interpretation of contrast with *the driver*. In other cases of

unacceptability in structures lacking parallelism, the focus falls on the auxiliary verb, evoking contrast of tense, aspect, mood, polarity, or a combination of both, which is evidenced by the occurrence of a more prominent pitch accent over this item. Thus, the author argues that the (lack of) parallelism between the elided structure and the antecedent structure is actually related to the focus structure and not just the syntactic structure or discursive coherence.

Kertz's (2010) analysis approaches, in a way, the work of Tancredi (1992) and Rooth (1993), by linking the licensing of the ellipsis to focus. For these authors, the ellipsis phenomenon is related to the syntax-phonology interface with a deaccentuation effect, as observed in studies on the effects of focus on prosody, in which pitch accents occur on the item in focus with the deaccentuation of the remaining propositional content. In Kertz's hypothesis, however, the reduction in acceptability in cases of non-syntactic parallelism is not related to the syntax *per se*, but rather the lack of topic/comment parallelism in a contrastive topic structure.

Although many types of ellipsis exist, sluicing (cf. (6)) and VP ellipsis (cf. (7)), for their similarity with regard to questions of discourse, are the most often dealt with types when it comes to analysis which aims at verifying the semantic or syntactic approach.

(6) John can play something, but I don't know what

(7) John can play the guitar and Mary can, too

In the two above examples, there is the need for some equivalent antecedent to be the object of some kind of parallelism. In the hybrid approach, however, the distinction between the different types of ellipsis plays a major role. According to Winkler (2005), for example, VP ellipsis is the result of passing through the first derivational cycle, as per Chomsky (2005), while stripping (8) and gapping (9) are the result of passing through two derivational cycles.

(8) John can play the guitar, {and Mary, too/and Mary as well/but not Mary}
John can play the guitar better than Mary

(9) John can play the guitar, and Mary the violin
John can play the guitar better than Mary the violin

This division into two classes of ellipsis is related to two different concepts of focus: informational and contrastive focus, with the occurrence of various computational processes in the grammar.

In Winkler's model (2005), the syntactic theory of information structure is crucial in the surface derivation of semantic interpretations, i.e., the information structure is considered a central subcomponent of LF, separating structures which do not require shifting from those requiring movement operations; only the latter would have an effect on the superficial semantic interpretation. Thus, syntax and information structure form the core component of grammar, working in parallel with the other components. The two cycles discussed comprise the derivation process in phases, in which cycle 1 operates automatically on the lowest stage sent to LF and which allocates the informational focus *in situ*, while cycle 2 checks the phase for displaced material.

With regard to the PF interface, the silence in the location of the ellipsis, in this model, is the result of a cost-effective division of labor between the phonological, semantic and pragmatic components for discourse-bound ellipsis (DBE), and between the syntactic and semantic components for the derivation of the sentence-bound ellipsis (SBE). The sentence-bound ellipsis is one that does not involve contrast--the focused element remains *in situ*; the discourse-bound ellipsis, on the other hand, is one that involves contrast, the focused element being elevated to a position higher up in the structure, Spec-FocP. The presence or absence of contrastive interpretation is crucial to distinguish between the two types of ellipsis. If the element must be moved to a higher position in which it checks its feature [+contrastive], it must leave traces in intermediate positions. This leads us to believe that, in this case, there is a syntactic structure sufficient to encompass those traces.

Brazilian Portuguese seems to provide interesting data regarding the role of the information structure in ellipsis licensing. Note the data set below:

- (10) a. *O acidente deveria ter sido reportado pelo motorista, mas ele não
 the accident should have been reported by-the driver, but did not
 [reportou o acidente]
 [reported the acidente]
- b. O acidente deveria ter sido reportado pelo motorista, mas ele não
 the accident should have been reported by-the driver, but he not

reportou

reported

‘The accident should have been reported by the driver, but he did not report’

In the above data, the negative item in the second sentence makes up the new information. Thus, the *não*, despite being the only new information, cannot occur without the presence of the verb, as can be observed in the data (10a-b). However, if the "subject-denial" set is interpreted as contrastive information, the *não* (*not*) can occur in isolation, as shown in the example below:

- (11) Dois pedestres foram feridos no acidente, mas o motorista não [foi ferido]
 two pedestrians were injured in-the accident, but the driver not (was injured)
 ‘Two pedestrians were injured in the accident, but the driver not [get hurt]’

Intuitively, the distinction between sentences (10) and (11) is in the contrastive interpretation observed in (11) but not in (10). In Teixeira de Sousa (2012), analyzing the negative structures [Neg VP], [Neg VP Neg] and [VP Neg], I argued that there are in BP three different lexical items with the same morphological form: 1) *não*₁, a sentential negation marker which always occurs in the pre-verbal position and which can be reduced to *num*; 2) *não*₂, present in the final position in the sentence in [Neg VP Neg] structures, which acts as a variable over time indicating the negation of a proposition; and 3) *não*₃ [VP Neg] structures, which functions as an external negation, analyzed in contrastive focus. In the analysis, *não*₁ in both [Neg VP] and [Neg VP Neg] structures has characteristics of a clitic, requiring a verb as host, while *não*₃, of contrastive focus, would be phonologically independent. According to this hypothesis, it may be said that while in (10) the structure is not contrastive and the negative element is the semantic negation *não*₁, in (11) we have the presence of the contrastive item *não*₃. This explanation is also supported in the work of Namiutti (2008). According to the author, in elliptical constructions such as in (10), the *não* cannot be reduced to *num*, and the presence of a tonal accent is essential:

- (12) a. * Dois pedestres foram feridos no acidente, mas o motorista *num*
 two pedestrians were injured in-the accident, but the driver not
- b. Dois pedestres foram feridos no acidente, mas o motorista *não*
 two pedestrians were injured in-the accident, but the driver not
 ‘Two pedestrians were injured in the accident, but the driver not’

If this interpretation of the data in (10) and (11) is correct, the distinction of acceptability in the two cases may indicate either that we have two different structures, which serves as evidence for the interpretation that there is structure in elliptical sentences; or that there is structure in one case but not the other, as in the hybrid approach.

Another interesting question concerns the derivation of these structures. In Teixeira de Sousa (2012), considering the phase hypothesis (CHOMSKY 2005, 2008), I argued different negative items have scope over different elements of the structure, or rather, they are derived at different stages. The *não₁* would have scope over the *vP* phase, *não₂* over the TP in the CP phase, and *não₃* over the item moved to the periphery to the left of the CP phase. The distinction between the structures in the phase hypothesis is still in accordance with the work of Winkler (2005). The author derives the ellipsis from a grammatical system in which syntactic derivation and interpretation occur in parallel. In this model, different types of ellipsis are derived by differing processes involving two cycles: the semantic and pragmatic components interact with the functional cycle, becoming essential for the derivation of discourse-bound ellipsis (DBE), while the syntactic-semantic interface is relevant to the sentence-bound ellipsis (SBE). This proposal deals with two different types of focus, the contrastive and the informational, with different operational processes in the grammar. The location of the informational focus would be in cycle 1, according to the *in-situ* hypothesis, while the location of the contrastive focus is in cycle 2, as explained by the dislocation hypothesis. In this sense, while the informational structural function of the SBE is to isolate the contrastive focus (narrow), the informational structural function of the DBE is to mark the elliptical material as anaphoric or given. Thus, when there is displacement, the constituent that is moved is interpreted as a focus/contrastive topic, and the location of the ellipsis needs syntactic representation; when there is no motion involved, as in the cases of VP ellipsis and VP-anaphora, the construction is explained by the pro forma feature.

Another important point about BP with respect to ellipsis is explored by Namiuti (2008). The author, taking up the work of Cyrino (1997) on the evidence for a syntax-prosody relationship in the regularity of accentuation proposed by João de Barros ², who wrote the first grammar of the Portuguese language, argues in favor of the hypothesis of obligatory accentuation in elliptical constructions in BP. For this, as in cases (13) and (14), in addition to the presence of the *não* and the article for licensing the ellipsis, the pitch accent is also essential.

- c. O João agrediu o Pedro, mas o Paulo **não**
 the João assaulted the Pedro, but the Paulo not
 ‘João assaulted Pedro, but **not** Paulo’
- d. Esses são **ôs** que peço
 these are those that ask-1P
 ‘These are **those** I ask for’ (NP elided *os cadernos (the notebooks)*)
 Ex. (39) and (41) Namiuti (2008)

To Namiuti, both the negation and the determiner cannot be prosodically deficient in cases of ellipsis, because these carry the accent of the phonological phrase. Such behavior would be opposite to that observed in non-elliptical constructions, in which the negative item and the determiner cannot carry tonal accent, as in sentences like “O Pedro *não* comeu.” (“Pedro *did not* eat”) and “A Ana comeu *a maçã*” (“Ana ate *the apple*”), since the lexical head of the phonological phrases *did not eat* and *the apple* are, respectively, the verb ‘eat’ and the NP ‘apple’. Although Namiuti has not performed any experimental study to test her hypothesis, her intuition, if proven, would demonstrate the role of PF in the licensing of the ellipsis in BP.

In the next section, I move on to the research methodology and, following that, the analysis of the results of corpus data.

² “Diálogo da Viciosa Vergonha” (DVV), “Diálogo em Louvor da Nossa Linguagem” (DLNL) e “Gramática da Língua Portuguesa” (GLP).

2. Methodology

I propose the survey and analysis of various types of ellipsis in BP in an attempt to verify the role of information structure in the licensing process. With this, I seek to answer the following questions:

- i. Is there prosodic evidence in favor of the deletion and/or pro-form hypothesis?
- ii. Is there a relationship between the informational role of the elements that make up the ellipsis and the acceptability restrictions?
- iii. If there is an influence of the informational structure on the restriction of certain cases of ellipsis, what is the role of this component?
- iv. What is the role of the PF component in the licensing of ellipsis?

Considering the above questions, I undertook to conduct a study of cases of ellipsis in BP, analyzing experimental data. The guiding hypothesis was that different ellipsis licensing conditions are related to the presence of syntactic structure or pro-forms, as advocated by the hybrid proposal. In order to test this hypothesis, since the information structure tends to be supermarked in the intonation production of sentences, I chose to perform acoustic analysis of elliptical sentences, trying to find different patterns of prosodic realization.

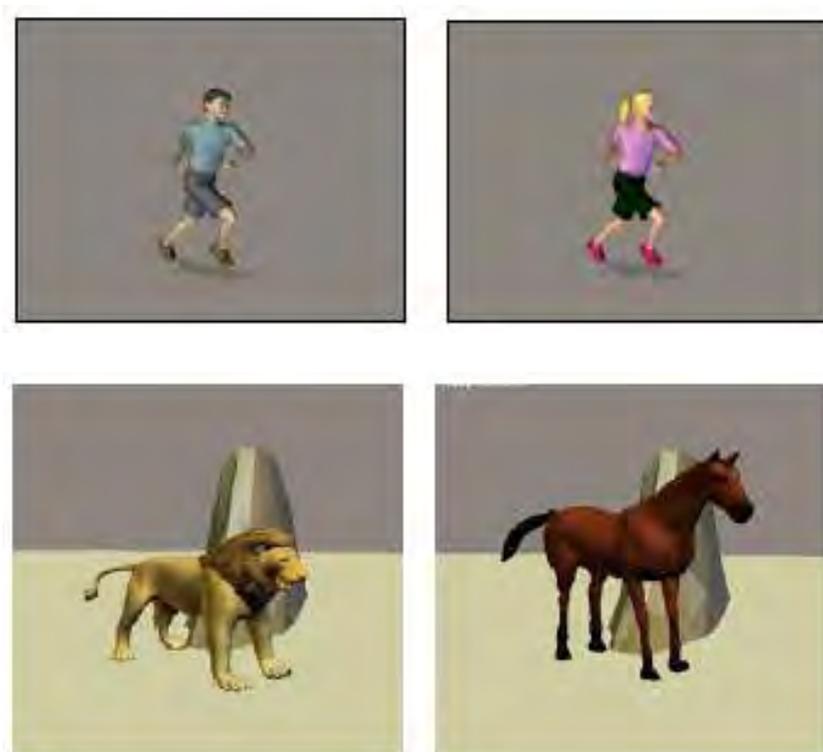
2.1. The Sample: Description

The data that comprise the corpus used in this study consist of recordings collected through controlled tests conducted with 12 female respondents aged between 18 and 35, all natives of the metropolitan region of Belo Horizonte/MG.

The recordings were made in a soundproof booth located in the Language Faculty of the Federal University of Minas Gerais (FALE/UFMG). The equipment and tools used for this task were an MXL model 8900 unidirectional microphone and the recording software Audacity 1.3 beta. The speakers remained seated during the recording, with microphones fixed near their mouths. Data were recorded directly into the computer through the program Audacity 1.3 beta and analyzed with the software *Praat* (Boersma, Weenink, 2010 version 5.2.11).

2.2. Methods and Procedures

The first stage consisted in gathering sentences that were complete and/or contained elided elements. Since my goal was to investigate the relationship between information structure and the occurrence of ellipsis, I turned to the Information Structure Reference Manual, produced by researchers at the University of Potsdam in Germany. I took from this manual some images specially created by researchers at the aforementioned university for studies involving information structure. The images, for the most part, were selected to make up a frame with two scenes, and between the first and second scene there was always a change; if, for example, in the first scene a girl was sitting, in the second there might be a boy in the same position (contrast of the subject). In some pictures, however, there was only one scene with two or more characters performing the same action with different objects; for example, there might be a scene in which a woman eats an apple and then a man eats a banana (contrast of the object). These images served to produce a situation in which it was possible to clearly distinguish between given and new elements and the relationship of contrast between one scene and another. See some of the images used:





The experiment devised for this study was structured as follows: initially, 28 images were selected from those collected in the reference manual and were laid out in a PowerPoint file, one at a time. Of these, six were not specifically related to information structure, but rather passivity versus agency, and were chosen to serve as distractions. The remaining 22 images were used in two situations in the experiment. Initially, I showed the images to the respondents, one at a time, and I asked the respondents to describe them; then, the respondents answered questions. With this process, I expected that the participants would use more complete sentences in the first case and more sentences with ellipses in the second. The informants acted as expected most of the time and produced more complete sentences in the first case and more sentences with ellipses in the second. With this, I was able to collect, for each image, either complete sentences or sentences with ellipses produced by the same speaker, which allowed us to perform a comparative analysis. In addition to the images, I selected phrases, complete and containing ellipses, to be read by the informants. The goal was to collect some VP ellipsis data, given that, in a pilot study, I found that the structure of the experiment, in using contrast elements, would lead to the production of a greater number of IP ellipses.

At the end of the experiment, I analyzed a total of 696 sentences, 372 complete and 324 with VP and IP ellipses.

Considering what the theory predicts with regards to information structure and movement of constituents in general and, in particular, in BP, I defined as IP ellipsis every sentence that did not present any verbal structure in the elliptical construction, and as VP ellipsis those in which only the lexical verb was not pronounced. Thus, even those sentences in which the object of the sentence appeared were considered IP

ellipses, due to the function of contrastive focus observed in the context, given that items marked by the feature [+contrast] are moved to a higher position in the syntactic structure, namely Spec-FocP. For example, we drew from the data the examples below for each of these two types of ellipsis:

- Stripping*
- Na primeira imagem, a menina está correndo e, na segunda, o menino.
in-the first picture the girl is running and in-the second the boy
'In the first picture, the girl is running, and, in the second, the boy.'
 - O João é simpático com todas as pessoas e a Maria também.
the João is friendly with all the people and the Maria too
'João is friendly with everyone and Maria too.'
- Gapping*
- A Ana tinha comido uma maçã e o Pedro uma banana.
the Ana had eaten an apple and the Pedro a banana
'Ana had eaten an apple and Pedro a banana.'
- VP ellipsis
- Ontem ele não tinha lido o artigo, mas hoje já tinha.
yesterday he not had read the article but today already had
'Yesterday he had not read the article, but today he had.'
 - O carro foi atribuído à Maria, mas os outros prêmios não foram.
the car was awarded to Maria but the other prizes not were
'The car was awarded to Maria, but the other prizes were not.'

In all recordings, the sentences were segmented into V-V units, from the start of a vowel to the beginning of the immediate following vowel. This choice was motivated by the observation of many scholars (DOGIL; BRAUN 1988, WONG; SCHREINER 2003, among others) of the relevance of CV transition both for the production and for the perception of speech. Also according to Barbosa (2006), all articulating and acoustic events that are candidate points of anchoring are located in the vicinity of the acoustic *onset* of a vowel.

2.3. Acoustic analysis

In prosody studies, it is common to assert that certain speech events such as focused constituents may find acoustic correlates, either through the establishment of breaks to separate statements, or through the production of emphasis on some constituent to highlight it, etc. Considering our hypothesis that there are differences in prosodic realization of elliptical structures in BP representing informational focus or contrastive focus, I went on to investigate the intonational characteristics of these structures. To this end, I conducted an experiment with the objective of describing the intonational characteristics of elliptical structures and relating them to the information structure.

In prosodic phonology (Cf. SELKIRK, 1984, 1995, NESPOR, VOGEL 1986), the prosodic structure underlying a sentence is hierarchically organized, the phonological segments are grouped into syllables (σ), syllables into feet (Σ), feet into prosodic words (ω), prosodic words into phonological phrases (P), phonological phrases into intonational phrases (I), and those last into utterances (U).

The units of prosodic hierarchy are defined based on mapping rules of syntactic XPs for the prosodic structure. In languages, per Selkirk (1984), phonological phrases are prosodic entities that are derived from syntactic XPs. Many prosodic domains are derived from syntactic XPs and are classified as phonological phrases (P). Selkirk (1984) establishes, then, the XP mapping Condition for P:

XP-to-P Mapping Condition

Mapping constraints relate XPs to phonological phrases, but do not relate XPs to other prosodic entities.

The theory of Selkirk (1995) on syntax-prosody mapping holds that the syntax determines the location of prosodic boundaries, which means that each head of a phonological phrase carries a pitch accent, while the information structure affects the location of the pitch accent. In the case of focus, for example, the element on which it falls carries a pitch accent, which consequently leads to the formation of a prosodic phrase with focus element as its head.

Féry & Ishihara (2010), on the other hand, propose a syntax and prosody interaction representation model that distinguishes phenomena such as prosodic phrasing mapped from syntax and the effects of information structure on prosody as

changes in the f0 register. So while in a sentence of broad focus (when the whole sentence is new information) prosodic boundaries are established according to the syntax; in the case of information structure (focus-given, for example), there would be no modification of borders, but rather a change in pitch record. So for Féry & Ishihara (2010), focus and givenness affect the F0 scale in certain prosodic domains, but do not directly affect the prosodic phrasing. The authors assume, as Truckenbrodt (1995) does, that focus is realized by prosodic prominence:

Focus Prominence

Focus is realized by prominence in the focal area.

The prosodic domain of the focus as expressed in (2) corresponds to its semantic scope, meaning that it is in its domain that the focus is interpreted and receives prosodic prominence. Thus, the domain contains the focused element and identifies the relevant presupposed information for the semantic interpretation of focus. It is a formal feature F that marks the focus and determines its scope. Thus, we see that the notion of givenness is also important for Féry & Ishihara's approach, which adopts the Schwarzschild (1999) hypothesis on the free allocation of F-marking. The authors present two restrictions, one stating that given information is un-F-marked, and another restricting to a minimum the number of F-markings:

- a. GIVEN: A constituent that is not F-marked is given.
- b. AvoidF: Do not F-mark.

Complementing the work of Schwarzschild, Féry & Ishihara hold that given elements are G(iven)-marked. This distinction is necessary because, according to Féry & Samek-Lodovici (2006), given constituents, when not focused, are deaccentuated:

Deaccentuate-Given

A given phrase is not prosodically prominent.

In Féry & Ishihara's hypothesis, then, information structure does not manipulate the boundaries of prosodic phrases, but does change the pitch registers, expanding or narrowing them. The effect of pitch accent manipulation is, according to the authors, the increase or reduction of height compared to the unmarked situation, considering its

status of focus or givenness. In the case of focus, the trend is to increase the topline, since the identification of a constituent as given causes the lowering of the topline of a given domain. The difference of this approach with respect to the others is that the relationship between the different parts of the sentence is modified, so the information structure would cause changes in the scale of the whole sentence rather than targeting only the most prominent accent pitch.

Since prosody appears to be an important feature for both the definition of prosodic boundaries and the marking of information structure, I undertook to carry out an acoustic analysis in order to verify features indicating prominence. The acoustic analysis was performed based on certain acoustic parameters considered in the analysis of emphasis, namely: breaks, melodic contour and initial and final F0 values.

Intonational analysis

In intonational phonology, intonation has its own phonological organization, interpreted as discrete sequences of tonal events realized by the f_0 curve. These events are essentially of two types: *pitch* accents and edge tones. To Ladd (1996), pitch accents are elements of intonational contour and serve as indicators of syllabic prominence, since they need to be associated with lexically stressed syllables, though they do not constitute syllabic prominence in themselves. Edge tones, on the other hand, mark the boundaries between the domains of melodic contours. Thus, the decomposition of intonation contours occurs in a sequence of tonal events, namely high (H) or low (L) targets associated with lexically stressed syllables (tonic syllables) or phrase boundaries.

The ToBI annotation system was developed initially for English and later adapted for other languages. For Brazilian Portuguese, in addition to ToBI, there is the DaTo (Dinamical Tones of Brazilian Portuguese) annotation system. This annotation system proposed by Lucente (2012) describes Focus in intonation according to the concept of dynamic contours, holding melodic intonation contours, the range of tonal variation and the specific alignment between the curve and the linguistic material to be melodic primordials. Given that the DaTo has been specially developed for Brazilian Portuguese, I chose to adopt it in my analysis.

Intonation contours in this system are specified by a static target, while the pitch accent is associated with the lexically stressed syllable of a prominent word. The gamut of tonal variations, on the other hand, delimits the pitch interval in which a target is implemented, considering the value of the target and the range of variation. The specific

alignment, central to this approach, is established, considering the existence of synchrony between movements that produce f_0 and articulatory movements that generate the spectral patterns.

Table 1 below shows DaTo notation with respect to the theory of intonational phonology:

ToBI	DaTo
<i>Pitch accent</i>	<i>Level contours</i>
L *	L
H*	H
	<i>Dynamic Contours</i>
L+H*	LH
L*+H	>LH vLH
H+!H*	LHL HLH
H+L*	HL
H*+L	>HL vHL
<i>Prasal Accents</i>	
L- H-	
<i>Edge Tones</i>	<i>Edge Tones</i>
L%	L
H%	H

Source: Lucente (2012)

Lucente (2012) divides the types of boundaries marking focus into ascending tones (LH, >LH, HLH), descending tones (HL, >HL, LHL) and edge tones (H, L). Regarding ascending tones, she holds that the descent of f_0 within syllables preceding the lexically accentuated syllable (mandatory for the realization of the subsequent movement of ascent) was interpreted by listeners as a change in the degree of focus, becoming thus less emphatic after the rise and more emphatic as the amplitude of the fall increases. The deaccentuation of syllables in a post-focal position also shows that the perception of focus is also associated with the opposition that occurs in unaccentuated syllables after the focus. Thus, the LH contour is presented as the default

tone, most frequent in realizing narrow focus; >LH corresponds to the same function but more emphatic; and HLH may be associated with contrast focus.

Still among the rising contours, Lucente (2012) adds vLH. This contour, called compressed rising, marks narrow focus between two peaks of f_0 where there is no space for, or physiological possibility of, achieving a higher pitch.

On the descending contours, it is said that, although they share a default movement (smooth rise of f_0), the alignment of these contours does not occur in a unified manner; HL, and >HL share the same characteristics, >HL coming late relative to the descent of f_0 ; LHL, on the other hand, has a milder contour declination pattern, triggered by several factors, including number of syllables.

Now that the methods and procedures used in the research have been described, I move on to analysis of the data and presentation of results.

3. Results collected

As described in previous sections, the intonational study of sentences with ellipses can contribute to studies on the subject. Considering the theories of ellipsis, I expected a pitch accent and deaccentuation to occur in elliptical structures that expressed contrast and standard prosodic realization in structures that corresponded to informational focus. This is also in accordance with experimental studies of phonology. Under Selkirk's (1995) theory, in syntax-prosody mapping, syntax determines the location of prosodic boundaries, which means that each phonological phrase head carries a pitch accent. Information structure, on the other hand, affects the location of the pitch accent when there is focus; for example, the element on which focus falls carries a pitch accent, which consequently leads to the formation of a prosodic phrase with the focal element as its head. This latter case describes exactly what Namiuti (2008) says regarding elliptical structures with negations or pronouns in BP, as shown in the previous section.

Féry & Ishihara (2010), like Selkirk, associate prosody and information structure; however, they put forth a model of representation of interactions between syntax and prosody that distinguishes phenomena such as mapping from syntax to prosodic phrasing and effects of information structure on prosody as changes in f_0 register. For these authors, prosodic effects of syntactic structure and information structure should be treated separately.

With respect to the prosodic effects of syntactic structure, Féry & Ishihara hold that the syntactic structure is mapped in prosody through prosodic phrasing, this

phrasing being recursive. In the case of sentences with broad focus, the formation of prosodic phrases and tonal pattern depend entirely on the morpho-phonological structure. While prosodic phrases have heads in the form of abstract positions, such as the realization of pitch accent, prosodic heads are not necessarily performed with pitch accents and can be expressed by duration, intensity or completely different parameters such as tonal and segmental changes. Thus, it is observed that the prosodic heads have correlates in metrical or hierarchical prosodic structure. This relevance of the different levels of phrasing is treated in the literature in different ways, such as, for example, phonological phrases and intermediate phrases or accent domains and rhythmic groups.

In my analysis, as described in the methodology section, I observed mainly breaks and intonational contours, as these are the main indicators of prosodic edges and prominence. To identify possible changes in intonation in the analyzed sentences, I considered the intonation pattern of neutral and focused sentences in BP, as described in the autosegmentally based work developed by Tenani (2002) and Fernandes (2007). According to these authors, the main feature of neutral sentences in BP is the presence of tones associated with phonological words. According to Fernandes, there is the presence of a HL pitch accent associated with the head of the last phonological phrase of the prosodic phrase, accompanied by a low edge tone (L). With respect to the presence/absence of a pitch accent, Fernandes noted that focused elements have the same pitch accent as they have in a neutral context, whereas sentences with focused subjects have no tones associated with phonological words, the accent prominence being associated with the right edge of the P, which contains the focused subject.

An analysis of the selected parameters revealed an interesting pattern. Recalling that the corpus was made up of both complete sentences and ellipses, I observed that complete sentences and those with gapping showed the same pattern of intonation. Sometimes, I observed the occurrence of a break to the right of the contrasted element; however, regardless of whether there was a break, the contrasted item was intonationally marked with focus contour in all data.

Figure 1 below illustrates a case of contrast in the action, induced by an image in which the two scenes are distinguished by the subject of the action, in the first case kicking a ball and in the second throwing it upwards.

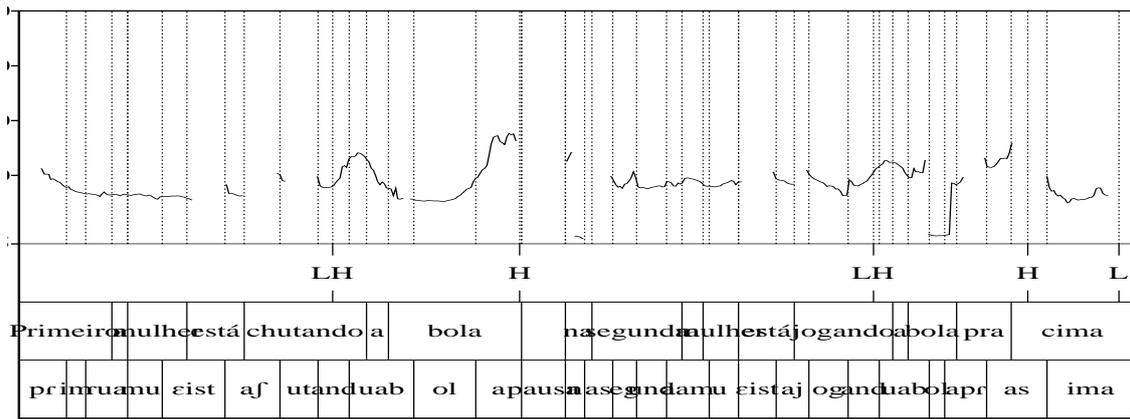


Fig. 1. Deaccentuation of propositional content, LH contour on the contrasting items and high edge contour in the utterance “Primeiro a mulher está chutando a bola; na segunda a mulher está jogando a bola pra cima”.

As can be seen in the figure above, in the two sentences, the propositional content was deaccentuated and an LH focus contour was present in the lexical verb, which carries the semantic information of the action, "kicking" (chutando) in the first sentence, which contrasts with "playing" (jogando) in the second. It is also evident that there was a declination in the f_0 register of the elements in the second sentence. For example, while the peak of the LH contour in the first sentence is at a frequency of 250Hz, in the second, the peak of the LH contour of "playing" is 230Hz. This reduction in the f_0 register, although small in some cases, occurred in nearly all the collected data. Figure 2 below illustrates the same pattern, this time with HL contour of the contrasting items, "table" (mesa) and "chair" (cadeira).

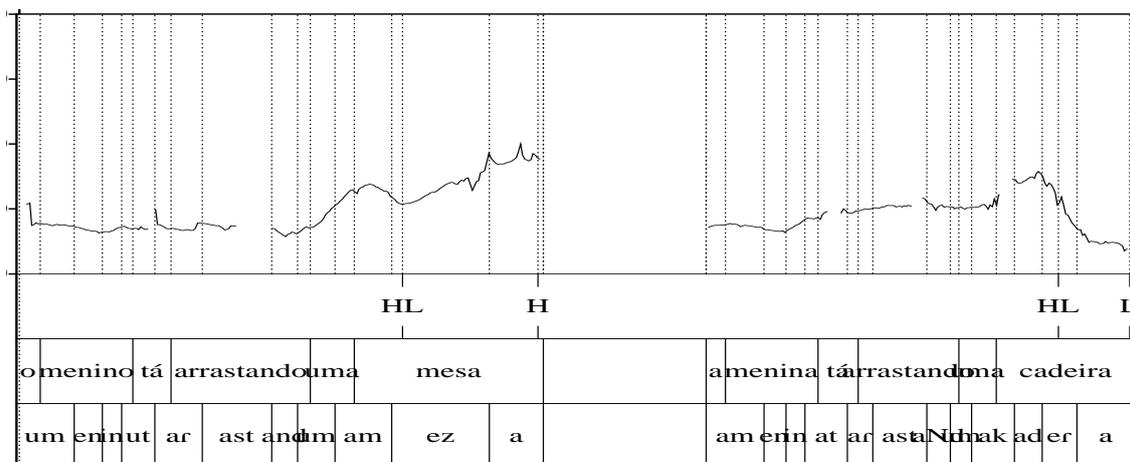
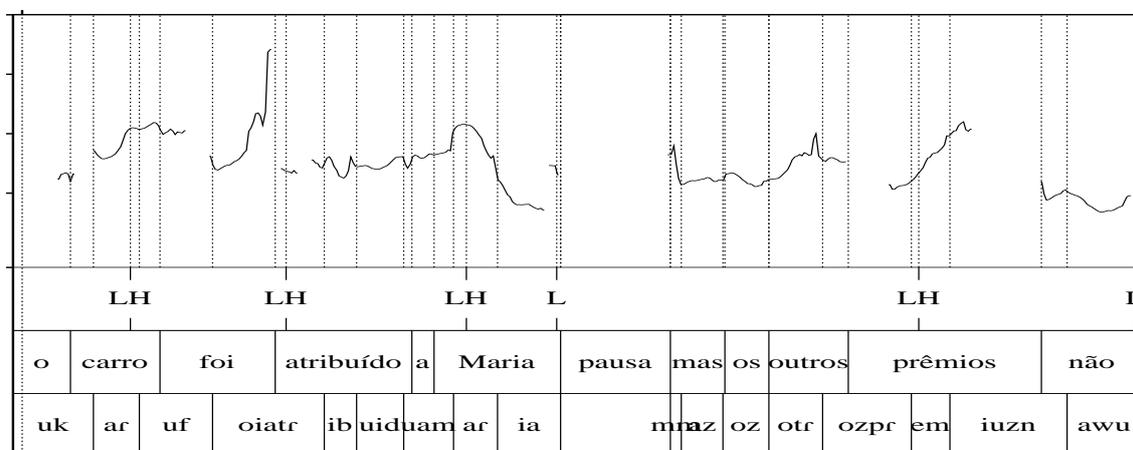


Fig. 2. Deaccentuation of propositional content and HL focus contour on the items "table" and "chair".

deaccentuation of propositional content. This same pattern was observed in other cases of stripping, as illustrated in Figure 5 below.



Fi. 5. Intonation pattern of neutral sentences with LH contour on all phonological phrases.

As we can see, the default intonation in cases of stripping is what is expected in neutral sentences, where there is tonal marking of all phonological phrases and low edge tones at the end of each utterance. This fact is unexpected, since both stripping and gapping would be the result of undergoing the same derivational cycle, according to the hybrid hypothesis. One possible way to explain this distinction is to suppose that in the case of stripping, the element that remains in the elided structure functions as a pro-form, an item that takes up the semantic content from the first sentence. This assumption takes into account the fact that this type of structure almost always includes the presence of items such as "also" and "no", which are not present in the "matrix" sentence and which seem to have the function of taking up the polarity of the previous sentence. Specifically on the "no", as mentioned, because it bears a focus function and it is not the result of movement, but of an external merge in the CP area. If this is correct, then, still taking the hybrid approach, since there is no movement, there is also no need to propose the existence of a syntactic structure. One possible problem for this analysis is the absence, in some cases, of the focalizing items "also" and "no."

Another example of the occurrence of stripping is shown below. The only difference is in the presence of high contour (H) over the "no." As there is no tone over the contrasting subject, it can be said that the subject and the focus marker are in the same phonological phrase.

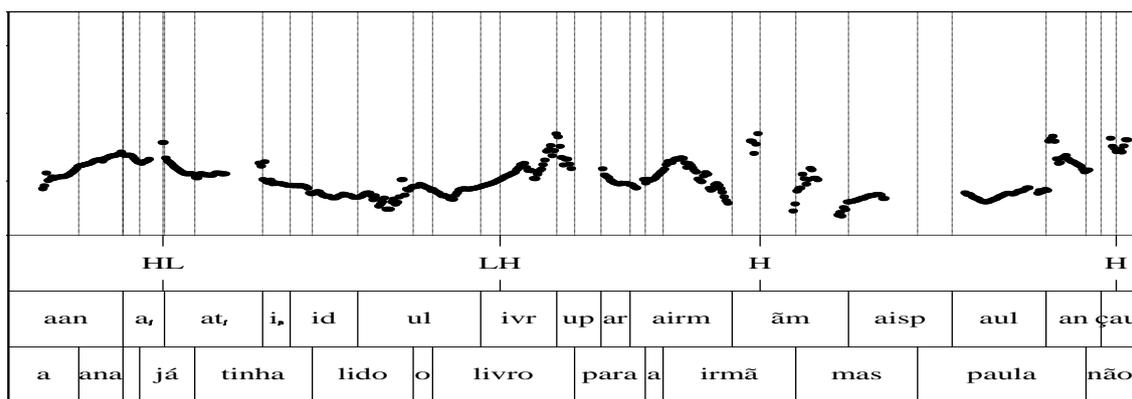


Fig. 6. Presence of tones over the phonological phrases of the first sentence and the "no" in the second.

The most interesting question regarding the data analyzed so far is that the focus contour appears over the contrasted item, which invariably occurs in the final position of the sentence. The main distinction between those pairs of complete sentences and gapping, on the one hand, and stripping, on the other, is the occurrence of deaccentuation in the former case but not the latter. Since deaccentuation happens both in complete sentences and in cases of gapping, it seemed consistent to interpret that, for this latter type of ellipsis, there is indeed a non-pronounced syntactic structure. Another argument in favor of this hypothesis is the relationship between interpretation and linear ordering; if the object is always interpreted contrastively and appears soon after the contrastive topic without the presence of a verbal structure, it is possible to conclude that this ordering is the result of the subject's movement to the topic and the object into the focus. If there is movement, there are necessarily traces occupying intermediate positions.

Also, the data on complete sentences seem to point in that direction. As described in the methodology section, the images used in the experiment induced the interpretation of contrast between one element in two scenes. So when I asked the informants to describe the images, many produced complete sentences. These statements, however, proved to be highly relevant to our analysis. Almost every time that informants produced complete sentences, they produced pauses between the item in focus and the rest of the sentence. Note the figure below:

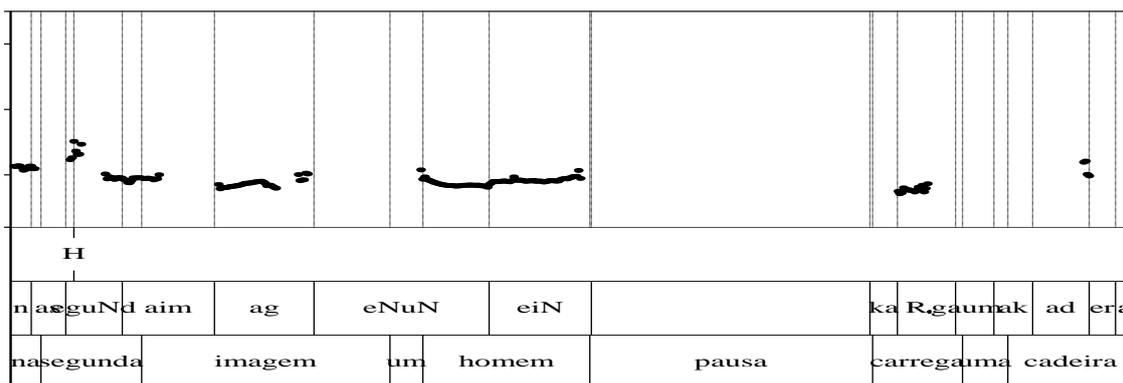


Fig. 7. Presence of a pause immediately after the contrasting item.

In the image that induced the production of the above sentence, there was a first scene in which a woman was carrying a chair and another in which a man was carrying the same chair. The informant, in describing the image, said: “na primeira imagem, uma mulher carrega uma cadeira e, na segunda imagem, um homem carrega uma cadeira”. (“In the first image, a woman carries a chair and, in the second image, a man carries a chair.”) Interestingly, when producing the second sentence, the speaker produced a pause between the subject that comprised the new information through the contrast between the scenes, and the rest of the sentence. The pause, which as I have stated was fairly frequent in the data, establishes the edge of a phrase; thus, it could be said that the subject of this sentence, “man” is not in the IP specifier position, but rather in the focus position. If this interpretation is correct, again, we have an item that needed to be moved, having to leave a trace in the initial position (i.e, Spec-IP) so that it can be interpreted also as the subject of the sentence. As seen in the above data, in all cases of ellipsis produced in those contexts, there was focus contour over the new item with respect to the previous sentence. If there is contour indicative of focus on the item not elided in the sentences with ellipsis produced by informants, then it is possible to apply the same analysis on the complete sentences and say that this item is in a focal position, and, in this case, I have necessarily to say that there is an unaccentuated syntactic structure in these constructions. If this is correct, I discard, at least in part, the semantic hypothesis.

As we have seen, the syntactic hypothesis cannot cover all of the possibilities of occurrence of ellipsis. For this reason, some authors defend the hybrid hypothesis. According to the latter, syntactic structure is present in some

cases and not in others. The idea is that a syntactic structure is only present in cases where it is necessary, i.e., those cases where there is movement of a constituent. To Winkler (2005), this difference is related to the different types of ellipsis: IP ellipsis would present a syntactic structure, since it involves movement of the constituent into focus; meanwhile, VP ellipsis would be better explained by the pro-form, as it does not obviate movement.

In the corpus collected by me, there were few clear data on VP focus; most occurred in data from read sentences, so I do not have sufficient evidence to make assertions in this regard. However, in most data, I observed the following prosodic structure: on the one hand, focus contour occurs over the new element, which does not necessarily occur in the sentence final position, as in cases of IP ellipsis; on the other hand, it presents edge pitches, as illustrated in Figure 8 below.

Thus, the VP ellipsis data, unlike those with IP ellipsis, do not suggest the interpretation that there was movement of a constituent, and, as predicted by the hybrid hypothesis, they can be analyzed as instances of pro-forms. The verification of this analysis, however, would require the collection of more VP ellipsis data.

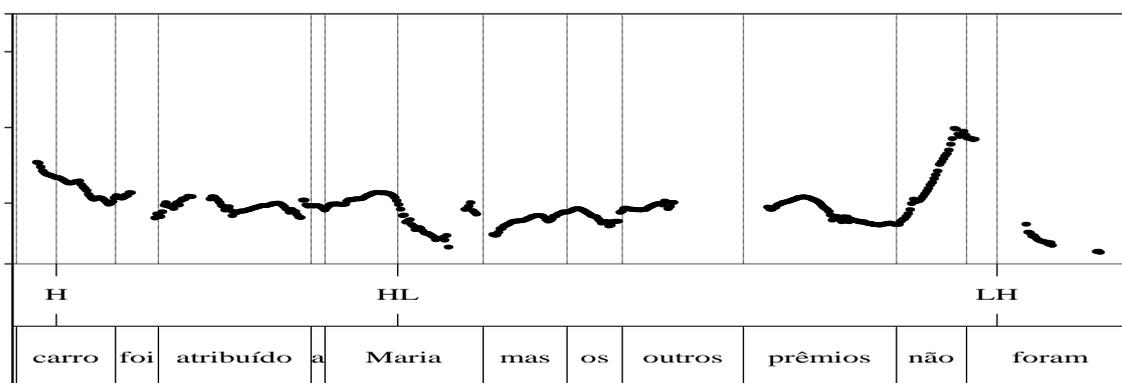


Fig. 8. Presence of pitch accent in the phonological phrases and LH contour in new information.

Although there are few VP ellipsis data, it is still interesting to note that, as in the cases of stripping, there was no deaccentuation of propositional content. If we can really consider deaccentuation a strong indication of the presence of unpronounced

syntactic structure, then these data are a further indication that, in the case of VP ellipsis, there really is no syntactic structure, but rather pro-forms.

In the analysis developed in this study, we saw that there are prosodic differences between the different types of ellipsis. On the one hand, we have the cases of gapping, which behave similarly to the pairs of complete sentences analyzed, i.e., in both cases, there was deaccentuation of propositional content and the presence of focus contour on the contrasting item. On the other hand, I observed that, in the cases of stripping and VP ellipsis, deaccentuation does not occur, the tones remaining on the phonological phrases. Thus, I propose that there is movement of phrases, with a consequent presence of syntactic structure in the gapping data and a presence of pro-forms in cases of stripping and VP ellipsis.

As stated, the VP ellipsis data in the corpus were few, being insufficient to support statements about this kind of ellipsis. I also emphasize that there are no studies that specifically deal with the prosodic behavior of pro-forms in linguistic sentences. I may predict, however, that these items do not cause syntactic restructuring, since they are typically concatenated in the position in which they are uttered. In this case, I reaffirm the proposition that there are no syntactic structures in stripping and VP ellipsis data.

The hypothesis that I reach at the end of our study is very close to what is proposed by the hybrid hypothesis, with the presence of syntactic structure in some types of ellipsis and pro-forms in others. Unlike the hybrid case, however, I suggest that there is a structure in cases of gapping and proforms in cases of stripping and VP ellipsis, which raises a problem in the formal treatment of such structures. In the hybrid hypothesis, the difference between IP ellipsis, referred to here as *stripping* and *gapping*, and VP ellipsis is in the undergoing of two or one derivational cycles, respectively. This explanation, however, does not account for the cases discussed here, since I distinguished the types of IP ellipsis, suggesting that in one case there is structure, and in the other there isn't. If my analysis is correct, it is necessary, then, to propose another way to derive these different types of ellipsis. Such an analysis would be extremely relevant, but beyond the scope of this study. Thus, I conclude with the belief that I have brought forth important issues for studies involving ellipsis in BP, emphasizing that there is still much to be studied on this topic.

4. Final remarks

My purpose in this study was to perform an acoustic analysis of sentences with ellipsis in BP as a way to verify the relationship between informational structure and licensing. Another objective was to determine which hypothesis, semantic or syntactic, better predicted BP data. As I have seen, there are differences in the prosodic structure depending on the type of ellipsis; namely, there is deaccentuation and the presence of focal contour of the contrasting item in cases of gapping and neutral intonation marking with focusing contour of the new element in cases of stripping and VP ellipsis.

From the results of the analysis, I propose that there is enough syntactic structure to cover the traces left by the items moved in gapping data and the presence of pro-forms in cases of stripping and VP ellipsis. Although this proposal resembles the hybrid hypothesis, it differs from the latter by distinguishing between stripping and gapping, which raises problems for the derivation of the difference between the different types of ellipsis. If for the hybrid hypothesis the difference is due to the undergoing of one or two derivational cycles, in our case, the distinction should be otherwise, given that stripping and gapping both undergo two derivational cycles.

Finally, I conclude that this study raised interesting questions that should be observed in future studies on ellipsis in BP. An important development was the establishment of an experiment that also considered the occurrence of pro-forms.

References

- BARBOSA, Plínio A. *Incurções em torno do ritmo da fala*. Campinas: Pontes Editora, São Paulo: Fapesp, 2006.
- CHOMSKY, N. *On phases*. Ms., MIT, 2005.
- CHOMSKY, N. “On Phases”. In. FREIDIN, R.; OTERO, C.; ZUBIZARRETA, M. L. (orgs). *Foundational issues in linguistic theory: essays in honor of Jean-Roger Vergnaud*. Cambridge, Mass: The MIT Press, 2008, p.133-166.
- CHOMSKY, N. ; LASNIK, H. “The theory of principles and parameters”. In.: JACOBS, J.; VON STECHOV, A.; STERNEFELD, W.; VENNEMAN, T. (eds.). *Syntax: An International Handbook of Contemporary Research*. Walter de Gruyter, 1993. P. 506-569.
- CYRINO, S.M. L. “A acentuação gráfica em João de Barros e elementos nulos no SN”. *Estudos Linguísticos e Literários*. N.19, 1997. P.189-196.
- DOGIL, G. & BRAUN, G. *The PIVOT model of speech parsing*. Viena, 1988.
- Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017.*

- FERNANDES, Flaviane. *Ordem, focalização e preenchimento em português: sintaxe e prosódia*. Campinas. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem/UNICAMP, Campinas, 2007.
- FÉRY, C.; ISHIHARA, S. “How Focus and Givenness shape prosody”. In: ZIMMERMANN, M.; FÉRY, C. (Eds.) *Information structure*. New York: Oxford University Press, 2010, p. 36-63.
- FÉRY, C.; SAMEK-LODOVICI, V. “Focus projection and prosody prominence in nested foci”. *Language*, v. 82, n. 1, 2006, p. 131-150.
- KEHLER, A. “Coherence and the resolution of ellipsis”. *Linguistic and Philosophy*, v. 25, 2000, p.533-575.
- KERTZ, Laura. *Ellipsis reconsidered*. Tese (Doutorado em Linguística). Universidade da Califórnia, San Diego, 2010.
- HARDT, D. *Verb Phrase Ellipsis: Form, meaning, and Processing*. Tese (Doutorado em Linguística). Universidade da Pensilvânia, Pensilvânia, 1993.
- HARDT, D.; ROMERO, M. “Ellipsis and structure of discourse”. *Journal of semantics*. v.21, 2004, p. 375-414.
- LADD, D. Robert . “Intonational phonology”. *Cambridge Studies in Linguistics* Cambridge: Cambridge University Press, v.79, 1996.
- LASNIK, H. “Pseudogappings puzzles”. In: LAPIN, S.; BENMAMOUN, E. *Fragments: studies in ellipsis and gapping*. Oxford/New York: OUP, 1999, p. 141-174.
- LOBECK, A. *Ellipsis: functional heads, licensing and identification*. OUP, 1995.
- LUCENTE, Luciana. Aspectos dinâmicos da fala e da entoação no português brasileiro. Campinas. Tese (Doutorado em Linguística) – IEL/UNICAMP, 2012.
- MERCHANT, J. *The syntax of silence. Sluicing, islands and the theory of ellipsis*. OUP, 2001.
- MERCHANT, J. “Fragment and ellipsis”. In: *Linguistics and Philosophy* v.27, 2004. p.661-738.
- MERCHANT, J. *Voice and ellipsis*. Ms. University of Chicago, 2007.
- MERCHANT, J. “An asymmetry in voice mismatches in VP-ellipsis and pseudogapping”. *Linguist Inquiry*, v. 39, 2008, p. 169-179.
- NAMIUTI, C. *Aspectos da história gramatical do Português: Interpolação, Negação e Mudança*. Tese (Doutorado em Linguística). Instituto de Estudos da Linguagem, Unicamp, Campinas, 2008.
- NESPOR, M.; VOGEL, I. *Prosodic phonology*. Dordrecht: Foris Publications, 1986.

- ROOTH, M. “Ellipses redundancy and reduction redundancy”. In.: BERMAN, S. & HESTVIK, A. (eds.). *The Stuttgart Workshop on Ellipsis*, 1993.
- SAG, I. “A note on verb phrase deletion”. *Linguistic Inquiry*, v. 7, 1976. P. 664-671.
- SCHWARZSCHILD, R. “GIVENness, Avoid F and other constraints on the Placement of the Focus”. *Natural Language Semantics*, v. 7, n. 2, 1999, p. 141-177.
- SELKIRK, E. O. “Sentence prosody: intonation, stress and phrasing”. In.: GOLDSMITH, J. (ed.). *The handbook of phonology theory*. Cambridge, Mass.: Blackwell, 1995. P. 550-569.
- SELKIRK, E. O. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press, 1984.
- TANCREDI, C. *Deletion, deaccenting and presupposition*. Tese (Doutorado em Linguística). MIT, Mass, 1992.
- TEIXEIRA DE SOUSA, L. *Sintaxe e interpretação de negativas sentenciais no Português Brasileiro*. Tese (Doutorado em Linguística). Instituto de Letras, Unicamp, Campinas, 2012.
- TENANI, L. *Domínio Prosódicos no Português*. Campinas. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Unicamp, Campinas, 2002.
- TRUCKENBRODT, H. *Phonological Phrases: Their relation to syntax, focus and prominence*. Tese (Doutorado em Linguística) MIT, Cambridge, Mass, 1995.
- TRUCKENBRODT, H. “On the relation between Syntactic Phrases and Phonological Phrases”. *Linguistic Inquiry*, n.30, 1999, p. 219-255.
- WINKLER, S. *Ellipsis and Focus in generative grammar*. Mouton de Gruyter: Berlin-Nova York, 2005.
- WONG, S.W. & SCHREINER, Christoph E. “Representation of CV-sounds in cat primary auditory cortex: intensity dependence”. *Speech comm*, n. 41, 2003, p. 93-106.

The Morphosyntax of Nominalization in Wayoro (Tupí): a preliminary approach

Antônia Fernanda de Souza Nogueira^a

ABSTRACT

This study aims to provide a preliminary account of the morphosyntax of nominalization in Wayoro (Tupian family), an endangered language which is situated in the state of Rondônia (Brazil). In this paper I describe some of the morphosyntactic and distributional properties of nouns and verbs. An additional objective of this paper is to examine the nominal and/or verbal properties of the constructions involving the morpheme {-p} ‘nominalizer’. The data suggest that there are two kinds of constructions involving the morpheme {-p}: a lexical, or VP nominalization, and a clause nominalization.

KEYWORDS: Wayoro language; morphosyntactic and distributional properties of nouns and verbs; nominalization

RESUMO

Esta pesquisa tem como objetivo fornecer uma abordagem preliminar sobre a morfossintaxe da nominalização em Wayoro (família Tupi), uma língua ameaçada localizada no estado de Rondônia (Brasil). Neste artigo, descrevo algumas propriedades sintáticas e distribucionais de nomes e de verbos. Outro objetivo deste trabalho é examinar propriedades verbais e/ou nominais de construções que envolvem o morfema {-p} ‘nominalizador’. Os dados sugerem que há dois tipos de construções relacionadas ao morfema {-p}: uma nominalização lexical ou de VP e uma nominalização sentencial.

PALAVRAS-CHAVE: língua Wayoro; propriedades morfossintáticas e distribucionais de nomes e verbos; nominalização

^a Universidade de São Paulo & Universidade Federal do Pará – Breves Campus, afernanda@ufpa.br

I would like to thank the Wajuru people and the community of the Ricardo Franco village (TI Rio Guaporé) for their hospitality, as well as the consultants for their patience in explaining the Wayoro language, and the Fundação Nacional do Índio – FUNAI (authorization 56/CGEP/08 and 119/AAEP/10). I would also like to thank Luciana R. Storto, Ana Vilacy Galucio, Ivan Rocha, the editors and the anonymous reviewers for their useful comments on earlier drafts of this paper. Any remaining errors are mine.

Introduction

Wayoro is a Tupian language of the Tuparian subfamily (MOORE; GALUCIO; GABAS JR, 2008), which is situated in the Brazilian state of Rondônia (TI Rio Guaporé, Guajará-Mirim, and Alta Floresta d'Oeste). There is a variety of indigenous names to denominate the language, such as Wajuru, Ayuru, Ajuru, Wayoró. However, speakers use Wayoro ([waj.o.'ro] or [ŋg^waj.o.'ro]) to refer to their language, which is one of the most endangered native languages in Brazil (MOSELEY, 2011), since it is only spoken by 3 (three) people. Furthermore, the new generations are not acquiring Wayoro as their first language. The data for this study were collected in fieldwork on the indigenous land Rio Guaporé (Terra Indígena Rio Guaporé).



Fig. 1. Map of location of TI Rio Guaporé, in Rondônia (Brazil). Adapted from FUNAI, 2017.

The main objective of this study is to describe some of the morphosyntactic and distributional properties of nouns phrases (NPs) and verbs phrases (VPs) in Wayoro and to investigate their behavior in constructions denominated nominalizations in this language.

The sentences that are used as examples in this paper represent data from the following categories: elicitation sessions, individual narratives, traditional stories, accounts of procedural tasks, personal recounts, and published material. The data were

mainly collected during various visits to the Ricardo Franco village (Guajará-Mirim, Rondônia) between 2008 and 2015.

This paper begins with a general characterization of noun phrases and verb phrases in Wayoro (sections 1 and 2). The subsequent section describes two different types of nominalizations in Mekens (another Tuparian language). Finally, in section 4, I describe and analyze constructions involving the Wayoro ‘nominalizer’ suffix {-p}, comparing it with its cognate in Mekens.

1. Noun phrases

Noun phrases (NPs) have the following distributional properties: NPs can be used as complements of verbs (1a), subjects of verbs (1a-b), complements of postpositions (1c). NPs can also be modified by adjectives (1c), by another NP (b), or by a numeral (1d).

(1) Distributional properties of NPs

a. NP used as a complement

[awi]	ka-t	yã ¹
father	eat-NFUT	mother

‘Mommy has eaten daddy.’ (*Text*)

b. NP used as a subject and NP modified by another NP

te-tera-t	[mberu	ngõ]
3COR-go-NFUT	capuchin.monkey	pet

‘The capuchin went.’ (*Text*)

c. NP modified by an adjective and NP used as the complement of a postposition

[Ugu akara]	mbiro	[txi-ndek	ere]
water big	have	1PL.INCL-house	LOC

‘There is a river in our village/house.’

¹ In this paper, I followed the orthographic conventions used to write Wayoro (NOGUEIRA, 2012). Examples originating from natural texts or published materials are followed by the citation of the source.

d. NP modified by a numeral

aramĩra	[mbogop	ndurut]	tuuwa-t
woman	child	two	give.birth-NFUT

‘The woman gave birth to twins.’

Concerning nouns in Wayoro, the bound pronominal markers refer to the possessor of possessive NPs, as in (2). The morpheme {te-} is the cognate of {se-} and {te-}, coreferential in Mekens and Akuntsu (GALUCIO, 2011; ARAGON, 2014), and {te-} reflexive in Tupari (ALVES, 2004).

(2) Bound pronominal markers as the possessor of possessive NPs

o-tak	‘my daughter (man-speaking-term) ’
e-tak	‘your daughter’
ndeke-tak	‘his daughter’
txi-tak	‘our daughter (inclusive)’
ote-tak	‘our daughter (exclusive)’
djat-tak	‘your (plural) daughter’
ndeat-tak	‘their daughter’

(NOGUEIRA, 2013, p.328)

ndop_i te_i-tak atinãn

‘the father_i is.stingy.about his_i daughter’

In Wayoro, OV is an inseparable unit. This is shown in the examples below (NOGUEIRA, 2011, p. 210-211). Examples (4a, 4b) show that the insertion of an intervening PP within the OV unit is ungrammatical, such as the PP *te-ndaup mē* ‘to his son’, headed by the postposition *mē* ‘to’, as in (4b).

(4) OV unit in Wayoro

- a. *te-ndaup mē* *ngwaykup* [*uwoy.tūkwa-p* *yõã-n*]_{ov}
 3COR-son to man fish.with.fishing.net-p give-NFUT

‘The man gave a fishing net to his son.’

- b. **ngwaykup* *uwoy.tūkwa-p* *te-ndaup* *mē* *yõã-n*
 man fish.with.fishing.net-p 3COR-son to give-NFUT

‘The man gave a fishing net to his son.’

2.2. Personal morphemes

An absolutive morphosyntactic alignment pattern of person morphology is used on verbs. The personal (free) pronouns are used for ergative arguments (A) and are also used optionally for S arguments with intransitive verbs. Prefixes are used for absolutive arguments (S/O) (see below 5-6)³.

In Tables 1 and 2, I list the inventory of the personal prefixes and personal pronouns:

³ Other authors have suggested that the language can also be analyzed as Nominative-Absolutive: “considering the distribution of the pronouns as a whole, this hybrid alignment found in the Tuparian languages is a clear instantiation of the Nominative-Absolutive alignment pattern proposed by Gildea and Castro Alves (2010, in press) for Cariban and Northern Jê Languages. The nominative pattern (free pronouns) co-occurs with the absolutive pattern (argument marking on the verb)” (GALUCIO; NOGUEIRA, in press).

Bound pronouns (prefixes/S and clitics/O arguments)		
	Singular	Plural
1 st person	<i>m- ~ o-</i>	<i>txi-</i> (inclusive)
		<i>ote-</i> (exclusive)
2 nd person	<i>e-</i>	<i>djat-</i>
3 rd person	<i>te-</i>	<i>te-</i> (Subject)
	<i>y- ~ ndeke-</i>	<i>y- ~ ndeat-</i> (Object)

Table 1. Bound pronouns

Free pronouns		
	Singular	Plural
1 st person	<i>On</i>	<i>txire</i> (inclusive)
		<i>ote</i> (exclusive)
2 nd person	<i>Em</i>	<i>djat</i>
3 rd person	<i>ndeke</i>	<i>ndeat</i>

Table 2. Free pronouns

The first person morpheme has three phonologically conditioned allomorphs in Wayoro: *o-* before consonant initial stem, *m-* before nasal(ized) vowel initial stem, and *mb-* before oral vowel initial stem. Also, the *o-* allomorph can be nasalized before a nasal consonant. There are some homophonous pronouns⁴, but they belong to a different paradigm, as we can see in Table 3.

⁴ These examples may be analyzed as a case of syncretism in pronominal forms. “A syncretism is found when morphemes composed of distinct sets of syn[tactic]sem[antic] features are realized with the same phonological exponent” (EMBICK, 2015, p. 25)

	Free pronouns (S and A arguments)	Bound pronouns (prefixes/S and clitics/O arguments)	
1PL.EXCL	ote	ote-	
2PL	djat	djat-	
3SG	ndeke	ndeke-~dj-~y-	te- (only S argum.) ⁵
3PL	ndeat	ndeat-~dj- ~y-	te- (only S argum.)

Table 3. Homophonous pronouns

The third person morpheme has four phonologically conditioned allomorphs: *ndeke-* and *ndeat-* before vowel initial stem, *dj-* before vowel initial stem, and *y-* before nasal(ized) vowel initial stem (see NOGUEIRA, 2011, p. 80-81). Also, the third person morpheme does seem to have an indefinite use.

(5) Paradigm of personal pronouns used in subject of transitive (A) function

on	dj=ipitkwa	‘I’ll throw something.’
en	dj=ipitkwa	‘You threw something.’
ndeke	dj=ipitkwa-t	‘He threw something.’
txire	dj=ipitkwa-t	‘We (incl.) threw something.’
ote	dj=ipitkwa	‘We (excl.) threw something.’
djat	dj=ipitkwa-t	‘You threw something.’
ndeat	dj=ipitkwa-t	‘They threw something.’

(6) Paradigm of bounds (clitics) pronouns used in object (O) function

aramĩra	o =tigaa-t	‘The woman painted me.’
aramĩra	e =tigaa-t	‘The woman painted you.’
aramĩra	ndeke =tigaa-t	‘The woman painted him.’
aramĩra	txi =tigaa-t	‘The woman painted us (incl.).’

⁵ The third person distribution suggests that the language also seems to have a tripartite alignment in which there is A≠S≠P. This suggestion is under examination.

aramĩra **ote**=tigaa-t ‘The woman painted us (excl).’

ote **djat**=tigaa-t ‘The woman painted you.’

aramĩra **ndeat**=tigaa-t ‘The woman painted them.’

- (7) Paradigm of bounds (prefixes) pronouns used in subject of intransitive verb (S) function

	m -engukwa-t	(on)	‘I’m sweating.’
	e -engukwa-t	(en)	‘You are sweating.’
(ndeke)	te -ngõyã-n		‘He is sitting.’
(txire)	txi -engukwa-t		‘We (incl.) are sweating.’
	ote -engukwa-t	(ote)	‘We (excl.) are sweating.’
(djat)	djar -engukwa-t		‘You are sweating.’
(ndeat)	te -ngõyã-n		‘They are sitting.’

S or O is always marked regardless of other features (semantic class of the intransitive verb, person hierarchy, etc.), as shown in examples (7) and (8).

- (8) Wayoro (NOGUEIRA, 2011, p. 159; SANTOS, 2010, p. 132)

a.	ngwaykup	o=toa-t	
	man	1SG=see-NFUT	
	‘The man saw me.’		
b.	ndeke=toa-p	nã	on
	3SG=see-p	FUT	1SG
	‘I’ll see her.’		

In addition, as shown in (9), the personal prefixes function as the object, while the free pronouns function as subject, relating to the 1st person singular and the 2nd person singular.

(9) Wayoro transitive verbs

- a. en o=tigaa-t
 2SG 1SG=body.paint-NFUT
 ‘You painted me.’
- b. on e=tigaa-t on⁶
 1SG 2SG=body.paint-NFUT 1SG
 ‘I painted you.’
- c. o=puruga-t en
 1SG=pierce-NFUT 2SG
 ‘You pierced me.’ (NOGUEIRA, 2011, p. 155)
- d. e=puruga
 2SG=pierce/take.virginity
 ‘Take your virginity.’ (Lit: ‘Pierce you.’)

Object (O) arguments encoded by personal prefixes do not co-occur with a free pronoun or a NP. As in (10a) and (10b), the object of the transitive verb *ombaa* ‘hit’ can be a third personal prefix (10a) or a NP (10b), but not both simultaneously.

(10) Personal prefix or NP as third person object

- a. y=ombaa-t on
 3=hit-NFUT 1SG
 ‘I hit them.’
 (NOGUEIRA, 2011, p. 68)

⁶ The repetition of the free pronoun is attested only in the 1st person singular as the subject of the transitive verb, but more work is needed to better comprehend this kind of data.

- b. ndaikut en [e-men ombaa]
 tomorrow 2SG 2SG-husband kill
 ‘Tomorrow, you will kill your husband.’ (*Texto*)

The ungrammatical examples below attest that personal prefixes cannot be preceded by co-referential free pronouns or NP in the same function. See examples (11a-b) and (11a’-b’).

(11) Ungrammatical data

- a. Irene o=tigaa-t
 Irene 1SG=body.paint-NFUT
 ‘Irene painted me.’
- a’. *Irene on o=tigaa-t
 Irene 1SG 1SG-body.paint-NFUT (Intended reading: ‘Irene painted me.’)
- b. mbogop amêko ombaa-t
 child jaguar/dog hit-NFUT
 ‘The child hit the dog.’
 (NOGUEIRA, 2011, p. 111)
- b’. *mbogop amêko y-ombaa-t
 child jaguar/dog 3-hit-NFUT (Intended reading: ‘The child hit the dog.’)

Note that, in (12), free pronouns cannot function as the object of a transitive verb.

(12) Free pronouns cannot function as the object of a transitive verb

- a. Irene o=tigaa-t
 Irene 1SG=body.paint-NFUT
 ‘Irene painted me.’
- b. *Irene on=tigaa-t
 Irene 1SG=body.paint-NFUT (Intended reading: ‘Irene painted me.’)

In the second person, the personal prefix {e-} co-occurs with the co-referential free pronouns {en}, but it has a very different meaning (13). The NP *Irene* is a vocative

Diadorim, Rio de Janeiro, Revista 19 — Volume Especial 2017.

and the clause is reflexive, where the personal prefixes can co-occur with co-referential free pronouns (see also NOGUEIRA, 2011, p. 181-217, 2013 for a discussion of reflexive constructions in Wayoro).

(13) Reflexive construction

- a. Irene en e=tigaa
 Irene 2SG 2SG=body.paint
 ‘Irene, you will paint yourself?’ (Impossible interpretation: ‘Irene painted you.’)
- b. aramĩra e=tigaa-t
 woman 2SG=body.paint-NFUT
 ‘The woman painted you.’

Intransitive verbs always occur with personal prefixes, as in (14). The prefixes cross-reference a NP or a (free) pronoun, if present.

(14) Wayoro intransitive verbs (NOGUEIRA, 2011, p. 70)

- a. o-tera-t (on)
 1SG-gosing- NFUT 1SG
 ‘I went.’
- a’. *tera-t on
 1SG-gosing-NFUT 1SG (Intended reading: ‘I went.’)
- b. mbogop te-enunkara-t
 child 3COR-breath- NFUT
 ‘The child breathed.’
- b’. *mbogop enunkara-t
 child breath-NFUT (Intended reading: ‘The child breathed.’)

In addition, note that the prefix {te-} ‘coreferential’ never co-occurs with a NP object, as in (15).

(15) {te-} ‘coreferential’ never co-occurs with a NP object (adapted from NOGUEIRA, 2011, p. 78)

- a. *amēko te-punkwa-t⁷
 jaguar/dog 3COR-shoot-NFUT
 (Intended reading: ‘(He) shot the dog.’)
- b. ndeke amēko punkwa-t
 3SG jaguar/dog shoot-NFUT
 ‘He shot the dog.’

This distribution, shown in (14), suggests that S argument personal prefixes are agreement morphemes in intransitive clauses, whereas the O personal pronouns are not, since they cannot co-occur with a full DP in object position. This distribution suggests that bound pronouns are functioning as objects rather than agreement. These pronouns can be analyzed as cliticized pronouns.

The same type of analysis was proposed for the Mekens language (also from the Tuparian branch) (GALUCIO, 2001, 2011), as in (16). In Mekens, in the intransitive VPs,

the person prefix shows person and number agreement with an overt NP subject, if present. [...] Transitive VPs are composed of a transitive verb and a direct object, which is realized either by an NP [...] or by a personal prefix [...], but not both at the same time [...]. The subject of transitive clauses is expressed by an NP (either nominal or pronominal) external to the VP. (GALUCIO, 2011, p. 28)

(16) Mekens: intransitive and transitive VPs

- a. roque se-er-a-t
 roque 3COR-sleep-TH.V-PST
 ‘Roque slept.’

⁷ The omission of third person, as in (i), is common in subject and object function.

(i) e=toa-t
 2SG-see-NFUT
 ‘(He) saw you.’

- b. ameko aose so-a-t
jaguar/dog man/person see-TH.V-PST
'The jaguar saw the man.'
- c. ameko i-so-a-t
jaguar/dog 3SG-see-TH.V-PST
'The jaguar saw him/her/it.'
- d. *ameko aose i-so-a-t
jaguar/dog man/person 3SG-see-TH.V-PSTCC(Intended reading:
'The jaguar saw the man.')

2.3. Tense and aspect

After the verb stem, the morpheme {-t ~ -n} is used to indicate non-future, as in (17).

(17) Non-future semantic of {-t} suffix

- a. e-ngora mǎyã-n on ega
2SG-music sing-NFUT 1SG yesterday
'Yesterday, I sang (a song) for you.'
- b. m-emburaa-t
1SG-fly-NFUT
'I'm flying.'
(NOGUEIRA, 2013, p. 330)
- c. djat-paga-t nen djat
2PL-get.drunk-NFUT INTERR 2PL
'Did you (plural) get drunk?' or 'Are you (plural) drunk?'
(NOGUEIRA, 2013, p. 330)

The use of the morpheme {-p}, followed by the morpheme {nã}, is used to refer to the future, as in (18a). In (18b), the non-future suffix {-t} is not permitted since there is a temporal adverb in the sentence: *ndaikut* 'tomorrow'.

(18) Future semantic of the use of {-p} followed by {nã}

- a. ndaikut ngwaikup te-tera-p nã
 tomorrow man 3COR-go-p FUT

‘Tomorrow, the man will go.’

- b. *ndaikut ngwaikup te-tera-t
 tomorrow man 3COR-go-NFU

(Intended reading: ‘Tomorrow, the man will go.’)

Following are the aspectual morphemes identified to date: {-rara} to indicate repetition and {-kw} to indicate repetition and intensification (plurality of event). Both morphemes can co-occur with morphemes that indicate tense, as we can see in (19).

(19) Examples of {-kw} and {-rara} aspectual suffixes

- a. m-amõy-kw-a-t on
 1SG-dance-PL-TH.V-NFUT 1SG

‘I’m dancing fast.’

(NOGUEIRA, 2011, p. 126)

- b. aramĩra-ian atiti ndet-kw-a-t
 woman-PL corn grind-PL-TH.V-NFUT

‘The women are grinding the corn very much.’

(NOGUEIRA, 2011, p. 137)

- c. ndat to-rara-t on
 snake see-REP-NFUT 1SG

‘I saw a snake again.’

- d. Tuero nder-ara-p nã on
 chicha grind-REP-p FUT 1SG

‘I’ll grind *chicha* (a fermented drink) again.’

According to Galucio (2001, 2014), /-ra/ is a repetitive suffix in Mekens, as in (20).

(20) Example of {-ra} repetitive suffix in Mekens

o-so-a-ra kot
 1SG-bath-TH.V-REP IM.FUT

‘I will bath (again).’

(GALUCIO, 2001, p. 92)

2.4. Transitivity

The causative morpheme {mõ- ~ õ-} adds an external argument to an intransitive verb, turning it into a transitive verb, as shown in (21) (see NOGUEIRA, 2011). The intransitive verb *era* ‘to sleep’ (21a) receives the ‘causative/transitivizer’ prefix {mõ-}, and the subject *o-mẽpit* ‘my child’ functions as the object in the derived verbal construction, as in (21b).

(21) Example of transitivity in Wayoro (NOGUEIRA, 2011)

- a. o-mẽpit te-era-t
 1SG-child 3COR-sleep-NFUT
 ‘My child slept.’
- b. aramĩra o-mẽpit mõ-era-n
 woman 1SG-child CAUS-sleep-NFUT
 ‘The woman made my child sleep.’

We have seen that clitic pronouns may encode objects in transitive verbs. The following examples (22a-b) illustrate that the derived transitive verb with {mõ- ~ õ-} can also have a clitic pronoun functioning as its object.

- (22) Personal prefixes as object of the derived transitive verb
- a. en o=mõ-era-n
2SG 1SG=CAUS-sleep-NFUT
'You made me sleep.'
- b. on e=mõ-era-n
1SG 2SG=CAUS-sleep-NFUT
'I made you sleep.'

3. Nominalization(s) in Tuparian family

In this section, I will briefly present Galucio's analysis of nominalization in Mekens, which will be relevant for the preliminary analyses of Wayoro nominalization, since they are sister languages. Galucio (2011) considers that Mekens has two kinds of nominalization: grammatical and lexical nominalization (SHIBATANI; MARKHASHEN, 2009 apud GALUCIO, 2011). The circumstantial nominalizer {-ap} refers to instrumental and locative nouns (lexical nominalization), as in (23):

- (23) Mekens
- a. otat poka-ap
fire burn/light-NMLZ
'match or lighter' (lit. tool (for) lighting the fire)
- b. iki ekwe-ap
water run-NMLZ
'rapids' (lit. part of a river where the current is swift)
- c. [o-to-ap] i-pagop
1SG-lie-NMLZ 3SG-new
'My hammock is new.'

In this kind of structure, “[...] person prefix [...] functions in the derived noun (phrase) as the possessor of the nominalization” (GALUCIO, 2011, p. 243), such as the prefix {o-} in (23c).

In event nominalization, the circumstantial nominalization combines with postpositions, thereby resulting in postpositional phrases that encode the adverbial modification (GALUCIO, 2011, p. 239-245). In (24), the nominalization functions as the object of the postpositional phrase, headed by the locative postposition *ese*.

(24) Mekens

[o-ib-ra-ab=ese]	tabir=eri	ka
1SG-come-REP-NMLZ=LOC	field=ABL	go/come
ki-po-e-mot-kwa		
1PL.INCL-hand-INTR-make-TR		

‘When I come back from the field, we will work.’

(Lit. at my coming back from the field, we will work.)

(GALUCIO, 2011, p. 243)

According to Galucio (2011, p. 244)

In lexical nominalizations, the nominalizer morpheme *-ap* directly follows the verb root, while in event nominalizing [...], the nominalizer appears after the thematic vowel and other inflectional morphemes. Furthermore, the event nominalizing maintains the syntactic properties (specifically the argument structure) of the verb, while the lexical nominalization holds only a morphological and semantic association to the verb root. (GALUCIO, 2011, p. 244).

Therefore, in Mekens, there are two different kinds of nominalization: grammatical and lexical nominalization, which are expressed through the same morpheme, {-ap}. As we will see below, in Wayoro, constructions with {-p} morpheme can refer to a nominalized VP or a clause nominalization.

4. Wayoro nominalizations with {-p}⁸

According to van Gijn, Haude and Muysken (2011, p. 10-11), nominalization is “probably the most common subordination strategy in South American languages”. There is *participant nominalization*, which functions as relative clauses; *event nominalization*, used for complementation, but also for relative and adverbial relations; and *participial nominalization*, in which the nominalizer also encodes a verbal property.

In this section, I will show that constructions with {-p} morpheme can refer to: (i) an instrumental or locative noun (as *ndoap* ‘hammock’ and *toap* ‘mirror’); (ii) an event nominalization. These constructions have different morphosyntactic properties and can also be classified as a nominalized VP or as an aspectual phrase (event or clause nominalization).

4.1. VP nominalization

Nominalized verbs, like NPs, can be modified by other nouns. For example, in the possessive construction in (25b), the intransitive verb *ndoa* ‘to lie down’ is nominalized by {-p}, generating the word for ‘hammock’, and it can subsequently be modified by the noun ‘foreigner’. As we saw above in section 1, Wayoro has possessor-possessed word order. In (25b), the nominalized verb is the possessed and in (25c) the possessor. Furthermore, we can also see that the nominalized verb can be modified by an adjective, as in (25d). In this example, *ndoap* ‘hammock’ is modified by the adjective *txuup* ‘wet’.

(25) Nominalized verbs modified by nouns and by adjectives

- a. E-ndoa-t nen
 2SG-lie-NFUT INTERR
 ‘Are you lying down?’
- b. [ngwerep ndoa-p] emõ ponã
 foreigner lie-NMLZ EMPH use
 ‘Just use the foreigner's hammock.’

⁸ The morpheme {-p} is realized as [-p] after an oral vowel, and as [-m] after a nasal vowel.

(Text)

c. [txi-ndoa-p] yam

1PL.INCL-lie-NMLZ rope

‘Our hammock cord.’

d. [o-ndoa-p] txuup

1SG-lie-NMLZ wet

‘My hammock is wet.’

In addition, like nouns, nominalized verbs can be the object of a postposition, as in (21). In this example, the intransitive verb *ngõyã* ‘to sit’ is nominalized. It has the meaning ‘mat’, and it is the complement of the postposition *tere*.

(26) Nominalized verb as complement of postposition

E-ngõyã [ngõyã-m tere]

2SG-sit sit-NMLZ on

‘Sit on the rush mat.’

As noted above, the intransitive verbs *ndoa* ‘to lie down’ and *ngõyã* ‘to sit’ can occur without the personal prefixes usually required by intransitive verbs. In the data, causative prefix, tense and aspect morphemes have not been found. A similar pattern is found with lexical nominalization of the transitive verbs *toa* ‘to see’ (27a) and *puruga* ‘to pierce’ (27b).

(27) Transitive nominalized verbs

a. o-[toa-p] mbiro

1SG-see-NMLZ have

‘There is my mirror.’

- b. [puruga-p] dj-akara pikam
 pierce-NMLZ 3-large deep
 ‘The hole is large and deep.’

Such lexical nominalizations cannot occur with sentential properties, as with the aspectual suffix *-rara* ‘repetition’, as shown in the ungrammatical examples (28b, d).

(28) Wayoro (Nogueira’s field notes)

- a. o-ndo-ap txuup
 1SG-lie-NMLZ wet
 ‘My hammock is wet.’
- b. *o-ndo-rara-p txuup
 1SG-lie-REP-NMLZ wet
- c. e-ngõy-rara-a ngõyã-m tere
 2SG-sit-REP-IMP sit-NMLZ on
 ‘Sit on the mat again.’
- d. *e-ngõy-a [ngõy-rara-m] tere
 2SG-sit-REP-IMP sit-REP-NMLZ on

We can, therefore, analyze such constructions as nominalized VPs, as in (29).

(29) Lexical nominalization: $[_{NP} [_{VP} ndoa] -p]$

4.2. Event or clause nominalization

Another kind of construction involving the $\{-p\}$ suffix was attested in the corpus. In Wayoro, these constructions can be complement and adverbial clauses that function as clausal arguments and adjuncts, respectively, as in (30) and (31). Despite the absence of the tense suffix $\{-t\}$ in the data, there are clausal properties, such as the causative prefix (which adds an external argument to an intransitive verb) and the aspectual suffix $\{-rara\}$ ‘repetition’. Also, there are always personal prefixes, if the verb is an intransitive verb. I will now call these constructions clause nominalizations, referring to the same type of phenomenon as event nominalization proposed for Mekens by Galucio (2011).

In (30a-b) below, we can see that the required personal prefix {m-} ‘1st person singular’ of the intransitive verb *mãyã* ‘to dance’ cannot be absent, such as in any intransitive clause. Personal pronouns, however, never appear. This suggests that there is no agreement among the data.

(30) Absence of personal prefix is not permitted in clause nominalization

- a. on [m-amõyã-m] nĩa-rom on
 1SG 1SG-dance-p want-NEG 1SG

‘I, I do not want to dance.’

- b. *on [amõyã-m] nĩa-rom on
 1SG 1SG-dance-p want-NEG 1SG

(Intended reading: ‘I, I do not want to dance.’)

Lexical nominalization, in contrast, permits the absence of a personal prefix, as we have seen in (21b) and (22).

In (31), the construction involving {-p} presents the aspectual suffix {-rara}, meaning ‘repetition’. In (31a), the complement clause functions as the object of the transitive verb. In (31b), the nominalization is used as the complement of the postposition *ere*, and the construction is interpreted as an adverbial clause.

(31) Occurrence of {-rara} aspectual suffix in clause nominalization

- a. [m-amõyã-rara-m] nĩa-rom on
 1SG-dance-REP-p want-NEG 1SG

‘I do not want to dance again.’ (*Text*)

- b. y-engu ate ka [nder-ara-p ere]
 3-*chicha* ? ingest grind-REP-p LOC

‘Nevertheless, (they) were drinking her *chicha* (a fermented drink), when (they) were making *chicha* again.’ (*Text*)

As illustrated by the examples above, these constructions involving {-p} allow the presence of an aspectual suffix.

Finally, causative/transitivizer morphology can also be found in these constructions. In the following examples, the verb *apokaya* ‘to arrive’ is intransitive, as we attest in (32a). This verb appears with the transitivizer prefix {mõ-} in the complement construction involving {-p} (32b).

(32) Causative/transitivizer morpheme in clause nominalization

- a. mbogop te-apokaya
 child 3COR-arrive
 ‘The children arrived.’ (*Text*)
- b. txi-mõ-apokaya-p toa-t on
 1PL.INCL-CAUS-arrive-p see-NFUT 1SG
 ‘I saw us arrive’. (*Text*)

In this way, clause nominalization in Wayoro is a construction that can have the functional heads AspP (aspectual morphemes) and vP (causative/transitivizer morpheme), which are sentential properties (see sections 2.3 and 2.4).

5. Final conclusions

My paper began with a presentation of some of the morphosyntactic properties of nouns and verbs in Wayoro. Then I described two kinds of constructions involving the morpheme {-p}, glossed as ‘nominalizer’ in both cases (NOGUEIRA, 2013, 2014). We have seen some differences between these two constructions. The first set of data has some properties of the nouns (lexical nominalization) and the second set of data has some sentential properties (aspectual nominalization), as in Wayoro’s sister language, Mekens.

It is interesting to note that the Karitiana language (also a Tupian language) morphologically differentiates a nominalizer {-pa} from an infinitive {-p} suffix (ROCHA, 2016). These morphemes suggest that the grammatical properties of nominalization and infinitive subordination are related to cognate morphemes in the Tupian family.

Van Gijn, Haude and Muysken (2011, p. 13) claim that “event nominalization in South America often allows for the retention of verbal morphology on the nominalized predicate”. One hypothesis to be tested is whether the event/clause nominalization sentences would be an infinitive phrase, since there are sentential properties, such as the causative prefix and aspectual suffix {-rara}. Infinitives are also considered nominal forms of the verb.

Abbreviations

1PL.INCL first person plural inclusive; 1SG first person singular; 2PL second person plural; 2SG second person; 3 third person; 3COR third person co-referential; 3SG third person singular; ABL ablative; CAUS causative; EMPH emphasis; FUT future; IM.FUT immediate future; IMP imperative; INTERR interrogative; INTR intransivizer; LOC locative; NEG negation; NFUT non-future; NMLZ nominalizer; PL plural; PST past; RED reduplication; REP repetition; TH.V theme vowel; TR transitivizer

References

- ALVES, P. M. *O léxico do Tuparí: proposta de um dicionário bilíngüe*. 2004. 286 f. Tese (Doutorado em Lingüística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista “Júlio de Mesquita”, São Paulo, 2004.
- ARAGON, Carolina Coelho. *A grammar of Akuntsu, a Tupian language*. Tese (Doutorado) - University of Hawai ‘I at Manoa, 2014.
- BRASIL: terras Indígenas, situação fundiária. Brasília: FUNAI, 2017. Escala 1:5.000.000. Disponível em: <http://mapas2.funai.gov.br/portal_mapas/pdf/terra_indigena.pdf>, acesso em 15/02/2017.

- GALUCIO, A. V. *The Morphosyntax of Mekens (TUPI)*. 2001. 253 f. Tese (Doctor of Philosophy) – Humanities School or Division, The University of Chicago, Chicago, 2001.
- GALUCIO, Ana Vilacy. Nominalization in the Mekens language. In: BRUNO, A. C et al. *Amerindia: la structure des langues amazoniennes II*. França: A.E.A, 2011. P. 238-260.
- GALUCIO, Ana Vilacy. Argument marking and reference tracking in Mekens. In: van GIJN, Rik et al. (Org.). *Information Structure and Reference Tracking in Complex Sentences*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2014. v. 1, p. 373-395
- GALUCIO, Ana Vilacy; NOGUEIRA, Antônia Fernanda. From Object Nominalization to Object Focus Construction: the innovative A-alignment in the Tuparian Languages (Tupian Family). *Journal of Historical Linguistics*, no prelo.
- MOORE, Dennis; GALUCIO, Ana Vilacy; GABAS JR., Nilson. O desafio de documentar e preservar as línguas amazônicas. *Scientific American Brasil – Amazônia, A Floresta e o Futuro (Brasil)*, São Paulo, n. 3, p. 36 – 43, 01 set. 2008.
- MOSELEY, Christopher (Org.). *Atlas of the World's languages in Danger*. 3. ed. Paris: <<http://www.unesco.org/culture/en/endangeredlanguages/atlas>>. Acesso em: 15 dez. 2011.
- NOGUEIRA, Antônia Fernanda. *Wayoro ãmẽto: fonologia segmental e morfossintaxe verbal*. 2011. Dissertação (Mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- NOGUEIRA, Antônia Fernanda. Sistemas ortográficos de línguas indígenas: a importância de aspectos sociolinguísticos. In: CONGRESSO INTERNACIONAL DE DIALETOLOGIA E SOCIOLINGUÍSTICA, 2., 2012, Belém. *Anais*. São Luís: EDUFMA, 2012. p. 400-415.

- NOGUEIRA, Antônia Fernanda. Descrição e análise do prefixo {e-} intr. da língua Wayoro (Ayuru, tronco Tupí). *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas*, Belém, v. 8, n. 2, p. 325-341, maio-ago, 2013.
- NOGUEIRA, Antônia Fernanda. Em busca das classes de palavras da língua Wayoro. *Linguística* (Rio de Janeiro), v. 10, p. 240-255, 2014.
- ROCHA, Ivan. *Não-finitude em Karitiana: subordinação versus nominalização*. 2016. Tese (Doutorado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- SANTOS, J. F. *Documentação da língua Wayoró: para a preservação das línguas indígenas amazônicas*. 2010. Dissertação (Mestrado) – Departamento de Ciências da Linguagem, Fundação Universidade de Rondônia, Guajará-Mirim, 2010.
- VAN GIJN, Rik; HAUDE, Khatarina; MUYSKEN, Piter. *Subordination in Native South American Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2011.

Some Observations on Word Order and Prosody in Karitiana Relative Clauses

Karin Camolese Vivanco^a

ABSTRACT

This paper aims to investigate possible prosodic differences in two word orders attested for object relative clauses in Karitiana (Tupi-Arikém). In an elicited production task, Vivanco (2014) observed that there are two possible word orders for object relative clauses in this language: OSV and SOV. Assuming the tenets of the Autosegmental Metrical Theory (Pierrehumbert, 1980, 2000, Ladd 2008), we revisit Vivanco's (2014) data and demonstrate that there is a prosodic difference between these two orders, namely, an obligatory L*+H pitch accent on the stressed syllable of subjects in SOV relatives. On the other hand, OSV object relative clauses do not obligatorily show any specific intonational contour.

KEYWORDS: Karitiana; relative clauses; word order; intonation

RESUMO

O objetivo do presente trabalho é investigar possíveis alterações prosódicas em duas ordens de palavras encontradas nas orações relativas de objeto da língua karitiana (Tupi-Arikém). Em um experimento de produção elicitada, Vivanco (2014) verificou que há duas ordens possíveis para relativas de objeto na língua karitiana: OSV e SOV. Assumindo a Teoria Autossegmental Métrica (Pierrehumbert, 1980, 2000, Ladd 2008), revisitamos os dados de Vivanco (2014) e mostramos que há uma diferença prosódica entre essas duas ordens: especificamente, orações relativas com a ordem SOV têm um evento tonal (um acento L*+H) na sílaba tônica do NP sujeito; relativas de objeto com a ordem OSV, em contrapartida, não exibem nenhum contorno entoacional obrigatório.

PALAVRAS-CHAVE: Karitiana; orações relativas; ordem de palavras; entoação

^a Universidade de São Paulo, karincv@usp.br

The research for this paper was financially supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grant numbers 2011/15927-7 and 2014/15141-1. We also thank the Karitiana community, Luciana Storto, Didier Demolin, Flaviane Svartsman-Fernandes, Maria del Mar Bassa Vanrell, Karolín Obert, Paula Armelin, Antônia Nogueira, Janayna Carvalho, Aline Benevides, Júlia Fernandez, and the audience of the Encontro Intermediário do GT-Teoria da Gramática, the 7th Conference on Indigenous Languages of Latin America and the 4th Workshop de Línguas Indígenas da USP. All remaining errors are my responsibility

Introduction

In Karitiana, there are two possible word orders for object relative clauses¹: OSV (example (1)) and SOV (example (2)):

(1) *OSV object relative clause*

Yn Ø-na-aka-t i-pyting-Ø [kinda'o sosy ti-'y]-ty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR [fruit armadillo TI-eat]-OBL
 'I want the fruits that the armadillo was eating'

(2) *SOV object relative clause*

Yn Ø-na-aka-t i-pyting-Ø [Luciana pykyp ti-pipāram]- aty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR [Luciana clothes TI-sew]- <e.v.>OBL
 'I want the clothes that Luciana sewed'

Variable word order is a characteristic of some of the world's languages (see Hale, 1983, 2013 and Legate, 2002 on non-configurationality). In these, factors other than the syntactic function seem to regulate serialization of NPs (Fanselow, 1990). Therefore, one question that arises in (1-2) is what factors could be involved in this variation of word order in Karitiana. In this paper, we examine possible prosodic differences in the relative clauses in (1) and (2), showing that intonation seems to be a differentiating factor between them.

The paper is organized as follows: in section 1, the necessary background information about the language and its relative constructions is provided. The theory of intonation assumed in this paper – the Autosegmental Metrical Theory – will be described in section 2, along with the inventory of the relevant tonal events in Karitiana. The phonetic analysis is developed in section 3, where it will be demonstrated that one of the word orders attested above (namely, the SOV in (2)) displays a fixed intonational contour. Finally, some concluding remarks are made in section 4.

¹ In object relative clauses, the head (*i.e.*, the relativized NP) is the direct object of the embedded verb.

1. Karitiana

Karitiana is a Tupian language of the Arikém family (Rodrigues, 1986), spoken by around 400 speakers². Most of them live in an indigenous reservation in Porto Velho, northwestern Brazil.

Landin (1982) was the first author to notice that Karitiana displayed a considerable amount of different word orders. In this regard, Storto (1999) detected a complementary distribution between matrix and embedded clauses: the former can be either verb-initial or verb-second, whereas the latter are always verb-final.

(3) *Verb-initial matrix clause*

Ø-pyry-‘a	saryt-yn	keerep	Gokyp
3-ASS-do	IND.EVID.-NFUT	in.the.old.times	sun

‘The sun was like this in the old days’ (STORTO, 2002, p.155)

(4) *Verb-second matrix clause*

Taso	Ø-na-oky-t	boroja
man	3-DECL-kill-NFUT	snake

‘The man killed the snake’ (STORTO, 2002, p.153)

(5) *Embedded clause*

[Taso	boroja	oky	tykiri]	Ø-naka-hyryp-Ø	õwã
man	snake	kill	PERFVE	3-DECL-cry-NFUT	child

‘When the man killed the snake, the child cried’
(STORTO, 1999, P. 121)

According to Storto, this complementary distribution is derived through a V2-effect in the language: in matrix clauses, the verb moves to C, while it stays *in situ* in

² Most speakers are fluent in both Karitiana and Portuguese.

embedded environments ³. In addition to word order, Storto shows that the occurrence of verb movement is also associated with verb morphology: as shown in examples (3-5), verbs surface with person, mood and tense morphology in matrix clauses; embedded verbs, on the other hand, do not exhibit such morphology.

There is also some evidence of SOV being the default word order. When asked to translate sentences such as “I saw Karin hitting Ivan” into Karitiana, speakers often give us constructions like (6). The OSV equivalent, with the reversed order of NPs, can only mean “I saw Ivan hitting Karin.” for most speakers:

(6) *Embedded clause*

Yn Ø-na-aka-t i-so’oot-Ø [Karin Ivan mĩ]-ty
 1s 3-DECL-COP-NFUT NOM-see-ABS.AGR. [Karin Ivan hit]-OBL
 ‘I saw Karin hitting Ivan’ (VIVANCO, 2014, P.13)

(7) *Embedded clause – reversed word order*

Yn Ø-na-aka-t iso’oot [Ivan Karin mĩ]-ty
 1s 3-DECL-COP-NFUT NOM-see-ABS.AGR. [Ivan Karin hit]-OBL
 ‘I saw Ivan hitting Karin’ (VIVANCO, 2014, P.14)

As can be seen in examples (5-7), embedded clauses in Karitiana also do not show complementizers of any sort.

Structurally, relative clauses are very much alike other embedded clauses⁴. There are no relative pronouns, resumptive pronouns or any other morphology specific

³ This pattern is possibly due to the lack of a CP layer in embedded clauses (see Storto, 1999).

⁴ There is, however, an important difference between complement clauses (including relative clauses) and adjunct clauses: only the latter shows the adverbial marking {-t} (Rocha, 2013). This morpheme is ungrammatical in relative clauses:

(i) *Relative clause grammatical without {-t}*

Yn Ø-na-otet-Ø [pikom õwã ti-m-pykyna ki]
 1 3-DECL-cozihar-NFUT [monkey child INV-CAUS-run ANTER]
 “I cooked the monkey that didn’t run from the child.” (ROCHA, 2013, p. 21)

(ii) *Relative clause ungrammatical with {-t}*

*Yn Ø-na-otet-Ø [pikom õwã ti-m-pykyna ki-t]
 1 3-DECL-cook-NFUT [monkey child INV-CAUS-run ANT.-ADVZ] (ROCHA, 2013, p. 21)

(10) *SOV object relative clause* [= (2)]

Yn Ø-na-aka-t i-pyting-Ø [Luciana **pykyp** ti-pipāram]- aty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR. [Luciana **clothes** TI-sew]- <e.v.>OBL
 ‘I want the shirt that Luciana sewed’

Externally-headed relative clauses never have their heads so deeply embedded. In fact, the head on the periphery of the clause is a defining property of externally-headed relatives, whereas internally-headed constructions are surrounded by the relative clause itself (for this reason, it is called ‘circumnominal relatives’ by De Vries, 2002).

Another argument pointing to the conclusion that relative clauses in Karitiana are internally-headed is that whenever there are oblique arguments in the embedded clause, they must stay on the periphery of the subordinate clause and before the head itself. Crucially, it is not possible for the head to stay on the periphery of the relative clause, as example (13) shows⁶:

(11) *Subject relative clause*

[‘ep ohynt taso aka] Ø-na-aka-t i-a-mĩ-t a-piri
 tree on.top man COP] 3-DECL-COP-NFUT NOM-PASS-hit-ABS.AGR. 2-by
 ‘The man on top of the tree was hit by you’

Literally: ‘The man that is on top of the tree was hit by you’

(Adapted from Everett, 2006, p. 384)

(12) *Subject relative clause*

Yn Ø-na-aka-t i-so’ oot-Ø [São Paulo-pip õwã aka-ty]
 1s 3-DECL-COP-NFUT NOM-see-ABS.AGR. São Paulo-in child be-OBL
 ‘I saw the child that was in São Paulo’

⁶ Another argument first pointed out by Storto (1999) is case-marking on the head. In these relative clauses, the head surfaces with the case-marking demanded by the embedded verb, not by the matrix verb. Therefore, there is a closer relation between the head and the embedded verb. This pattern also strongly resembles the behavior of internally-headed in some of the world’s languages (see Vivanco, 2014)

(13) *Ungrammatical subject relative clause*

*Yn Ø-na-aka-t i-so' oot-Ø [õwã São Paulo-pip aka-ty].
 1s 3-DECL-COP-NFUT NOM-see-ABS.AGR. child São Paulo-in be-OBL
 Intended: I saw the child that was in São Paulo'

It has been reported that some languages with internally-headed relatives show variation in the position of the head (Basilico, 1996). Consider the following data of Mojave, taken from Munro (1976):

(14) *Relative clause with head in situ*

[Masahay ahvay ʔ-ay-n^Y-č] ʔahot-m [Mojave]
 girl dress 1-give-DEM-SUBJ good-TNS
 "The girl I gave the dress to is nice"
 'The dress I gave to the girl is nice' (Munro, 1976, p. 198)

(15) *Relative clause with fronted head*

[Ahvay masahay ʔ-ay- n^Y-č] ʔahot-m [Mojave]
 dress girl 1-give-DEM-SUBJ good-TNS
 'The dress I gave to the girl is nice' (Munro, 1976, p. 198)

In these examples, one can see that the head 'ahvay' can surface in two different positions, namely [O_{GOAL} O_{THEME} V] and [O_{THEME} O_{GOAL} V]. This variation is similar to what has been observed for Karitiana. As it was discussed at the beginning of this paper, object relative clauses have two possible word orders, OSV and SOV. In the latter, the head is embedded within the relative clause itself, whereas it stays on the periphery of the clause in OSV relatives:

(16) *OSV object relative clause* [= (1)]

Yn Ø-na-aka-t i-pyting-Ø [kinda'o sosy ti-'y]-ty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR. [fruit armadillo TI-eat]-OBL
 'I want the fruits that the armadillo was eating'

- (17) *SOV object relative clause* [= (2)]
 Yn Ø-na-aka-t i-pyting-Ø [Luciana **pykyp** ti-pipãram]-aty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR. [Luciana **clothes** TI-sew]-<e.v.>OBL
 ‘I want the piece of clothes that Luciana sewed’

In sum, it was argued in this section that the sentences in (16) and (17) are internally-headed relative clauses and that there is a variation between the OSV and SOV word orders in object relatives. The next step is to analyze whether or not there could be any other difference between them besides word order.

2. Theoretical background

2.1. Autosegmental metrical theory of intonational phonology

The property of sound that is most relevant for intonation is pitch, as intonation is manifested through it (Ladefoged, 2003). Pitch can be defined as follows:

Strictly speaking, pitch is an auditory property – something you hear. It is not an acoustic property – an aspect of the sound wave that you can measure. From a practical point of view when discussing the pitch of the voice, it can usually be said to be the rate at which vocal fold pulses recur, and thus the fundamental frequency of the sound wave. (Ladefoged, 2003, p.75)

Since it is a perceptual property, it is not possible to measure pitch itself. Nevertheless, it is possible to access its acoustic correlate: the fundamental frequency (F0), measured in (Hz). Intonation analyses thus look at this property, graphically represented as a curve by several computer programs.

According to Nespor and Vogel (2007), Intonational Phrases (henceforth IntP) are commonly defined as the relevant prosodic domain for intonational contours, being delimited by pauses and/or lengthening and also by the presence of boundary tones (see

the discussion in Cruttenden (1997) on the definition of Intonational Phrases). Importantly, the borders of an IntP do not necessarily coincide with the syntactic limits of clauses. IntPs can be restructured according to several factors, such as speech rate, formality, etc., generating prosodic constituents non-isomorphic to the ones created by the syntax (Nespor; Vogel, 2007).

Our analysis is built upon the Autosegmental Metrical Theory of Intonational Phonology as proposed by Pierrehumbert's (1980, 2000) and Ladd (2008). One of its basic tenets is to view the intonational contour as a product of underlying target tones. Therefore, it is not exactly the whole form of the tune that matters, but some specific tonal events aligned with particular points of the segmental chain.

According to this theory, this underlying sequence can be described with an inventory of only two target tones, low (L) and high (H). Importantly, these tones are strictly relational: a certain tone is classified as either high or low in relation to its preceding tones, and not because it is produced within a certain range of frequency.

Pierrehumbert (1980) proposes three types of tonal events: **pitch accents**, **phrase accents**, and **boundary tones**. Pitch accents would be the tonal movements occurring on stressed syllables of words within an utterance. In the notation system proposed by her, pitch accents are indicated by a star, such as H* or L*.

Whenever these tonal excursions involve more complex movements, such as a downward movement of F0 followed by a peak, these accents can be described through bitonal pitch accents (e.g., L*+H or L+H*). In this case, the star would indicate the relative alignment of one tone in relation to the stressed syllable (Pierrehumbert, 2000).

There are still two other types of tonal events that are usually restricted to the right boundary of IntPs: phrase accents and boundary tones. Both occur after the last pitch accent of an IntP - also known as the **nuclear pitch accent**. Boundary tones are the tonal events that occur at the end of an Intonational Phrase, and phrase accents control the transition between the nuclear pitch accent and the boundary tone (e.g., specifying whether the transition will have an ascending or a descending pitch)⁷.

⁷ Additionally, phrase accents can also be used to mark the boundaries of intermediate Intonational Phrases, an intonational domain smaller than the Intonational Phrase described here (PIERREHUMBERT; BECKMAN, 1986). These constituents also have tonal events at their boundaries (as well as pre-boundary lengthening). Since they involve a more subtle perception of intonational

Crucially, there is a one-to-one relationship between boundary tones and IntPs: any IntP must have a boundary tone at its end, and boundary tones signal the end of IntPs.

To transcribe tonal events, we adopt the ToBI conventions (Hirschberg; Beckman, 1994), unless otherwise specified.

2.2. Intonational grammar of Karitiana

Authors such as Storto (1999), Storto and Demolin (2005), and Everett (2006, 2008) described the intonational grammar of Karitiana in more detail. These authors, however, often assumed different premises from the ones assumed here. In this section, we highlight their main findings and discuss how these can be described in our terms.

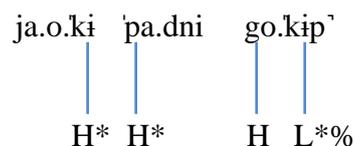
Firstly, these authors assume an isomorphic relationship between sentences and Intonational Phrases, so each sentence will correspond to only one intonational domain. As seen above, we admit the possibility of breaking down a sentence into more than one IntP (*i.e.*, restructuring). Ultimately, this difference means that in our analysis, there are more boundary tonal events than in the previous literature on Karitiana.

Storto (1999) was the first author to describe the interaction between the last pitch accent and the end of a declarative sentence in Karitiana. According to her analysis, its intonational grammar would assign H* tones to all stressed syllables in a given prosodic domain. Since declarative sentences always end in a low pitch (a fact also noticed by Everett, 2008), she proposes an L boundary tone that attaches to the last stressed syllable of the sentence. This would make the H tone previously assigned to float and to attach to the preceding syllable, generating the high-low pattern of the final parts of declarative sentences.

Consider how her system would work for an utterance like ‘Iaoky padni Gokyp’⁸:

contours, we chose not to indicate them at this point. However, future research may show that some of the tonal events described here may actually be a combination of pitch accents and phrase accents (see fn. 9).

⁸ Here, only the treatment of the nuclear pitch accent and of the boundary tones are discussed. Nevertheless, Storto (1999) and Storto and Demolin (2005) propose other rules to account for the contour of ‘iaokyp’ and ‘padni’ as well.

(18) *First step – assignment of H**(19) *Second step – assignment of L% and dislocation of H*

[based on example 13b of STORTO and DEMOLIN, 2005, p.19]

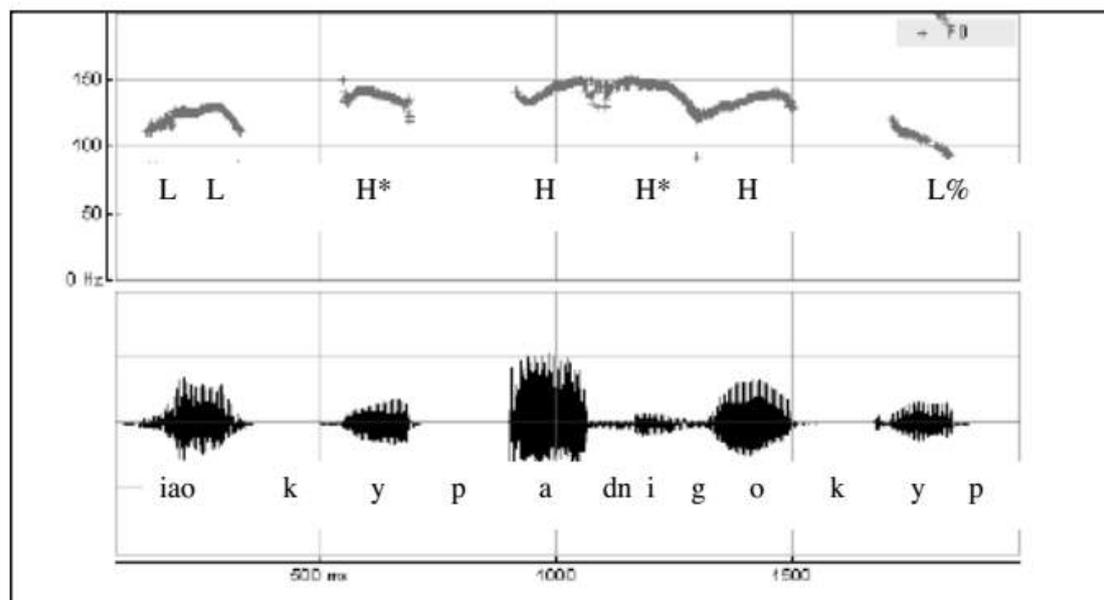


Fig. 1. Tonal analysis of the utterance 'Iaoky padni gokyp' ('Gokyp was not killed') according to Storto and Demolin (2005, p.18)

Storto's system is incompatible with the premises exposed here, as (nuclear) pitch accents and boundary tones are independent of each other in the Autosegmental Metrical Theory. In this paper, we chose to analyze Karitiana according to AM's assumptions, but the issue of how nuclear pitch accents and boundary tones could interact is a topic that deserves future investigation in the Karitiana literature⁹.

⁹ In English, at least, there is evidence that nuclear pitch accents and phrase accents/boundary tones have to be distinct tonal events, as almost all the combinations between them occur.

Storto's observation about the tonal excursions at the end of a declarative sentence can be captured by positing two distinct tonal events: a nuclear pitch accent $H+!H^*$ and a boundary tone $L-L\%$. The $H+!H^*$ would comprise an early aligned H tone that occurs in the pre-stressed syllable(s) and a downstepped $!H$ tone aligned with the stressed syllable. This tonal movement can also be captured by positing an $H+L^*$, even though this pitch accent is not assumed in Hirschberg and Beckman's (1994) ToBI guidelines for English. It is reasonable to propose that the tone target aligned with the stressed syllable is an L^* , as the downward movement often approaches the bottom of the speaker's range.

After this pitch accent, there is a sharp downward movement in the pitch track and the voice frequently becomes creaky. We take this as an indication of an L- phrase accent and $L\%$ boundary tone, as this downward movement is similar to the $L-L\%$ contours identified for English.

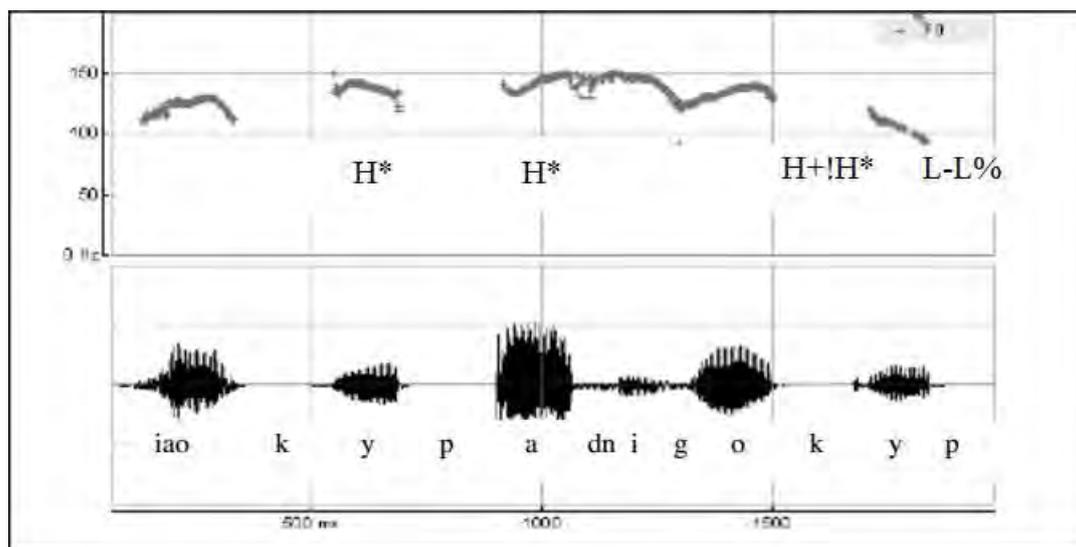


Fig. 2. Our analysis of the utterance 'Iaoky padni gokyp.'

Since the previous literature paired each sentence with only one Intonational Phrase, any tonal excursions before the end of a sentence would be either analyzed as pitch accents (as in Storto, 1999 and Storto; Demolin, 2005) or as nothing at all (as in some of Everett's 2008 examples). In this paper, we recognize that boundary tones can occur before the end of a sentence – in this case, restructuring of the IntP has taken place.

This pattern occurs in several biclausal sentences. These are frequently restructured into two smaller IntP, as can be detected by a pause between them. In this environment, a contour different from the H+!H L-L% arises. This new pattern is illustrated in the example below, where the matrix clause ‘Yn naakat iso’oot...’ (‘I saw’) ends differently from the embedded clause ‘... Maria gok amangaty.’ (‘...Maria planting the manioc.’):

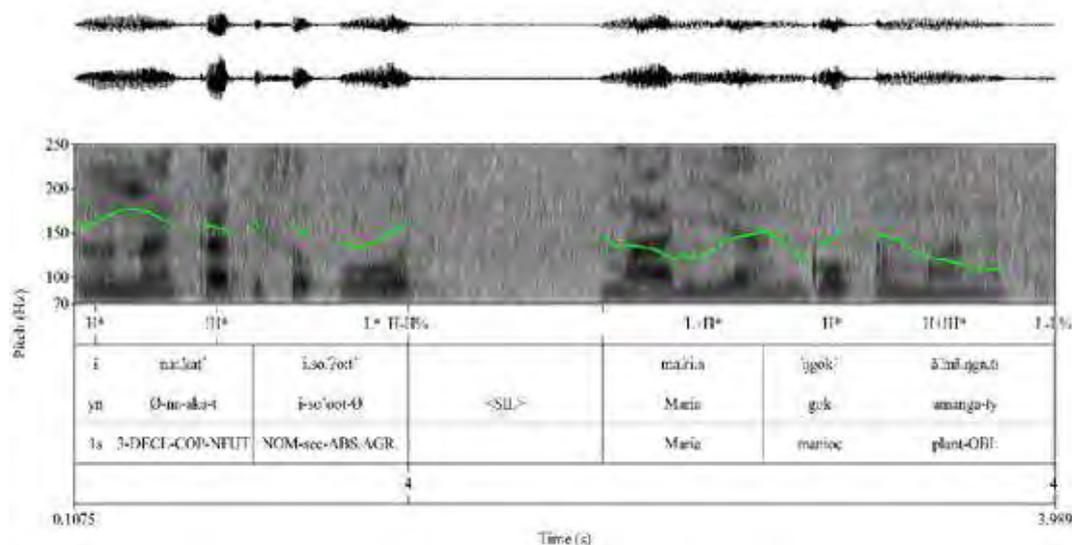


Fig. 3. Waveform and F0 for ‘Yn naakat iso’oot Maria gok amangaty.’ (‘I saw Maria planting the manioc’) – Speaker E

First, the F0 is low on the last stressed syllable, indicating the presence of an L* nuclear pitch accent. Besides, what happens after this low tone is strikingly different from the aforementioned contour: there is a rise right after the L*, indicating the presence of an H- controlling the last portion of the sentence and an H% boundary tone.

The boundary tone itself is a little difficult to see in most cases, but we can consider it an H% because of the sharp rise format. Below there is another example of this contour on the matrix clause ‘Yn naakat ipyting...’¹⁰:

¹⁰ As the F0 drops slightly at the very end of the Intonational Phrase, it is possible to analyze it as an H-L% as well. However, this could also be an effect of a segmental perturbation.

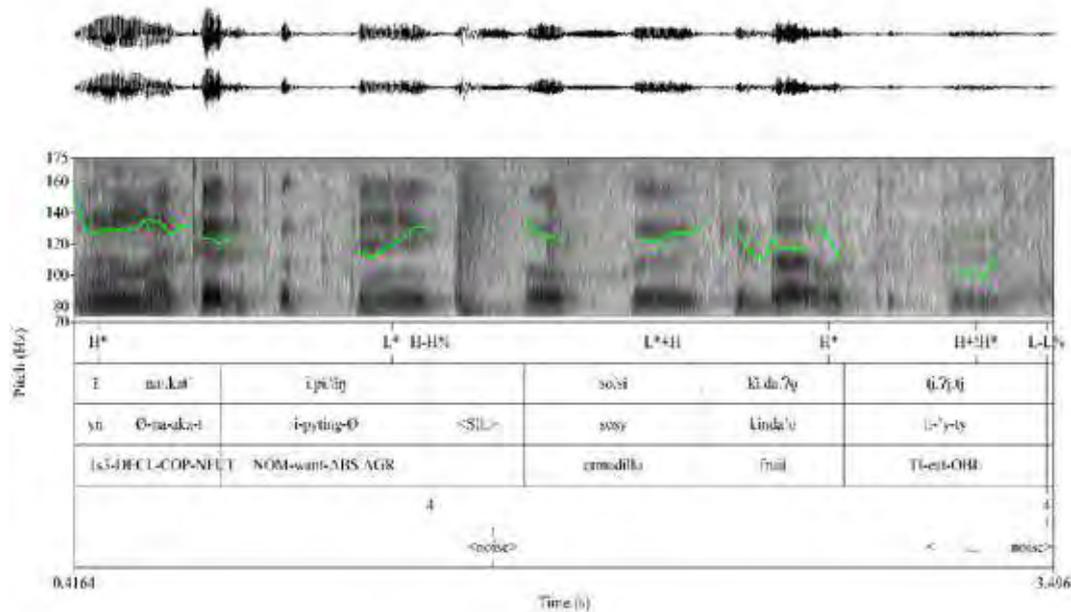


Fig. 4. Waveform and F0 for 'Yn naakat ipyting sosy kinda'o ti'yty.' ('I want the fruits that the armadillo was eating') – Speaker R_OBJ2

Crucially, these examples also illustrate that assuming the theoretical background of the Autosegmental Metrical Theory, the inventory of pitch accents in Karitiana becomes much more diverse than what has been assumed in previous accounts.

2.3. Summary

In this section, we reviewed some of the main aspects of the Autosegmental Metrical Theory of Intonational Phonology and some pitch accents, phrase accents, and boundary tones of Karitiana under this account. So far, the inventory of tonal events would be as follows:

(20) *Pitch accents of Karitiana*

- H*
- L*
- L+H*

- L*+H
- H+!H* (or a H+L*)

(21) *Phrase accents and boundary tones of Karitiana*

- H-H% (for matrix clauses of bi-clausal sentences)
- L-L% (end of declarative sentences)

In the next section, the major intonational differences between OSV and SOV relative clauses will be discussed. As will be seen, we argue that there seems to be a strong relation between one of these word orders (SOV) and the presence of a pitch accent L*+H.

3. Some observations on order variation in object relative clauses

As exposed at the beginning of this paper, our goal is to describe phonological differences between different word orders found in object relative clauses. In this section, one of the major phonological differences between these two types of relative clauses – intonation – will be presented.

All the data analyzed here comes from a production task detailed in Vivanco (2014) and Storto, Vivanco & Rocha (to appear), which was based on a protocol designed by Labelle (1990) to elicit relative clauses. In this experiment, speakers should always choose one of two pictures depicting identical characters or objects that could only be differentiated through the action of another element in the scene. Informants should say to the researcher which picture was chosen:

(22) Real extract taken from Vivanco's (2014) experiment

Researcher: There are two shirts here. This one was sewed by Ana and this other one was sewed by Luciana. You have to choose one of them and tell us which one was chosen.

Speaker: Yn Ø-na-aka-t i-pyting-Ø [Ana pykyp ti-pipāram]-aty
 1s 3-DECL-COP-NFUT NOM-want-ABS.AGR. [Ana clothes TI-sew]-<e.v.>OBL
 ‘I want the shirt that Ana sewed’

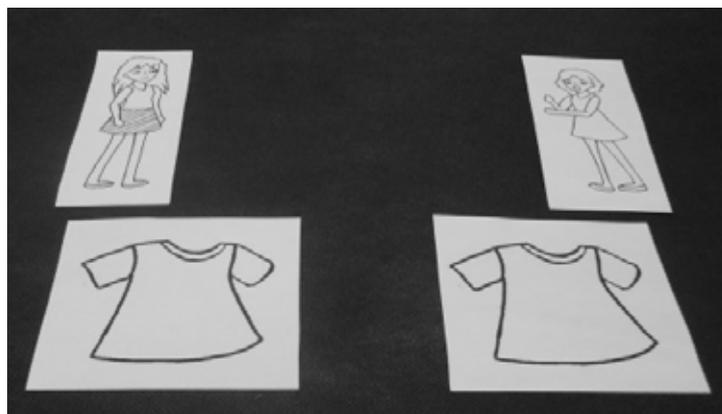


Fig. 5. Pictures used in the experiment

All contexts followed the pattern in (22). The only thing that changed throughout the experiment was the elements involved in the scene (*e.g.*, Luciana, Ana, man, woman, etc.) and the type of action described (sewing, eating, building, etc.).

The number of sentences used for phonetic analysis is 14 OSV relative clauses and 10 SOV relative clauses. These sentences were produced by 4 different speakers (3 males and 1 female) and collected at a sampling rate of 44100Hz. These four speakers produced both SOV and OSV word orders, but we occasionally considered additional data from two other speakers who only produced one of them.

The description of the situation in which SOV and OSV relative clauses were produced lead us to conclude that at least in this case, major contextual modifications do not seem to induce word order variation. Hence, this makes us wonder what the differences between these word orders could be.

Our working hypothesis is that there are phonological differences between these word orders, especially because prosody has been described to play a role in word order variation in many languages (see Antonyuk-Yudina; Mykhaylyk, 2013 on scrambling in Ukrainian, for example). The prosodic property investigated is pitch modulations on

stressed vowels and the relevant comparison here is between NPs occupying the first position of the relative clause (*i.e.*, objects of OSVs and subjects of SOVs):

(23) Comparison made in the phonetic analysis

OSV: REL[**Object** Subject Verb]

SOV: REL[**Subject** Object Verb]

In order to detect possible pitch variations, we measured the fundamental frequency (F0) of the relevant segments using the free software Praat (Boersma; Weenink, 1992-2017). Only vowels were measured and these were always stressed and short, as Karitiana has a phonemic distinction between short and long vowels (see Storto, 1999).

Given that IntPs are the domain of intonation, they will be the prosodic constituent relevant for our analysis of pitch. Just like other languages, the crucial properties to identify and delimit an intonational phrase in Karitiana are the presence of (1) pauses, (2) lengthening and (3) boundary tones.

The object relative clauses considered here usually formed their own intonational phrase, since a pause is almost always introduced between the matrix clause ‘Yn naakat ipyting’ (‘I want...’) and the relative clause itself. Besides, these relatives were always marked by an H+!H* nuclear pitch accent and a L-L% phrase accent and boundary tone. As seen above, this is the characteristic pattern of declarative sentences. In Figure 6, one can see the pause breaking up two intonational phrases in sentence (2). In the IntP corresponding to the matrix clause, the nuclear pitch accent on the syllable [tĩŋ] is an L*, and the end of this prosodic domain has an H-H% tune. The embedded clause forms its own IntP with its own contour: a H+!H* on [pã] and a L-L% at the end ¹¹:

¹¹ As the reader will notice in Figures 6-14, some F0 tracks display large and sudden jumps. These are in fact abrupt or irregular changes in frequency which the computer software tries to recognize. They usually occur at the end of an intonational phrase with an L-L% tone, as the speakers’ voice gradually becomes creaky.

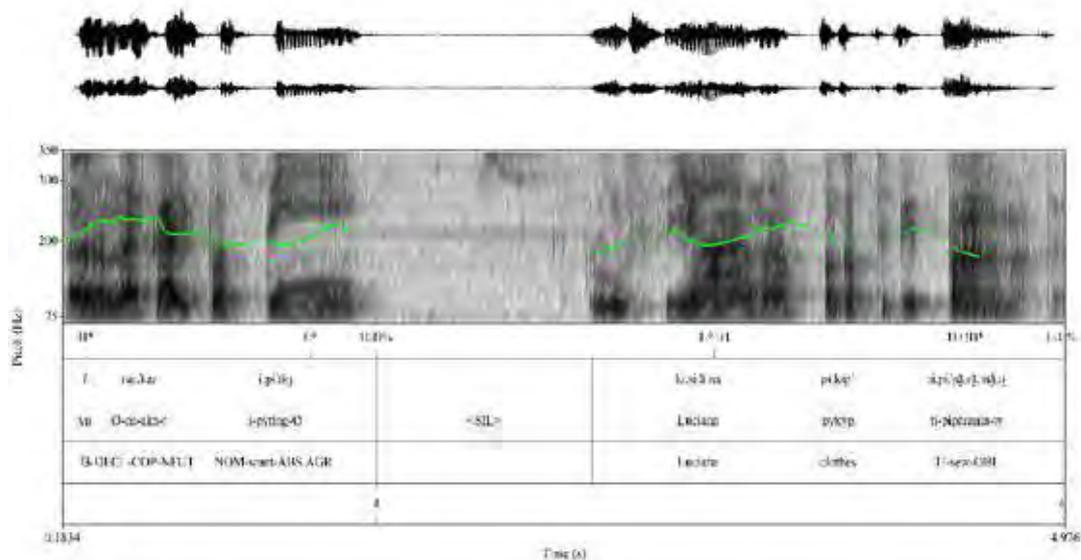


Fig. 6. Waveform and F0 for ‘Yn naakat ipyting Luciana pykyp tipipãramaty.’ (‘I want the clothes that Luciana sewed’) – Speaker S_OBJ7

Returning now to the variation under discussion, there seems to be a phonological difference between relative clauses with OSV and SOV word order. Intonation-wise, OSV and SOV differ in the sense that only the latter has a **marked and fixed pitch pattern**. Specifically, SOV relatives obligatorily show an L*+H accent on the stressed vowel of subjects, which can be identified in the F0 track as a valley or plateau (the L tone) followed by an ascending line (the H tone)¹².

We provide below four examples to demonstrate this pattern. It is necessary to remember that assuming the tenets of Autosegmental Metrical Theory, the most important thing here is not the absolute values of frequency reached during the production of the stressed vowel. Therefore, it does not matter whether the intonational contour is a smooth line or an abrupt valley-rise. What matters here is that the underlying pattern L*+H is recurrent in all of these examples¹³.

¹² It is also possible to break up this pitch accent into a combination of a pitch accent and a phrase accent. Most cases of SOV object relative clauses seem to include lengthened syllables, which could indicate the presence of an intonational boundary. In this case, this tonal movement could be also analyzed as an L* H- pattern (or perhaps even an L* H-H% or an L* H-L%).

¹³ Following the ToBI conventions, the location of pitch accents (including bitonals) is placed in the middle of the stressed vowel. In some cases, this does not correspond exactly to the place with the lowest or the highest F0 in the pitch track.

Starting with Figure 7, this pattern can be observed on the stressed vowel [ã] of ‘Luciana’, the subject of the relative clause: the F0 track first falls (the tone L*) and then starts to rise on the next syllable (the late-aligned tone H):

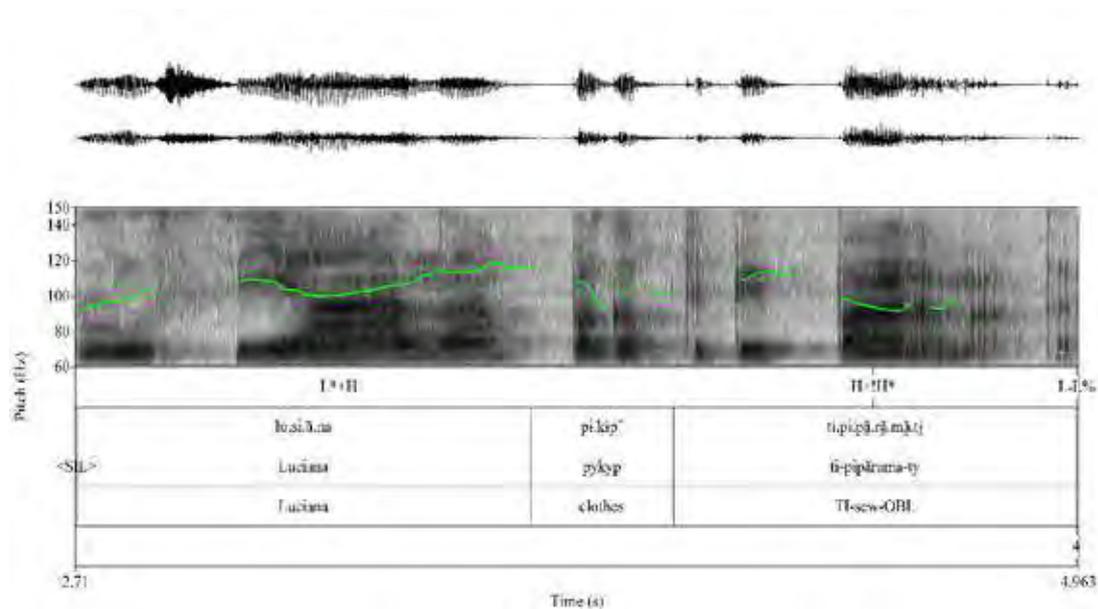


Fig. 7. Waveform and F0 for ‘... Luciana pykyp tipipãramaty.’ (‘... clothes that Luciana sewed’) – Speaker S_OBJ7

This example is also illustrative because the stressed syllable is the penultimate. Most Karitiana nouns have stress on the last syllable (Storto, 1999). Therefore, in most examples the relative alignment of tones in some bitonal pitch accents can be blurred. For this reason, cases like Figure 7, in which the subject has a different stress placement, show the alignment patterns more clearly¹⁴.

Figure 8 also shows another word with the penultimate stress as the subject, ‘Ana’. One can see that the low part of the tone (the L*) is aligned with the first stressed syllable [ã], whereas the late-aligned H extends onto the post-stressed syllable [nã].

¹⁴ ‘Luciana’ possibly has a different pattern because it is a loanword from Portuguese. The name ‘Ana’ from Figure 8 is also a loanword.

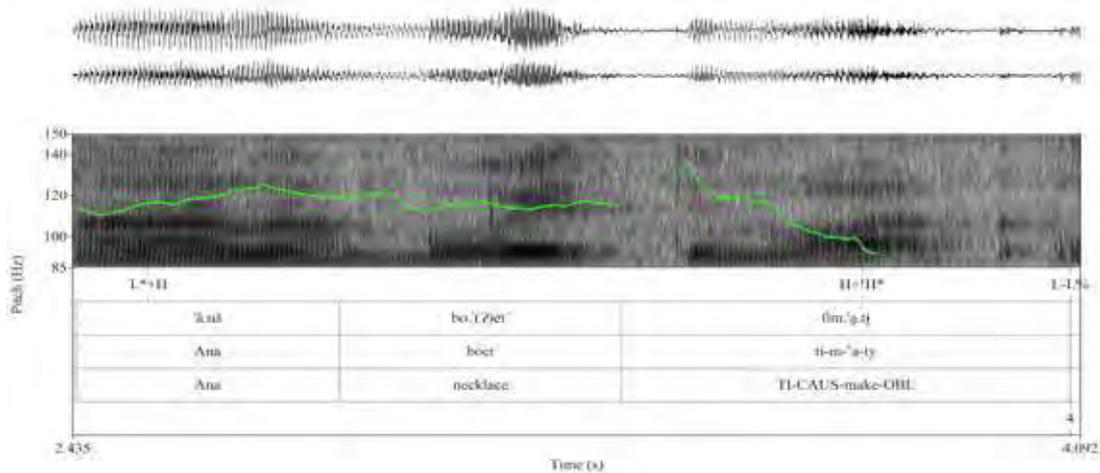


Fig. 8. Waveform and F0 for ‘... Ana boet tim’aty.’ (‘... the necklace that Ana made’) – Speaker R_OBJ9

In Figure 9, the subject of this object relative clause is ‘soty’ ([so.ʔi]). One can see a brief valley followed by an ascending line on the stressed syllable [ʔi]. The fact that there is a small peak when the F0 starts to rise is unimportant, because what truly matters here is the upward movement following the L tone, indicating the presence of an underlying H. Even though both target tones seem to occur on the same vowel (as [so.ʔi] has its stress on the last syllable), the L* is perceptually more prominent.

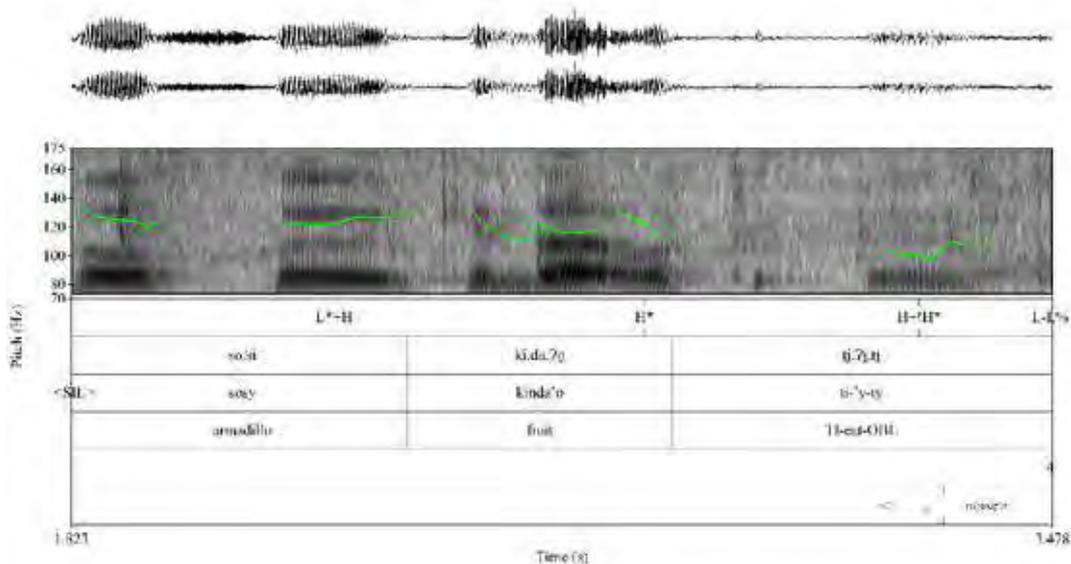


Fig. 9. Waveform and F0 for ‘... soty kinda’o ti’yty’ (‘... the fruits that the armadillo was eating’) – Speaker R_OBJ2

Finally, one can see in Figure 10 the L* more clearly, as it forms a deep valley in the F0 track aligned with the stressed syllable [so] of the subject ‘jonsõ’. The ascending line is somewhat irregular, but as in the example above, the general upward movement is what counts as an H tone:

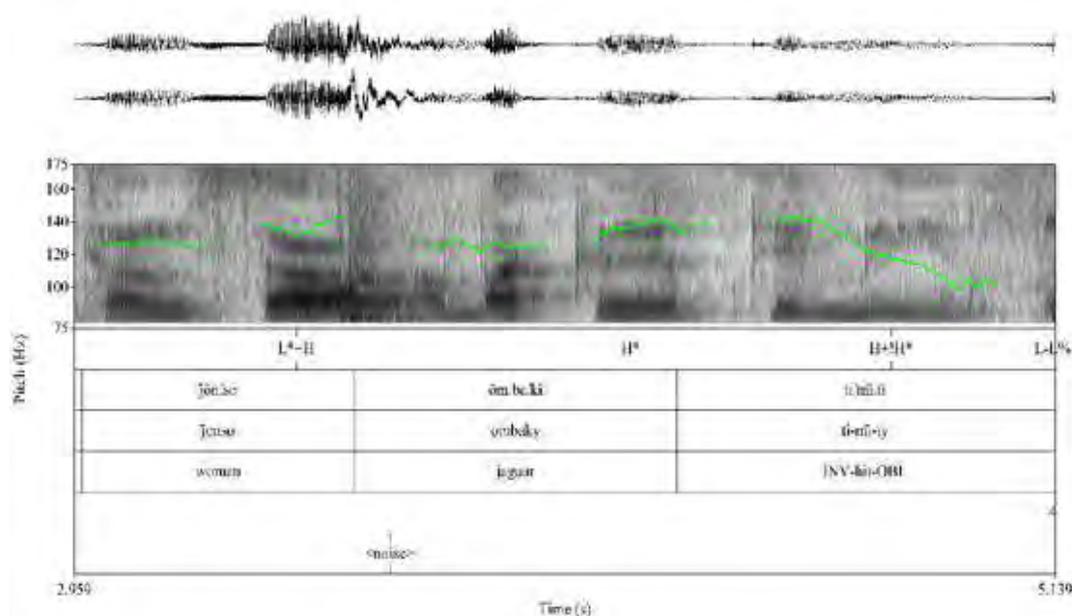


Fig. 10. Waveform and F0 for ‘... jonso ombaky timĩy.’ (‘... the jaguar that the woman hit’) – Speaker F_OBJ8

All these examples show that SOV object relatives have a fixed intonation, since all F0 tracks display an obligatory L*+H on the stressed vowel of subjects. On the other hand, the same pitch pattern is not obligatory in OSV object relative clauses. In this word order, the form of the contour on stressed vowels of objects is much freer than in SOVs. In Figures 11 and 12, for example, there is a clear rise in the F0 track, creating a clear peak different from the valley-rise pattern observed in SOVs.

For instance, Figure 11 shows an OSV object relative. In the first element of this clause – the object ‘kinda’o’ –, one can see a clear peak aligned with the stressed syllable [ʔo]. This pattern indicates the presence of an H*, and it generates a different tune from the aforementioned L*+H.

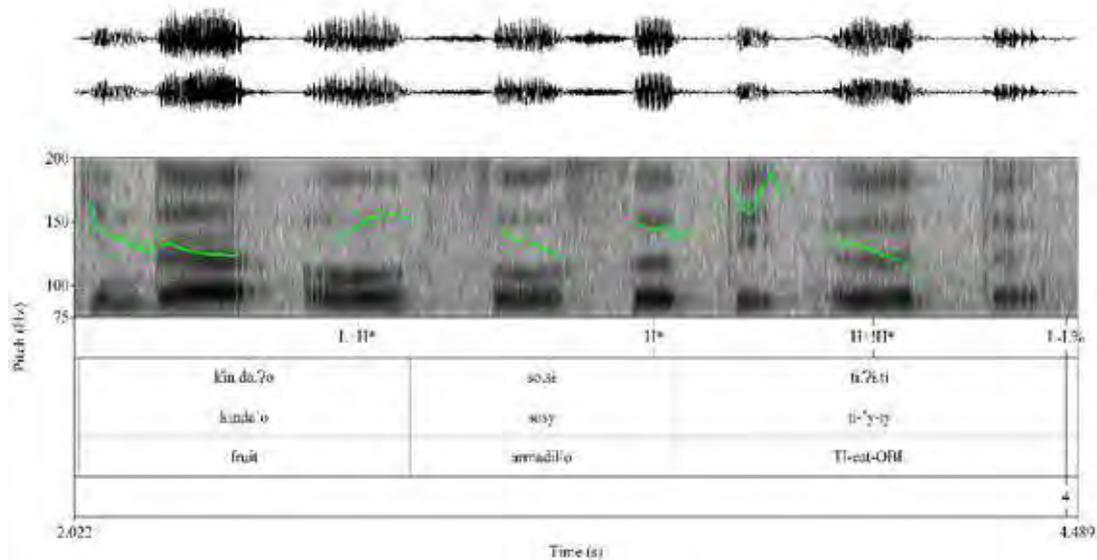


Fig. 11. Waveform and F0 for ‘... kinda'o sosy ti'yty’ (‘...the fruits that the armadillo was eating.’) – Speaker I_OBJ2

In the Figure 12, there is another example of an H* on the object of this relative clause. In this case, one can notice a peak on stressed syllable [bi] of ‘ambi’, indicating the presence of a target H*.

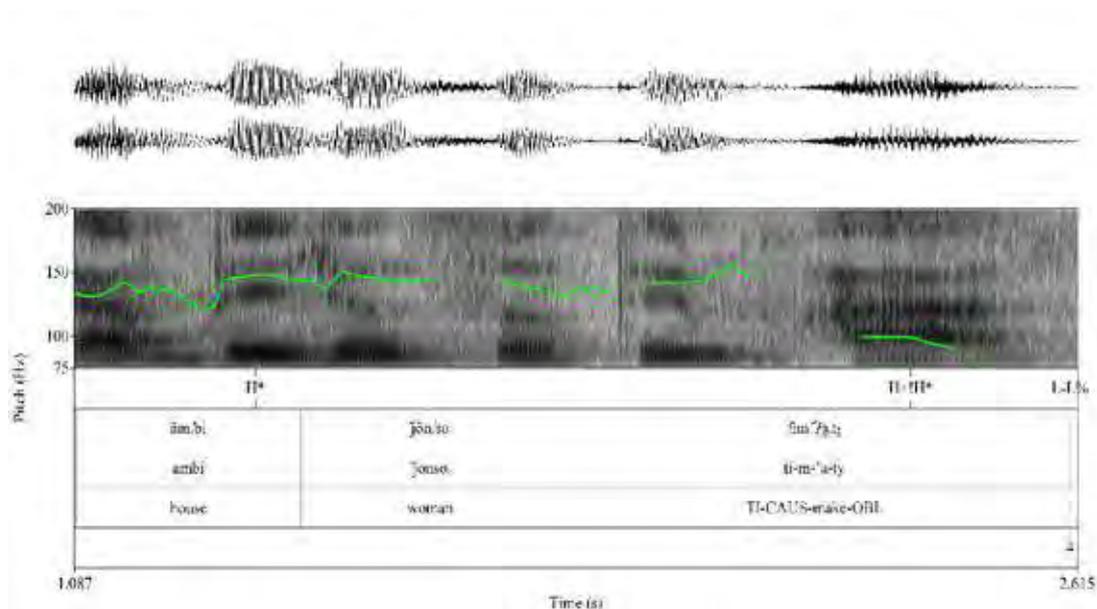


Fig. 12. Waveform and F0 for ‘... ambi ʒõso tim'aty’ (‘...the house that the woman made.’) – Speaker F_OBJ6

In another example, one can see a rising F0 beginning on the stressed syllable [i] of ‘irip’, indicating the presence of an H*.

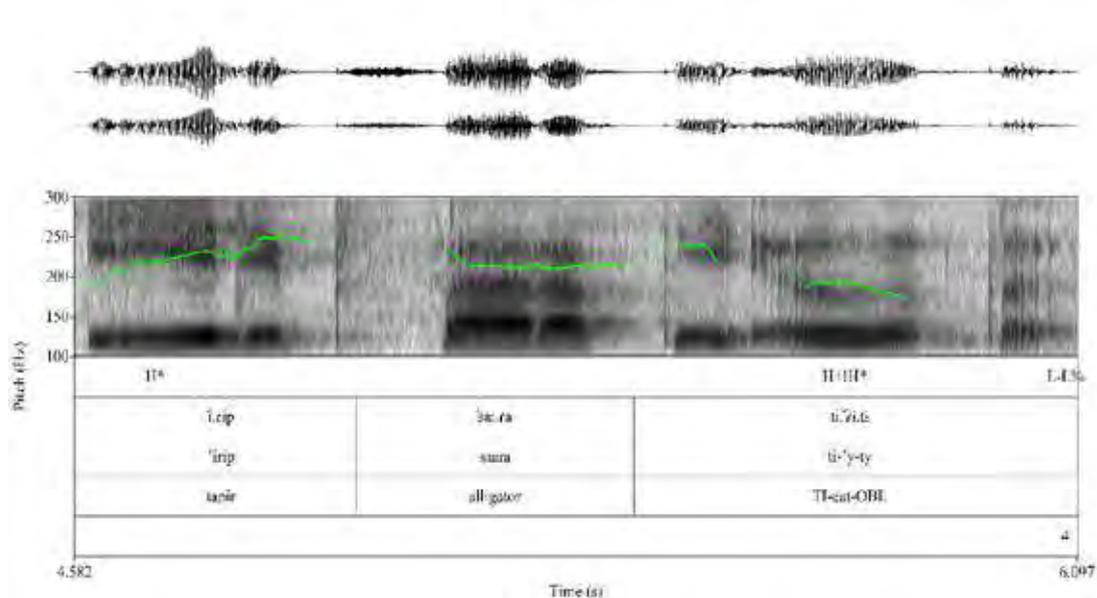


Fig. 13. Waveform and F0 for ‘... ‘irip saara ti’yty’ (‘... the tapir that the alligator ate.’)
– Speaker S_OBJ4

Finally, it is worth mentioning that some examples do seem to display some kind of L*+H pitch accent on objects of OSV relative clauses. Figure 14 shows one of these: the object ‘irip’ has a downward movement of F0 followed by a rise, and the L tone seems to be more prominent in this case. However, one can notice a pause between the object ‘irip’ and the subject ‘asori’, so the rise in this case would be probably due to an IntP boundary marked with H-H% ¹⁵:

¹⁵There are other similar cases which do not seem to have an audible pause. Nevertheless, the stressed syllable is long, and this lengthening could indicate the presence of a boundary.

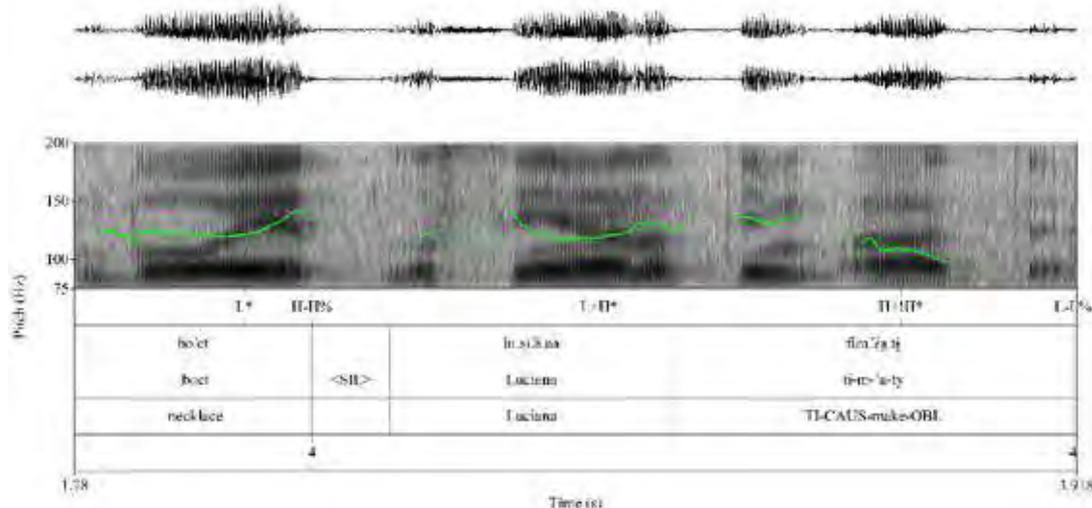


Fig. 14. Waveform and F0 for ‘... boet Luciana tim’aty’ (‘...the necklace that Luciana made.’) – Speaker I_OBJ9

Summing up, there is a tonal event that seems to distinguish OSV and SOV relative clauses. Therefore, we are now able to answer (at least partially) the question introduced at the beginning of this paper: given the variation in word order depicted in (1-2), what are the factors involved in it? So far, we are able to say that one of them is prosody, which differentiates these constructions by means of an obligatory L*+H tune on the stressed syllable of subjects in SOV relative clauses.

4. Final remarks

In this paper, we showed that object relative clauses with OSV and SOV word orders differ intonation-wise, in the sense that only the latter has a marked contour (namely, an L*+H pitch accent) on the stressed syllable of subjects. However, there are still some considerations that need to be made.

First of all, the observations made here need to be tested in future research. Given that there are only ten recorded examples of SOV object relatives, it could be the case that the specific tune in these constructions is an artifact of the low number of available data.

Besides, even though there seems to be a correlation between word order variation and a fixed tune, the exact nature of it remains somewhat unclear. If our conclusion is correct, then one of the word orders permitted in object relative clauses has a special prosody. However, it remains to be clarified whether it is the word order that is triggering a special contour (and if so, how exactly it works).

Another point that still needs attention is the (possible) relationship between the L*+H contour and contrastive focus. As seen in section 4, subjects of the object relative clauses were being contrasted in the experiment; thus, all sentences produced by our speakers can be roughly translated as “I want the clothes that Luciana (and not Ana) sewed”. Therefore, it could be the case that L*+H is the tune for contrastive focus in Karitiana, and that this pitch accent is merely being extended to relative clauses. However, it would still be unclear why this tune does not obligatorily arise on subjects of OSV object relatives, as they were equally contrasted in the experiment. Additionally, such a line of investigation needs to be pursued along with an in-depth discussion of how intonation of contrastive focus works in Karitiana, a topic still under-investigated at this point.

Finally, even though this is a pilot study on intonation and word order variation in Karitiana, we think that the major findings of this paper shed some light onto this topic. Furthermore, it also helps us to better understand the problem of word order variation by investigating how intonation might play a major role in languages with this phenomenon.

Abbreviations

<e.v.>	epenthetic vowel
1	1 st person agreement
1s	1 st person singular pronoun
2	2 nd person agreement
2p	2 nd person plural pronoun
2s	2 nd person singular pronoun

3	3 st person agreement
3s	3 rd person singular pronoun
ABS.AGR.	absolute agreement
ADVZ	adverbializer
ANT	anteriority marking
ASS	assertive mood
CAUS	causativizer
COP	copula
DECL	declarative mood
FUT	future
INT.COP.	interrogative copula
NFUT	non-future
NOM	nominalizer
OBL	oblique
OSV	object-subject-verb word order
PERFVE	perfectivity marking
REL	relative clause
SOV	subject-object-verb word order
TI	{ti-} - inverse voice morpheme
WH-	wh- element

References

- ANTONYUK-YUDINA, Svitlana; MYKHAYLYK, Roksolona. “Prosody of Scrambling”. *Proceedings of NELS 40*, 2013.

- BASILICO, David. “Head position and internally headed relative clauses”. *Language* 72 (3), 1996.
- BOERSMA, Paul; WEENINK, David (2017). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.25, retrieved 11 February 2017 from <http://www.praat.org/>
- CRUTTENDEN, Alan. *Intonation*. Cambridge: Cambridge University Press, 1997.
- EVERETT, Caleb. *Patterns in Karitiana: Articulation, perception, and grammar*. Thesis (Ph.D in Linguistics). Rice University, Houston, 2006.
- EVERETT, Caleb. “Constituent focus in Karitiãna”. Unpublished ms., 2008.
- FANSELOW, G. “Scrambling as NP-Movement”. In: GREWENDORF, G. & STERNEFELD, W. (orgs.) *Scrambling and barriers*. Amsterdã/Filadélfia: John Benjamins publishing company, 1990.
- HALE, Kenneth. “Warlpiri and the Grammar of Non-Configurational Languages”. *Natural Language & Linguistic Theory* 1, 1983.
- HALE, Kenneth. “On nonconfigurational structures”. In: MARÁ CZ, Lázló; MUYSKEN, Pieter. (eds.). *Configurationality: The Typology of Asymmetries*. Foris: Dordrecht, 1989.
- LADD, D. Robert. *Intonational phonology*. Cambridge: Cambridge University Press, 2008.
- LABELLE, Marie. “Predication, wh-movement, and the development of relative clauses”. *Language acquisition* 1, 1990.
- LANDIN, David. “An outline of the syntactic structure of Karitiana sentences”. In: DOOLEY, Robert (org.) *Estudos sobre línguas Tupi do Brasil*. Brasília: SIL, 1984.
- LANDIN, Rachel. “Word order variation in Karitiana”. *Arquivo Linguístico* 149, 1982.
- LEGATE, Julia. *Warlpiri: Theoretical Implications*. Thesis (Ph.D in Linguistics). Massachussets Institute of Technology, Cambridge, 2002.
- MUNRO, Pamela. *Mojave syntax*. New York: Garland Publishing, 1976.

- NESPOR, Marina; VOGEL, Irene. *Prosodic phonology*. Berlin: Walter de Gruyter, 2007.
- PIERREHUMBERT, Janet. *The phonology and phonetics of English intonation*. Thesis (Ph.D in Linguistics). Massachusetts Institute of Technology, Cambridge, 1980.
- PIERREHUMBERT, Janet. “Tonal elements and their alignment”. In: HORNE, Merle. *Prosody: Theory and Experiment: Studies Presented to Gosta Bruce*. Dordrecht: Kluwer, 2000.
- BECKMAN, Mary E.; PIERREHUMBERT, Janet B. “Intonational structure in Japanese and English”. *Phonology* 3, 1986.
- ROCHA, I. *A Estrutura Argumental da Língua Karitiana: desafios descritivos e teóricos*. Dissertation (Master’s in Linguistics). Universidade de São Paulo, São Paulo, 2011.
- ROCHA, I. *Estrutura argumental dos verbos em orações subordinadas em karitiana: valência verbal e sua interação com núcleos funcionais de aspecto* Grant report. Universidade de São Paulo, São Paulo, ms., 2013.
- RODRIGUES, Aryon. *Línguas Brasileiras: para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola, 1986.
- STORTO, Luciana. *Aspects of a Karitiana Grammar*. Thesis (Ph.D in Linguistics). Massachusetts Institute of Technology, Cambridge, 1999.
- STORTO, Luciana. “Algumas Categorias Funcionais em Karitiana”. In: *Línguas Indígenas Brasileiras: Fonologia, Gramática, e História. Atas do I Encontro Internacional do Grupo de Trabalho sobre Línguas Indígenas da ANPOLL I*, 2001.
- STORTO, Luciana. “Relativas de núcleo interno em Karitiana”. *Caderno de Resumos da ABRALIN 2003*. Rio de Janeiro: ABRALIN, 2003.
- STORTO, Luciana. “Caso e Concordância nas Línguas Tupi”. *Estudos Lingüísticos* 24, 2005.

STORTO, Luciana; DEMOLIN; Didier. “Pitch Accent in Karitiana”. In KAJI, S. *Cross Linguistic Studies of Tonal Phenomena*. Tokyo University of Foreign Affairs. Japan, 2005.

STORTO, Luciana; VIVANCO, Karin; ROCHA, Ivan. “Multiple Embedding of Relative Clauses in Karitiana”. To appear.

VIVANCO, Karin. *Orações relativas em karitiana: um estudo experimental*. Dissertation (Master’s in Linguistics). Universidade de São Paulo, São Paulo, 2014.

Absence of Evidence is Not Evidence of Absence: the Pirahã Case

Cilene Rodrigues^a

ABSTRACT

After Everett's (2005), the theoretical basis of Generative Grammar has been severely attacked. The mere possibility of existing a language with no self-embedding was taken by many people interested in linguistics be a conclusive proof that the notion of universal grammar is a spurious theoretical construct. However, the arguments for the unavailability of self-embedding in Pirahã are mostly based on the absence of evidence for self-embedding in the language. This calls for a serious discussion of the validity of scientific arguments within linguistics, as absence of evidence is not the same as evidence of absence. Unlike many of the language analyzed throughout this volume, Pirahã is a language with almost no morphological clues to its syntactic structure. Thus, its syntax may not be transparent, and relying on Pirahã superficial simplicity (E-language) as a way of measuring its underlying complexity (I-language) can be a serious *faux pas*. In this paper, we show that once one investigates the syntax-semantics interface, focusing on interpretative dependencies, evidence for self-embedding in the language are found. Thus, the beauty of the Pirahã grammar as a systematic recursive system is expressed in the cognitive abstract processes of its speakers.

KEYWORDS: Pirahã; syntax; self-embedding; semantic-dependencies

RESUMO

Desde Everett (2005), as bases teóricas da Gramática Gerativa têm sido severamente atacadas. A mera possibilidade de existir uma língua sem auto-encaixamento foi entendida por muitas pessoas interessadas em linguística como uma prova cabal de que a noção de Gramática Universal é um construto teórico espúrio. No entanto, os argumentos para a não-existência de auto-encaixamento em Pirahã apoiam-se na indisponibilidade de evidências contrárias. A aceitação desse tipo de argumentação indica a necessidade de discussão sobre a validade de argumentos com peso científico dentro da linguística formal, pois ausência de evidência não é evidência de ausência. Diferentemente de muitas das língua analisadas neste volume, Pirahã apresenta poucas pistas morfológicas sobre sua estrutura sintática. Portanto, a sintaxe Pirahã não é transparente, e confiar na simplicidade da língua (linguagem externa) para fazer julgamentos sobre a complexidade da gramática (linguagem interna) pode ser um passo falso. Neste artigo, mostramos que quando consideramos a interface sintaxe-semântica, focando em dependências interpretativas, encontramos evidências bastante robustas de auto-encaixamento em Pirahã. A beleza da gramática Pirahã,

^a PUC-Rio, crodrigues@puc-rio.br

como um sistema computacional sistematicamente recursivo se revela, portanto, nos processos cognitivos abstratos realizados por seus falantes.

PALAVRAS-CHAVE: Pirahã; sintaxe; auto-encaixamento; dependências semânticas

Introduction

This volume contains many papers about morphological agreement and its relation to syntactic processes in different languages. Consider, as an example, the verbal agreement in (1). [Datum from Camargos, this vol.: 1, example (3)]

- (1) ne-pytywà Tentehar a'e
 2SG-help Tenetehára 3
 ‘The Tenetehára helped you’

(1) constitutes a positive evidence for object-verbal agreement in Tenetehára. On the other hand, the contrast in (2) provides us with negative evidence for the hypothesis that the object can also control agreement on the morpheme that follows the verb. That is, (2) shows that only the external argument (the subject in (2)) can trigger on agreement the post-verbal morpheme. [Data from Camargo, this vol.: 12, examples (19) & (20)]

- (2) a. ne-r-exak rakwez kwarer ka'a r-upi a'e ri'i
 2SG-INV-see UDPAST boy forest OBL-in 3 EM
 ‘The boy certainly saw you in the forest’
- b. *ne-r-exak rakwez kwarer ka'a r-upi ne ri'i
 2SG-INV-see UDPAST boy forest OBL-in 2 EM
 ‘The boy certainly saw you in the forest’

That is, in our field, statements and hypotheses about a specific E-language or I-language can be proven (or disproven) on the basis of positive and negative evidence.

However, since there are around 7.000 languages in the world, and some of them are still understudied, being spoken only by monolinguals, it may not be easy to prove (or disprove) statements or hypotheses about unfamiliar languages, given that accessing their speakers' I-language might turn out to be difficult. Thus, the grammar of less known native languages can be a research challenge and might be subject to either incomplete analyses or misanalyses. In this paper, we consider Pirahã, a language spoken in the Brazilian Amazon by about 400 hundred natives, mostly monolinguals, as an example of research challenge in formal linguistics.

Although many aspects of the Pirahã grammar is still obscure to us, Pirahã became a famous language after Everett's (2005) claim that it has a non-recursive grammar, disallowing self-embedding altogether. In this paper, we wish to reopen the discussion, presenting new fieldwork data, showing that, contrary to Everett's claim, self-embedding seems possible and productive in Pirahã.

In contrast to the many languages that compose this volume, Pirahã is a language with almost no morphological clues to its underlying syntactic structure. It exhibits no verbal or nominal agreement, and almost no functional categories. Prepositions with no semantic contributions (e.g. *of*), determiners and complementizers seem to be all absent or morphologically impoverished. Thus, a good description of the Pirahã grammar can help us understanding less-transparent, abstract syntactic processes, as well as the interfaces between syntax and the other components of the grammar.

The take-away lesson from the discussion on recursion in Pirahã is that: (a) differences among languages are not unbound; (b) absence of evidence is not evidence of absence.¹ That is, given all that we know about human language and cognition, especially in comparison with other species,² we should not expect to find a radically different language, a language that would differ from others in profound ways. If we

¹ This aphorism is also known as *argument from ignorance* (*argumentum ad ignorantiam*, where ignorance is understood as lack of contrary evidence) and it basically states that propositions built on assumptions like *P is true because it hasn't been proved false* or *P is false because it hasn't been proved true* can be fallacious. They should be considered non-fallacious only if we can presume that our knowledge base is complete. That is, only when we have an optimal understanding of X (X being the object under investigation), we can safely assume propositions about X to be either true or false in face of absent empirical evidence. Also, propositions based on arguments from ignorance are problematic because they shift the burden of proof to the person who is questioning them. If one claims that X does not exist because evidence for X was not found, one's critics will unfairly receive the call of duty, becoming responsible for finding evidence for the existence of X.

² See Premack (2007), Hauser and Fitch (2002), Fitch (2010)

claim to have found a language like that, we must provide strong empirical evidence supporting our claims. Not finding empirical data against our hypothesis is not enough to conclude that it is right, especially when we are working with a grammar (i.e. psychological structured representation of a body of linguistics knowledge) that cannot be easily accessed.

The paper is organized as follows: in section 1, we go through the debate about the (non)existence of recursion/self-embedding on Pirahã. In sections 2 and 3, we present fieldwork evidence for self-embedding in this grammar. Section 4 is devoted to the conclusions.

1. (Non-)Recursion in Pirahã: a summary of the debate

Hauser, Chomsky & Fitch (HCF - 2002), understanding recursion as (internal and external) Merge, starts with the assumption that recursion is universal, being the core mechanism behind the engine of Grammar (NFL- Narrow Faculty of Language). That is, Merge is the sole combinatorial operation that allows human beings to go beyond iteration of symbols, forming complex hierarchical linguistic structures in pairing sounds and meaning. The emergency of this operation within the cognitive system, a consequence of some rewiring in the brain, is taken to be the great leap forward in human evolution (Chomsky 2010, Bolhuis et al. 2014, Berwick and Chomsky 2015). Thus, merge as a recursive operation is unique to humans, but universal within the species, providing us all with linguistic structures that can be complex and unbounded in length, and might as well be responsible for some of our high level cognitive abilities, such as theory of mind, the powerful skill of simulating the mind of others (de Villiers, 2007).

Everett (2005) understood the term recursion as synonymous of representations with self-embedded constituents (e.g. sentences within sentences, nominal expressions within nominal expression etc.),³ and claimed that Pirahã is a non-recursive language, disallowing self-embedding altogether.

³ The term recursion has been used in different ways. First, it can be easily confused with iteration, which also gives us infinity. More elaborated analyses follow a line of reasoning stemming from mathematics, understanding recursion as induction, frequently exemplified with the successor function used in Peano's axioms to define the natural numbers ((i) 1 is in \mathbf{N} , (ii) If n is in \mathbf{N} , then $(n+1)$ is in \mathbf{N}). The successor

This grammatical gap was explained by Everett as the result of a highly constrained culture, which was resumed as being a “here and now” way of living. The author, thus, concluded that Pirahã constitutes counter-evidence against the universality of the NFL defined by HCF solely in terms of recursion. But clearly, Everett’s argumentation does not hit the nail on the head, as HCF were not talking about self-embedding, but rather about Merge as a generalized transformation mechanism that intermingles with lexical insertion, feature valuation and dislocation. Nevins, Pesetsky and Rodrigues (2009), in a reply to Everett, presented a triple contestation of Everett’s claims: First, they correctly pointed out Everett’s misunderstanding around the term recursion. Second, they observed that many of the constraints on self-embedding noticed by Everett are also true of other languages (e.g. ban on relatives), and, of course, it is not clear that these other languages are subject to the same cultural constraint Everett claimed for Pirahã. Third, they argued that data previously published by Everett do not support the conclusion that self-embedding is unavailable in Pirahã.

More recently, Futrell et al. (2016) did a computer analysis of a syntactic corpus of Pirahã, which were composed by stories originally collected and translated into English by Daniel Everett and Steve Sheldon. These stories/texts were broken into sentences (forming a total of 1.149 sentences) and each sentence was parsed and searched for self-embedding structures. The conclusion of their work is that there is not undisputed

function is understood to be recursive because its induction step ((ii)) is a procedure that takes its own output as its input, creating a loop that can go on forever (see Tomalin 2007, Di Sciullo & D. Isac 2008, Corballis 2011). However, recursion can also be defined in terms of computability, where a recursive procedure is one that builds up upon a hierarchy of deferred operations, accumulating unfinished objects. This forces the computation to be carried out in a given order. To exemplify this, consider the natural number 4 (four). At a certain point of the computation, the procedure used to generate 4 reaches the following derivational step $((((1)+1)+1)+1)$, which is a hierarchy of unfinished/open computations, and these computations have to be closed in a given order, with the procedure moving from the most inner layers to the outer ones. As pointed out in Epstein and Hornstein (2005) and Lobina (2010), recursion in terms of computability might align better with the minimalist notion of generalized transformation. In the present paper, we will not consider the multiple definitions of recursion any further (Rodrigues forthcoming, for a more detailed discussion). For our present purpose, distinguishing self-embedding representations from recursive procedures is sufficient. Syntactic representations containing self-embedded constituents are recursive only in the sense that they are generated by a recursive procedure. These representations are, therefore, one (but not the only one) possible outcome of application of a recursive procedure, which is taken to be the operation Merge within the minimalist program. As pointed out by Pinker and Jackendoff (2005), within linguistics, we must tease apart recursive structures from recursive processes.

evidence for self-embedding in Pirahã, as the corpus does not contain any morpheme that undisputedly marks self-embedding.

The main criticism to this research is that it assumes that the combinatorial properties of I-language are to be evident in E-language. That is, it starts with the assumption that all languages are morpho-phonologically transparent with respect to this underlying syntactic structure. However, this assumption is not warranted. For example, in languages like Portuguese and English, scope interactions among quantified nominal expressions may affect the logic representation of a sentence without modifying its morpho-phonological representation. Therefore, a linguist may not find evidence for this type of interaction by searching at the E-language level (e.g. corpora searching). Finding evidence for this type of interaction requires a step towards abstraction, as they are to be found on the interface between syntax and semantics. Thus, in general, when trying to uncover the underlying syntactic structure of a given language, we are expected to go beyond E-language, considering also processes that might be observed only at the I-language level. Hence, Futrell et al.'s computer analysis informs us only that Pirahã does not seem to have overt morphemes used exclusively to mark self-embedding in .⁴

In what follows, we will discuss new pieces of evidence for self-embedding within VPs and PPs in Pirahã. The evidence was found in the interface syntax-semantics, when we considered interpretative dependencies among constituents.⁵

2. Obligatory control

Most of the languages we know well exhibit what is called Control (Rosenbaum 1967, Postal 1970, 1974, Hornstein (2001), Landau 2001, Rodrigues 2004). Control can be obligatory (3) and non-obligatory (4).

- (3) John wants to leave the meeting right now
- (4) Leaving the meeting right now would be a disaster

⁴ But see Sauerland (2010) for evidence that Pirahã might mark subordination through tonal variation.

⁵ For time reasons, self-embedding within DPs will not be discussed here. See Salles (2015) for new data from fieldwork on possessive DPs.

Since these sentences involve two predicates with independent external theta-role, control is standardly analyzed as involving a null pronoun (PRO) in the subject position of the non-finite predicate:

- (3') John wants PRO to leave right now
- (4') PRO leaving the meeting right now would be a disaster

However, obligatory controlled and non-obligatory controlled PRO behave quite differently in the syntactic-semantic interface: while Obligatory controlled PRO is subject to principle A of Binding Theory, behaving like an anaphor, non-obligatory controlled PRO behave like a pronoun, obeying, thus, Principle B of the Binding theory (Chomsky, 1981, 1986). Thus, in obligatory control, differently from non-obligatory control, there is an interpretative dependency between PRO and the matrix subject.⁶ Given that anaphors are licensed only in structural configurations in which they are bound by a c-commanding antecedent (in accordance with Principle A), it follows that in sentences like (3) PRO must be c-commanded by *John*, the antecedent. Therefore, control is informative about how meaning is composed by the grammar, as the denotation of PRO is syntactic constructed. It is also informative about the syntactic structure itself. In obligatory control, the non-finite predicate must be embedded within the matrix VP, so that PRO can be bound by *John*. That is, the two predicates in (3) are not paratactically conjoined. If they were, PRO would not be controlled by the matrix subject. That is, obligatory control involves predicate self-embedding, predicate within predicate. For this reason, obligatory control can be used as a diagnostic for structures with self-embedded predicates. Knowing that, Rodrigues et al. (forthcoming) did fieldwork in Pirahã looking for obligatory control. The data we found show that obligatory control occurs in Pirahã and can optionally trigger word order changes, which makes evident at that the controlled predicate is subordinated under the matrix VP.

⁶ There is also an interpretative dependency in terms of tenses. The tense of the non-finite clause is interpreted in function with the matrix tense (Wurmbrand, 2001).

Everett (1986, 1991) describes Pirahã as an SOV language, which can turn into SVO if the object is a heavy constituent, as shown in (5) and (6). In (5), the object is a light DP and the SOV word order is maintained. In (6), there is a heavy DP in object position, which triggers SVO order. [Data from Everett, 1986:202]

(5) ti xíbogi ti-baí

I milk drink-INTNSF

‘I really drink milk’

(6) tiobáhai koho- ái hiab -a tomati gihió-

child eat- ATELIC- NEG- REMOTE tomato bean

kasí páiai táí páiai

name also leaf also

‘(The) children do not eat tomatoes or beans or leaf(y vegetables)’

This change in word order caused by the heaviness of the object is commonly called heavy-NP shift (Ross, 1967), a syntactic transformation that shifts the position of the object to the end of the sentences due to phrasal phonological constraints,⁷ and it occurs in many languages, including Portuguese: [Data from Brazilian Portuguese]

(7) a. Eu dei aquela camiseta para o Arturo

I gave that t-shirt to the Arturo

⁷ Heavy-NP shift might be the result of stylistic rules (Ross, 1967), but it has been studied as a syntactic phenomenon as well. Pesetsky (1995) analyzes it as rightward adjunction to the VP. In contrast, Larson (1998) proposes that it results from leftward movement of everything inside the VP except a heavy-NP. Kayne (1994), following Larson’s analysis, adds that light-NPs, as opposed to heavy-NPs, move leftwards to a higher specifier position. For the purpose of the present paper, any of these analyses can be adopted.

- b. Eu dei para o Arturo aquela camiseta dos Beattles que ele tinha me
 I gave to the Arturo that t-shirt of-the Beattles that he had me
 pedido
 asked
 ‘I gave Arturo the that Beattles T-shirt he had asked me for’

Given that it also occurs in Pirahã, it might as well be responsible for the shift from SOV to SVO that Everett (2005) observed in sentences containing sentential objects, such as (8):

- (8) kohoibiihai hi gái- sai hi hi xogi-hiab- iig- á gáih
 Kohoibiihai 3 say-NOMLZR 3 3 want-NEG- CONT-REMOTE that
 ‘Kohoibiihai said (that) he's not wanting that’

This analysis was not considered by Everett (2005). Instead, he suggests an analysis of (8) as a case of parataxis, according to which the second sentence is juxtaposed to the first one. Thus, he argues, the SVO order observed in complex sentences is positive evidence for the lack of self-embedding in Pirahã: if the second sentence were embedded under the first one, the expected word order would be SOV in (8). However, this conclusion seems a bit premature, as questions about a possible shift caused by the heaviness of the sentential complement was not even raised. Actually, this is not difficult to test. If Pirahã sentential complements cause a word order shift due to its phonological heaviness, we predict that light embedded sentences will not cause this type of shift, maintaining the canonical SOV order. This is exactly what is reported in Rodrigues et al. (forthcoming) for sentences displaying obligatory control. In obligatory control, both SVO and SOV seem to be possible. [Data from Rodrigues et al., forthcoming]

- (9) a. ti ogabagai kapiiga kagakai [SVO]
 I want paper study
 ‘I want to study’
- b. ti ogabagai tiisi ikohaipiha
 I want fish eat
 ‘I want to eat fish’
- (10) a. ti kapiiga kagakai ogabagai [SOV]
 I paper study want
 ‘I want to study’
- b. ti tiisi ikohaipiha ogabagai
 I fish eat want
 ‘I want to eat fish’

In (9), as well as in (10), the dependent predicates (*paper study* and *fish eat*) are semantically understood as the complement of the desiderative verb (*want*) and these predicates assign an external theta-role to an unpronounced external argument, which is interpreted as obligatory coreferent with matrix subject (*I*). Hence, these are *bona fide* cases of obligatory control, and, in harmony to what happens in other language, in both (9)-(10) an interpretative dependency is observed, as the subject of the lower predicate is interpreted in function with the matrix subject. Therefore, given what we know about this type of dependency, we should analyze (9) and (10) under the assumption that the second predicate is embedded predicate is embedded under the first one.

The optional word order in (9)-(10) is also crucial to our present discussion. It shows that there is a clear alternative to Everett’s account of (8). Sentential complements might trigger a SOV-to-SVO shift just because sentences are heavy constituents. In order to justify an analysis of (10) as containing two syntactically independent, juxtaposed sentences, one has to unorthodoxly claim that these sentences involve three syntactically independent pieces, as in sketched in (11). Note that an analysis like (11) is not empirically justified and it is actually a buck-passing-game because at some linguistic level (probably at the semantic-pragmatic component) the three pieces in (11) has to be stitched together such that the pronoun *I* is interpreted as the subject of the desiderative predicate *want* and the predicate *paper study* is

understood as the complement of want.⁸ In other words, saying that Pirahã is a non self-embedding language is somehow misleading, as it does not tell us how the semantics of the sentences works.

- (11) [ti] [kapiiga kagakai] [ogabagai]
 I paper study want

As discussed in Rodrigues et al., one possible counter-argument to the line of reasoning we're developing here is that we do not really know the morphological status of what we are calling dependent predicates in (9) and (10). These can actually be nominalized predicates, resulting from some morphological process of compounding. If so, *kapiiga kagakai* and *tiisi ikohaipiha* would better be translated into English as nominal: *paper-studying* and *fish-eating*, respectively, and, therefore, neither (9) nor (10), are structures containing self-embedding. Be that as it may,⁹ other pieces of data suggest that an unified analysis of sentential complements as nominal compounds cannot be easily attained. Consider, for example, the sentence in (12).

- (12) ti kapiiga kagakai sogabagai Kapoogo
 I paper study want Kapoogo
 'I want Kapoogo to study'

In (12), the DP *Kapoogo* is the subject of *kapiiga Kagakai* 'paper study'. Thus, (12) shows that desiderative verbs can take non-controlled, full sentences as its complement. Since Pirahã does not morphologically marks Case, we do not know yet whether (12) is an exceptional case marking configuration or not, but, it is evidence that

⁸ According to Rodrigues et al.'s informant (Iapohen Pirahã - a native speaker), the sentence in (i) is not acceptable in Pirahã. Hence in (9)-(11), the desiderative verb S-selects the dependent predicate.

- (i) * ti ogabagai
 I want

⁹ To evaluate this alternative analysis for (9) and (10), we would need to collect in Pirahã data like (i), in which the complement of the controlled predicate is a complex DPs, containing another predicate.

- (i) I want to eat the fish that you cooked yesterday.

its embedded predicate may not be a nominalized predicate. An important syntactic property of (12) is that only the embedded VP (*kapiiga kagakai*) can appear between the matrix subject and the matrix VP, the embedded subject (*Kapoogo*), must be occur after the matrix verb, as the unacceptability of (13) indicates. [Datum from Rodrigues et al., forthcoming]

- (13) *ti Kapoogo kapiiga kagakai sogabagai
I Kapoogo paper study want

The contrast between (12) and (13) support the conclusion that only light verbal complements can appear between the subject and the verb, triggering the SOV order. When the verbal complement is a full CP or TP, as in (8) and (12), it is too heavy to appear before the verb. Thus, it is either spell-out after the verb or will be a split constituent, as in (12).

Another strong piece of evidence in favor of self-embedding in Pirahã is (14), in which the complement of the matrix desiderative verb is another desiderative verb that, in its turn, take the non-finite predicate *kapiiga kagakai* as its complement. That is, multiple self-embedding seems possible in Pirahã. [Datum from Rodrigues et al., forthcoming]

- (14) ti kapiiga kagakai ogabagai sogabagai
I paper study want would-like
'I would like to want to study'

Given (14), there is no alternative, but accepting the availability of self-embedding in Pirahã. An analysis of obligatory controlled embedded predicates as nominalizations would lead us to the conclusion that there are two nominizations in (14), one embedded within the other!

In sum, Everett's claim that the SVO order observed in sentences with sentential complements is positive evidence for the ban on self-embedding in Pirahã makes the wrong predication with respect to control configurations. In obligatory control, both SVO and SOV order are possible. The licensing of SOV order in these configurations provides us with evidence that the controlled predicates is syntactic and semantically embedded within the matrix predicate. This is in accordance with the interpretative dependencies observed in obligatory control.

3. Prepositional phrases

If we do a *picture matching* experiment with native speakers of English and Portuguese, in which the task is to point to one of the pictures below (figure 1) after hearing a sentence, we expect them to point to the second picture of the second row after hearing (15). In contrast, if they hear (16), we foresee that they will point to the first picture in the second row.

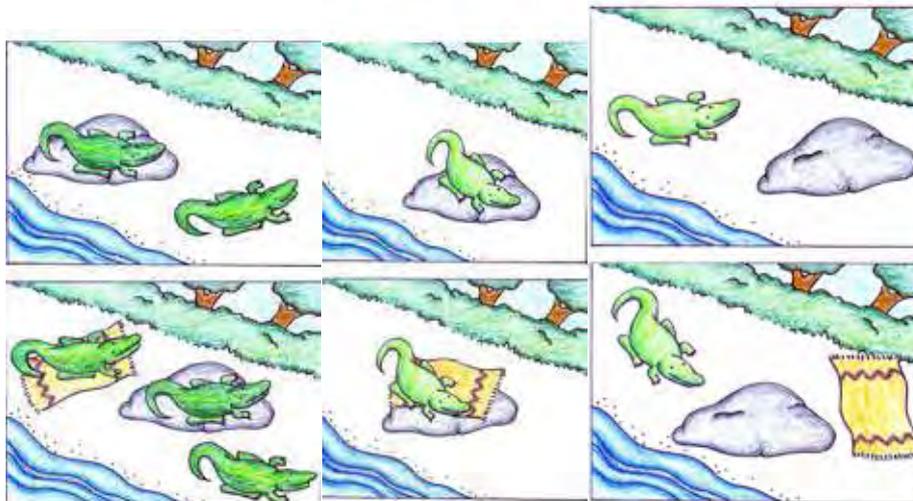


Figure 1: From Sândalo et al. (forthcoming)

- (15) a. Alligator on the mat on the stone on the sand
 b. Jacaré no tapete na pedra na areia
- (16) a. Alligator on the mat, on the stone, on the sand
 b. Jacaré na tapete, na pedra, na areia

In (15), we have a cascade of self-embedded prepositional phrases. Semantically, it means that there is an interpretative dependency, according to which the reference of each DP node is defined *top-down*,¹⁰ the reference of *alligator/jacaré* is build upon the reference of *mat/tapete*, whose reference is defined based on the reference of *stone/pedra*, whose reference is defined upon the reference of *sand/areia*. These reference dependencies are a reflex of the underlying syntactic structure of (15), in which the DP *alligator/jacaré* contains the PP *in the mat/no tapete*, the DP *mat/tapete* contains the PP *in the stone/na pedra*, and the DP *stone/Pedra* contains the PP *na areia*.¹¹ This is different from what we have in (16), which is a coordination of prepositional phrases, and the reference of each DP is defined independently.

The difference between (15) and (16) is not only in syntax and semantics, but also in processing. Maia (forthcoming) observes that coordinations like (16) are more accessible default forms. However, the early spontaneous use of sentences like (15) in acquisition suggests that these structures are available in children and adult grammar. Also, the acceptability of (17) in Kaingang, a native language spoken in the south of Brazil, indicates that structures with multiple self-embedded PPs are available in many languages, despite their processing issues. [Datum from Sândalo et al, forthcoming]

- (17) Kākénh tá runja kãki lata ki krêkufár vyn kÿ pó ki
 canoe on bucket inside can in fish grab thn rock in
 krêkufár rē fi
 fish near put
 ‘Grab the fish in a can inside a bucket in the canoe then put (it) near the fish in
 the rock’

If we accept the ban on self-embedding in Pirahã, we predict that native speakers of this language do not license interpretative dependency of this sort, allowing only coordination readings. This prediction does not born out, however. Sândalo et al.

¹⁰ Notice that these semantic building blocks can also be assembled bottom-up as in (i). However, for some speakers, (i) harder than (14) to process as a self-embedding structure.

- (i) a. Alligator in the sand in the stone in the mat
 b. Jacaré na areia na pedra no tapete

¹¹ See Perez (forthcoming) for discussion on interpretative dependencies in structures with self-embedding.

(forthcoming) did a picture matching pilot experiment using the same set of the pictures in figure 1. The pilot was done with a young native monolingual speaker of Pirahã called *Ioá Pirahã*.¹² The results demonstrate that at least this speaker has no problem in producing and interpreting sentences involving multiple self-embedded PPs.

In the first part of the experiment, each individual picture of figure 1 was introduced to the speaker, and he was asked to describe it. As a result, sentences like (18) were elicited. Notice that the sentences in (18) are potentially ambiguous, describing either pictures with coordination of entities (i.e. pictures 1 of the first and second rows) and pictures with interpretative dependencies (i.e. pictures 2 of the first and second rows).

- (18) a. koxoahai bege apo xaxai apo
alligator floor on stone on
- b. koxoahai bege apo xaxai apo tahoasi apo
alligator floor on stone on mat on

In the second part of the experiment, we reversed the roles. One of the experiments pronounced the elicited sentences out loud and the participant had to point to the picture that matched the sentence he heard.

The result of the pilot suggests that speakers of Pirahã are able to produce, process, comprehend and differentiate ambiguous prepositional phrases. Ioá consistently paired sentences like (18) with the pictures involving interpretative dependencies, as illustrated in (19). [Data from Sândalo et al, forthcoming]

¹² In this experimental task, we were helped by Augusto Diarroi, a non-native speaker of Pirahã, who has some knowledge of the language. Augusto's father is Pirahã. Both Augusto and Ioá live in Pequiá, a Pirahã village, located in the district of Humaitá/Amazon.

(19)

<u>Paired picture</u>	<u>Target sentence</u>
	→ koxoahai bege apo xaxai apo alligator floor on stone on 'Alligator on the stone on the floor'
	→ koxoahai bege apo xaxai apo tahoasi apo alligator floor on stone on mat on 'Alligator on the mat on the stone on the floor'

Importantly, as shown in (20), in the second part of the experiment, when asked to describe the pictures the experimenter pointing to, Ioá modified the sentences in (18), introducing *piai* a coordinative particle translated as *also* by Everett (1990))¹³ every time the experimenter pointed to a picture involving coordinated entities. [Data from Sândalo et al., forthcoming]

(20)

<u>Paired Picture</u>	<u>Target sentence</u>
	→ koxoahai bege apo xaxai apo piai alligator floor on stone on also 'Alligator on the floor (and) on the stone also'
	→ koxoahai bege apo xaxai apo (piai) tahoasi apo piai alligator floor on stone on also mat on also 'Alligator on the floor, on the stone, (and) on the mat also'

¹³ 'pí(x)ái conjunction. Free form. And, to join or add, also' (Everett, 1990:57)

These results clearly show that Pirahã speakers are capable of differentiating structures with coordination from structures with semantic dependencies. Thus, it might be that the sentences in (18) not ambiguous, allowing only readings with interpretative dependencies. At any rate, we can concluded from this results that structures containing self-embedded PPs are available in Pirahã; speakers have at least preference for treating (18) as containing self-embedded PPs; and a possible structural ambiguity is resolved by inserting an overt coordinator *piai*, forcing a coordinative reading.

4. Conclusion: *absence of evidence is not the same as evidence of absence*

Everett's (2005) arguments for disputing the universality of *merge*, a syntactic mechanism that builds representations recursively (Hauser, Chomsky & Fitch, 2002), illustrate a misunderstanding, both theoretically and empirically. First, Everett's definition of recursion as synonymous of self-embedding is a simplification. Given the technical definition of recursion, a language might have a recursive combinatorial system, although it may disallow self-embedding altogether. In addition, as we have shown above, the fact that Everett, as well as Futrell et al (2016), did not find any unambiguous occurrence of self-embedding in the data they analyzed does not lead to the conclusion that Pirahã is non-self-embedding language. The assumption that all the abstract properties of I-language have to be transparently manifested in the corresponding E-language is not warranted. An utterance might not easily reveal properties of its correspondent underlying syntactic structure. As Di Sciullo (2002) puts it, E-complexity is not a reliable way of measuring I-complexity. When we probed for structures with interpretative dependencies, such as controlled sentential complements and cascades of prepositional phrases, Pirahã speakers provides clear evidence that their I-language is not different in terms of recursion (and self-embedding). Therefore, what Pirahã shows us is that the aphorism *absence of evidence is not evidence of absence* is a general truth.

References

- BERWICK, R. and N. CHOMSKY. 2015. *Why only us: language and evolution*. Cambridge, MA: The MIT Press.

- BOLHUIS, J., I. TATTERSALL, N. CHOMSKY, R. BERWICK. 2014. How could language have evolved? *PLoS Biol* (12)(8).
- CAMARGOS, Q. (This volume) *Exploring agreement displacement from the IA to the EA in the Tenetehára language (Tupí-Guaraní family)*
- CHOMSKY, N. 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.
- CHOMSKY, N. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger Scientific.
- CHOMSKY, N. 2010. Some simple evo devo theses: how true might they be for language? In R. K. LARSON, V. DEPREZ, H. YAMAKIDO (eds) *The Evolution of Human Language*. Cambridge: Cambridge University Press.
- CORBALLIS, M. C. 2011. *The Recursive Mind: The Origins of Human Language, Thought, and Civilization*. Princeton, New Jersey: Princeton University Press.
- DE VILLIERS, J. (2007). The interface of language and theory of mind. *Lingua* 117(11): 1858-1878.
- DI SCIULLO, A. M. and D. ISAC. 2008. The asymmetry of merge. *Biolinguistics* 2(4): 260-290.
- EPSTEIN, S. and N. HORNSTEIN (2005) Letter on ‘the future of language’. *Language* 81: 3-6
- EVERETT, D. L. 1986. Pirahã. In Desmond C. Derbyshire and G. K. Pullum (eds.) *Handbook of Amazonian languages*. Berlin: Mouton de Gruyter. 200–326
- EVERETT, D. L. 1991. *A língua Pirahã e a teoria da sintaxe: descrição, perspectivas e teoria*. Campinas, São Paulo: Editora da UNICAMP.
- EVERETT, D. L. 2005. Cultural constraints on grammar and cognition in Piraha: another look at the design features of human language. *Current Anthropology* 46. 621-646.
- FITCH, W. T. 2010. *The evolution of language*. Cambridge, Cambridge University Press.
- FUTRELL, R., L. STEARNS, D. L. EVERETT, S. T. PIANTADOSI and E. GIBSON. 2016. A corpus investigation of syntactic embedding in Pirahã. *PLoS ONE* 11(3).
- HAUSER, M. D, N. CHOMSKY, T. FITCH. 2002. The faculty of language. What is it, who has it, and how did it evolve. *Science, New Series*. 298; 5598: 1569-1579.

- HORNSTEIN, N. 2001. *Move! A Minimalist theory of construal: Generative syntax*. United Kingdom: Blackwell.
- KAYNE, R. S. 1994. *The antisymmetry of syntax*. Cambridge, MA: The MIT Press.
- LANDAU, I. 2001. *Elements of control: Structure and meaning in infinitival constructions*. Dordrecht: Kluwer
- LARSON, R. K. 1988. On the double object construction. *Linguistic Inquiry* 19: 335-91.
- LOBINA, D. J. 2010. Recursion and Linguistics: an addendum to Marcus Tomalin's Reconsidering Recursion in Syntactic Theory. *Interlingüística XX*.
- PINKER, S. and R. Jackendoff. 2005. The faculty of language: what's special about it? *Cognition*, 95: 201-236.
- MAIA, M. (forthcoming) A computational efficiency principle in action in the processing of recursively embedded PPs in Brazilian Portuguese and in Karajá. In L. AMARAL, M. MAIA, A. NEVINS, T. ROEPER (eds) *Recursion in Brazilian languages and beyond*. Cambridge: Cambridge University Press.
- NEVINS, A., D. PESETSKY and C. RODRIGUES. 2009. Pirahã exceptionality: a reassessment. *Language* 85. 355-404.
- PESETSKY, D. 1995. *Zero syntax: experiencers and cascades*. Cambridge, MA: The MIT Press.
- PINKER, S. and R. JAKENDOFF. 2005. The faculty of language: what's special about it? *Cognition* 95: 201-236.
- POSTAL, P. 1974. *On Raising*. Cambridge, MA: The MIT Press
- PREMACK, D. 2007. Human and animal cognition: continuity and discontinuity. *PNAS - Preceedings of the National Academy of Science of the United States*. 28; 104(35): 13861-13867.
- RODRIGUES, C. 2004. *Impoverished morphology and A-movement out of Case domains*. PhD dissertation, University of Maryland, College Park.
- RODRIGUES, C. (forthcoming) *Recursão: O que é e quem tem*. Manuscrito, PUC-Rio.
- RODRIGUES, C., R. SALLES and F. SANDALO (forthcoming). Word order in obligatory control: evidence for self-embedding in Pirahã. In L. AMARAL, M.

- MAIA, A. NEVINS, T. ROEPER (eds) *Recursion in Brazilian languages and beyond*. Cambridge: Cambridge University Press.
- ROSENBAUM, P. S. 1967. *The grammar of English predicated complement constructions*. Cambridge, MA: The MIT Press.
- ROSS, J. 1967. *Constraints on variables in syntax*. PhD dissertation, MIT.
- SANDALO, F., C. RODRIGUES, L. AMARAL, M. MAIA, T. ROEPER. A Pilot Study on Pirahã Postpositional Phrases. In L. AMARAL, M. MAIA, A. NEVINS, T. ROEPER (eds) *Recursion in Brazilian languages and beyond*. Cambridge: Cambridge University Press.
- SAUERLAND, U. (Forthcoming). False speech reports in Pirahã. a comprehension experiment. In L. AMARAL, M. MAIA, A. NEVINS, T. ROEPER (eds) *Recursion in Brazilian languages and beyond*. Cambridge: Cambridge University Press.
- TOMALIN, M. 2007. Reconsidering recursion in syntactic theory. *Lingua* 117, 1784-1800.
- SALLES, R. 2015. *Understanding recursion and looking for self-embedding in Pirahã: the case of possessive constructions*. MA thesis, PUC-Rio.
- WURMBRAND, S. 2001. *Infinitives: restructuring and clausal structure*. Berlin: Mouton the Gruyter.

Exploring Agreement Displacement from the Internal to the External Argument in the Tenetehára Language (Tupí-Guaraní Family)

Quesler Fagundes Camargos^a

ABSTRACT

This article aims to describe and examine the verbal agreement system in the Tenetehára language (of the Tupí-Guaraní linguistic family). We assume the hypothesis that the agreement displacement phenomena – which are sensitive to person hierarchies – come from the mechanism of Agree, that operates on articulated ϕ -feature structures in cyclic syntax (Rezac, 2003; Béjar, 2000ab, 2003; Béjar; Rezac, 2009). We explore such agreement displacement in order to understand its syntactic and morphological character and its parameterization in Tenetehára. The analysis of the target language shows that cyclicity and locality derive a preference for agreement control by the internal argument, rather than by the external. Furthermore, the articulation of the probe derives when the agreement control displaces – in terms of cyclic syntax – to the external argument, which is sensitive to the following person hierarchy: $1 > 2 > 3_{[+foc]} > 3_{[-foc]}$ (Duarte, 2007). In sum, when the resulting syntactic configurations are submitted to Transfer, properties of the morphological component further parameterize the outcome. Thus, the agreement displacement phenomenon in Tenetehára characterizes at least three classes of derivations corresponding to direct, inverse and direct-inverse contexts.

KEYWORDS: Tenetehára (Tupí-Guaraní); agreement displacement; person hierarchy; cyclicity

RESUMO

Este artigo tem o objetivo de descrever e examinar o sistema de concordância verbal na língua Tenetehára (da família linguística Tupí-Guaraní). Assumimos a hipótese de que os fenômenos de deslocamento de concordância – que são sensíveis às hierarquias de pessoa – surgem a partir dos mecanismos da operação Agree, que operam sobre as estruturas articuladas de traço- ϕ na sintaxe cíclica (Rezac, 2003; Béjar, 2000ab, 2003; Béjar; Rezac, 2009). Exploramos este deslocamento de concordância para entender seu caráter sintático e morfológico e sua parametrização em Tenetehára. A análise dessa língua mostra que a ciclicidade e a localidade derivam preferencialmente por uma concordância controlada pelo argumento interno, ao invés do argumento externo. Além disso, a articulação da sonda

^a Assistant Professor of Linguistics at the Department of Intercultural Education of the Federal University of Rondônia (DEINTER/UNIR). PhD in Linguistics from the Federal University of Minas Gerais (POSLIN/UFMG). Member of the *Grupo de Pesquisa em Educação na Amazônia* (GPEA) and of the *Laboratório de Línguas e Culturas Indígenas* (LALIC/UNIR). E-mail: queslerc@gmail.com

I would like to thank two anonymous reviewers from Revista Diadorim for their comments and constructive critiques, that contributed greatly to improving the paper. I would like to extend my sincere thanks to the Tenetehára people for their invaluable assistance with my field research. Special thanks are due to Cintia Guajajára. All remaining errors are, of course, my own. An earlier version of this paper was presented at the Federal University of Minas Gerais in 2015, during the *Encontro Intermediário do Grupo de Trabalho em Teoria da Gramática* (GTTG/ANPOLL).

deriva quando a concordância muda, em termos de sintaxe cíclica, para o argumento externo, o qual é sensível à seguinte hierarquia de pessoa: $1 > 2 > 3_{[+foc]} > 3_{[-foc]}$ (Duarte, 2007). Em suma, quando as configurações sintáticas resultantes são submetidas ao mecanismo de transferência (Transfer), as propriedades do componente morfológico parametrizam o resultado final. Assim, o fenômeno de deslocamento de concordância na língua Tenetehára exibe pelo menos três classes de derivação, as quais correspondem aos contextos direto, inverso e direto-inverso.

PALAVRAS-CHAVE: Tenetehára (Tupí-Guaraní); deslocamento de concordância; hierarquia de pessoa; ciclicidade

Introduction

In the Tenetehára¹ language (of the Tupí-Guaraní linguistic family), a *portmanteau* agreement morpheme is one that tracks features from two nuclear arguments. As the example² in (1) indicates, the agreement morpheme {*uru-*} spells out the first person feature from the subject and the second person feature from the object. It is important to note that this *portmanteau* morpheme is distinct from the agreement morpheme that cross-references the first person subject, as in (2), and also from the morpheme that cross-references the second person object, as it can be seen in (3).

- (1) *uru-pytywà* *ihe*
1SG.2SG-help 1SG
“I helped you”
- (2) *a-pytywà* *Tentehar* *ihe*
1SG-help Tenetehára 1SG
“I helped the Tenetehára”
- (3) *ne-pytywà* *Tentehar* *a’e*
2SG-help Tenetehára 3
“The Tenetehára helped you”

¹ Tenetehára belongs to the Tupí-Guaraní family, Tupí Stock (Rodrigues, 1985). It is located in the northern region of Brazil and spoken by two indigenous groups: the Tembé and the Guajajára (Duarte, 2007). For a detailed analysis of the morphosyntax of Tenetehára, see Duarte (1997, 2003, 2007, 2012), Castro (2007, 2017), and Camargos (2013, 2017).

² The following abbreviations are used in the glosses: 1: first person; 2: second person; 3: third person; DEO: deontic modality; EA: external argument; EM: epistemic modality; EXCL: exclusive; FOC: focus; FUT: future; IA: internal argument; INCL: inclusive; INV: inverse marker; OBL: oblique case; PL: plural; SG: singular; UDPAST: unattested distant past.

The goal of this paper is to answer how and where *portmanteau* agreement is formed in Tenetehára grammar. Furthermore, we intend to answer why and how the verb agrees with the external argument, as in (2), while on the other hand, it agrees with the internal argument, as can be seen in (3).

The paper is organized as follows: section 2 describes the relevant data used to investigate the agreement displacement in Tenetehára. In Section 3, we introduce the basics of the theoretical framework adopted here, exploring in detail the hypothesis that the sensitivity of agreement displacement phenomena to person hierarchies comes from the mechanism of Agree, which operates on articulated ϕ -feature structures in a cyclic syntax (Rezac, 2003; Béjar, 2000ab, 2003; Béjar; Rezac, 2009). Sections 4 and 5 investigate the agreement system in Tenetehára, which, in terms of cyclic syntax, generates three natural classes of derivations for transitive clauses: direct context, inverse context and direct-inverse context. Finally, Section 6 concludes the investigation.

1. The relevant data

This section aims to provide the reader an overview of grammatical facts regarding the agreement displacement phenomena that are sensitive to person hierarchies. First of all, in Tenetehára, subject and object nominal phrases do not exhibit morphological case marking. Moreover, there are three sets of person markers used to encode the syntactic functions carried out by the verbal arguments. The first set corresponds to the so-called direct context, where the external argument controls agreement, as can be seen below:

- (4) a. *a-exak ka'i ka'a r-upi ihe*
 1SG-see monkey forest OBL-in 1SG
 “I saw a monkey in the forest”
- b. *uru-exak ka'i ka'a r-upi ure*
 1EXCL-see monkey forest OBL-in 1EXCL
 “We saw a monkey in the forest”

- c. *xi-exak ka'i ka'a r-upi zane*
 1INCL-see monkey forest OBL-in 1INCL
 “We saw a monkey in the forest”
- d. *ere-exak ka'i ka'a r-upi ne*
 2SG-see monkey forest OBL-in 2SG
 “You saw a monkey in the forest”
- e. *pe-exak ka'i ka'a r-upi pe*
 2PL-see monkey forest OBL-in 2PL
 “You saw a monkey in the forest”
- f. *w-exak ka'i ka'a r-upi a'e (wà)*
 3-see monkey forest OBL-in 3 PL
 “He saw a monkey in the forest”
 “They saw a monkey in the forest”

As indicated in the following examples, the second set of person markers corresponds to the so-called inverse context (Payne, 1994), where the internal argument controls agreement:

- (5) a. *he-r-exak Tentehar a'e*
 1SG-INV-see Tenetehára 3
 “The Tenetehára saw me”
- b. *ure-r-exak Tentehar a'e*
 1EXCL-INV-see Tenetehára 3
 “The Tenetehára saw us”
- c. *zane-r-exak Tentehar a'e*
 1INCL-INV-see Tenetehára 3
 “The Tenetehára saw us”
- d. *ne-r-exak Tentehar a'e*
 2SG-INV-see Tenetehára 3
 “The Tenetehára saw you”
- e. *pe-r-exak Tentehar a'e*
 2PL-INV-see Tenetehára 3
 “The Tenetehára saw you”

- f. *upaw pira Tentehar h-exak a'e (wà)*
 all fish Tenetehára 3-see 3 PL
 “All the fish, the Tenetehára saw it”

In Table 1, we summarize the first and second sets of verbal agreement. Notice that the second column shows the personal pronouns, which occupy the syntactic positions of subject and object. The third column displays the markers that refer to the external arguments. Lastly, the fourth column presents the agreement prefixes that indicate the internal argument.

	Pronouns	First set (EA)	Second set (IA)
1st person, singular	ihe	a-	he-
1st person, exclusive	ure	uru- ~ oro-	ure-
1st person, inclusive	zane	xi-	zane-
2nd person, singular	ne	re-	ne-
2nd person, plural	pe	pe-	pe-
3rd person	a'e (wà)	u- ~ o- ~ w-	i- ~ h-

Table 1. First and second sets of agreement

The third set of person markers corresponds to the *portmanteau* agreement, where the external and the internal arguments control agreement, as can be seen below:

- (6) a. *uru-exak ka'a r-upi ihe*
 1SG.2SG-see forest OBL-in 1SG
 “I saw you_(sg) in the forest”
- b. *uru-exak ka'a r-upi ure*
 1EXCL.2SG-see forest OBL-in 1EXCL
 “We saw you_(sg) in the forest”
- c. *apu-exak ka'a r-upi ihe*
 1SG.2PL-see forest OBL-in 1SG
 “I saw you_(pl) in the forest”

- d. *urupu-exak ka'a r-upi ure*
 1EXCL.2PL-see forest OBL-in 1EXCL
 “We saw you_(pl) in the forest”

Table 2 summarizes the third set of verb agreement.

EA → IA	Third set (<i>portmanteau</i>)
I → you _{SG}	uru-
we _{EXCL} → you _{SG}	uru-
I → you _{PL}	apu-
we _{EXCL} → you _{PL}	urupu-

Table 2. Third set of agreement

2. Cyclic agreement

From a descriptive perspective, Duarte (2007) analyzes agreement in the Tenetehára language using person hierarchies, as in (7).

- (7) 1st person > 2nd person > 3rd person [+FOC] > 3rd person [-FOC]
 (> means “more prominent than”)

As schemed in (7), the choice of which argument will be Agreed with is an independent component of ϕ -agreement. In such approach, ϕ -agreement is treated as a uniform phenomenon that depends on the choice of the target.

Based on Harley & Ritter (2002), Béjar (2003) and Béjar & Rezac (2009), we propose that Tenetehára uses cyclic verbal agreement. That is, agreement takes place in a cyclic way through a list of arguments. Béjar & Rezac (2009, p. 39) claim that “we interpret the core pattern, where IA agreement bleeds EA agreement, to mean that the relevant ϕ -probe is on the v head and so has only the IA in its search space at first”, as can be seen in the morphosyntactic representation in (8).

- (8) [_{VP} F_{EA} [v +AGR [_{VP} V F_{IA}]]]

Following Béjar & Rezac (2009), we assume that the sensitivity of agreement displacement phenomena to person hierarchies is possible because the mechanism of Agree operates on articulated ϕ -feature structures in a cyclic way. According to the authors, the cyclicity and the locality derive a preference for agreement control by the internal argument. Consequently, articulation of the probe determines when the agreement controller cyclically displaces to the external argument. We will see that this system characterizes three classes of derivations corresponding empirically to direct, inverse and direct-inverse contexts.

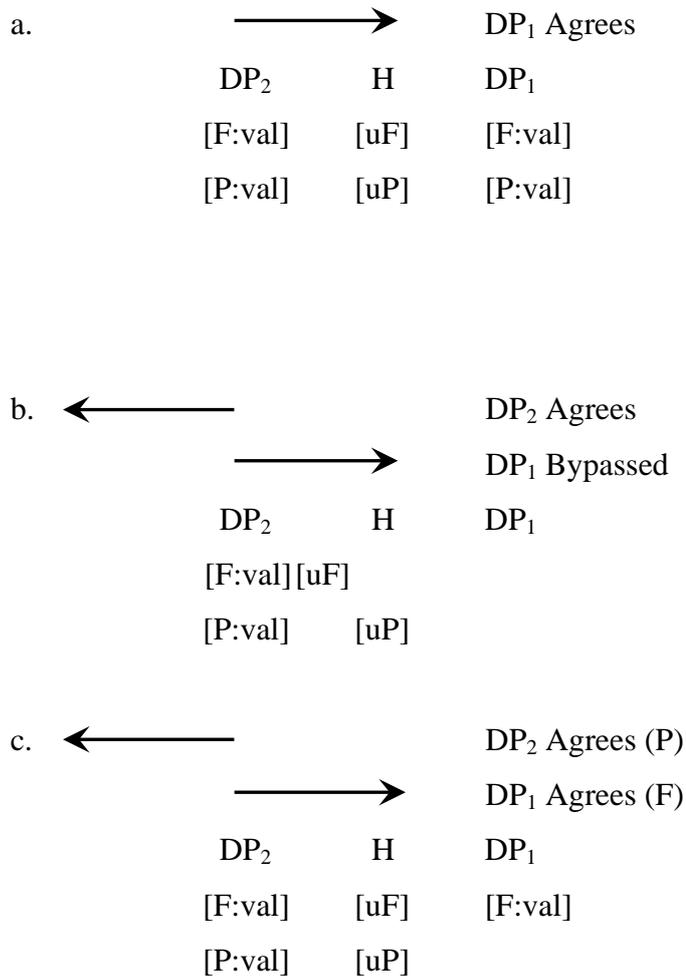
In addition, we will adopt the Béjar & Rezac (2009, p. 47) approach, according to which the ϕ -features permit us “to distinguish individual ϕ -values by representing them as subsets of a single feature structure”. This means that the person hierarchy sensitivity to agreement displacement can be modeled by the following facts:

- (9) a. Matching of a proper subset of the features of a probe by a goal leaves an active residue able to match another goal
- b. Different cross-linguistic person hierarchy sensitivities follow from different articulations of the probe
- (Béjar; Rezac, 2009, p. 47)

According to Béjar & Rezac (2009), the pattern of agreement displacement presents a preference for the internal argument as the controller, which is superseded by an external argument if the internal argument does not suffice to check all segments of a language’s characteristic probe (Rezac, 2003; Béjar, 2000ab, 2003).

The internal argument will fail to Agree for a particular feature [uF] or [uP] of such an articulated ϕ -probe when it lacks a matching [F] or [P]; thereby [F] or [P] on the external argument can then be the goal of Agree. Therefore, the full agreement can be controlled by the internal argument, as in (10a), and bypassed by the internal argument in favor of control by the external argument, as can be seen in (10b). In the last situation, (10c), the agreement on H is controlled by [F] on the internal argument and by [P] on the external argument.

(10) *Cyclic Expansion* (adapted from Béjar; Rezac, 2009, p. 42)



According to Béjar & Rezac (2009, p. 42), “one such system is developed by Harley & Ritter (2002) for morphological ϕ -features, which is extended to the ϕ -features visible to Agree, following Béjar (2000ab, 2003)”. Accordingly, the ϕ -feature bundle is structured into subgroups that include semantic entailment relations and natural classes. Therefore, all persons include some shared features. In addition, first and second persons are specified as discourse participants and so grouped into a natural class. Finally, first and second persons are differentiated from one another by a feature on the first person, distinguishing it as the speaker (Béjar; Rezac, 2009, p. 42).

3. Agreement paradigm shift

Taking into account that “morphological derivations must directly reflect syntactic derivations (and vice versa)”³ (Baker, 1985, p. 375), we assume that the agreement pattern in Tenetehára is characterized by a single-head agreement, which can be controlled by one or two nuclear arguments. Furthermore, spell-out is sensitive to the person feature value on both agreeing arguments, leading to the characterization of such systems as sensitive to person hierarchies. According to Béjar & Rezac (2009, p. 36), the fundamental principles that enter into account are:

- (11) a. Intervener-based locality (Rizzi, 1990), relativized to features (Chomsky, 1995): Agree for a feature [F] is sensitive only to other elements with [F]
 b. A fine-grained approach to cyclicity, where every syntactic operation defines a cycle and thus a potential feeding-bleeding relationship (Rezac, 2003)
 c. A fine-grained approach to ϕ -features (specifically person or ϕ -features), and especially ϕ -probes, associating with each person value (ϕ -value) a different feature structure and thus a different locality class (Béjar, 2003)

In the Tenetehára language, these mechanisms will generate three natural classes of derivations for transitive clauses: direct context, inverse context and direct-inverse context, as we can see below:

Inverse context

- (12) a. *he=r-exak ka'a r-upi a'e* (3 → 1 = 1)
 1SG-INV-see forest OBL-in 3
 “He/she saw me in the forest”
 b. *he=r-exak ka'a r-upi ne* (2 → 1 = 1)
 1SG-INV-see forest OBL-in 2SG
 “You saw me in the forest”

³ According to Baker (1985, 1988), the order of affixes reflects the order in which the associated syntactic operations apply.

- c. *ne=r-exak ka'a r-upi a'e* (3 → 2 = 2)
 2SG-INV-see forest OBL-in 3
 “He/she saw you in the forest”

Direct context

- d. *a-exak ka'a r-upi ihe* (1 → 3 = 1)
 1SG-see forest OBL-in 1SG
 “I saw him/her in the forest”
- e. *ere-exak ka'a r-upi ne* (2 → 3 = 2)
 2SG-see forest OBL-in 2SG
 “You saw him/her in the forest”

Direct-inverse context

- f. *uru-exak ka'a r-upi ihe* (1 → 2 = 2)
 1SG.2SG-see forest OBL-in 1SG
 “I saw you in the forest”
- g. *uru-exak ka'a r-upi ure* (1 → 2 = 2)
 1EXCL.2SG-see forest OBL-in 1EXCL
 “We saw you in the forest”

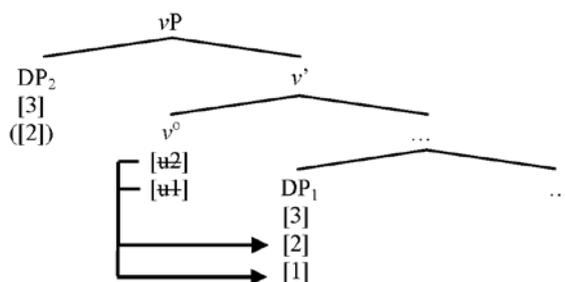
In order to implement the theoretical proposal developed above, we will now show how the cyclic Agree mechanism derives the basic pattern of agreement displacement in Tenetehára in terms of the following person hierarchy: [1>2>3]. To simplify the explanation, only the person feature will be considered. The number feature will be ignored. The relevant data is given in (12). Note that the verbal prefix cross-references the person of the external argument when it is more highly specified than the internal argument.

Let us start with inverse context, in which the internal argument checks all the probe’s features that it can match. In this situation, the core π -probe of v does not Agree with the external argument. See the examples repeated below.

- (13) a. he=*r-exak* *ka'a* *r-upi* *a'e* (3 → 1 = 1)
 1SG-INV-see forest OBL-in 3
 “He/she saw me in the forest”
- b. he=*r-exak* *ka'a* *r-upi* *ne* (2 → 1 = 1)
 1SG-INV-see forest OBL-in 2SG
 “You saw me in the forest”

As the derivation in (14) indicates, the internal argument checks all the segments of the probe in the first cycle. Therefore, the second cycle is totally unnecessary, as the probe cannot Agree with the external argument anymore.

(14) *First Cycle*

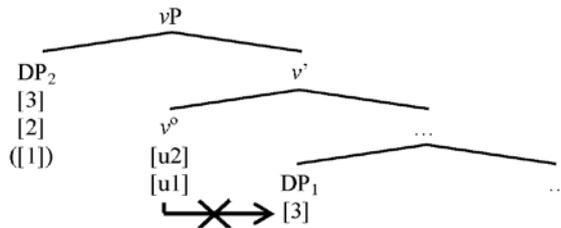


In direct context, the external argument is more highly specified than the internal argument. Hence, after trying, but failing, to Agree with the internal argument, the probe Agrees for its unchecked segments with the external argument. In this situation, the core π -probe of v Agrees only with the external argument, since the segments of the internal argument could not control the agreement. See the examples repeated below.

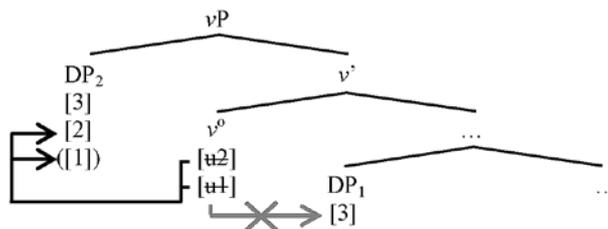
- (15) a. a-*exak* *ka'a* *r-upi* *ihe* (1 → 3 = 1)
 1SG-see forest OBL-in 1SG
 “I saw him/her in the forest”
- b. ere-*exak* *ka'a* *r-upi* *ne* (2 → 3 = 2)
 2SG-see forest OBL-in 2SG
 “You saw him/her in the forest”

As the derivation in (16) exhibits, the internal argument cannot check the segments of the π -probe in the first cycle. For this reason, the π -probe has to be assigned a value on the second cycle (EA > IA), so that it Agrees for its unchecked segments with the external argument. Note that articulation of the probe derives when the agreement control displaces, in terms of cyclic syntax, to the external argument.

(16) a. *First Cycle*



b. *Second Cycle*

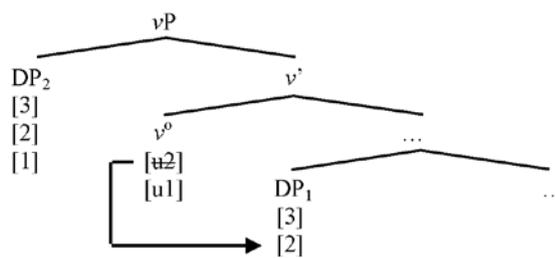


Finally, in direct-inverse context, the external argument is more highly specified than the internal argument. Hence, after the characteristic probe has Agreed as fully as possible with the internal argument, it Agrees for its unchecked segments with the external argument. In this situation, the core π -probe of v Agrees with both the internal argument and the external argument, for different segments, as we can see in the following repeated examples.

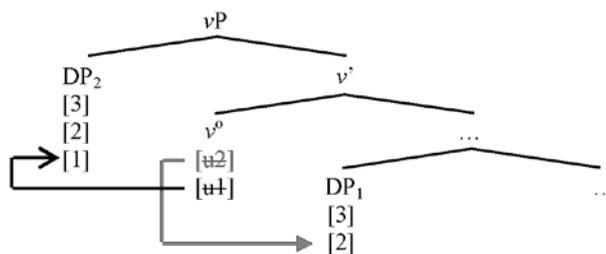
- (17) a. uru-*exak* *ka'a* *r-upi* *ihe* (1 → 2 = 2)
 1SG.2SG-see forest OBL-in 1SG
 “I saw you in the forest”
- b. uru-*exak* *ka'a* *r-upi* *ure* (1 → 2 = 2)
 1EXCL.2SG-see forest OBL-in 1EXCL
 “We saw you in the forest”

As the derivation in (18) reveals, the internal argument partially checks the segments of the π -probe in the first cycle. After that, it Agrees for its unchecked segments with the external argument. As a consequence, the outcome is a *portmanteau* morphology, which arises when features of more than one syntactic terminal (in this case, internal argument and external argument) are spelled-out by a single vocabulary item.

(18) a. *First Cycle*



b. *Second Cycle*



In sum, this analysis of the Tenetehára language demonstrates that cyclicity and locality derive a preference for agreement control by the internal argument. Additionally, articulation of the probe derives when the agreement control displaces, in terms of cyclic syntax, to the external argument, which is sensitive to the following person hierarchy: $1 > 2 > 3_{[+foc]} > 3_{[-foc]}$ (Duarte, 2007).

4. Agreement in the C/TP-domain

It could be proposed that v is responsible for agreement with the internal argument, whereas a higher head, probably T or C, Agrees with the external argument. However, this is not supported by Tenetehára data. The examples given above show that there is an agreement displacement paradigm, suggesting that we are dealing with just one ϕ -probe that oscillates between two controllers. This means that there is just one slot for

agreement. Furthermore, the preference for agreement with the internal argument is evidence that this ϕ -probe has to be low in the structure (i.e. within the ν P shell).

It is also important to observe that the Tenetehára language displays a second agreement slot, which is not a verbal affix (possibly because the verb does not move to TP, according to Duarte, 2012). In terms of ϕ -features, this head can only be controlled by the external argument. We propose that this agreement is in the C/TP-domain because it is next to, for example, the expression of modality and evidentiality, as can be seen below:

- (19) *ne-r-exak rakwez kwarer ka'a r-upi a'e ri'i*
 2SG-INV-see UDPAST boy forest OBL-in 3 EM
 “The boy certainly saw you in the forest”
- (20) **ne-r-exak rakwez kwarer ka'a r-upi ne ri'i*
 2SG-INV-see UDPAST boy forest OBL-in 2SG EM
 “The boy certainly saw you in the forest”

In the C/TP-domain, there is also an agreement in terms of number feature. As the examples demonstrate (21)-(22), the number of the subject is marked at the end of the sentence (the singular is not marked, though). What is surprising is that the head in the C/TP-domain can also Agree, in terms of number feature, with the internal argument, as can be seen in (23), (24) and (25).

- (21) *w-exak kwarer tata a'e wà*
 3-see boy fire 3 PL
 “The boys saw fire”
- (22) *ne-pytywà kwarer a'e wà*
 2SG-help boy 3 PL
 “The boys helped you”
- (23) *w-exak Tukàn kwarer a'e wà*
 3-see Tukàn boy 3 PL
 “Tukàn saw the boys”
- (24) *a-exak ka'i ka'a r-upi ihe wà*
 1SG-see monkey forest OBL-in 1SG PL

- “I saw monkeys in the forest”
- (25) *ere-zuka-putar ka'i ne wà nehe*
 2SG-kill-FUT monkey 2SG PL DEO
 “You will kill the monkeys”

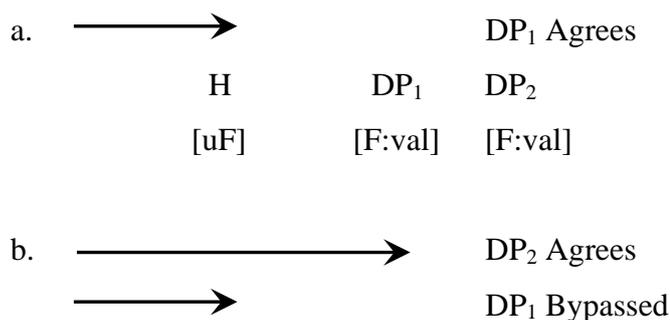
One such system is identified by Harley & Ritter (2002) and Béjar & Rezac (2009) for φ -features, which we extend to the number feature in terms of cyclic agreement. Furthermore, we propose that the number agreement happens in a cyclic way through a list of arguments whose morphosyntactic representation is:

- (26) [CP ... #P+AGR [VP F_{EA} [VP V F_{IA}]]]

In line with Béjar & Rezac (2009), this paper shows that the sensitivity of agreement displacement phenomena to number arises from the mechanism of Agree operating on articulated number feature structures in a cyclic syntax. Additionally, the locality derives a preference for agreement control by the external argument. Accordingly, articulation of the probe determines when the agreement controller cyclically displaces to the internal argument.

The external argument will fail to Agree for a particular feature [uF] of such an articulated number probe when the external argument lacks a matching [F]; thereby [F] on the internal argument can then be the goal of Agree. Therefore, the agreement can be controlled by the external argument (see (27a)) and bypassed by the external argument in favor of control by the internal argument, as in (27b).

- (27) *Cyclic Expansion* (adapted from Béjar; Rezac, 2009, p. 42)



H	DP ₁	DP ₂
[uF]	[F:val]	

5. Final remarks

In the Tenetehára language, a *portmanteau* agreement morpheme is one that codifies features from two nuclear verbal arguments. This *portmanteau* morpheme is distinct from the agreement morpheme that cross-references the subject and from the morpheme that cross-references the object. From a descriptive perspective, Duarte (2009) analyzes agreement in this language using person hierarchies: $1 > 2 > 3_{[+foc]} > 3_{[-foc]}$. In addition, the choice of which argument will be agreed with is an independent component of ϕ -agreement.

Following Béjar & Rezac (2009), we assumed that the sensitivity of agreement displacement phenomena to person hierarchies is possible because the mechanism of Agree operates cyclically on articulated ϕ -feature structures. According to the authors, the fact that the derivation unfolds in cyclic and local fashion derives a preference for agreement control by the internal argument. Accordingly, articulation of the probe determines when the agreement controller cyclically displaces to the external argument. We've seen that this system characterizes three classes of derivations that correspond empirically to direct, inverse and direct-inverse contexts.

References

- BAKER, Mark. "The Mirror Principle and morphosyntactic explanation". *Linguistic Inquiry* 16 (3), pp. 373-415, 1985.
- BAKER, Mark. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press, 1988.
- BÉJAR, Susana. "Economy, cyclicity and markedness in Georgian verbal agreement". *GLOW Newsletter* 44, pp. 18-19, 2000a.
- BÉJAR, Susana. "Structural markedness in formal features: Deriving interpretability". *Revue Québécoise de Linguistique* 28, pp. 47-72, 2000b.

- BÉJAR, Susana. *Phi-syntax: A theory of agreement*. 2003. Tese (Doutorado em Linguística) – University of Toronto, Toronto, 2003.
- BÉJAR, Susana; REZAC, Milan. “Cycle Agree”. *Linguistic Inquiry* 40, pp. 35-73, 2009.
- CAMARGOS, Quesler Fagundes. *Estruturas causativas na língua Tenetehára: uma análise minimalista*. 2013. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2013.
- CAMARGOS, Quesler Fagundes. *Aplicativização, causativização e nominalização: uma análise unificada de estruturas argumentais em Tenetehára-Guajajara (Família Tupí-Guaraní)*. 2017. Tese (Doutorado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2017.
- CASTRO, Ricardo Campos. *Interface morfologia e sintaxe em Tenetehára*. 2007. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2007.
- CASTRO, Ricardo Campos. *Morfossintaxe Tenetehára (Tupí-Guaraní)*. 2017. Tese (Doutorado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2017.
- CHOMSKY, Noam. *The Minimalist Program*. Cambridge, MA: MIT Press, 1995.
- DUARTE, Fábio Bonfim. “Tenetehára: a predicate fronting language”. *Canadian Journal of Linguistics* 57 (3), pp. 359-386, 2012.
- DUARTE, Fábio Bonfim. *Análise gramatical das orações da Língua Tembé*. 1997. Dissertação (Mestrado em Linguística) – Instituto de Letras, UnB, Brasília, 1997.
- DUARTE, Fábio Bonfim. *Estudos de Morfossintaxe Tenetehára*. Belo Horizonte: UFMG, 2007.
- DUARTE, Fábio Bonfim. *Ordem dos constituintes e movimento em Tembé: minimalismo e anti-simetria*. 2003. Tese (Doutorado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2003.

- HARLEY, Heidi; RITTER, Elizabeth. “Person and number in pronouns: A feature-geometric analysis”. *Language* 78, pp. 482-526, 2002.
- PAYNE, Doris L. “The Tupí-Guaraní inverse”. In.: HOPPER, P.; FOX, B. (org.). *Voice: form and function*. Amsterdam: John Benjamins, 1994, pp. 313-340.
- REZAC, Milan. “The fine structure of cyclic Agree”. *Syntax* 6, pp. 156-182, 2003.
- RIZZI, Luigi. *Relativized Minimality*. Cambridge, MA: MIT Press, 1990.
- RODRIGUES, Aryon Dall’Igna. “Relações internas na família lingüística Tupí-Guaraní”. *Revista de Antropologia*, 27/28, pp. 33-53, 1985.

Stative Morpheme in Shimakonde, an anticausative morpheme?

Ronaldo Rodrigues de Paula^a

ABSTRACT

This paper aims to describe the syntax of the constructions that present the verbal extensions {-ik-} and {-uk-} and their allomorphs in Shimakonde, a Bantu language classified as P23 in the Guthrie classification (GUTHRIE 1967-71). This language is spoken in the northern regions of Mozambique and Tanzania. These verbal extensions are reported in literature under the labels of stative, impositive, pseudo-passive, neuter, and quasi-passive (DOKE, 1947; SATYO, 1985; MCHOMBO, 1993; DUBINSKY & SIMANGO, 1996; BENTLEY & KULEMEKA, 2001; LIPHOLA, 2001; NGUNGA, 2004; KHUMALO, 2009; LEACH, 2010; LANGA, 2013). The addition of the {-ik-} or {-uk-} morphemes to the verb structure usually demotes or suppresses the external argument, turning a basically transitive predicate into an intransitive one. This paper aims to investigate in Shimakonde if alternations from a dyadic to a monadic predicate, through the use of one of the aforementioned morphemes, are instances of the phenomenon known in literature as causative/anticausative alternation (HASPELMATH, 1987, 1993; LEVIN & RAPPAPORT HOVAV, 1992, 1995; NAVES, 1998, 2005; VAN HOUT, 2004; OLIVEIRA, 2011; KALLULLI, 2007). In order to do so, I analyze the grammatical role of this morpheme with two Shimakonde native consultants from different Mozambique districts (Mocimboa da Praia and Montepuez). The fieldwork activities consisted of translations of sentences from Portuguese to Shimakonde, testing the grammaticality of the proposed sentences. In order to examine the data that were collected, I adopted the Alexiadou, Anagnostopoulou and Schäfer (2006) refinement of the verbal categories by Levin & Rappaport Hovav (1992, 1995). One of the results obtained is that the verbal extensions display an atelic reading (giving rise to stative interpretation) or a telic reading (giving rise to anticausative or passive interpretation). To account for the different interpretations in these constructions, I propose distinct associations between Asp head and Voice head in accordance with Kratzer (1996), Pylkkänen (2002), van Hout (2004), and Oliveira (2010).

KEYWORDS: Shimakonde; causative/anticausative alternation,; stative morpheme; Voice head; Asp head

^a PhD candidate of Programa de Pós Graduação em Estudos Linguísticos, Faculdade de Letras – Universidade Federal de Minas Gerais – UFMG. Beneficiary of scholarship funded by CAPES – ronaldorodriguesdepaula@gmail.com

RESUMO

Este artigo almeja descrever a sintaxe de sentenças que apresentam as extensões verbais {-ik-} e {-uk-} e seus alomorfes em *Shimakonde*, uma língua banto de código P-23 na classificação de Guthrie (GUTHRIE 1967-71), falada mais proeminentemente ao norte de Moçambique e Tanzânia. Essas extensões verbais são reportadas na literatura sob muitos rótulos, tais como extensão estativa, impositiva, pseudo-passiva, neutra, quasi-passiva (DOKE, 1947; SATYO, 1985; MCHOMBO, 1993; DUBINSKY & SIMANGO, 1996; BENTLEY & KULEMEKA, 2001; LIPHOLA, 2001; NGUNGA, 2004; KHUMALO, 2009; LEACH, 2010; LANGA, 2013). A adição de {-ik-} ou {-uk-} à estrutura verbal geralmente remove ou suprime o argumento externo, transformando um predicado transitivo em intransitivo. O principal objetivo deste artigo é investigar se a alternância de predicados de diádicos para monádicos pelo uso dos morfemas mencionados são instâncias do fenômeno conhecido na literatura como alternância causativa/anticausativa (HASPELMATH, 1987, 1993; LEVIN & RAPPAPORT HOVAV, 1992, 1995; NAVES, 1998, 2005; VAN HOUT, 2004; OLIVEIRA, 2011; KALLULLI, 2007; etc.). Para tanto, analisou-se as principais funções e características deste morfema com dois falantes nativos de *Shimakonde* de diferentes distritos moçambicanos (Mocimboa da Praia e Montepuez). O método utilizado foi trabalho de campo, que consistiu da tradução de sentenças do Português para o *Shimakonde* e pelo teste de gramaticalidade de sentenças sugeridas. Adotou-se o refinamento da proposta de Levin & Rappaport Hovav (1993, 1995) sobre categorias verbais, feito por Alexiadou, Anagnostopoulou and Schäfer (2006) na investigação dos dados coletados. A principal característica da extensão é a habilidade de apresentar tanto uma leitura atélica (gerando uma interpretação estativa) ou télica (gerando uma leitura anticausativa ou passiva). Para dar conta dessas diferentes realizações sintáticas, foram propostas diferentes associações entre os núcleos AspP e VoiceP em termos de Kratzer (1996), Pyllkänen (2002), van Hout (2004) e Oliveira (2010).

PALAVRAS-CHAVE: Shimakonde; alternância causativa/anticausativa; morfema estativo; Núcleo Voice; Núcleo Asp

Introduction

This paper aims to describe the syntactic behavior of constructions that present the verbal extension known as stative {-ik-} and stative-separative {-uk-} (LIPHOLA, 2001; LEACH, 2010) in Shimakonde, a Bantu language spoken in the northern region of Mozambique and Tanzania, and labeled as P23 according to the Guthrie classification (GUTHRIE, 1967:71). These verbal extensions are also reported in literature under many labels, such as impositive, pseudo-passive, neuter-passive, quasi-passive, non-agentive-passive, impositive, and neuter (DOKE, 1947; SATYO, 1985; MCHOMBO, 1993; DUBINSKY & SIMANGO, 1996; BENTLEY & KULEMEKA,

2001; LIPHOLA, 2001; NGUNGA, 2004; KHUMALO, 2009; LEACH, 2010; LANGA, 2013).

The data presented in this paper were collected during fieldwork activities involving two Shimakonde native speakers from different Mozambique districts (Mocimboa da Praia and Montepuez). The methodology consisted of translations of sentences from Portuguese to Shimakonde and testing the grammaticality of the proposed sentences. The addition of this verbal extension to the verbal structure usually suppresses the external argument, making an inherent dyadic predicate turn into a monadic one. Hence, this paper mainly aims to investigate if the verbal extension in question could be used to trigger the verbal valence alternation known in literature as Causative/Anticausative (HASPELMATH, 1987, 1993; LEVIN & RAPPAPORT HOVAV, 1992, 1995; NAVES, 1998, 2005; VAN HOUT, 2004; OLIVEIRA, 2011; KALLULLI, 2007).

This phenomenon is characterized by the expression of a usual transitive verb as intransitive, with the internal argument taking the position of grammatical subject. This can be instantiated by the following examples:

- (1) The boy broke the window/The window broke

Causative/Anticausative alternation is expressed in different ways among the languages of the world. Commonly, one of the alternates is marked in a more morphological manner (HASPELMATH, 1993). Thus, in Shimakonde, the stative morpheme could work as an anticausative morpheme, as the following data suggest:

- (2) a. Mwánà à-ndì-shím-à nnángò
 NC1-child NC1-PERF-close-FV NC3-door¹
 “The child closed the door”
- b. Nnángò ù-ndì-shím-ík-à
 NC3-door NC3-PERF-close-STA-FV
 “The door closed”

¹ Abbreviations: NC = noun class; STA = stative morpheme; FV = final vowel; PASS = passive morpheme; PERF = perfective morpheme; STAS = stative-separative morpheme; SEP = separative morpheme

Note that the internal argument in (2a) is the grammatical subject in (2b) and the verbal extension {-ik-} is added to the construction.

This paper is organized in the following sections: Section 1 explores the main characteristics of the phenomenon reported in the literature; Section 2 describes the main characteristics of the {-ik-} and {-uk-} morphemes in Shimakonde. Studies in other Bantu languages, such as Chichewa (MCHOMBO, 1993; DUBINSKY & SIMANGO, 1996), Swahili (SEIDL & DIMITRIADIS, 2003) and Ndebele (KHUMALO, 2009), were also consulted. Section 3 examines the distinct syntactic structures of the Voice and Asp heads; Section 4 concludes this paper.

1. Causative/Anticausative alternation

The inherent ability of some verbs to switch their valence from transitive to intransitive is known in literature as Causative/Anticausative alternation. A prototypical example of this kind of valence alternation can be illustrated by the verb ‘to break’, as in the following examples:

- (3) a. The boy broke the window
b. The window broke

Studies in many languages (HASPELMATH, 1987, 1993; CROFT, 1990) indicate that there is a basic form and a derived one, which is marked in a more morphological manner, i.e. if the basic form is causative, the anticausative form will be marked, and vice-versa, as exemplified by the following data taken from Haspelmath (1993, p.89).

- (4) a. *Russian: anticausative derived from causative*

Causative:	rasplavit´	
	“melt”	(Transitive)
Anticausative:	rasplavit´-sja	
	“melt”	(Intransitive)

b. (*Khalkha*) *Mongolian: causative derived from anticausative*

Anticausative	xajl-ax	
	“melt”	(Intransitive)
Causative	xajl- uul -ax	
	“melt”	(Transitive)

Observe that the morphologically marked form of the verb ‘to melt’ in Russian is the intransitive form (4a). On the other hand, in Khalkha, the morphologically marked form of the verb ‘to melt’ is the transitive form (4b). This parametric variation seems to demonstrate that there is no specific direction from causative to anticausative or vice-versa. Croft (1990, p. 60) points out that “the more typically the change of state requires an external agent, the more likely the causative type will be unmarked”. In Brazilian Portuguese, according to Cançado & Amaral (2010), the morphological marking of the derived intransitive forms is carried out by the inchoative morpheme ‘*se*’. Verbs of transitive basic use accept the inchoative form with the presence of that morpheme. According to Oliveira (2011), there are verbs in Portuguese, such as ‘*amadurecer*’, that already have an inchoative marking in their root form; in this case, this is expressed by the verbalizer ‘*-ecer*’. This fact then explains why sentence (5b) becomes ungrammatical if the inchoative morpheme ‘*se*’ occurs in the structure. Compare the examples below.

- (5) a. O vaso quebrou-se
 “The pot broke”
- b. *A banana amadureceu-se
 “The banana ripened”

Anticausatives differ from passives in some important ways. According to Kallulli (2007), who examined the properties of passives and anticausatives in English, Albanian and Modern Greek, the passives select an external argument with an agentive theta role (in English introduced by the preposition *by*), whereas the anticausatives select an external argument with a causal theta role (in English preceded by the preposition *from*). This distinction is reinforced by adverbials and complementary clauses where passives and anticausatives can be associated. Passive clauses can co-occur with adverbials that denote agentivity or clauses that denote an idea of purpose, as shown by the following data:

- (6) a. The boat was sunk deliberately
 b. The boat was sunk to collect the insurance

Nevertheless, anticausatives become ungrammatical if adverbs or clauses that imply agentivity are adjoined to the predicate:

- (7) a. *The boat sunk deliberately
 b. *The boat sunk to collect the insurance

Several studies suggest that there is a correlation between anticausatives mapped as unaccusatives (PERLMUTTER, 1978) in syntax with the telic aspect. The telic aspect presents a punctual reading of a finite particular event, that is, a specific culmination point. The atelic aspect, however, does not have a defined point of culmination of the event.

The telicity or atelicity features are determined in a compositional manner, according to VP-specific settings. Consider the examples in (8).

- (8) a. John read the book in two hours

- b. John read the catalog for 5 minutes

Examples (8a) and (8b) are in the Past Perfect Tense, representing events that have already happened. However, (8a) and (8b) differ with respect to the culmination point of the event. In (8a) the event has reached a state of completeness. The construction implies that John finished the book. The predicate in (8a) is telic. Notwithstanding, in (8b), the event does not imply culmination, as John read the catalog for a specific amount of time. At the same time, the sentence itself does not specify the completeness of the event, that is, it does not imply that the entire catalog was read. Hence, (8b) is atelic.

Many researchers claim that there is a link between telicity and direct objects. From Dutch language data, van Hout (2004) proposes that the semantic notion of telicity figures as a syntactic entity. According to the researcher, the correlation between telicity and direct objects in the minimalist program (CHOMSKY, 1995) would be captured by what she calls the event feature check. The telicity would be checked in the direct object position, that is, in Spec-AgrOP. Like other elements within VP that determine the type of verbal predicate event, lexical-syntactic mapping is sensitive to aspectual properties of any VP in which a DP appears. For a telic reading, an object cannot be absent in the sentence. However, it is not only the transitivity expressed in the verbs that is required to encode telicity. The nature of the direct object also plays an important role in the derivation. Following the work of Verkuyl (1972) and Krifka (1989), van Hout argues that only direct objects denoting a specified amount allow telic reading. On the other hand, objects denoting homogeneous mass and indefinite plural terms, for example, do not allow for telicity.

Along these lines, van Hout proposes three interactions between telic or atelic features in dyadic sentences. In relation to the syntactic configuration, the presence of the telic feature can generate direct objects. However, the presence of the atelic feature generates oblique objects. Another distinction is linked to the semantics of the noun phrase in the object position. If the object is quantified, it enables telic reading; if the object is unquantified, it enables atelic reading. If there is no direct object, as in unergative clauses, there is no telicity. If there is an object in a transitive clause, but it is not quantified, telicity will not be possible. Thus, the telicity feature is checked by the

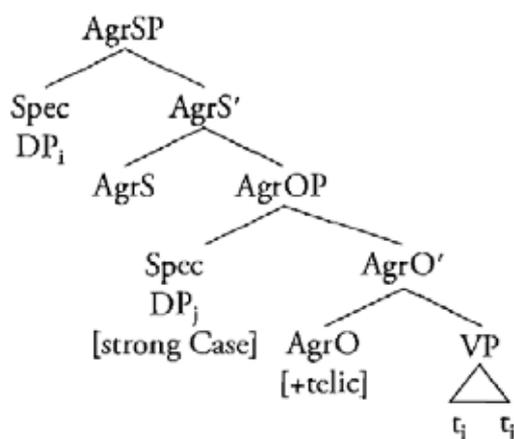
relationship between the object and the head AgrO (Strong Case). In atelic constructions, the objective noun phrase remains within the VP in the basic position of the object (Weak Case) or it is assigned the oblique case.² This is shown in the following examples and their respective syntactic trees (VAN HOUT, 2004. p. 65-68).

(9a) Telic (Strong Case)

Chaartje heeft *urenlang/ in 10 minuten een spekulaasje/ twee spekulaasjes gegeten

Chaartje has *hours-long/ in 10 minutes a ginger-cookie/ two ginger-cookies eaten

“Chaartje ate a ginger cookie/ two ginger cookies *for hours/ in 10 minutes”



(9b) Atelic (Weak Case)

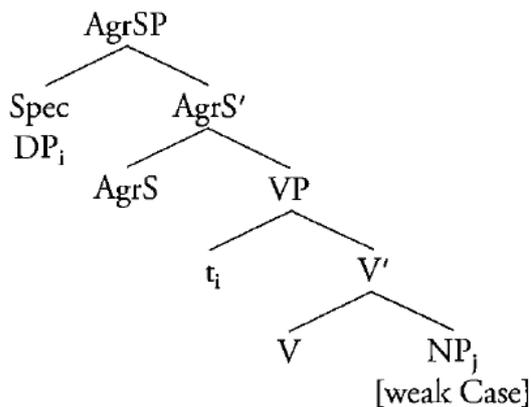
Chaartje heeft urenlang/ *in 10 minuten spekulaas/ spekulaasje gegeten

Chaartje has hours-long/ *in 10 minutes gingerbread/ ginger-cookies eaten

“Chaartje has eaten gingerbread/ ginger cookies for hours/ *in 10 minutes”

² Telicity Checking:

A telic event-type feature is checked via Specifier-Head agreement in AgrOP. It triggers movement of a noun phrase to the Specifier of AgrOP. The predicate's event-type properties must be compatible with the telic feature. (van HOUT, 2004, p.67)

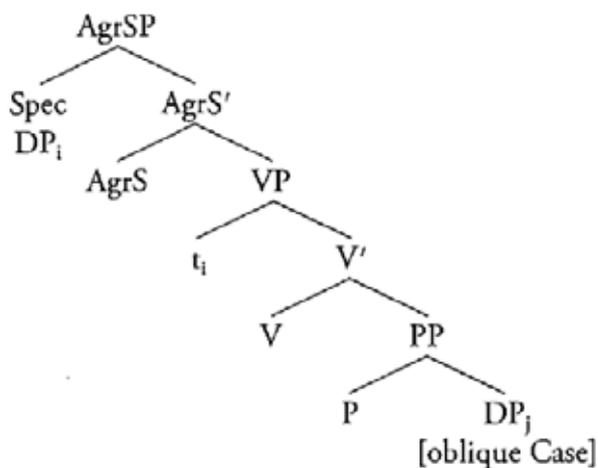


(9c) Atelic (Oblique Case)

Chaartje heeft van het spekulaas gegeten

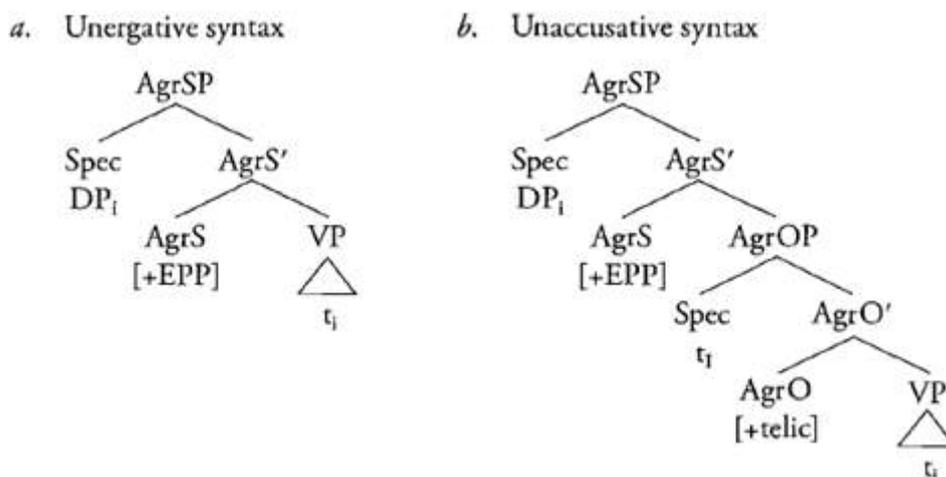
Chaartje has from the ginger-cookie eaten

“Chaartje ate from the ginger-cookie”



The feature checking also explains the difference between unergative and unaccusative verbs. Van Hout claims that monadic predicates become unaccusatives because, after checking their telicity in AgrOP, the objects cannot remain in situ in order to fulfill EPP. In turn, atelic monadic predicates are inherently unergatives, as shown in the syntactic trees below (VAN HOUT, 2004, p. 61):

(10) Syntax of unergative and unaccusative verbs



Oliveira (2011), who studied anticausative constructions in Portuguese, following Hale & Keyser (1993) and Salles (2007), proposes that an aspectual component in V is responsible for the syntactic distinction in double object constructions of verbs of change of location, such as *splash* and *smear*. Compare the following examples (HALE & KEYSER, 1993, apud OLIVEIRA, 2011, p.68-71):

(11a) The pigs splashed mud on the wall / Mud splashed on the wall

(11b) We smeared mud on the wall / *Mud smeared on the wall

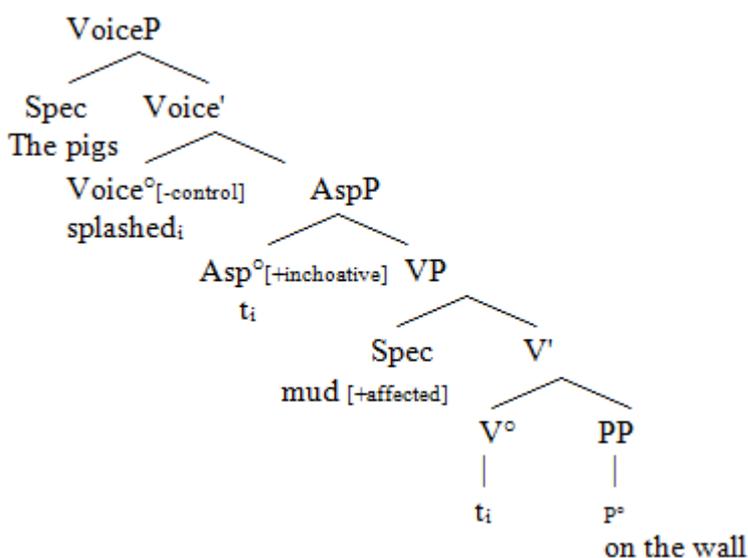
Oliveira claims that the following factors are at play for verbs of change of location to allow for alternation:

- (i) An Asp^o head above VP must be projected
- (ii) This head must carry out the aspectual feature [+inchoative] and the DP projected in [Spec-VP] must show the semantic propriety [+affected]

(iii) The event that holds those aspectual features must be necessarily [+telic] (OLIVEIRA 2011, p.68)³

Oliveira (2011) also claims that *splash* verb type predicates have an external argument with an indirect agent theta role, i.e. without the property [+control], as opposed to the *smear* verb types. She points out that similar syntactic operations occur with inchoative verbs derived from adjectives. The following syntactic trees were proposed by Oliveira (2011, p.68-71):

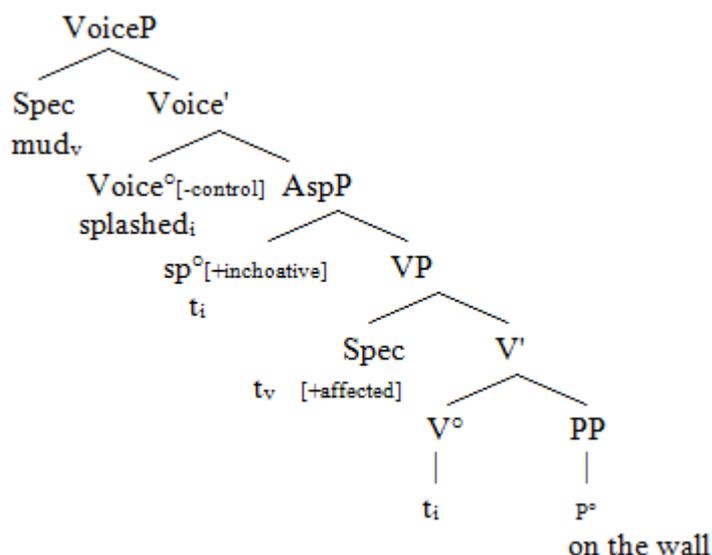
(12) a. ‘*splash*’ verb types



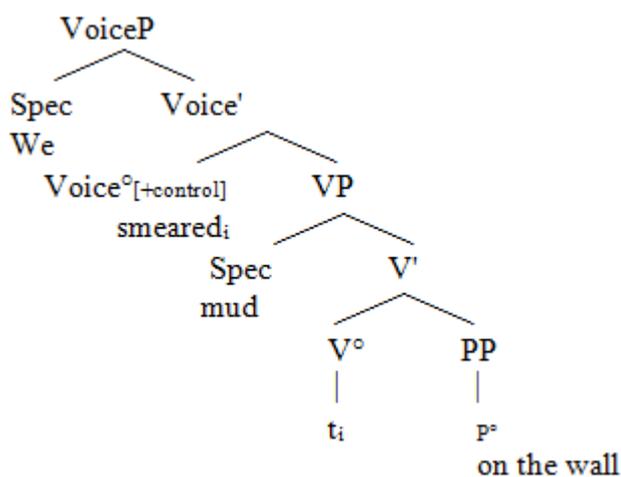
³ From the original:

- (i) possibilidade de haver um núcleo aspectual Asp° acima da projeção VP;
- (ii) esse núcleo portar o traço aspectual [+inchoativo] e o DP projetado em [Spec-VP] apresentar a propriedade semântica [+afetado];
- (iii) e o evento que carrega esses traços aspectuais ser, necessariamente, [+télico]. (OLIVEIRA, 2011, p.68)

b. ‘splash’ verb types, inchoative form:



c. ‘smear’ verb types:



There are pervasive restrictions among the languages that prevent certain kinds of verbs from triggering causative/anticausative alternation. Many studies have been devoted to the understanding of what kinds of verbs can display this phenomenon and for what reasons. In this paper, I adopt the proposal by Levin & Rappaport Hovav (1992, 1995), with further refinement by Alexiadou, Anastoupoulou and Schäfer (2006).

Levin and Rappaport Hovav (1992) suggest that verbs are divided into two distinct classes: verbs of internally caused eventualities and verbs of externally caused eventualities.

Verbs of internally caused eventuality are predicates where some property inherent to the argument of the verb is 'responsible' for bringing about the eventuality. This class of verbs is not necessarily agentive, that is, they do not necessarily select an agent DP. As well as verbs that imply volition (e.g. play, smile), verbs that express inherent qualities of objects (e.g. shine, glitter) also occur in this class.

On the other hand, according to Levin & Rappaport Hovav, the externally caused eventuality verbs "inherently imply the existence of an external cause with immediate control over bringing about the eventuality denoted by the verb: an agent, an instrument, a natural force, or a circumstance." (LEVIN & RAPPAPORT HOVAV, 1992, p.50). That is, the presence of an agent, an instrument, a natural force, or a condition is necessary with these verbs, as is illustrated in the examples below:

- (13) The wind opened the door
The storm devastated the village
The assailant murdered hostages
The journalist wrote his column

Even if some of these verbs can be used intransitively, it is clear that they could not appear without an external cause. A generalization that one may propose about this class is that only verbs of externally caused eventuality can participate in causative/anticausative alternation. Thus, intransitive verbs that causativize are those that can occur in an externally caused eventuality. According to the researchers, externally caused eventuality verbs that do not allow detransitivization are those that only accept an intentional agent as a subject (e.g. murder, write, build, remove). On the other hand, regarding alternating verbs, Levin & Rappaport Hovav claim that:

“(...) what characterizes the class of alternating verbs is a complete lack of specification of the causing event. Thus, the fact that a wide variety of subjects are possible with the alternating is just a reflection of the fact that the causing event is left completely unspecified. Therefore, we can reformulate the condition sanctioning detransitivization: an externally caused verb can leave its cause argument unexpressed only if the nature of the causing event is left completely unspecified.” (LEVIN & RAPPAPORT HOVAV, 1995, p.107)

Alexiadou, Anagnostopoulou & Schäfer (2006), using data from Greek, German and English, expand the classification by Levin & Rappaport Hovav into four verbal categories. They differ from each other according to Voice, which is the head responsible for introducing the external argument. Voice, according to the authors, is related to the grammatical features of agentivity and Manner. In anticausatives, Voice can be totally absent, or an agentless Voice [-AG] can be projected, selecting an implicit causal argument.

Internally caused verbal roots (e.g. to bloom, to wither) combine solely with the Cause head (PYLKKÄNEN, 2002). As a result, the Voice head is not present. These internally caused verbal roots cannot be caused by an external argument. In contrast to Levin & Rappaport, the authors do not consider unergative predicates as causative. Therefore, they cannot be internally caused. Direct causative forms are not expected for these types of verbs.

Externally caused verbal roots (e.g. to destroy, to kill) require an external argument and therefore the presence of Voice. These verbs are subjected to parametric variation among languages regarding the type of Voice head that they can combine with. Unlike agentive roots, they can be associated with [-AG] Voice head, allowing verbal alternation.

Agentive verbal roots (e.g. to build, to murder) are externally caused and also agentive. They appear only in contexts where Agent Voice Head [+AG] is projected, and therefore cannot form anticausatives.

Cause unspecified verbal roots (e.g. to break, to open) do not specify the type of causality involved. Hence, those verbs can appear with or without an external argument. This kind of verb alternates in a prototypical manner.

For the purpose of my analysis, I considered only the cause unspecified verb roots, which prototypically allow alternation, and agentive verbal roots, associated with the Voice head [+Ag], which prototypically do not allow alternation. In the next section, I will focus on these two types of roots in Shimakonde.

2. The {-ik-} and {-uk-} morphemes in Shimakonde

Leach (2010) calls the {-ik-} verbal extension ‘stative’ and the {-uk-} verbal extension ‘stative-separative’. In Shimakonde, the difference in distribution between the stative verbal extension and stative-separative verbal extension is that the latter can only occur in verbs that have been modified by the separative/reversive extension {-ul-} or that end in *ul*. Consider the following examples taken from Leach (2010, p.123):

- (14) a. kú-shím-a
 NC15-close-FV
 “To close”
- a’. kú-shím-ík-a
 NC15-close-STA-FV
 “To be closed”
- b. kú-shím-úl-a
 NC15-close-REV-FV
 “To open”
- b’. kú-shím-úk-a
 NC15-close-STAS-FV
 “To be opened”

Note that in sentence (14b), the verb received the separative extension {-ul-}; therefore, the use of {-uk-} was required to make the stativized form in (14b’). It is also

important to note that, due to vowel harmony, the verbal extension {-ik-} can be realized as {-ek-} and the verbal extension {-uk-} can be realized as {-ok-}. (LIPHOLA, 2001; LEACH, 2010), as shown in the examples below:

- (15) a. kú-tót-a
 NC15-sew-FV
 “To sew”
- b. kú-tót-ék-a (LIPHOLA, 2001, p.148)
 NC15-sew-STA-FV
 “To be sewn”
- (16) a. kú-bámól-a
 NC15-destroy-FV
 “To destroy”
- b. kú-bám-ók-a (LEACH, 2010, p.121)
 NC15-destroy-STAS-FV
 “To be destroyed”

The most salient feature of both morphemes is to express the idea of stativity to the verb, that is, to express the resulting final state. Consider the following examples:

- (17) a. kù-dáng-ék-à
 NC15-build-STA-FV
 “Be built”

- (17) b. Ìng'ándè í-ndí-dàng-ék-à
 NC9-house NC9-PERF-build-STA-FV
 “The house was built” (in a built state)
- (18) a. kw-ándík-ík-à
 NC15-write-STA-FV
 “Be written”
- b. Ìbàlúgwà í-ndy-ándìk-ík-à
 NC9-letter NC9-PERF-write-STA-FV
 “The letter was written” (in a written state)
- (19) a. kú-lúm-ík-à
 NC15-bite-STA-FV
 “Be bitten”
- b. shépò shí-ndí-lùm-ík-à
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit was bitten” (in a bitten state)
- (20) a. kù-tùmb-úk-à
 NC15-break-STAS-FV
 “Be broken”
- b. Shìlóngò shì-ndí-tùmb-úk-à
 NC7-bowl NC7-PERF-break-STAS-FV
 “The bowl was broken” (in a broken state)
- (21) a. kù-shím-ík-à
 NC15-closed-STA-FV
 “Be closed”

- b. Nnángò ù-ndí-shím-ík-à
 NC3-door NC3-PERF-close-STA-FV
 “The door was closed” (in a closed state)
- (22) a. kú-tám-ék-à
 NC15-crack-STA-FV
 “Be cracked”
- b. Nándè ù-ndì-tám-ék-à.
 NC3-branch NC3-PERF-crack-STA-FV
 “The branch was cracked” (in a cracked state)

Besides the stative reading, these morphemes also allow for potential reading of the verb meaning, (as in: to break, to be breakable), as can be noted in the following examples:

- (23) a. Shìlóngò shá-kú-túmb-úk-à nà ínyúndù
 NC7-bowl NC7-CN15-break-STAS-FV with NC9-hammer
 “The bowl is breakable with the hammer”
- b. Ìbàlúgwà yá-kw-ándík-ík-à nà ìlápì
 NC9-letter NC9-CN15-write-STA-FV with NC9-pencil
 “The letter is writable with a pencil”

The stative and potential readings are also found in Chichewa, Swahili and Ndebele, as the examples below indicate:

Chichewa (DUBINSKY & SIMANGO, 1996)

- (24) a. Nyemba zi-na-li zo-phik-ik-a
 beans NC-PERF-be NC-cook-STA-FV
 “The beans were cooked./ cookable”
- b. Mbale zi-na-li zo-sw-ek-a
 plates NC-PERF-be NC-break-STA-FV
 “The plates were broken./ breakable”

Swahili (DRIEVER, 1976, apud SEIDL & DIMITRIADIS, 2003)

- (25) a. Msichana a-me-vunj-a kikombe
 girl NC1-PERF-break-FV cup
 “The girl broke the cup”
- b. Kikombe ki-me-vunj-ik-a
 cup NC-PERF-break-STA-FV
 “The cup is broken/ breakable”

Ndebele (KHUMALO, 2009)

- (26) a. In-kukhu ya-quny-w-a (ngengqamu)
 NC9-chicken NC9-cut-PASS-FV with a knife
 “The chicken was cut (with a knife)”
- b. In-kukhu ya-qum-ek-a (ngengqamu).
 NC9-chicken NC9-cut-STA-FV with a knife
 “The chicken was cuttable (with a knife)”

Additionally, the morphemes {-uk-} and {-ik-} may encode the phenomenon of causative/anticausative alternation, functioning as a morphological marking of anticausativization. Consider the following examples:

- (27) a. Ìmépò ì-ndì-tùmbúl-à shìlóngò
 NC9-wind NC9-PERF-break-FV NC7-bowl
 “The wind broke the bowl”

- b. Shìlóngò. shì-ndì-tùmb-úk-à
 NC7-bowl NC7-PERF-break-STAS-FV
 “The bowl broke”
- (28) a. Ìngwélè í-ndì-tém-à nándè
 NC10-monkey NC10-PERF-crack-FV NC3-branch.
 “The monkey cracked the branch”
- b. Nándè ù-ndì-tám-ék-à.
 NC3-branch NC3-PERF-crack-STA-FV
 “The branch cracked”

The impossibility of the association of anticausatives with agentive oriented adverbs and purpose clauses is confirmed in constructions with the morphemes {-ik-} and {-uk-}. Compare the examples below:

- (29) a. Shìlóngò. shì-ndì-tùmb-úk-à *nàmáđì / *nàmú Lucas / nà ìmépò
 NC7-bowl NC7-PERF-break-STAS-FV *deliberately /*by Lucas /with the wind
 “The bowl broke *deliberately/ *by Lucas/ with the wind”
- b. Shìlóngò v-à-ndì-tùmbúl-à nàmáđì / nàmú Lucas/*nà ìmépò
 NC7-bowl NC2-PERF-break-FV deliberately /by Lucas /* with the wind
 “The bowl was broken deliberately/ by Lucas/ *with the wind” (Passive form)⁴

Note that, if the morpheme {-uk-} appears in the verb, as in (29a), agentive arguments cannot be adjoined to the predicate, nor can adverbs denoting agentivity (*namu Lucas/ namady*). On the other hand, these constructions can associate with causal arguments (*na imeepo*).

⁴ There are two passive constructions in Shimakonde: one with the passive morpheme {-igw-} and the other, such as in example (29b), which forms with the noun class 2 subject morpheme and which is equivalent to the third person plural. For more information about passives in Shimakonde, see Paula (2015).

The same pattern occurs in Chichewa (MCHOMBO, 1993; DUBINSKY & SIMANGO, 1996), Swahili (SEIDL & DIMITRIADIS, 2003) and Ndebele (KHUMALO, 2009), as shown by the following data:

Chichewa (DUBINSKY & SIMANGO, 1996)

- (30) a. *Chitseko chi-na-tsek-ek-a mwadala
 door NC-PERF-close-STA-FV deliberately
 “The door closed deliberately”
- b. Chitseko chi-na-tsek-edw-a mwadala
 door NC-PERF-close-PASS-FV deliberately
 “The door was closed deliberately” (Passive form)

Swahili (SEIDL & DIMITRIADIS, 2003)

- (31) a. *Kikombe ki-me-vunj-ik-a na msichana
 cup NC-PERF-break-STA-FV by girl
 “The cup broke by a girl”
- b. Pili a-li-pig-w-a na Juma
 Pili NC1-PERF-hit-PASS-FV by Juma
 “Pili was hit by Juma” (Passive form)

Ndebele (KHUMALO, 2009)

- (32) a. *isi-valo sa-val-ek-a ngu Thabo/ngabomo
 NC7-door NC7-close-STA-FV by Thabo/deliberately
 “The door closed by Thabo / deliberately.”
- b. isi-valo sa-val-w-a ngu Thabo/ngabomo
 NC7-door NC7-close-PASS-FV by Thabo/deliberately
 “The door was closed by Thabo / deliberately” (Passive form)

As the data above suggest, predicates formed with cause unspecified verb roots with the stative morpheme, such as *kutumbula*, “break”, *kushima*, “close” and *kutema*, “crack”, form structures that resemble anticausatives. However, agentive verb roots also form grammatical constructions when the stative morpheme is present in the verbal structure, as the examples below indicate:

- (33) a. kù-dáng-ék-à
 NC15-build-STA-FV
 “Be built”
- b. Ìng'ándè í-ndì-dàng-ék-à
 NC9-house NC9-PERF-build-STA-FV
 “The house was built”
- (34) a. kw-ándík-ík-à
 NC15-write-STA-FV
 “Be written”
- b. Ìbàlúgwà í-ndy-àndìk-ík-à
 NC9-letter NC9-PERF-write-STA-FV
 “The letter was written”
- (35) a. kú-lùm-ík-à
 NC15-bite-STA-FV
 “Be bitten”
- b. shépò shí-ndì-lùm-ík-à
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit was bitten”

If the static morpheme actually encodes anticausativity, either Shimakonde constructions differ dramatically from the examples of other languages or this verbal extension encodes other types of morphosyntactic phenomena in the language.

Another issue arising from the data is that the constructions in question can offer atelic reading when expressing a stative interpretation. But in some cases, they can also convey telic reading, thereby giving rise to a passive interpretation when occurring with an agentive verbal root, such as *kwandika* “write”, as in the following example:

- (36) Ìbàlúgwà ì-ndy-àndík-ík-à dī-ngwípì mú-dí-núkútà
 NC9-letter NC9-PERF-write-EST-VF NC10-few NC18-NC10-minutes
 “The letter was written in few minutes”

When the external argument is not adjoined to the sentence, such constructions seem to be interchangeable with passive constructions. Consider the examples in (37):

- (37) a. Nángù pàngùwikílè ìbàlúgwà ì-ndy-àndík-ík-à (*nàmú Lucas)⁵
 I arrived NC9-letter NC9-PERF-write-STA-FV
 “The letter was written when I arrived”
- b. Nángù pàngùwikílè ìbàlúgwà ì-ndy-àndík-ígw-à (nàmú Lucas)
 I arrived NC9-letter NC9-PERF-write-PASS-FV
 “The letter was written when I arrived”

In order to understand the phenomenon, it is necessary to define the contexts in which a dynamic reading is possible, at the expense of the stative reading.

In Shimakonde, the {-ndi-} morpheme, which encodes the perfective past, can refer to two different past events: the recent past and the remote past. A native speaker can tell the difference between them by tone assignment. In the recent past, the high vowel is assigned with a low tone. In turn, in the remote past, the high vowel is assigned with a high tone. One of the issues with Shimakonde orthography is that the tone assignment is not marked. Thus, in written texts, it is not easy to distinguish situations in which the perfective morpheme {-ndi-} refers to the remote past or to the recent past.

⁵ Regarding the examples in (37), the consultant was asked what he understood about the presented information’ i.e. if the sentences indicate that the speaker testified to the writing of the letter, (which would lead to a telic interpretation of the sentence) or if the speaker did not testify to the writing of the letter, just finding it done (which would lead to an atelic interpretation of the sentence). In these examples, the speaker testifies that the event is happening. These examples are from the Mocimboa da Praia consultant.

That distinction has become essential for the understanding of the phenomenon in question.

The crucial difference between the interpretability of a stative or anticausative/passive clause when the {-ik- -uk-} morpheme is present seems to be related to the tense of the clause. In other words, if it is in the recent past, it leads to one interpretation, but if it is in the remote past, it leads to another. The stative lexical aspect differs from the remaining classes proposed by Vendler (1967) because it does not express a dynamic event. Thus, stative events cannot answer the question “what happened?” because this question entails that something has changed from one state to another. For this reason, this question was asked to the consultants checking if it could be answered with sentences with {-ik-} {-uk-} extensions in both unspecified cause roots and with agentive roots. With the unspecified cause roots, this question yielded the same outcome for both consultants.

- (38) a. Shìtàndèkè nyàmàni? (Remote past)
 “what happened?”
- b. *Shìlóngò shì-ndí-tùmb-úk-à
 NC7-bowl NC7-PERF-break-STA-FV
 “The bowl was broken” (In a broken state) (stative interpretation)
- c. Shìlóngò shì-ndì-tùmb-úk-à (Recent past)
 NC7-bowl NC7-PERF-break-STA-FV
 “The bowl broke” (Dynamic interpretation)

Note that the question ‘Shìtàndèkè nyàmàni?’ cannot be answered in the remote past, only in the recent past, which shows that only the recent past entails a dynamic reading. Nonetheless, when agentive roots were tested, the outcome showed a parametric variation.

For the Mocimboa da Praia consultant, the same phenomenon occurs when the stative morpheme is present in agentive verb roots. If the sentence is in the remote past

tense, the reading will be stative and therefore atelic. However, if it refers to recent past, the predicate will present a telic reading, thus having a passive interpretation (probably due to the fact that this kind of verb root implies an implicit external agentive argument).

- (39) a. *Shitàndèkè nyàmàni?* “what happened?”
- b. **Shépò shì-ndí-lùm-ík-à* (Remote past)
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit was bitten” (in a bitten state) (Stative interpretation)
- c. *Shépò shì-ndì-lùm-ík-à* (Recent past)
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit was bitten” (Dynamic interpretation)

Conversely, for the consultant from Montepuez, agentive verb roots remain atelic and therefore stative both when they appear in the remote past and in the recent past. Compare the readings provided in the examples in (39), repeated below as (40):

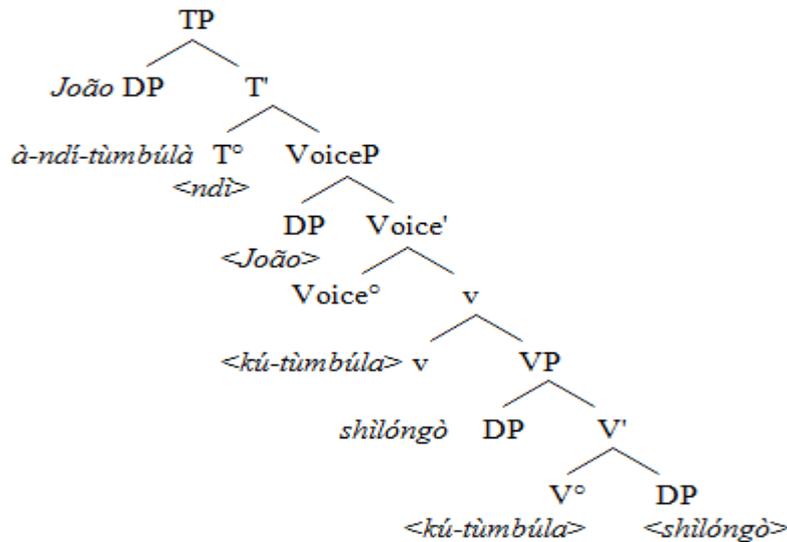
- (40) a. *Shitàndèkè nyàmàni?* “What happened?”
- b. **Shépò shì-ndí-lùm-ík-à* (Remote past)
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit was bitten” (in a bitten state) (Stative interpretation)
- c. **Shépò shì-ndì-lùm-ík-à* (Recent past)
 NC7-fruit NC7-PERF-bite-STA-FV
 “The fruit is bitten” (in a bitten state) (Stative interpretation)

According to the data provided by the consultant from Montepuez, it is possible to assume that, in his dialect, the constructions with stative morpheme are sensitive to the type of verbal root with which they associate, when they occur in recent past. In this case, cause unspecified verb roots have telic reading, leading to an anticausative interpretation. On the other hand, agentive verb roots show atelic interpretation both in remote past and in recent past. Therefore, in this regard, these constructions resemble the proposals made for anticausatives in other languages, which are possible only in unspecified cause verbal roots.

3. Aspect and Voice

From the data presented in previous sections, we hypothesize that {-uk-} {-ik-} morphemes completely remove the external agentive argument. Furthermore, they change the characteristics of Voice. Dynamic Voice becomes stative Voice (KRATZER, 1996). If we take a split view concerning voice head and cause head (PYLKKÄNEN, 2002; LEGATE, 2014), we can assume that the stative morpheme also eliminates the functional cause head, ($= v_{\text{cause}}P$), leaving only the resulting event of the causativization in the structure. In (41) we have a dyadic sentence that will be stativized in (42).

- (41) João àndítumbúlà shìlóngò
 “João broke the bowl”

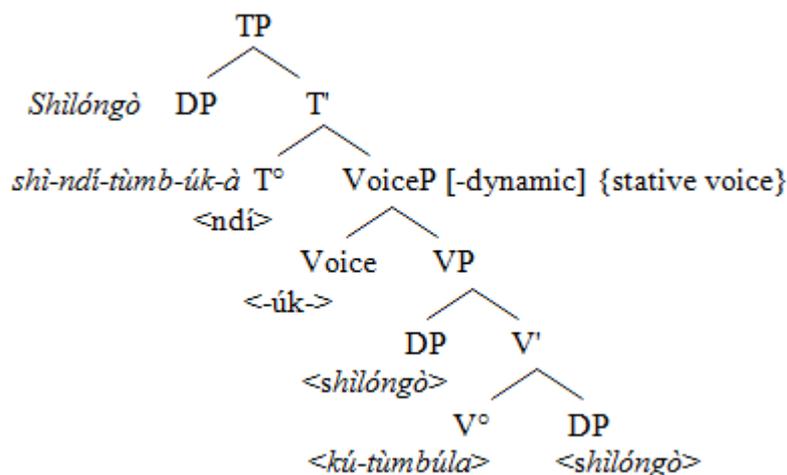


In (42) there is an example of the syntactic structure, when a stative morpheme is inserted and the tense is in the remote past. In these cases, the structure shows an atelic interpretation:

(42) *Shilóngò shìndítùmbúkà.*

“the bowl was broken” (in a broken state)

(Remote past)



The structure in (42) gives rise to a stative interpretation and hence the Cause head (v) is not present. The Cause head may be reinserted in the structure if an

Aspectual telic head is present. Thus, Voice would be dynamic again, but it would still impose restrictions on external agentive arguments.

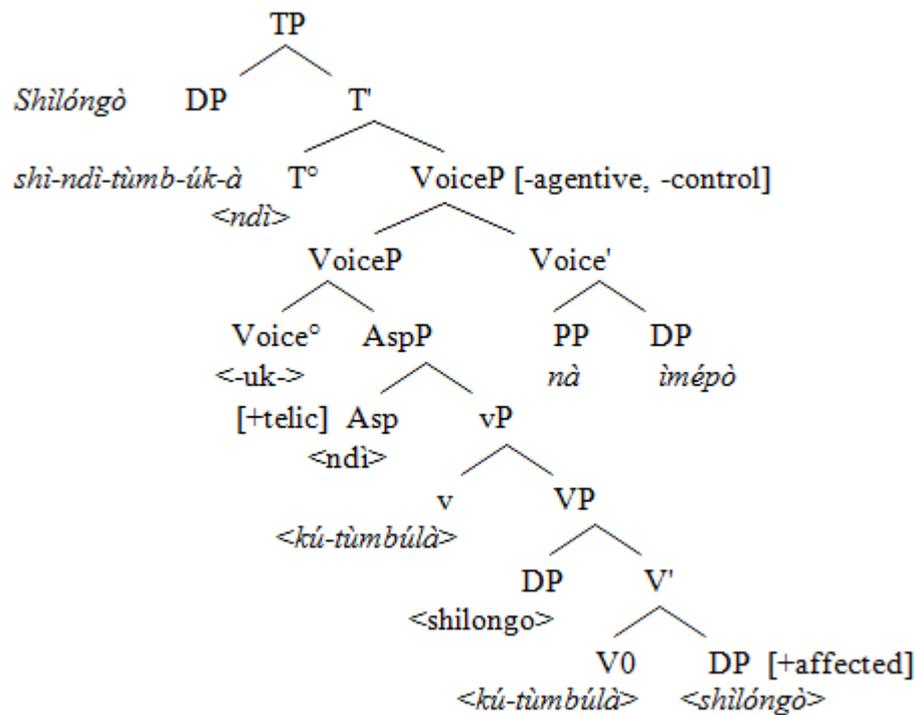
At this point, there is variation among the dialects spoken by the consultants. For the variant spoken by the Montepuez consultant, in which telic reading is not possible in agentive roots in such situations, I propose that the aspectual head is projected below the Voice head. Following the conditions proposed by Oliveira (2011), I will propose that, in the Montepuez variant, stative Voice could yield anticausative interpretation only if:

(I) There is an Aspectual head above the VP projection with the aspectual feature [+telic] (performed by the -ndi- morpheme in a low tone, representing recent past) and dominated by VoiceP

(II) There is a DP in [Spec-VP] with a semantic property [+affected]. In this situation, the aspectual feature [+telic] changes the stative voice back into the dynamic voice and it becomes Voice (-AG) (ALEXIADOU, ANAGNOSTOPOULOU & SCHÄFER, 2006). This derivation implies an unspecified cause argument which may be reintroduced in a PP. If the Voice head is generated by an agentive verbal root, that is, a verbal root with the feature [+control], the projection of Asp head will be blocked, since this verbal root yields Voice (+AG) (Adapted from OLIVEIRA, 2011, p. 68)

By adopting these conditions, the structure would derive the following syntactic tree:

- (43) Shilóngò shìnditùmbúkà nà ìmépò
 “The bowl broke with the wind” (Recent past)



This type of syntactic structure is restricted to unspecified cause verb roots, that is, verbs that do not have the feature [+control]. This is due to specificities in voiceP.

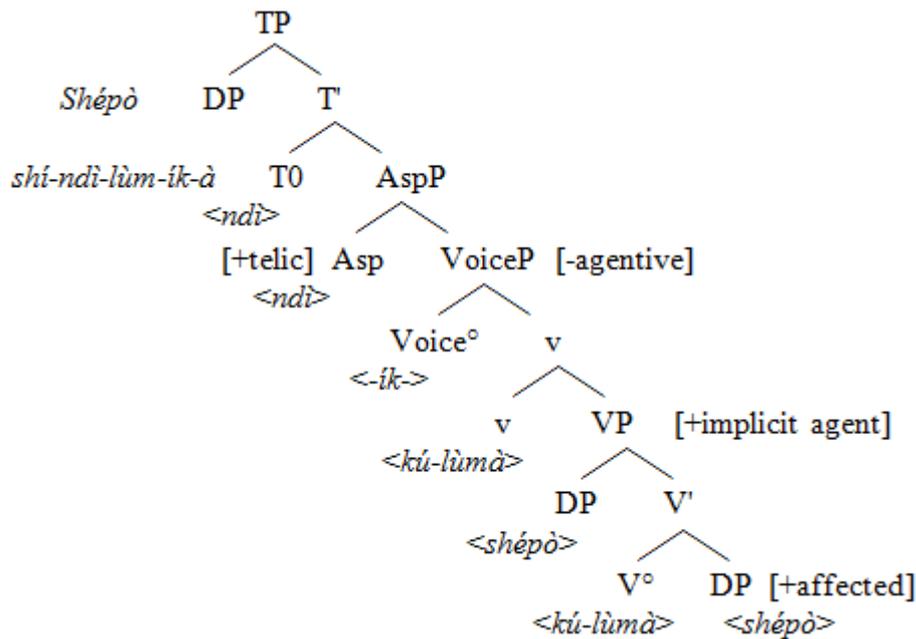
Notwithstanding, the situation is different for the variant of the Mocimboa da Praia consultant. As we have seen, in his variant, constructions with the stative morpheme in agentive verb roots, that is, verbs with the feature [+control], also give rise to telic interpretation in recent present clauses, forming structures that I refer to as pseudo passives. My proposal is that this variation is related to the location of AspP projection in the functional spine of the clause.

In the Montepuez variant, AspP head is projected below VoiceP and is conditioned by specific characteristics of this head. Conversely, in the Mocimboa da Praia variant, AspP is projected above VoiceP, at a later stage of derivation, so it is not conditioned by specificities in Voice, causing a passive reading for verbs that imply an implicit agent, such as bite, write, build, etc. This is represented in the structure below:

(44) Shépò shíndìlùmíkà

“The fruit was bitten”

(Recent past)



4. Final Remarks

In summary, this paper shows that the stative morpheme expresses an idea of stativity of the verb in the remote past tense. In Shimakonde, another feature of the stative morpheme is that it eliminates the agentive external argument. However, for verbs in the recent past tense, such constructions with the {-ik-} and {-uk-} morphemes can still present a telic aspectual head and a causal adjoined argument PP. In this scenario, for the variant spoken by the consultant from Montepuez, the aspectual head dominated by voiceP is projected, but it is not possible with [+AG] heads. With agentive roots, the stative morpheme will converge the dynamic Voice into stative Voice. Notwithstanding, if there are no control restrictions on Voice, as in cause unspecified roots, an aspectual telic head can be projected (this aspect head is characterized by the perfective morpheme in a low tone) and v (or Cause) is maintained in the structure. If this head is projected, dynamic Voice will not converge into stative Voice. The combination of aspectual head [+telic] with the causative head is what provides the anticausative interpretation and licenses an adjoined PP with the thematic role of cause. Otherwise, if there are control restrictions in Voice, the aspectual telic head cannot be projected and the resulting structure will be atelic. Thus, the Voice will remain stative.

On the other hand, in the variant spoken by the consultant from Mocimboa da Praia, VoiceP and AspP are projected at different stages of derivation. Similarly to the Montepuez variant, the stative morpheme converges dynamic Voice into stative Voice and eliminates *v* (cause) from the structure. In a later stage of derivation, the aspectual telic head can be projected above Voice. In this scenario, stative voice switches back again to dynamic Voice and *v* (or cause) is projected once again into the structure. The telic aspect head would not be restricted by the feature [+control] on Voice, since it is projected above this head and it is not dominated by it. As external arguments are generated on Voice head, the stative morpheme keeps restricting the addition of agentive PPs in the structure. If Voice is originally derived from an unspecified cause verb root or, in other terminology, having the feature [-control] (such as break or open), the interpretation will be anticausative, since these roots do not denote an implicit agentive argument in the structure. If the root is agentive or [+control] (such as write, bite or build), despite the impossibility of agentive PPs occurring in the structure, the interpretation will be passive, since these roots denote an implicit agentive argument.

Finally, one may assume that the stative morpheme has many functions in Shimakonde, giving rise to the emergence of the stative, anticausative and passive readings.

References

- ALEXIADOU, Artemis.; ANAGNOSTOULOPOULOU, Elena.; SCHÄFER, Florian. “The properties of anticausatives crosslinguistically”. In. FRASCARELLI, Mara. (ed.) *Phases of Interpretation*, Berlin: Mouton. 2006.
- BENTLEY, Mayrene.; KULEMEKA, Andrew. *Chichewa*. Munich: Lincom Europa. 2001.
- CANÇADO, Márcia.; AMARAL, Luana. “Representação Lexical de verbos incoativos e causativos no português brasileiro”. *Revista da Abralín*, v.9, n.2, pp. 123-147, 2010.
- CHOMSKY, Noam. *The Minimalist Program*. Cambridge, Mass.: MIT Press. 1995.
- CROFT, William. “Possible verbs and the structure of events”. In: TSOHATZIDIS, Savas L. (ed.) *Meaning and Prototypes*, London: Routledge. 1990, pp. 48-73.

- DOKE, Clement. M. *Textbook of Zulu Grammar*. Cape Town: Longmans, Green & Co. LTD. Fourth Edition, 1947.
- DRIEVER, Dorothea. *Aspects of a case grammar of Mombasa Swahili*. Hamburg: Hamburger Philologische Studien. 1976.
- DUBINSKY, Stanley.; SIMANGO, Silvester. “Passive and Stative in Chichewa: Evidence for Modular Distinctions in Grammar”. *Language* 72: pp.749–781, 1996.
- GUTHRIE, Malcolm. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Farnborough: Gregg International, 1967-1971. 4 v. (21, 25).
- HALE, Kenneth.; KEYSER, Samuel. Jay. *Aspect and the syntax of argument structure*, MIT, 1993.
- HASPELMATH, Martin. *Transitivity Alternations of the Anticausative Type*. Cologne: Institut Für Sprachwissenschaft, Universität zu Köln [Arbeitspapiere, N.F. 5], 1987.
- HASPELMATH, Martin. “More on the typology of inchoative/causative verb alternations”. In: COMRIE, Bernard.; POLINSKY, Maria. (eds). *Causatives and transitivity*. Amsterdam/Philadelphia: John Benjamins Publishing Co, 1993, pp. 87-121.
- KALLULLI, Dalila. “Rethinking the passive/anticausative distinction”. *Linguistic Inquiry*. pp. : 770-780, 2007.
- KHUMALO, Langa. “The passive and stative constructions in Ndebele”. *Nordic Journal of African Studies* 18(2): P. 154–174. 2009.
- KRATZER, Angelina. “Severing the External Argument from its Verb”. In: ROORYCK, Joban.; ZARING, Laurie. (eds.). *Phrase Structure and the Lexicon*. Dordrecht: Kluwer Academic Publishers, 1996.
- KRIFKA, Manfred. *Nominalreferenz and Zeitkonstitution: Zur Semantik von Massentermen, Pluralter- men und spektklassen*. Munich: Wilhelm Fink Verlag. 1989.
- LANGA, David. *Morfologia do verbo em Changana*: Maputo: CEA. 2013.

- LEACH, Michael. Benjamin. *Things Hold Together: Foundations for a systemic treatment of verbal and nominal tone in Plateau Shimakonde*. Universiteit Leiden. 435pp. (Doctoral dissertation). 2010.
- LEGATE, Julie. Anne. “Voice and v: Lessons from Acehnese”. *Linguistic Inquiry Monograph 69*, MIT Press, Cambridge, MA. 2014.
- LEVIN, Beth.; RAPPAPORT HOVAV, Malka. “Preliminary analysis of causative verbs in english”. *Língua 92*, p.35-77. 1992.
- LEVIN, Beth.; RAPPAPORT HOVAV, Malka. “Unaccusativity: At the Syntax-Lexical Semantics Interface”. *Linguistic Inquiry Monograph 26*, MIT Press, Cambridge, MA. 1995.
- LIPHOLA, Marcelino. *Aspects of phonology and morphology of Shimakonde*. Ohio: Ohio state University. Dissertation. 2001.
- MCHOMBO, Sam. A. “Formal analysis of the stative construction in Bantu”. *Journal of African Languages and Linguistics 14*, pp. 5-28, 1993.
- NAVES, Rozana. Reigota. *Aspectos sintáticos e semânticos das estruturas com verbos psicológicos*. Dissertação (Mestrado) – Universidade de Brasília, Brasília, 1998.
- NAVES, Rozana. Reigota. *Alternâncias sintáticas: questões e perspectivas de análise*. Tese (Doutorado) – Universidade de Brasília, Brasília, 2005.
- NGUNGA, Armindo. *Introdução à Linguística Bantu*. Livraria Universitária. UEM: Maputo. 2004.
- OLIVEIRA, Maria. José. *Aspecto: Núcleo licenciador da alternância incoativa no PB*. Belo Horizonte: Universidade Federal de Minas Gerais. Dissertação. 2011.
- PAULA, Ronaldo. Rodrigues. De *Aspectos de Morfossintaxe Shimakonde*. Belo Horizonte: Universidade Federal de Minas Gerais. Dissertação. 2015.
- PERLMUTTER, David. M. “Impersonal passives and the unaccusative hypothesis”. *BLS*, v.4, pp. 157-189 , 1978.
- PYLKKÄNEN, Liina. *Introducing Arguments*. Tese (Doutorado em Linguística) – Departamento de Linguística e Filosofia do Massachusetts Institute of Technology, Cambridge, 2002.

- SALLES, Heloisa. Maria. Moreira. Lima. “O aspecto e a alternância dativa”. In: *Nos Domínios do Verbo*, Curitiba. Livro de Resumos - Nos Domínios do Verbo. 2007.
- SATYO, Sizwe. *Topics in Xhosa Verbal Extension*. Ph.D. Dissertation. University of South Africa. 1985.
- SEIDL, Amanda.; DIMITRIADIS, Alexis. “Statives and reciprocal morphology in Swahili”. In: SAUZET, Patrick.; ZRIBI-HERTZ, Anne. (eds.). *Typologie des langues d’afrique et universaux de la grammaire*. Paris: L’Harmattan, 2003, pp. 239-284.
- VAN HOUT, Angeliek. “Unaccusativity as Telicity Checking”. in. ALEXIADOU, Artemis.; ANAGNOSTOPOULOU, Elena.; EVERAERT, Martin. (eds.). *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*. Oxford University Press. New York, 2004, pp. 60-83.
- VENDLER, Z. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press, 1967.
- VERKUYL, Henk. J. *On the Compositional Nature of Aspect*. Dordrecht: D. Reidel. 1972.