

# Anaphoric relations processing in European Portuguese

Armanda Costa, Gabriela Matos, Paula Luegi FLUL/CLUL  
AMLAP - Architectural and Mechanisms for Language Processing, Barcelona, September 7-9

## 1. Abstract

In the framework of grammatical and referential processing research in European Portuguese (EP), we conducted an eye-tracking study to examine the cognitive costs of anaphoric relations' processing involving null elements, taking into account different types of nominal null categories and sentence complexity. While reading non-manipulated texts, the subjects dealt with complex syntactic structures presenting anaphoric chains whose head either co-refers with a *pro* or binds a WH-variable or a NP-gap. Analysing the reading times of contexts with these three types of empty categories, we found a slight advantage in the reading time of the WH-variables and NP-gaps over the reading time of *pro*. We attributed this fact to the higher cost in processing *pro*, which has a higher referential value.

## 2. Aims

- To evaluate the processing costs of referential chains with different grammatical properties.
- To verify the differences brought about by the three types of empty categories: the null pronominal (*pro*) base generated, and the WH- and NP-gaps resulting from movement.

## 3. Null categories in grammar and referential processing

### 3.1. Referential processing in EP

Theories of hierarchy and accessibility in reference assignment predict that processing strategies are held up by the phonetic, morphological and syntactic forms of the anaphoric expression that works as a cue to search the antecedent. The head of the referential chain may have different levels of activation due to its structural position, its thematic role, the linear position in the sentence, or its pragmatic status as topic.

When the antecedent is Topic, Agent, Subject, and in first position, it displays a set of conditions that makes it the best referent for an anaphoric expression that occurs in a lower position in the sentence or in the text (Costa, Matos & Faria 1998; Costa, Faria & Kail 2004; Costa 2005).

Studies on EP show that, when co-reference is at stake, an antecedent with syntactic or discursive prominence favours the choice of a subject *pro*, instead of an overt pronominal, in accordance with an economy strategy recalling the Avoid Pronoun principle (1)-(2). This is also in agreement with the preference for minimal forms to point out prominent antecedents (The accessibility marking scale, Ariel 1996; Centering Theory, Grosz, Joshi & Weinstein 1995).

- (1) O João não cumprimentou o Pedro porque *pro* estava com a sua ex-namorada.  
John didn't greet Peter because [-] was with his ex-girlfriend

- (2) O João não cumprimentou o Pedro porque ele<sub>wh</sub> estava com a sua ex-namorada.  
John didn't greet Peter because he was with his ex-girlfriend

In this study, we contrast how the is assigned to different minimal forms, the null ones, taking into account their syntactic status: null pronominal (*pro*), and WH- and NP-gaps as a consequence of movement.

### 3.2. European Portuguese – some grammatical properties

EP exhibits interesting conditions for the study of processing anaphoric relations since verbal agreement licenses and identifies the content of the *pro* in subject position (Mateus et al. 2003). This opens the range of null categories that may occur in the subject position of a finite clause: *pro*, bound variables, and NP-gaps.

#### Null categories occurring in Subject position in EP

##### *pro*

- Base generated null pronominal, merged in the argument position to which it is related.
- It occurs in the Subject position of root and subordinate tense clauses in Null Subject Languages and in Inflected Infinitive clauses in EP.
- It presents intrinsic referential content or occurs as an expletive.
- In a co-reference relation, it involves two independent referential elements (the antecedent NP and *pro*) in a composite/complex referential chain.

(i) *pro* denoting entities, either definite, (3) or arbitrary (4):

- (3) *pro* lemos esse livro.  
(we) read that book

- (4) *pro* é escusado *pro* dizê-lo.  
it is useless (someone) saying it

(ii) *pro* as an expletive pronoun (5):

- (5) *pro* parece que *pro* lemos esse livro.  
(it) seems that (we) have read that book

(iii) *pro* in anaphoric relations with a co-referring NP (6):

- (6) O homem levantou-se do degrau quando *pro* ouviu correr os ferrolhos.  
the man got up from the step when (he) heard opening the bolts

##### WH-gaps

- Null category arising from the movement of a WH-phrase to a sentence left peripheral position (Non-Argument Movement).
- It occurs in a WH-chain headed by the moved WH-phrase, from which it does not have independent content.
- It is interpreted as a variable bound by the displaced WH-phrase.

- (7) O homem [WH que] [TP [WH -] queria um barco ]  
The man who wanted a boat

##### NP-gaps

- Null category arising from the movement of a nominal category to an argument position (Argument Movement).
- It occurs in a NP-chain headed by the displaced NP, and presents no intrinsic reference.
- Its content is established by the displaced NP that locally binds it.

- (8) O homem<sub>i</sub> foi recebido [NP -] pelo rei.  
The man was received by the king.

## 4. Experiment

### 4.1. Hypothesis

- The less referential the anaphoric null category is, and the more minimal the distance or structural complexity between the antecedent and the null element are, the lower the cognitive costs in processing are.
- Considering *pro*, WH- and NP-gaps, and the stronger referential value of the first one, we predict that processing costs will be higher, in terms of reading time, on contexts where the *pro* gaps occur.

### 4.2. Procedure

The eye-movements of 20 Portuguese adult native speakers were registered with the R6-HS model of ASL (at a 240 Hz rate) while they read an excerpt of the novel *O conto da linha desconhecida* written by José Saramago. <http://www.ufsrgs.br/tramse/pead/textos/saramago.pdf>

The text, which had 809 words, was presented on a screen divided in 9 different slides.

After calibrating the equipment, participants had to read an introductory slide, with biographical information about the writer, to prepare them for the task and also to check whether the equipment was well calibrated.

At the end the participants had to complete a reading comprehension task, with questions about the text.

### 4.3. Dependent variables

Target sentences (presented below) were divided into different regions of analyses and the eye-movement behavior on each one was analyzed. In every region, we have examined: Total Reading Time, Number of Fixations, Mean Fixation Duration, Total Reading Time (in milliseconds – ms) spent on a region divided by its number of Character (TRT/Char).

Since the regions of interest vary a lot in length, because they are natural contexts, not manipulated in any form, we will only present the results of the last measure: TRT/Char.

## 4.4. Materials

We selected contexts for analysis in the text, and, within them, we delimited some regions (sign-posted in the data by a vertical bar) that incorporate the target anaphoric expression. The left border of each region matches the boundary of the clause, which begins with an anaphoric expression. This anaphoric expression is both Subject and (non-marked) Topic of the clause and co-refers with an antecedent presented in a previous clause.

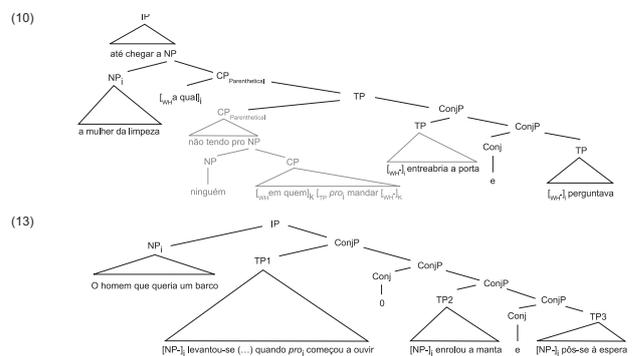
In (9) to (13) we present the contexts and regions analysed (regions in grayscale have not been considered):

- (9) | Esta chamava o terceiro, [CP que, [P [wh-] mandava o primeiro-ajudante, [CP que, [P [wh-] mandava o segundo]]] |  
this called the third one, who ordered the first-helper, who in turn ordered the second (one)
- (10) | até pro chegar à mulher da limpeza, [e qual, [P [wh-] não tendo ninguém [CP em quem [pro] mandar]] |  
until it arrived at the domestic-help, who, not having anybody (who) to order  
[P [wh-] entreabria a porta e [[wh-] perguntava].  
half-opened the door and asked
- (11) | ao primeiro-secretário, [P [wh-] O qual, [P [wh-] pro escusado seria dizer], [P [wh-] passava a encomenda ao segundo-secretário]] |  
to the first-helper who, who no need to say who, [P [wh-] passava a encomenda ao segundo-secretário]] |  
until it arrived at the domestic-help, who replied yes or no according to her mood
- (13) | [CP O homem [CP que, [P [wh-] queria um barco]]], [TP [wh-] levantou-se do degrau da porta | [quando pro começou a ouvir correr os ferrolhos]. |  
the man who wanted a boat got up from the step of the door when (he) started to hear (to) open the bolts.  
[TP [wh-] enrolou a manta] e [TP [wh-] pôs à espera].  
rolled up his cloak and waited

In these data occur:

- Sentence coordination presenting Across-the-Board effects in WH-chains (9) and NP-chains (13).
- Parenthetical clauses disrupting the WH-chain, interspersed between the WH-Phrase and the gap, (9), (10) and (11).

We will draw up simplified representations for the sentences (10) and (13):



## 4.5. Results

The next Table presents the Mean of Total Reading Time per Character in all regions with WH-gap, parenthetical WH-gap, NP-gap, *pro* and arbitrary *pro*:

Referential expression	Num Regions	TRT/Char
WH-gap	4	49 ms
WH-gap (parenthetical)	3	53 ms
NP-gap	3	45 ms
Co-referential <i>pro</i>	4	61 ms
Arbitrary <i>pro</i>	1	50 ms

## 5. Discussion

The differences between the values presented in the Table above may not have any statistical significance, but there are certain trends that are worth noting:

- The TRT/Char in regions with WH- and NP-gaps is lower than in contexts with *pro*:
  - WH- and NP-gaps vs. *pro*: 49ms vs. 61ms
- Regions with NP gaps have a lower TRT/Char than WH-gaps:
  - NP-gap vs. WH-gap: 45ms vs. 51ms
- Arbitrary *pro* (only one case) has a lower TRT/Char than co-referential *pro*:
  - Arbitrary *pro* vs. co-referential *pro*: 50ms vs. 61ms.
- In conditions containing WH-chain with interposed parenthetical material between the WH- displaced phrase and its gap, there is an increase in the reading time when compared with contiguous conditions:
  - 49ms vs. 53ms

## 6. Conclusion

This preliminary study produced results which deserve to be explored in future research:

- Anaphoric relations involving Subject gaps are not uniform with respect to their cognitive costs:
  - anaphoric relations with co-referential *pro* are more costly than with movement gaps;
    - Co-referential *pro* is the slowest form to convey an anaphoric relation: as a (covert) pronoun, it has intrinsic referential content and is locally free (P&P: principle B). Therefore its processing requires additional grammatical and discursive costs.
  - WH-gaps are more costly than NP-gaps.
  - WH-fillers activate search strategies for a gap, and therefore the link is more automatic and efficient than when there is a *pro*; the filler-gap strategy (Frazier 1987) seems to be useful in processing this type of anaphoric relationship;
  - NP-gaps seem to elicit different behaviour in the processor, perhaps because of their lack of referential value or the absence of any cue signalling the chain to the processor.
- Our results allow us to propose that any hierarchy scale of anaphoric elements aiming to explain the anaphoric processing should include other null elements besides *pro*, and should consider that they do not have all the same processing costs.

## References

- Ariel, M. (1996). Referring expressions and +/- coreference distinction. In T. Fretheim & J. Gundel (Eds.), *Reference and accessibility*. Amsterdam: John Benjamins Publishing Company.
- Costa, A. (2005). *Processamento de frases em Português Europeu*. Lisboa: Fundação Calouste Gulbenkian.
- Costa, A., Faria, I. H. & Matos, G. (1998). Competitive information sources in referential ambiguity resolution. *Proceedings of 5th International Congress of the ISAPL*, Porto.
- Costa, A., Faria, I. F., Kail, M. (2004). Semantic and Syntactic Cues' Interaction on Pronoun Resolution in European Portuguese. In Branco, McEnery & Mitkov (Eds.) DAARC 2004, *Proceedings of 5th Discourse Anaphora Resolution Colloquium*. Lisboa: Ed. Colibri, 45-50.
- Costa, M. A. e Luegi, P. (2009). Complexidade linguística e processamento referencial. In Dermeval da Hora (Org). *Anais do VI Congresso Internacional da ABRALIN*. João Pessoa.
- Fodor, J. D. (1993). Processing empty categories: a question of visibility. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing*. Hillsdale: Lawrence Erlbaum Associates, Publishers.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5. 519-559
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1-76.
- Grosz, B., Joshi, A. & Weinstein, S. (1995). *Computational Linguistics*, 21 (2), 203-225.
- Mateus, M. H., M., Brito, A. M., Duarte, I., Faria, I. H., Frota, S., Matos, G., Oliveira, F., Vigário, M., Villalva, A. (2003). *Gramática da Língua Portuguesa*. Lisboa: Caminho.