A corpus based analysis on causal relations in European Portuguese

Joana Aguiar (CEHUM-UM)
State of the art

Corpus based analysis of linguistic phenomena

- Lack of (stratified) corpus
- Lack of frequency studies
- Lack of a sociolinguistic approach on the interpretation on the frequency of occurrence
- Lack of studies focused on syntactic variation phenomena
A variable is the abstract representation of two or more linguistic variants (Meyerhoff, 1999)
Structure

Causal Relations

Methodology

Results and discussion

Conclusion
Causal Relations
Causal Relations (tripartite distinction, Sweetser 1990)

- **Real World**
  - (1) O João partiu a perna porque caiu das escadas. 
    *John broke his leg because he fell from the stairs.*

- **Explicative**
  - (2) O João não almoçou, porque a cozinha está limpa. 
    *John didn’t have lunch, because the kitchen is clean.*

- **Speech act modifier**
  - (3) Vai almoçar, porque retomamos os trabalhos às 13h. 
    *Eat something, because we start working at 1p.m.*
Causal relations: syntactic structures

Real Cause

Copulative Coordination

- (4) O João caiu e partiu a perna.
  - John fell **and** broke his leg.

Central adverbial clauses (Haegeman, 2004, 2009, e.o.)

- *por, por causa de + inf, devido a + inf.* (Lopes, 2004)
  - (5) O João partiu a perna **porque** caiu.
  - John broke his leg **because** he fell. (not for any other reason)
Explicative Cause

- Juxtaposition

- **Supplementation** (Peres & Mascarenhas, 2006; Lobo, 2013, Mendes, 2013)
  
  *Pois, porque (explicative), que, logo, por isso, assim, portanto, etc.*

  (6) **O João é um bom aluno, pois tem sempre bolsa de mérito.**
  
  John is a good student, **for** he always has a scholarship.

- **Subordination**
  
    
    *como, visto que, posto que, dado que, uma vez que, já que, etc.*

    (7) **O João é um bom aluno, visto que tem sempre bolsa de mérito.**
    
    John is a good student, **since** he always has a scholarship.

- **Gerund and participial clauses**

  (8) **Sentindo-se adoentada, foi para casa mais cedo.** (Lopes, 2004, p. 49)
  
  **Felling** sick, she went home earlier.
Speech act modifier

- Juxtaposition
  
  (9) Traz-me um copo de água. Estás de pé e tudo.
  
  *Bring me a glass of water. You are standing and everything.*

- Supplementary Clauses (*porque, que*)
  
  (10) Traz-me um copo de água, que não me apetece levantar.
  
  *Bring me a glass of water, because I don’t want to get up.*

- Peripheral adverbial clauses (*já que, visto que, uma vez que*)
  
  (11) Traz-me um copo de água, já que estás de pé.
  
  *Bring me a glass of water, since you are standing.*
Methodology
Corpus

- 120 argumentative texts written upon request and 48 blog entries.
  - ±250 words each/ 48.509 words total

- Written by 84 European Portuguese speakers, stratified according to gender, education and age.
- The themes and the number of words per texts were controlled to avoid biased results (Aries & Johnson, 1983; Tannen, 1990; Herring & Paolillo, 2006)
The variables

External factors
• age
• sex
• level of education

Internal factors
• Syntactic structure
• connector
• position

Statistics: SPSS, Qui-Square test.
Research questions

- To what extent does syntactic representation of causal relations vary?
- What is the importance of social variables in the clarification of the variation observed?

Hypotheses

- More real cause relations in younger informants (easier to process)
- More subordination in the younger informants (prototypical structure to establish causality).
- More postposed adverbial clauses in women’s texts (Mondorf, 1996; 2002).
Results and Discussion

- 554 tokens

- Overall frequencies:
  - Sex
    - Masc. 49% / Fem. 51%
  - Causal Relation
    - Real 28,2 % / Explicative 67,1 % / Speech act modifier 4,7 %
  - Structure
    - Subordination 34,5% / Supplementation 36,1% / Juxtaposition 25, 3% / Coordination 4,1%
How does syntax convey causality?

Figure 1- Crosstab syntactic structure * causal relation.
Causal Relation and level of education
(cohort with age on 1CEB and 2 CEB)

Younger informants > real cause.

Older informants > explicative cause.

Real cause relations are easier to process when compared to explicative relations (Noordman & Blijzer, 2000: 38).

Figure 2- Argumentative subcorpus - Crosstab causal relation * level of education ($\chi^2(12)=26.550$, p = .009)
Figure 3- Blog subcorpus - Crosstab causal relation * level of education. ($\chi^2(3)=11.869, \ p=.008$)

- Higher level of education > % of explicative cause relations
- Higher % of speech act modifier cause
  - Are blogs more dialogic?
Subordination and supplementation > younger informants.
- The syntactic strategies to express causal relations are still under development (Lopes, 2004).
- In younger informants, *porque* (because) is more frequent:
  - prototypical conjunction to establish causality (Lopes, 2004);
  - common in school communication contexts (Diessel, 2004; Diessel & Hetterle, 2011).

- Adults with 3CEB and secondary tend to use a broader variety of connectors.
Similar results were found in the subcorpus of blog entries.

Figure 5 - Argumentative subcorpus – Crosstab syntactic structure * level of education.

Figure 6 - Blog subcorpus – Crosstab syntactic structure * level of education.
Position of subordinated clauses

<table>
<thead>
<tr>
<th>Position of adverbial clauses</th>
<th>Argumentative</th>
<th>Blogs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>21 16,7%</td>
<td>8 12,3%</td>
<td>29 15,2%</td>
</tr>
<tr>
<td>Final</td>
<td>105 83,3%</td>
<td>57 87,7%</td>
<td>162 84,8%</td>
</tr>
<tr>
<td>Total</td>
<td>126</td>
<td>65</td>
<td>191</td>
</tr>
</tbody>
</table>

Table 1- Position of adverbial clauses according to the subcorpus

- Overall tendency for the final position (default position in Portuguese (Decat, 1995; Paiva, 1998) and in many other languages (Ford, 1993, 1994; Diessel, 2001; 2005). This reflects the iconic order consequence-cause/affirmation-explanation (Neves, 1997).
- Obrigatory initial position of *como*
  - *Como não te vi, fui-me embora* ‘As I didn’t see you, I left’
Preposed adverbial causal clauses:
- work as guides and shifters of the discourse (Ford, 1993).  
- express “high commitment toward the truth of the proposition expressed” (Mondorf, 2002:166)
- require more planning and a higher cognitive effort (Ford, 1993).

Education* Position
- Anteposition > in informants with more years of schooling (except in argumentative texts written by informants with a degree (0%))

Gender*(ante)Position
- Higher % in argumentative texts written by female informants [20-40], 12 years of schooling, unlike predictions (Mondorf, 2002)
Conclusion

- This work sheds some light on the frequency of occurrence of the causal relations and the syntactic structures:

- Speech act modifier causality is more frequent in blogs.

- The frequency of occurrence of explicative causality increases with age and level of education.

- The frequency of occurrence of juxtaposition increases with age and level of education.

- Subordination is more common in the younger informants (also with less years of schooling).

- In EP the majority of the adverbial clauses that convey causality are postposed. The default position is more frequent in younger informants (also with less years of formal education).
References


Obrigada!

This work is part of the ongoing PhD Thesis entitled: *Mechanisms of clausal connection: the importance of the social variables.*

Research funded by a PhD grant from Fundação para a Ciência e Tecnologia (SFRH / BD/64349/2009) and Fundo Social Europeu—Programa Operacional Potencial Humano (POPH) do QREN.

ijoana.aguiar@gmail.com
http://uminho.academia.edu/JoanaAguiar
**Texts written upon request (120 texts / 25.720 words)**

<table>
<thead>
<tr>
<th>Education level</th>
<th>ongoing</th>
<th>completed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4th grade</td>
<td>9th grade</td>
</tr>
<tr>
<td>Age Interval</td>
<td>&lt; 10</td>
<td>10-12</td>
</tr>
<tr>
<td>Gender</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>No Inf.</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**Blog entries (48 texts / 22.789 words)**

<table>
<thead>
<tr>
<th>Education level</th>
<th>11th-12th grade</th>
<th>BA/MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Interval</td>
<td>20-45</td>
<td>&gt; 45</td>
</tr>
<tr>
<td>Gender</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>No Inf.</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>