A novel argument for the universality of parsing principles

Nino Grillo (nino.grillo@gmail.com)
João Costa (jcosta@fcsh.unl.pt)
Centro de Linguística da Universidade Nova de Lisboa
FCSH-UNL
25th CUNY Conference on Human Sentence Processing, CUNY, March 16, 2012

1 Introduction

• Asymmetries of attachment preferences for Relative Clauses (across languages and structures),
• Problem for Universality of parsing principles,
• Previously unnoticed grammatical distinction: the availability of Pseudo-Relatives,
• Hypothesis: once PRs are ruled out, Local Attachment is Universal,
• Results from 3 novel experiments and from previous work support the hypothesis,
• Concluding remarks

Asymmetries in Attachment Preferences

Variation in attachment preferences with Relative Clauses (RCs) across languages, (Cuetos & Mitchell 1988):

(1)  a. Someone shot the maid, of the actress, that was standing on the balcony
    b. Alguien disparó contra la criada, de la actriz, que, estaba, en el balcón

And across syntactic structures (see e.g. Hemforth, Fernández, Clifton, Frazier, Konieczny & Walter submitted, among others):

(2) NOMINALS  Gibson et al. (1996)
    a. The lamp near the painting of the house that was damaged by the flood
    b. la lámpara cerca de la pintura de la casa que fue dañada en la inundación

The maid of the actress that was sitting on the balcony is blonde

These findings are at odds with uniform Local Attachment preference found for other structures in the same languages (e.g. PPs) (Phillips & Gibson 1997);

They lead to question the universality of parsing principles, in particular of Right Association (Kimball 1973) / Late Closure (Frazier 1978) / Recency (Gibson 1991) / Merge Right (Phillips 1996);2

They pose serious problems to theories of acquisition and processing (Fodor 1998a,b).

Several syntactic (type of P, position of complex NP, Nominal vs. Clausal context) and prosodic (length of RC, lengthening of tonic syllable in NP2, duration of prosodic breaks) factors have been shown to influence attachment and

Several accounts have been proposed to explain these variations, e.g. the Tuning Hypothesis (Brysbaert & Mitchell 1996), Construal (Gilboy et al. 1995; Frazier & Clifton 1996), Predicate Proximity (Gibson et al. 1996), Anaphoric Binding (Hemforth et al. 1998, 2000b,a; Konieczny & Hemforth 2000), Implicit Prosody (Fodor 1998a,b) we will not discuss them here (see Fernández 2003; Augurzky 2005, for discussion).

Our main goal is to show the relevance of a previously unnoticed grammatical factor, which might be held responsible for attachment asymmetries both within and across languages.

2 It’s all about that

The following implicit assumption is omnipresent in the literature on RC Attachment:

English that = Spanish que = Italian che = Dutch die etc.

Assuming identity at the grammatical level put all burden for the explanation of variation on parsing. This assumption, however, is wrong:

English that ≠ Spanish que ≠ Italian che ≠ Dutch die etc.

One way in which Cs differ across languages is in whether they allow Pseudo Relatives, and if so, what variety.

---

2Late Closure: When possible, attach incoming lexical items into the clause or phrase currently being processed.
2.1 Pseudo Relatives

Pseudo Relatives and RCs are string identical, but they differ drastically in their structural and semantic properties:\footnote{On Pseudo Relatives see: Radford (1975); Graffi (1980); Burzio (1981, 1986); Kayne (1981); Taraldsen (1981); Declerck (1981, 1982); McCawley (1981); Auwera (1985); Guasti (1988, 1992, 1993); Rizzi (1992); Raposo (1989); Cinque (1992); Barros de Brito (1995); Labelle (1996); Rafel (1999); Côté (1999); Koenig & Lambrecht (1999); Koopman & Sportiche (2010), among others.}

(4) a. \(\checkmark\) Ho visto Gianni che correva / He visto a Juan que corría / J’ai vu Jean qui courait
b. *I saw John that ran
c. I saw John running

(5) **ADJUNCT PRS**

a. Ho incontrato Gianni che correva
   \(I\) \hspace{1em} hit \hspace{1em} Gianni that ran
   \(I\) \hspace{1em} *met Gianni, while he, was running
b. La foto di Gianni che corre é molto bella
   \(\ast\) A foto do João \hspace{1em} que corre é muito linda
   \(\ast\) The picture of Gianni that runs is very beautiful
   \(\ast\) The picture of Gianni running is very beautiful

PRs in some Languages (e.g. Italian, but not Portuguese) can appear in all environments in which Small Clauses can (see Cinque 1992).

(6) **COMPLEMENT SMALL CLAUSES**

Non sopporto Gianni e Mario \[vestiti così / that smoke in my house\]
I can’t stand Gianni and Mario \[dressed like that / \ast that smoke in my house\]

(7) **ADJUNCT SMALL CLAUSES**

Gianni lasciò la stanza \[ubriaco / che era ancora sotto l’effetto dell’alcohol\]
Gianni left the room \[drunk / \ast that was still under the effects of alcohol\]

(8) \[
\begin{array}{c}
V' \\
vedo \\
SC \\
\text{NP}_1 \\
\text{Gianni} \\
\text{CP} \\
\text{Spec} \\
\text{C'} \\
\text{che} \\
\text{IP} \\
\quad \quad e\ corse\ verso\ casa
\end{array}
\]
As shown in (8), in the environment of e.g. perceptual verbs, PRs behave just like Small Clauses; i.e. they project as a complement of V and not as a modifier of N, as RCs would. Table 1 lists a few distinctions between PR and RCs.

<table>
<thead>
<tr>
<th>Property</th>
<th>RCs</th>
<th>PRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refers to individuals</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Available w. objects</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Available w. Rel. Pronouns</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>NP modifier</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Available w. Proper Names</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Available in SC environments</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>VP modifier</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Aspectual restrictions</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>Refers to propositions</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1: RCs and PRs

For our present goals, two distinctions between RCs and PRs are crucial: i. the relation between C and the “subject”, and ii. the relation with the matrix V. We will argue that these distinctions are responsible for variation in attachment preferences, both across languages and syntactic structures.

a. \[ V' \]

\[ V \]

\[ DP_1 \]

\[ NP_1 \]

the son

of

\[ PP \]

\[ DP_2 \]

\[ NP_2 \]

the doctor

b. \[ V' \]

\[ V \]

\[ DP_1 \]

\[ the \]

\[ NP_1 \]

son

\[ N' \]

\[ PP \]

of

\[ DP_2 \]

\[ the doctor \]

that ran

\[ CP \]
3 Variable Syntax, Uniform Parsing

1. Variation in Attachment Preferences, both across and within languages, reduces to the availability of PRs:

2. When PRs are available, everything else being equal, they will be preferred over RCs because of 'Minimal Attachment’, PRs require fewer nodes than RCs. 45

3. When PRs are projected, “High Attachment” (i.e. attachment to DP1) is forced.

4. Corollary:
   - All attachment is local (PRs attach locally to closest VP)
   - Right Attachment/Late Closure/Recency are universal parsing principles

Further Predictions

1. Universal Low Attachment with genuine restrictive RCs,

---

4 Minimal Attachment: Attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language.

5 Our choice of Minimal Attachment is dictated mostly by ease of presentation. Our present goal is to establish that certain parsing preferences are universal. Their characterization and origin, as Late Closure and Minimal Attachment or otherwise, is not our primary concern here. What’s crucial to the present point is that something like Minimal Attachment seems to be at stake in this context, as in other contexts in which the principle has been used to explain the observation that, in ambiguous contexts, restrictive relatives are not the preferred parse in the absence of a context supporting the relevant presupposition. These findings can be interpreted under an economy principle restricting the depth of syntactic structure or operating over e.g. presuppositions. It seems to us that both statements would favor a small clause reading.
2. High Attachment whenever PRs are available,

3. High Attachment with Acc-ing construction in English (I saw the son of the doctor running) and Prepositional Infinitive Constructions (PIC) in Portuguese (vi o filho do medico a correr), i.e. string ambiguous between reduced RCs and correlates of PR.

<table>
<thead>
<tr>
<th>Language</th>
<th>Attachment</th>
<th>PRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Low</td>
<td>X</td>
</tr>
<tr>
<td>German</td>
<td>Low/High*</td>
<td>X</td>
</tr>
<tr>
<td>Romanian</td>
<td>Low</td>
<td>X</td>
</tr>
<tr>
<td>Basque</td>
<td>Low</td>
<td>X</td>
</tr>
<tr>
<td>Russian (?)</td>
<td>High</td>
<td>X</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Variation</td>
<td>Variation</td>
</tr>
<tr>
<td>Galician</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Dutch</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>French</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Serbo-Croatian</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Japanese</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Korean</td>
<td>High</td>
<td>✔</td>
</tr>
<tr>
<td>Greek</td>
<td>High</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 2: Attachment Preferences and PR availability.

Explaining Crosslinguistic Variation As shown in Table 2, with the exception of Russian, there is an almost perfect correspondence between crosslinguistic variation in PR availability and Attachment preferences.

Explaining Variation across syntactic structures The characteristic asymmetry in attachment preference disappears in all the syntactic environments detailed in the examples below, i.e. Low Attachment preference in displayed by BOTH English and Spanish speakers. Crucially, PRs are NOT available in any of these environments:

(10) subjects (Hemforth et al. submitted)
   a. The maid of the actress (that was) sitting on the balcony is blonde
   b. La criada de la actriz que estaba sentada en el balcón es rubia
   c. cf. The maid of the actress sitting on the balcony is blond

(11) nominals (Gibson et al. 1996)
   a. The lamp near the painting of the house that was damaged by the flood
   b. La lámpara cerca de la pintura de la casa que fue dañada en la inundación

---

6PRs, in fact, are available in NP context in some languages (e.g. Italian but not Spanish or Portuguese) but exclusively with event-introducing nominals, e.g. la foto di Gianni che corre é bella / that picture of John running is beautiful. See Rafel (1999) for discussion, and see below for a novel experiment which makes use of this distinction.
(12) NP WITH NP  
(De Vincenzi & Job 1993)
a. Qualcuno ha sparato alla governante con l’attrice che stava seduta in balcone 
b. Someone shot the maid with the actress (that was) sitting on the balcony

(13) UNAMBIGUOUS RELATIVE PRONOUNS  
(Fernández 2003, p. 31) 
Vi al hijo del médico el cual estaba en el balcón 
I saw the son of the doctor who was on the balcony

4 New experimental evidence

In the following 3 questionnaires we used different grammatical constraints to selectively manipulate the availability of PRs: position of complex NP (Center Embedding / Right Branching), position of extraction (subject / object), type of V / N (event introducing / states)

4.1 Experiment 1 - PRs and subject restriction

Method  Questionnaire; Participants: (N=30) Italian native speakers (only one subject performed at chance on the fillers’ Questions [46.5%] and was excluded from the analysis).

Materials and Design  2x2 crossing Position (right branching [RB] vs. center embedding [CE]) and ExtractionSite (subject vs. object). 20 sets of target sentences (4 versions each); 80 fillers. Meaning was kept constant using passives in the A and C condition.

(14) Stimuli
a. PR CONDITION: RB / SUBJECT  
il dottore ha chiamato il figlio del signore che veniva attaccato dai poliziotti 
the doctor called the son of the man that was attacked by the policemen 
b. NO PR CONDITION: RB / OBJECT  
Il dottore ha chiamato il figlio del signore che i poliziotti avevano attaccato 
the doctor called the son of the man that the policemen had attacked 
c. NO PR CONDITION: CE / SUBJECT  
il figlio del signore che veniva attaccato dai poliziotti ha superato la prova 
the son of the man that was attacked by the policemen passed the test 
d. NO PR CONDITION: CE / OBJECT  
il figlio del signore che i poliziotti avevano attaccato ha superato la prova 
the son of the man that the policemen had attacked passed the test

→CHI ERA ATTACCATO? A. FIGLIO B. SIGNORE

7Conditions allowing PRs are underlined

8Thanks to Colin Phillips for suggesting to use passives.
• Counterbalanced materials (each subject only sees one version of each sentence) and questions (50% of the questions with NP1 answer first)

• PC using Linger software developed by Doug Rodhe (http://tedlab.mit.edu/dr/Linger) OR presented on Excel spreadsheet.

• Comprehension questions following EACH sentence

<table>
<thead>
<tr>
<th></th>
<th>RB</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>56%</td>
<td>44%</td>
</tr>
<tr>
<td>Object</td>
<td>33%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table 3: Percentage of High Attachment Preferences

**Results and Analysis**  Mixed Model Regression analysis in open-source code R with the lme4 library (Bates and Sarkar, 2007)

→ Significant effect of position (t value = 4.447): more High Attachment for RB than CE.

→ Significant interaction position*extractionsite (t value = -2.641): more High Attachment for subject extraction than object extraction IN RB only!

**Discussion**  High Attachment preference is observed only when PRs are available i.e. limited to subject extraction in RB cases.

Low attachment preference in all unambiguous RCs conditions (object extraction in RB and both subject and object extraction in CE)

**4.2 Experiment 2 - PRs and event introducing Vs**

**Method**  Questionnaire; (N=13) Portuguese native speakers.

**Materials and Design**  2x2 crossing *Type*(PR and noPR) and *Position* (RB vs. CE). 24 sets of target sentences (4 versions each), 80 fillers

(15)  **Stimuli**

a.  *PR, Right Branching*
   O João viu o filho do medico que estava a correr
   John saw the son of the doctor that was running

b.  *noPR, Center Embedding*
   A foto do filho do medico que estava a correr é muito linda
   The picture of the son of the doctor that was running is very nice

c.  *noPR, Right Branching*
   O João vive com o filho do medico que estava a correr
   John lives with the son of the doctor that was running
d. **noPR, Center Embedding**
   A moto do filho do medico que estava a correr é muito linda
   The motorbike of the son of the doctor that ran is very nice

   A. o filho corre
   B. o medico corre

<table>
<thead>
<tr>
<th></th>
<th>RB</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>70.5%</td>
<td>48.7%</td>
</tr>
<tr>
<td>no PR</td>
<td>47.4%</td>
<td>37.1%</td>
</tr>
</tbody>
</table>

Table 4: Percentage of High Attachment Preferences

**Results and Analysis**

- Mixed Model Regression analysis in open-source code R with the lme4 library (Bates and Sarkar, 2007)
  - Significant effect of type (t value = 3.147): High Attachment in PR > noPR.
  - Significant effect of position (t value = -4.008): more High Attachment for RB than CE

Languages differ in allowing PRs in “picture of” NPs, while they are fine in Italian, they are not allowed in Spanish, Catalan, and Portuguese. We are now running the same experiment in Italian, we predict we’ll find High Attachment preferences also in condition B.

- Next: We want to show that it is indeed the availability of the Small Clause reading and not position what determines attachment preferences, with this in mind we use again the verb / noun types manipulation with Prepositional Infinitive Constructions in Portuguese.

  i. PR (complement) availability through verb type: event-introducing vs. states-introducing Verbs (e.g. see vs. live with)
  ii. PR (adjunct) availability in subject position through N type: event-introducing (*picture of*) vs. states-introducing Nouns (*car of*).

**4.3 Experiment 3 - Beyond that: SCs and attachment**

Same method, procedure and (adapted) stimuli from previous experiment, 20 Portuguese native speakers participated in the experiment. All *que*-clauses were replaced with Prepositional Infinitive Constructions (PIC), which correspond to English *-ing* Small Clauses.

**4.3.1 Stimuli**

(16) a. SC, Right Branching
   O João viu o filho do medico a correr
b. **SC, Center Embedding**  
A foto do filho do médico a correr é muito linda

c. **noSC, Right Branching**  
O João vive com o filho do médico a correr

d. **noSC, Center Embedding**  
A moto do filho do médico a correr é muito linda  
A. o filho corre  
B. o médico corre

### Results

<table>
<thead>
<tr>
<th></th>
<th>RB</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>72.9%</td>
<td>69.7%</td>
</tr>
<tr>
<td>no PR</td>
<td>37.8%</td>
<td>49.1%</td>
</tr>
</tbody>
</table>

Table 5: Percentage of High Attachment Preferences

**Analysis of attachment preference**
- Mixed Model Regression analysis in open-source code R with the lme4 library (Bates and Sarkar, 2007)
- Significant effect of type (t value = 6.65): High Attachment in SC > noSC.
- Significant type*position interaction (t value = -2.05): more High Attachment for RB than CE.

The stimuli in condition A and B used here are in fact string ambiguous between a PIC and a reduced RC. The PIC interpretation should be preferred under Minimal Attachment. Once the string is parsed as a PIC, the only possible argument for it is the highest NP.

### 5 Conclusions

- We have shown, on the basis of both previous and original results that PRs availability modulates attachment both across languages and syntactic structures,
- Speakers of those languages that allow for PRs in the relevant contexts have been reported to prefer High Attachment, while speakers of languages that disallow PRs in those same contexts prefer Low Attachment,
- Moreover, within the same language, whenever PRs are not available, Low Attachment preferences are observed,
- On these bases we have argued that once PRs are taken into the equation, the Universality of Parsing principles of locality can be stated once again.

#### 5.1 Future Tasks

Huge amount of crosslinguistic work:  
At the theoretical level: a thorough comparison of PRs is needed, e.g. how similar/different are romance PRs from each other and from e.g. Japanese /
Korean perceptual constructions? What are these distinctions based on? Long tradition of syntactic analysis of Romance PRs coupled with semantic analysis of perceptual constructions, see e.g. Shimoyama 1999 and Kim 2009 treatment of IHRCs and perceptual constructions as E-type pronouns modifiers. Experimentally: comparing availability of PRs and attachment preferences: i.e. how strong is the generalization proposed here?

- Systematically manipulate PRs availability in different languages using e.g.:
  * Tense mismatch,
  * Modals *Ho visto il figlio del medico che poteva correre* (NO PR) / *I saw the son of the doctor that could run*,
  * Relative Pronouns,
  * *what I saw is x that . . . vs. who I saw . . . contexts*,
  * *que vs. qui* in French,
  * Matrix verb type in Japanese and Korean, e.g. *catch* (IHRC) vs. *see* (perceptual construction).

- PRs and Prosody
  * PRs have a special prosody, just like SCs (*John saw / / Mary running by herself* vs. *#John saw Mary / / running by herself; #John saw / / Mary running by herself* vs. *John saw Mary / / running by himself*),
  * To what extent previous results showing (or not showing) effects of prosody (e.g. length of breaks, length of relative clause) might have been influenced by the manipulation of PRs? *Ho visto Mario che correva* (PR) / *era corso a casa* (NO PR).

- PRs and Frequency
  * Frequency based approaches have not made distinction between RCs and PR, but the two structures should be counted separately
  * (notice that this might actually be good news for frequency based approaches).

**Acknowledgments:** This research is part of the project 'Syntactic and lexical factors in processing complexity’ funded by the Fundação para a Ciência e a Tecnologia with the research grant PTDC/CLE-LIN/114212/2009 awarded to Nino Grillo. We gratefully acknowledge the FCT contribution.

For extensive comments and suggestions, and for help with stats, thanks to Andrea Santi. Thanks to Keir Moulton for letting me know that the structures I had “discovered” were called pseudo relatives and for pointing me to the relevant literature and discussing it with me. To Bruno Fernandes (the
idea to use *picture of NPs* in experiment 2 and 3 is Bruno’s) and Margarida Tomaz for support, suggestions, comments. For useful comments and for help with judgments/literature on PRs in different languages thanks to: Giorgos Spathas, Colin Phillips, Michael Wagner, Janet Dean Fodor, Yosef Grodzinsky, Maria do Carmo Lourenço-Gomes, Petra Augurzky, Ad Neeleman, Klaus Abels, Junko Shimoyama, Mina Sugimura, Natalia Slioussar, Gueorgui Hristovsky, Donka Mangatcheva, Natalia Slioussar, Anca Sevcenco, Boban Arsenijević, Berit Gehrke, Hans van de Koot, Axiotis Kechagias, Larraitz Zubeldia.
References


