Animacy and the Status of the Referent in the Choice of Postverbal Complement Ordering: The Case of French

Juliette Thuilier

1 Université Paris Sorbonne & Alpage (INRIA - Paris Diderot)
juliette.thuilier@paris-sorbonne.fr
http://www.linguist.univ-paris-diderot.fr/~jthuilier/

12 novembre 2012

Typologie : syntaxe et phonologie
Journée scientifique Labex EFL
Auditorium de l’INALCO
Introduction

In French, postverbal constituents are rather freely ordered (Blinkenberg, 1928; Abeillé and Godard, 2000)

(1)  

a. Paul a donné [une fleur]NP [à Marie]PP  
b. Paul a donné [à Marie]PP [une fleur]PN  

Paul gave a flower to Marie

⇒ What are the factors influencing the postverbal ordering?  
⇒ Focus on two factors: animacy and the status of the referent (given vs. new)  
⇒ Quantitative approach based on corpus data and questionnaires
1 Previous works
   On French
   On other languages

2 Corpus study
   Data and studied variables
   Statistical modeling

3 Questionnaires
   Animacy
   The status of the referent

4 Discussion and conclusion
Previous works on French

To our knowledge, neither corpus studies nor experimental works

1. General preference for *Direct Object before Indirect Object* (Blinkenberg, 1928; Berrendonner, 1987)

2. Weight (Blinkenberg, 1928; Berrendonner, 1987) and lightness (Abeillé and Godard, 2004, 2006) : *short before long*

   (2)  
   a. Paul offrira [à son fils]PP [le livre dont je t’ai parlé]NP  
   b. Cet endroit fait [peur]NP [aux enfants]PP  

3. Ambiguity avoidance (Blinkenberg, 1928; Berrendonner, 1987)

   (3)  
   a. Il sut parler [à la France]PP [le langage qui convenait]NP  
   b. Il sut parler [le langage qui convenait]NP [à la France]PP
Previous works on French

4 Verb meaning (Schmitt, 1987a,b) : for a subset of verbs, the order of
the complements is determined by the semantics of the verb
(4) préférer NP à NP, faire de NP NP, remplacer NP par NP,
troquer NP contre NP...
literally : to prefer NP to NP, to make of NP NP, to replace NP with NP,
to exchange NP for NP

5 Definiteness : definite before indefinite (Berrendonner, 1987)

6 Status of the referent : given before new (Blinkenberg, 1928;
Berrendonner, 1987)
(5) (Qu’est-ce que Moscou a envoyé au mouvement des cent-un ?)
Moscou a envoyé au mouvement des cent-un un message
d’appui

7 Animacy not mentioned
Previous works on other languages

**Animacy** : *animate before inanimate*

1. Corpus studies
   - Dative alternation in English (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2010) : inanimate recipient favors prepositional *to*-dative construction
     
     \[
     \text{theme} \prec \text{recipient}_{\text{inanimate}} \quad \text{prepositional construction}
     \]
     
     \[
     \text{recipient}_{\text{animate}} \prec \text{theme} \quad \text{double-object construction}
     \]

   - Subject and Object ordering in German (Kempen and Harbusch, 2004) : for instance, in their data,
     
     \[
     \text{SUJ} \prec \text{IO}_{\text{inanimate}} \quad (93\%) \quad \text{SUJ} \prec \text{IO}_{\text{animate}} \quad (54\%)
     \]

2. Experimental works : convergent results for several languages
   - Greek : Subject-DO (Branigan and Feleki, 1999)
   - Japanese : Subject-DO (Tanaka et al., 2011)
   - Spanish : Subject-DO (Prat-Sala and Branigan, 2000)
Previous works on other languages

The status of the referent:

1. Definiteness: *definite before indefinite*
   - English dative alternation (Collins, 1995; Bresnan et al., 2007):
     * indefinite recipients favor prepositional construction
     * indefinite themes favor double-object construction
     
     \[
     \text{theme}_{\text{definite}} \prec \text{recipient}_{\text{indefinite}} \quad \text{prepositional construction}
     \]
     
     \[
     \text{recipient}_{\text{definite}} \prec \text{theme}_{\text{indefinite}} \quad \text{double-object construction}
     \]

   - Subject and Indirect Object ordering in German (Bader and Häussler, 2010):
     in the *Middlefield*, indefinite subjects tend to follow definite IOs
     
     \[
     \text{IO}_{\text{definite}} \prec \text{SUJ}_{\text{indefinite}} \quad (83\%)
     \]
     
     \[
     \text{IO}_{\text{definite}} \prec \text{SUJ}_{\text{definite}} \quad (44\%)
     \]

2. Given referent before new referent
   - Dative alternation and Heavy NP Shift in English (Arnold et al., 2000)
     * HNPS: besides weight, newness of the NP favors the HNPS (PP \prec NP_{new})
     * DA: constituents of equal length
     
     \[
     \text{recipient}_{\text{new}} \prec \text{theme}_{\text{given}} \quad (15\%)
     \]
     
     \[
     \text{recipient}_{\text{given}} \prec \text{theme}_{\text{new}} \quad (60\%)
     \]

   - Preference for given-new ordering hypothesized to be a language universal (Clark and Clark, 1978)
1 Previous works
   - On French
   - On other languages

2 Corpus study
   - Data and studied variables
   - Statistical modeling

3 Questionnaires
   - Animacy
   - The status of the referent

4 Discussion and conclusion
Corpus study

The database

- 1434 sentences containing a verb followed by two subcategorized complements (V NP PP or V PP NP)
- extracted from
  - two newspapers corpora (French Treebank and Est-Républicain)
  - two oral corpora: public radio corpus (ESTER) and spontaneous speech (C-ORAL-ROM)
- 181 verbal lemma
- Average 70.4% preference for NP-PP order
## Corpus study

### Studied variables

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td><strong>defNP</strong> : definiteness of the NP (definite vs. indefinite)</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td><strong>defPP</strong> : definiteness of the PP</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td><strong>animNP</strong> : animacy of the NP (animate vs. inanimate)</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td><strong>animPP</strong> : animacy of the PP</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td><strong>proNP</strong> : pronominality of the NP (pronoun vs non-pronoun)</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td><strong>proPP</strong> : pronominality of the PP</td>
</tr>
<tr>
<td><strong>7</strong></td>
<td><strong>NP-PP</strong>&lt;sub&gt;LEN&lt;/sub&gt; : relative length (number of words of NP − number of words of PP)</td>
</tr>
<tr>
<td><strong>8</strong></td>
<td><strong>V-PP</strong>&lt;sub&gt;COLLOC&lt;/sub&gt; : V-PP collocation as <code>mettre en relief</code> (to foreground), <code>mettre en lumière</code> (to bring sth in light), <code>prendre en compte</code> (to take into account)</td>
</tr>
<tr>
<td><strong>9</strong></td>
<td><strong>VERB+SEM</strong> : Verbal lemma + its semantic class in context (annotated according to dictionary of Dubois and Dubois-Charlier (1997))</td>
</tr>
</tbody>
</table>
Statistical modeling

**Logistic regression** (Agresti, 2007)
- modeling the behavior of a binary variable as a function of several predictive variables
- complement ordering = binary variable
  - NP-PP order = 0
  - PP-NP order = 1

**Logistic function**
Function for which values can be interpreted as conditional probabilities

\[
\pi_{\text{PPNP}} = \frac{e^{\beta X}}{1 + e^{\beta X}}
\]  
(1)

where
- \( \pi_{\text{PPNP}} \) = probability for the order PP-NP
- \( \beta = \) regression coefficients \( \alpha, \beta_0 \ldots \beta_n \)
- \( X = \) predictive variables \( X_0 \ldots X_n \)
Corpora and verbal lemma as sources of variation in the data

Verb + semantic class (cf. Schmitt, 1987a,b)

<table>
<thead>
<tr>
<th>Database</th>
<th>ajouter U</th>
<th>vendre D</th>
<th>donner D</th>
<th>mettre L</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP-PP order</td>
<td>70.4%</td>
<td>22.2%</td>
<td>36.7%</td>
<td>73.6%</td>
</tr>
</tbody>
</table>

Corpus

<table>
<thead>
<tr>
<th>Database</th>
<th>FTB</th>
<th>ER</th>
<th>CORAL</th>
<th>ESTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP-PP order</td>
<td>70.4%</td>
<td>67.8%</td>
<td>69.1%</td>
<td>73.6%</td>
</tr>
</tbody>
</table>

How to capture the characteristics of each group of data?

⇒ mixed-effects model:

- the idea: besides the general model (fixed-effects), each group of data receives its own coefficient (random-effects)
- each corpus and each verb has a coefficient capturing its individual behavior
Multifactorial statistical modeling

Random effects

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus</td>
<td>0.24245</td>
<td>0.49239</td>
</tr>
<tr>
<td>Verb + Sem</td>
<td>1.24298</td>
<td>1.11489</td>
</tr>
</tbody>
</table>

Number of obs: 1434, groups: verb + sem, 253; corpus, 4

Fixed effects

|                  | Estimate | Std. Error | Std. Error | P(>|z|)   |
|------------------|----------|------------|------------|----------|
| Intercept        | -1.4269  | 0.2879     | -4.955     | 7.22e-07 *** |
| NP-PPLen         | +2.6891  | 0.1565     | 17.183     | <2e-16 *** |

- the other variables do not have a significant effect on the probability of PP-NP ordering ($\chi^2(7) = 8.44$, $p = 0.30$)

⇒ the corpora study suggests that **animacy** and **definiteness** do not affect the relative ordering of verbal complements in French
Given vs. new in a subset of data

- Sample of 166 sentences with $-2 < \text{NP-PP} \text{LEN} < 2$ (in order to minimize the effect of length)
- 79.5% of NP-PP order
- Annotation of the referents of NP and PP according to the categories of Prince (1981)
  - new: when a speaker first introduces an entity into the discourse [...] we may say that it is new (Prince, 1981, p 235)
  - evoked (=given): if some NP is uttered whose entity is already in the discourse model [...] it represents an evoked entity (Prince, 1981, p 236)
- 2 variables:
  - $\text{NP status}$: the referent of the NP is new or given
  - $\text{PP status}$: the referent of the NP which is complement of the preposition is new or given
- Statistical modeling of NP-PP order using mixed-effects logistic regression: $\text{statusNP}$ and $\text{statusPP}$ do not have a significant effect on the probability of PP-NP ordering

$\Rightarrow$ the corpora study suggests also that given vs. new opposition does not affect the relative ordering of verbal complements in French.
1 Previous works
   - On French
   - On other languages

2 Corpus study
   - Data and studied variables
   - Statistical modeling

3 Questionnaires
   - Animacy
   - The status of the referent

4 Discussion and conclusion
Animacy

- Animacy effect does not show up in corpus study
- It could be due to uncontrolled correlation between this factor and other variables (length and verbal lemma)
- Questionnaire set up in order to study this factor controlling correlations

Hypothesis

Speakers tend to prefer animate before inanimate ordering

Method

Participants

- 38 native speakers of French, recruited at University Paris Diderot

Items

- 16 sentences containing ditransitive verbs
- NP and PP: same length, and both definite
- sample of verbs with different preferences estimated in the corpus
Method

Items

- inanimate NP
- equal number of items with animate and inanimate PP
- both order presented to the participants
- 22 fillers
- the 38 items were randomly ordered in each questionnaire

Procedure

- Acceptability judgment task
- Each continuation has been judged using a scale ranging from 1 to 5 (1 = not acceptable; 5 = fully acceptable)
- Each participant has a different questionnaire (ordered differently)

Il faut que les Israéliens maintenant, dans les prochaines semaines, dans les prochains mois

A. donnent les réponses précises à ces questions.  
B. donnent à ces questions les réponses précises.
Results

- Expected result: preference for PP[animate]-NP[inanimate] ordering
- Participant’s judgments show an overall preference for NP-PP order (Mean = 4.06, standard error = 1.07) (Figure 1)
- Participants slightly prefer items where PP is animate, but not significantly (Figure 2)
- Interaction between order and animacy: participants tend to prefer PP-NP ordering, when PP is animate (Figure 3)
Results

- Acceptability judgments modeled using mixed-effects linear regression:
  - subjects and items as random effects
  - order, animacy of PP and interaction between both variables as fixed effects

- Both random effects are significant (for subject $\chi^2(1) = 88.703, p < 0.0001$, for items $\chi^2(1) = 11.265, p < 0.001$)

- One fixed effect is significant: \textbf{order}

- Neither animacy nor its interaction with order are significant factors for predicting the judgment made by the participants ($\chi^2(2) = 3.95, p = 0.14$)

⇒ Initial hypothesis not confirmed, no conclusion possible
⇒ Neither corpus study nor questionnaire show an effect of animacy on postverbal constituent ordering
The status of the referent

- The status of the referent (definiteness/newness) does not show up in the corpus study
- Questionnaire set up in order to study this factor controlling correlations

Hypothesis

Speakers tend to prefer *given before new* ordering

Method

Participants

- 28 native speakers of French, recruited at University Paris Sorbonne

Items

- 16 sentences containing ditransitive verbs
- NP and PP: same length, and both inanimate
- indefinite and new NP
- equal number of items with given and new PP
- 8 verbs with different preferences estimated in the corpus
  - for each verb, one context with given PP, another with new PP
**Method**

**Items**
- both order presented to the participants
- 17 fillers
- the 33 items were randomly ordered in each questionnaire

**Procedure**
- Acceptability judgment task
- Each continuation has been judged using a scale ranging from 1 to 10 (1 = not acceptable; 10 = fully acceptable)
- Each participant has a different questionnaire (ordered differently)

---

*En Camargue, seuls les flamants roses peuvent donner à un paysage monotone une pointe de relief.*

1 □ □ □ □ □ □ □ □ □ □

10

*donner une pointe de relief à un paysage monotone.*

□10

*De nombreuses questions se posent à propos de la situation économique du pays. Il faut que les candidats maintenant donnent à ces questions des réponses appropriées.*

1 □ □ □ □ □ □ □ □ □ □10

*donnent des réponses appropriées à ces questions.*

1 □ □ □ □ □ □ □ □ □ □10
Results

- Expected result: preference for PP\([\text{given}]\)-NP\([\text{new}]\) ordering
- Participant’s judgments show an overall preference for NP-PP order (Figure 1)
- No preference between given and new referent (Figure 2)
- Interaction between the order and the status of the referent: no effect of the interaction on judgment (Figure 3)
Results

- Observations confirmed by the modeling of acceptability judgments
- Acceptability judgments modeled using mixed-effects linear regression:
  - subjects and items as random effects
  - order, status of the PP and interaction between both variables as fixed effects
- Only one random effect is significant: Subject (Items not significant \( \chi^2(1) = 0.70, p = 0.38 \))
- One fixed effect is significant: order
- Neither the status of the referent nor its interaction with order are significant factors for predicting the judgment made by the participants \( \chi^2(2) = 0.40, p = 0.82 \)

⇒ Initial hypothesis not confirmed, no conclusion possible
⇒ Neither corpus study nor questionnaire show an effect of the status of the referent on postverbal constituent ordering
1 Previous works
  - On French
  - On other languages

2 Corpus study
  - Data and studied variables
  - Statistical modeling

3 Questionnaires
  - Animacy
  - The status of the referent

4 Discussion and conclusion
Results concerning animacy and the status of the referent

- Using quantitative study on corpus and acceptability questionnaires
- No visible effect on postverbal complement ordering

⇒ No indication that French postverbal complement ordering is affected by these factors, contrary to other languages

How to explain these observations?

Methodological issues (questionnaires)

- Noise: the long contexts introducing the items and the varied lexicon used might introduce noise that hides the expected effect
- Number of participants: we could eventually detect an effect by adding participants to our observations
- Weakness of the effect: the effect may be very weak and the experimental protocol may not be sensitive enough to detect this effect
- Metalinguistic task: direct methods could have more chances to show up the expected effect (like sentence recall tasks used in Tanaka et al., 2011)
How to explain these observations?

**Linguistic characteristics**

- In other languages, most word order phenomena include the Subject
  - Subject is typically *animate* and *given*
- In English and German at least, verb dependent ordering involves pronominal constituents
  - Pronouns are prototypically *given* and can be *animate*
  - In French, most of the pronouns appear as clitics before the verb

⇒ These observations could explain the lack of effect/or the weak effect of animacy and the status of the referent in French postverbal complement ordering

**Perspectives**

- To use other experimental tools in order to explore the effect of animacy and the status of the referent on postverbal complement ordering in French
- To investigate other variables using quantitative methods on corpora in order to better understand which factors influence the postverbal ordering in French


