Ordering preferences for postverbal complements in French

Juliette Thuilier, Anne Abeillé, Benoît Crabbé

ORDERING PREFERENCES FOR POSTVERBAL COMPLEMENTS IN FRENCH

Juliette Thuilier[13], Anne Abeillé[24] and Benoît Crabbé[23]
[1]Université Paris Sorbonne, France
[2]Université Paris Diderot, France
[3]Alpage INRIA research unit, France
[4]LLF CNRS research unit, France

1 Introduction

In French, postverbal constituents are rather freely ordered: the verb has a fixed position and, after it, the ordering of the constituents is generally considered to be fairly flexible (Blinkenberg 1928, Abeillé & Godard 2006, a.o.). In this paper, we focus on the order of the postverbal NP and PP complements, which can generally be permuted, as in example 1:

1. (a) Pierre fonce dans la nuit porter [la bonne nouvelle] [à sa fiancée] (Est Républicain)
   (b) Pierre fonce dans la nuit porter [à sa fiancée] [la bonne nouvelle]
   'Pierre runs in the night to-bring the good news to his fiancée'

Most previous studies on complement ordering in French (Blinkenberg 1928, Berrendonner 1987, Schmitt 1987a, b, Abeillé & Godard 2006…) have not been underpinned by large corpus studies, contrary to existing studies for other languages (as in Arnold *et al.* 2000 or Wasow 2002 for English). Our goal here is to study the different factors conditioning the relative ordering of complements of ditransitive verbs by using a quantitative approach based on corpus and experimental data, along the same lines as Wasow (2002), Bresnan *et al.* (2007), and Bresnan and Ford (2010).

Previous works on heavy NP shift (*bring the wine we ordered to the table vs bring to the table the wine we ordered*) or dative alternation (*give John a book vs give a book to John*) in English have showed that one can draw generalizations concerning preferences on the basis of corpora, and that these preferences are the result of the interaction of heterogeneous factors.

More generally, previous works on word order (Behagel 1909, Hawkins 1994…) have established that there are general patterns in the

alignment of verbal arguments across languages: animate, pronominal and definite constituents tend to precede inanimate, non-pronominal and indefinite ones. Furthermore, in SVO languages, short constituents tend to appear before long ones.

This paper aims to show which preferential constraints influence the choice between V-NP-PP or V-PP-NP complement order in French, taking into account the general tendencies observed in other languages. For this purpose, we have extracted and annotated relevant data from two newspaper corpora (the French Treebank which is based on the newspaper *Le Monde*, and the *Est Républicain* corpus) and two spoken corpora (the public radio corpus ESTER and the French part of C-ORAL-ROM), which lead us to quantitative results and statistical modelling. We have also used two psycholinguistic questionnaires in order to collect the acceptability judgments of native speakers.

2   State of the art

Some work has already been done on the ordering of complements in French. Blinkenberg (1928) observes a general preference to order the direct before the indirect object (example 2) and also that several factors may favor the reverse order: the length of the object (especially if it is a complement clause as in 3), and in literary style, the fact that the indirect object is repeated or known or the wish to avoid a wrong PP attachment as in 4a (where postposing the PP may induce a wrong attachment to *convenir* 'to suit' instead of *parler* 'to speak'):

  2. J'aurais besoin de demander [un service] [à votre compagnon].
      'I would need to ask some help to your companion'

  3. Il avait promis [à mon frère] [qu'il reviendrait le lendemain].
      'He had promised to my brother that he would come-back the next-day

  4. (a) Il sut parler [à la France] [le langage qui convenait] (Bertrand, Louis XIV, p. 285)
      'He knew how to speak to France the language which was convenient'
      (b) # Il sut parler le langage qui convenait à la France

Berrendonner (1987) proposes several ordering factors, which are supposed to be communicatively helpful (planning for the speaker or comprehension for the hearer): short before long, focus (or new) before ground (or given), and ambiguity avoidance. Given before new orders a

definite complement before an indefinite one as in 5, and a non contrastive complement before a contrastive one as in 6:

5. Et c'est justement ce qui suscite [chez les gros consommateurs d'électricité] [une grogne qui tourne à la fronde]
   'And this is precisely what raises among the big electricity consumers a grunting turning into a revolt'

6. Solon donna au peuple les droits civils, non les droits politiques
   'Solon gave to-the people the civil rights, not the political rights'

Abeillé and Godard (2004, 2006) propose that lightness also plays a role, in the opposite direction from weight: light complements (bare quantifiers, bare nouns) precede phrasal ones, as in 7 and 8, unless they are made heavier as in 9:

7. (a) Cet endroit fait peur aux enfants
   'this place gives fear to-the children'
   (b) *Cet endroit fait aux enfants peur

8. (a) Paul expliquera tout à son fils
   'Paul will-explain everything to-his son'
   (b) ?Paul expliquera à son fils tout

9. (a) Cet endroit fait aux enfants [vraiment très peur].
   'this place gives to-the children very much fear'
   (b) Paul expliquera à son fils [absolument tout]
   'Paul will-explain to his son absolutely everything'

Schmitt (1987a, b) is the first to our knowledge to take the verb meaning into account. For 12 verbs (such as *joindre* 'to join', *lier* 'to tie', *préférer* 'to prefer', *remplacer* 'to replace'…), he observes a tendency to follow what he considers to be a logical or temporal ordering, placing the complement denoting the initial state before the complement denoting the final state (in J-P. Sartre's novel *Les Mots*). This results in a fixed ordering for verbs such as: *faire de SN1 SN2* 'make out-of NP1 NP2' and *remplacer SN1 par SN2* 'replace NP1 by NP2', as in 10a and 10b, where the final state (or object) is placed after the initial state (or object). But he also observed a fixed ordering for *préférer SN1 à SN2* 'prefer NP1 over NP2', as in 10c where the final choice is placed before the first choice, so this logical or chronological tendency is not so clear. He also observes that some complements form a collocation with the verb and are thus adjacent

to it (*servir de modèle à quelqu'un* 'serve as a model to someone', *mettre en marche quelque chose* 'set something in motion'):

10. (a) faire de l'exception la règle
    'make out-of the exception the rule'
    (b) remplacer les bruits de ma vie par des inscriptions ineffaçables
    'replace the noises of my life by unerasable inscriptions'
    (c) préférer le devoir au plaisir
    'prefer duty to pleasure'

For other languages, previous work on word order (e.g. Wasow 2002, Bader and Häussler 2010) has established that there are general patterns in the alignment of verbal arguments across languages: animate, pronominal and definite constituents tend to precede inanimate, non-pronominal and indefinite ones. Furthermore, in SVO languages, short constituents tend to appear before long ones (what Behagel 1909 calls *das Gesetz der wachsenden Glieder* 'the law of growing elements'), whereas the reverse is true in SOV languages, where long elements tend to precede short ones (Hawkins 1994). Although English has a general tendency to order the direct before the indirect object, the alternation called "heavy NP shift" favors the reverse order, where a long (or heavy) NP is placed after a (shorter) PP complement, as in 11. In German, complement ordering in the Mittelfeld is flexible, but pronominal complements precede non pronominal ones, as in 12a and 12b:

11. The waiter brought [to the table] [the wine we had ordered] (from Wasow 2002)

12. (a) Der alte Mann hat ihm das Buch geschenkt (from Bader and Häussler 2010)
    literally 'the old man has him the book offered'
    (b) Der alte Mann hat es seinem Sohn geschenkt
    literally 'the old man has it his son offered'

Another factor is semantic connectedness. As Behagel wrote in 1932: "what belongs close together conceptually also gets placed close together". Wasow (2002), in the Aligned Hansard corpus, observes more PP NP orderings when the PP forms a collocation with the verb, as in: John took [into account] [his personal acquaintances].

Bresnan *et al.* (2007) and Bresnan and Ford (2010) present the results of several quantitative and experimental studies of the dative alternation in different varieties of English. For American English, they use the spoken corpus Switchboard, which manifests an overall tendency to prefer the

double object construction (*give someone something*) over the prepositional construction (*give something to someone*). They show that this general tendency is correlated with a high proportion of instances of the verb *give* in the corpus, and that other ditransitive verbs (like *bring*) may present the opposite tendency, as in example 13 where the prepositional construction is preferred (with theme and recipient of equal length and both definite and non pronominal). They also show that several factors are statistically significant in determining the preferred ordering: a definite, pronominal, animate, short recipient tends to precede an indefinite, non-pronominal, inanimate, longer theme (thus favoring the double object construction, as in example 14),

13. (a) I had to go out and find some wood and put up some kind of a structure to house that pony, because he brought the pony to my children (Switchboard corpus)
(b) ?[…] because he brought my children the pony.

14. (a) I told him if you want to give me a present for Christmas, give me a backpack. (Switchboard corpus)
(b)? […] give a backpack to me.

Contrary to previous studies where the different factors were not clearly set apart, Bresnan *et al.* (2007) and Bresnan and Ford (2010) were able, using multifactorial models, to measure their respective correlations and to show their significance independently of one another. They also performed psycholinguistic experiments asking subjects to judge the naturalness of one construction over the other, or to continue a given sentence after the verb (given a certain context). They observe the same preferences as in their corpus studies.

More generally, psycholinguistic studies have shown a tendency to order definite, animate, pronominal arguments before indefinite, non-animate, non-pronominal ones. This may favor the choice of the passive over the active construction, as in 15a preferred over 15b in Japanese (Tanaka *et al.* 2011); this may favor object left dislocation, as in 16a preferred over 16b in Spanish (Prat-Sala & Branigan 2000).

15. (a) ryoshi-o booto-niyotte hakobareta
'the-fisherman(subject) by-the-boat was-carried'(passive)
(b) booto-ga ryoshi-o hakonda
'the-boat(subject) the-fisherman(object) carried' (active)

16. (a) A la mujer la atropelló el tren
'the woman(object) the train(subject) hit her'

(b) El tren atropelló a la mujer
'the train hit the woman'

In these studies, which are often devoted to subject and complement ordering, it is sometimes difficult to set apart the factors influencing thematic role or syntactic function assignment from those influencing mere constituent ordering. But it is striking that the same factors seem to affect the ordering of nominal conjuncts (which share the same grammatical function and the same thematic role) as shown in Branigan *et al.* (2007). For example, people prefer a short conjunct before a long one (as in *Paul and the next door neighbors* preferred over *the next door neighbors and Paul*), and as far as animacy is concerned, they prefer an animate conjunct before an inanimate one as in *the fisherman and the boat*, preferred over *the boat and the fisherman*.

Our goal is to use similar methods (quantitative corpus based and experimental psycholinguistic) to test whether the same factors are also in play for complement ordering in French.

3   Corpus description

The database of this study consists of 1434 sentences containing a ditransitive verb followed only by a subcategorized NP and a subcategorized PP. These sentences contain 182 different verb lemmas and were extracted from 4 corpora: the French Treebank (henceforth FTB) which comprises about 20 000 sentences from the newspaper *Le Monde* (1989-1994) fully annotated and manually validated for syntax purposes (Abeillé *et al.* 2003, Abeillé *et al.* 2004); the Est-Républicain corpus (henceforth ER) which comprises 148 millions words and has been tagged for part-of-speech and lemmatized by Seddah *et al.* (2012); the radio corpus ESTER-2 (henceforth ESTER) with extracts from *France Inter*, *France Info*, *France Culture* and *Radio Classique* 2003-2004 (Galliano *et al.* 2006) and the French part of the spoken corpus C-ORAL-ROM (henceforth CORAL) (Cresti & Moneglia 2005).

3.1   Extraction and annotation of the data

The sentences containing a ditransitive verb followed by an NP and a PP were automatically extracted from the functionally annotated part of the French Treebank (Abeillé *et al.* 2003, Abeillé *et al.* 2004). This part of the data comprises 339 sentences with 146 different verbs. As for ER,

ESTER-2 and CORAL corpora, we manually selected the sentences using the verbal lemmas to access the intended pattern. In order to make the three samples as representative as possible, we respected the proportions of verbs found in the FTB. In the ER part of our database, there are 782 sentences with 150 different verbal lemmas. The ESTER part is composed of 65 verbal lemmas representing 204 sentences. Finally, in the CORAL part, there are 42 verbal lemmas representing 110 sentences.

In order to investigate the role of the factors we are interested in, we annotated each extracted sentence for these factors, capturing relevant information with 11 variables that are summarized in Table XX-1.

Name	Description	Possible values
ORDER	attested order of verbal complements	*NP-PP or PP-NP*
DEFNP	definiteness of the NP	*definite* or *not*
DEFPP	definiteness of the PP	*definite* or *not*
PRONP	pronominality of the NP	*pronoun* or *not*
PROPP	pronominality of the PP	*pronoun* or *not*
ANIMNP	animacy of the NP	*animate* or *not*
ANIMPP	animacy of the PP	*animate* or *not*
NP-PPLEN	length of NP – length of PP (number of words)	value of the difference (log scale)
PREPNOUN	the SP consists of a preposition + a noun	*true* or *false*
VERBSEM	the verbal lemma + the semantic class	
CORPUS	corpus	*ER, FTB, ESTER or CORAL*

Table XX-1: The studied variables

First, we encoded the attested order of the complements in each sentence, with a binary variable ORDER, taking the value *NP-PP* or *PP-NP*.

Then, we collected information about the NP and PP complements: definiteness (DEFNP and DEFPP) and pronominality (PRONP and PROPP). A complement is considered as definite either if it contains a definite, possessive or demonstrative determiner, or if it consists of a proper noun. Otherwise, it is marked as indefinite. Pronominality is defined by the presence of personal, demonstrative, possessive or indefinite pronoun.

In order to take into account the effect of weight and lightness, we used a measure of length, which is computed as the NP length in words minus the PP length in words (NP-PPLEN)[1]. In the case of oral corpora, we considered the transcription as a good approximation of the number of words produced by the speaker and thus counted the repetition of a word as one word (e.g. the PP "*en en marche*" 'on the march' contains 3 words)[2].

We also included the internal structure of the PP with the variable PREPNOUN. The variable takes the value *true* if the PP of the sentence consists in a preposition followed by a bare noun (e.g. *en place* 'in place'); otherwise its value is *false*.

Finally, we manually annotated two semantic features: the semantic classes of the verbs in context and the animacy of the referent of the NP and the PP. More details about the annotation procedure and the inter-annotator agreements are given in Thuilier and Danlos (2012). The semantic classification of the verb is based on the dictionary *Les Verbes du Français* (Dubois & Dubois-Charlier 1997). This is a hand-written resource containing 25 610 verbal entries, representing 12 310 verbs classified according to their syntactico-semantic properties. We used the broader level of classification, which corresponds to the following 14 semantic classes:

17. Generic classes of Dubois and Dubois-Charlier
    i.   **C**: communication
    ii. **D**: donation/privation
    iii. **E**: entrance/exit
    iv. **F**: to hit/to touch
    v.   **H**: physical condition/behavior
    vi. **L**: locative
    vii. **M**: movement in place
    viii.**N**: to provide/to remove
    ix. **P**: psychological verbs

---

[1] Based on corpus studies, Wasow (1997, 2002) showed that measures in terms of words or in terms of number of syntactic nodes are almost equivalent in explaining constituent ordering in English. Postulating that is also the case in French, we used the simpler measure: the number of words.

[2] In her dissertation, Thuilier (2012) showed that, in order to account for the relative order of postverbal complements in French, number of words is a better measure than number of syllables. This can be explained by the fact that the effect of weight is partly determined by the complexity of the constituents and the number of syllables gives fewer indications about the complexity of a constituent than the number of words.

x. **R**: achievement/setting up
xi. **S**: to grab/to grip/ to own
xii. **T**: transformation/change
xiii. **U**: to combine/to bring together
xiv. **X**: auxiliary verbs

Each verb has been annotated in context using the online version[3] as annotation guidelines. Note that the resource has not been conceived for an annotation task. Thus, uses of verbs found in corpora do not systematically correspond to lexical entries in the dictionary. For example, the database contains occurrences of the verb *mettre* 'to put' employed with predicative PPs, as in *mettre en valeur* 'to emphasize', that are not listed in the dictionary. In those cases, we approximated the meaning of the verb in context with the closer lexical entry available. With this semantic annotation, we created the VERBSEM variable, which is the concatenation of the verbal lemma and its semantic class in context. The annotated data contain 253 different values for this variable. The VERBSEM variable is used to capture different usages of the same lemma, as in example 18.

18. (a) ADRESSER C: Mme Audinot […] a **adressé** ses félicitations aux élèves et aux professeurs qui les ont encadrés
'Ms Audinot offered her congratulations to the students and the professors who supervised them'
(b) ADRESSER E: Les personnes […] marqueront leur intérêt en **adressant** leur chèque à la mairie de Vaufrey ou de Glère
'People will show their interest by sending their cheque to the town council of Vaufrey or Glère'

The animacy of the referent of the NP and the PP has been annotated exploiting the same 9 categories as Zaenen *et al.* (2004) used for English[4]. We created two other variables which encode the animacy of the verbal complements: ANIMNP and ANIMPP. In order to reduce the sparseness of the data, we merged the 9 possible categories into a binary distinction: animate (human, animal, organization) vs. inanimate.

Finally, we included as a variable the corpus from which the sentence was extracted (CORPUS) in order to observe if there is variation between corpora.

We grouped this information together in the form of a datatable, where each line corresponds to a sentence and each column to an annotated

---

[3] http://rali.iro.umontreal.ca/Dubois/
[4] The 9 categories are: human, animal, organization, concrete, non-concrete, machine, vehicle, place and time.

variable. We observe an average 70.4% preference for NP-PP order. The next sections are devoted to the description of the effect of the 10 predictive variables on the ORDER variable.

3.2   Descriptive results

3.2.1   Pronominality

In the database, pronominal PPs are headed by personal pronouns (19), demonstrative pronouns (20), or indefinite pronouns (21).

19. […] ramènent à **eux** toutes les activités (CORAL)
    'bring back to them all the activities'

20. […] séparons le cas de la Syrie de **celui** de l'Iran (ESTER)
    'let separate the case of Syria and the one of Iran'

21. […] donner des des idées à **certains** (ESTER)
    'give ideas to some people'

As for NPs, they only consist of demonstrative (22) or indefinite pronouns (23), because personal pronouns are obligatorily realized as clitics before the verb.

22. on met **ça** au four (CORAL)
    'we put this in the oven'

23. […] incitait **chacun** au mutisme (FTB)
    'conduced each one to silence'

There are very few occurrences of pronominal constituents: 2.2% of the NPs and 3.3% of the PPs. Among the pronominal occurrences, there is no clear tendency indicating that pronominal constituents tend to directly follow the verb. Note that in studies concerning English or German, where the preference for pronominal before non-pronominal has been established, pronominality is generally defined more restrictively: indefinite pronouns are excluded. Pronominality is thus linked to given and familiar material. In the database, only 8 PPs are headed by a personal pronoun and 6 of them are adjacent to the verb (75% of pronominal PP before NP). If we add demonstrative *ça*, *cela*, *celui-ci* and its variants to the personal pronouns, we observe that 100% (among 9 occ.) of pronominal NPs and 78,6% (among 14 occ.) of pronominal PPs are

adjacent to the verb. The tendency observed in English and German seems to be borne out. Thus, it can be the case that indefinite pronouns do not pattern as demonstrative and personal pronouns. Moreover, demonstratives and indefinites can be modified by a clause or a PP, as in example 20. This makes the constituent heavier and favours the postposition of the pronoun-headed constituent.

3.2.2 Animacy

In the datatable, animate referents tend to appear adjacent to the verb. Among the 379 animate PPs, 65.6% appear with the NP-PP order, which is less than the general proportion (70.4%). Animate NPs are occurring with this order in 79.1% of the cases (148 NPs). These proportions are significantly different from the general proportion of NP-PP order: for the NP, $\chi^2 = 7.36$ (DF = 1, p<0.01); for the PP, $\chi^2 = 10.6$ (DF = 1, p<0.01). Moreover, in the context of sentences with animate PP and inanimate NP, we observe 66% of NP-PP order. This shows that there is a slight preference for *animate before inanimate* ordering in our data.

3.2.3 Definiteness

Definiteness seems to have the expected influence on verbal complements ordering. First, there is a higher proportion of NP-PP order with definite NPs: 70,4% in the entire datatable, against 72.7% with definite NPs ($\chi^2 = 5.88$, DF = 1, p<0.02). By contrast, definite PPs occur with NP-PP order in 71.4% of the datatable and this proportion is not significantly different from the general one. Nevertheless, when definite PPs co-occur with indefinite NPs, we observe 66.8% of NP-PP order, Thus there is a slight tendency towards PP-NP ordering with definite PP, which complies with the *definite before indefinite* tendency, observed in English and German.

3.2.4 Length of NP and PP

In a SVO language as French, constituents tend to be ordered according to the law of growing constituents (Behagel 1909, Hawkins 1994), that is shorter and less complex constituents tend to appear before longer and more complex ones. According to the NP-PPLEN variable, 70.2% of our data comply with the general principle, *short before long*: in 67.3% of the

sentences with NP-PP order, the PP is longer than the NP; and in 76.9% of
the sentences with PP-NP order, the NP is longer than the PP.

3.2.5   V-PP collocation

In some sentences, the verb and the PP make up collocational sequences,
as *mettre en relief* 'foreground', *mettre en lumière* 'bring sth in light'.
Despite the collocation, the PP can be realized before or after the NP
complement, as shown in examples 24 and 25.

> 24. […] qui essaie de **mettre en relie**f cet instrument (ER)
>   'who tries to foreground this instrument'

> 25. Quoi de mieux qu'un lustre pour **mettre** l'orgueil **en lumière** (ER)
>   'What could be better than a ceiling light in order to bring pride in light'

As observed by Wasow (2002) for English, the particular semantic
connection between the verb and the PP seems to favor the PP-NP
ordering. In fact, when they appear in a collocation, the two-word PPs
occur directly after the verb in 69.2% of the cases, while only 42.6% of the
non-collocational two-word PPs occur in PP-NP order. The datatable
contains only 26 sentences with collocational PPs (i.e. 1.8% of the data).

3.3   *Variations in the data*

The proportion of NP-PP order (70.4%) varies significantly according to
the variables CORPUS and the VERBSEM.

First, the corpus is a source of variation since only 67.8% and 69.1% of
the data show NP-PP order in respectively FTB and ER corpora, whereas,
this order shows up 73.6% and 76.4% of respectively CORAL and ESTER
data.

Second, we observe strong preferences depending on the identity of the
verb, as the four examples of the Table XX-2 show. The verbs *ajouter* 'to
add' and *vendre* 'to sell' favour PP-NP order, while *réduire* 'to reduce'
and *mettre* 'to put' have a preference for the opposite order.
Besides the variation among verbal lemmas, some verbs show different
behaviours according to their semantic class in context. For instance,
mettre has three possible uses: *mettre L* that corresponds to the locative
sense of the verb (26), *mettre D* that refers to a donation meaning (27) and
mettre R that has a sense of realization (28).

	ajouter 'to add' (10)	*vendre* 'to sell' (30)	*réduire* 'to reduce' (20)	*mettre* 'to put' (74)
NP-PP	30%	36.7%	65%	74.3%
PP-NP	70%	63.3%	35%	25.7%

Table XX-2: The variable ORDER according to 4 verbs with the number of occurrences in brackets

26. *mettre L*: […] parce que on on va **mettre** l'en l'enfant dans une position initiale de méfiance par rapport aux autres. (ESTER)
    'because we will put the child in an initial position of wariness of the others'

27. *mettre D*: […] **mettre** son action au service de l'espoir. (ER)
    'to place his action at the service of hope'

28. *mettre R*: C'est l'AVDAM […] qui **met** en oeuvre ce dispositif […]. (ER)
    'It is AVDAM who carries out this operation'

Similarly, the verb *réduire* can be used in two different contexts, as shown in examples 29 and 30. In 29, the verb is used with an indirect complement introduced by the preposition *à* 'to' and its meaning contains the idea of simplifying, whereas in example 30, the complement is headed by the preposition *de* 'by' and the verb means 'to lower'.

29. *réduire E*: […] un "monde qui **réduit** l'espace international à la liste d'un annuaire téléphonique". (FTB)
    'a world who reduces the international space to the list of a telephone directory'

30. *réduire M*: […] l'épidémie de pneumopathie **réduirait** la demande mondiale de 300 mille barils de pétrole par jour en Asie. (ESTER)
    'the epidemic of chronic obstructive pulmonary disease would reduce the world demand by 300 thousand barrels of oil a day in Asia'

Table XX-3 shows that each of these uses has different preferences concerning the complement ordering. *Réduire E* and *mettre L* have a strong preference for NP-PP order while *réduire M* and *mettre R* favor the opposite ordering.

VERBSEM	NP-PP order	VERBSEM	NP-PP order
mettre D (4)	75%	*réduire E* (7)	85.7%
mettre L (51)	88.2%	*réduire M* (13)	53.8%
mettre R (19)	36.8%		

Table XX-3: The variable ORDER according to five different values of VERBSEM with the number of occurrences in brackets

The variation among the data observed in the previous section raises at least one question: are the effects picked out using proportions (length, animacy, definiteness, V-PP collocations) due to general factors playing a role in the complements ordering phenomenon? Or, can these effects be reduced to verbal idiosyncrasies or corpus specificities? In order to shed light on this question, we need a tool that allows taking into account simultaneously the variation among the data and the general factors. Moreover, in the descriptive results, the effect of each variable is measured independently. One can imagine that looking at the effects of the all variables at the same time could reduce the effect of a variable: for instance, if animate referents tend to be shorter than inanimate ones, the effect of animacy could be reduced to weight[5].

Then, in order to have a better picture of the phenomenon by taking into account simultaneously all the variables as well as the variation in the data, we used a multifactorial statistical model, called mixed-effects logistic regression.

4   Multifactorial statistical modelling

The statistical modelling of relative order of postverbal complements has been done using the mixed-effects logistic regression (Agresti 2007; Gelman & Hill 2006). This statistical tool allows one to model the behaviour of a binary variable. More precisely, in our case, it estimates the probability that a PP-NP order will be chosen as a function of the predictive variables presented in the previous section.

---

[5] Looking simultaneously the effect of the all variables could also allow to observe interaction of variables: for example, it could be the case that animate and definite constituents are more likely to appear adjacent to the verb when both characteristics are combined. We will not develop this point in the following sections.

One advantage of the mixed-effects logistic regression model is that it is predictive, in the sense that one can build a model on a set of data and use this model to predict the choice between NP-PP and PP-NP on new data. This way, we can evaluate how well the model generalizes from the training set.

The construction of the model consists in estimating the coefficients that are associated with each variable. Each coefficient can be interpreted as the preference of its variable: in the case of a variable having only positive values, a positive coefficient indicates a preference for PP-NP, and negative one a preference for NP-PP. Besides the predictive variables, also called fixed-effects, mixed-effect models are able to take into account the variation in the data by means of random-effects. In our case, the variables VERBSEM and CORPUS are the random effects in order to model the verbal idiosyncrasies and the corpus specificities. Each value of these variables constitutes a group in the data, which is assigned a randomly varying normally distributed effect in the model. Thus, associating each value of the random effects with a specific coefficient accounts for the different behaviours according to corpora and verbal uses that we described in section 2.3

Using our datatable, we built a model with 8 fixed-effects (DEFNP, DEFPP, PRONP, PROPP, ANIMNP, ANIMPP, NP-PPLEN[6] and PREPNOUN) and 2 random-effects (VERBSEM and CORPUS).

To make the model more compact, we removed the variables that did not have a significant effect on the probability of PP-NP ordering[7]. The only remaining fixed-effect is NP-PPLEN. This means that pronominality, definiteness, animacy and collocational V-PP do not significantly contribute to predict the relative ordering of complements, contrary to what descriptive results suggested. Concerning the variable PREPNOUN, the small number of data presenting collocational V-PP sequences can explain the lack of effect. As for definiteness and animacy, the modelling shows that by taking into account the disambiguated lemma, the corpus and the relative length, these factors have no effect on complements' relative ordering in French. The corpus model is presented in Figure XX-1. For each random effect, the standard deviation of the normal distribution is given; for the fixed-effect, we present the estimate

---

[6] We use the log transformed values of the lengths of NP and PP in order to reduce the effect of outliers. The exact value of NP-PPLEN is thus: log(length of NP) – log(length of PP). More details about logarithmic transformation can be found in Howell (2009: 338-340).

[7] Eliminating the non-significant fixed-effects has been done using the likelihood ratio test. For more details about this method, see Howell (2009: 156-157).

coefficient (Estimate) and the p-value[8] testifying that the coefficient associated with the variable is significantly different from 0 (i.e. the variable has a significant effect).

Formula: ORDER ~ NP-PPLEN + (1 | CORPUS) + (1 | VERBSEM)

Random effects:

Groups	Std.Dev.
VERBSEM	1.24
CORPUS	0.24

Fixed effects:

|	Estimate	P-value
(Intercept)	-1.24	7.22e-07
NP-PPLEN	+2.77	< 2e-16

Figure XX-1: The corpus model

The interpretation of the coefficient associated with LENNP-PP depends on its own value: if NP-PPLEN has a positive value, the variable votes for PP-NP order, whereas, if its value is negative, it favours NP-PP order. Thus, thanks to the multifactorial modelling, the corpora study suggests that animacy, definiteness and pronominality do not affect the relative ordering of verbal complements in French. In order to better understand why animacy and definiteness are not significant, we present the graphics in Figure XX-2. The plot on the right represents the proportion of PP-NP ordering as a function of the relative length of both complements. The points correspond to the sentences with animate NPs, whereas the crosses refer to the sentences containing inanimate NPs. The two lines are the regression lines that best summarize the scatter graphs of the two types of NPs (solid line for animate NPs and dotted line for inanimate ones). The layout of these curves shows that, when the length is taken into account, the behaviour of animate NP and animate PP is very similar. The plot on the left of Figure XX-2 represents the same kind of information for animate and inanimate PP. Both these graphics illustrate the result of the modelling: animacy does not significantly participate in predicting PP-NP order, since other variables are simultaneously used to model the phenomenon. The same observations hold for definiteness.

---

[8] The statistical test used to obtain the p-value is the Wald test. More details about this test can be found in Agresti (2007: 11-13).

The corpus study provides one main piece of information: only three variables (NP-PPLEN, VERBSEM and CORPUS) are useful to explain the complement ordering in the sample we studied. However, this kind of study has a weakness: on the one hand, as it is multifactorial, we can include many factors in the same model, but on the other hand, we cannot easily decorrelate those factors. As shown by the Figure XX-2, animacy is correlated to length in the data and the effect of the former is overridden by the effect of the latter. The corpus study does not allow controlling these correlations.
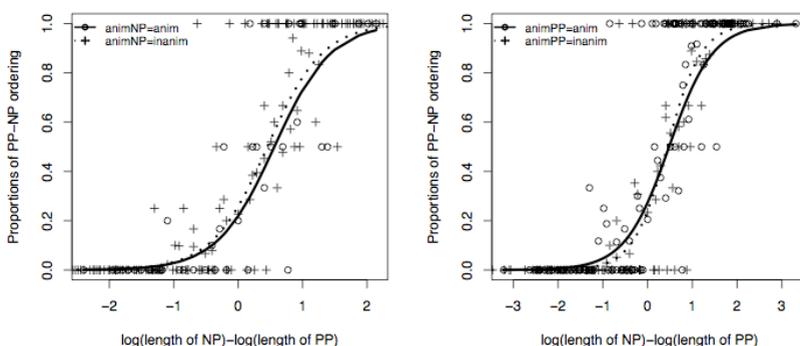
Figure XX-2: Scatter plots representing the proportion of PP-NP order as a function of the relative length, with regression lines that correspond to animate and inanimate NPs (right), and animate and inanimate PPs (left)

5   Psycholinguistic experiments

Correlations may hide some relevant factors. In our case, the corpus study reveals that the weight is a key factor. Can we directly conclude that pronouns, animate referents and definite constituents are not a significant factor for predicting the word ordering of the complements? Obviously not. To test for their significance, we should be able to neutralise the weight factor. This is the purpose of our first pilot experiment. Since relative length plays a key role in our corpus model it may blur the relevance of other factors.

5.1 Pilot study

Since relative length has emerged as a major effect, we created a pilot questionnaire study neutralizing the relative length variable, to enhance the chance of other factors to appear significant. We extracted 23 sentences from our corpora with complements of equal length, as in example 31. The sentences used for the experiment where sampled from the corpus under the constraint that the length of both complement was equal and that 72% of those sentences had NP-PP ordering as observed in the corpus.

31. Pierre fonce dans la nuit porter [la bonne nouvelle] [à sa fiancée] (ER)
    'Pierre runs in the night to-bring the good news to his fiancée'

We tested 25 participants (Université Paris Diderot students) for preferences between NP-PP and PP-NP continuations using a 5-point Likert scale, with each order as an endpoint on the scale:

32. La Poste, d'une part, France Télécom, d'autre part, disposent désormais d'une autonomie financière. Ce changement de statut va
    A. contraindre les deux exploitants publics à plus de rigueur.
    B. contraindre à plus de rigueur les deux exploitants publics.
           A ☐☐☐☐☐B
    'La Poste, one the one hand, France Telecom, on the other hand, have financial autonomy from now on. This change of status will: force both the public operators to be more rigorous'

The test sentences were randomized with distractor items, where the subjects were proposed continuations with and without ellipsis (also taken from corpora).

We coded the results from 1 = strong PP-NP preference to 5 = strong NP-PP preference, with 3 = no preference, 4 weak NP-PP preference and 2 = weak PP-NP preference.

The study confirms the overall preference for NP-PP order (with average rating 3.5). We built a linear mixed model to predict the ratings, with the same predicting variables as in the corpus study (minus complement relative length, plus subject as random effect). In this model, NP definiteness becomes significant ($p < 0.02$), but this is difficult to interpret: NP definiteness appears to favour PP-NP order, which is not what is found in English and other languages where definite tend to precede indefinite (this tendency being often reinterpreted as given precedes new). In our questionnaire, 11 items (out of 23) had definite NPs

and these definite NPs were most of the time combined by collocational PPs (4 items) or definite PPs (5 items as in example 31), and only 2 items had indefinite PPs (as in example 32), so we did not test definiteness *per se*. As in the corpora study pronominality and animacy effects remain non significant. Compared to English and German, the lack of a pronominality effect can be explained by the fact that French has a different strategy (preverbal cliticization) for ordering pronominal arguments. But the lack of an animacy effect is a major surprise, which we tried to confirm with an experiment.

5.2   The animacy experiment

Our pilot study was designed to further observe the data with the weight being neutralized. Its purpose was mostly to shed light on potential factors that might have been hidden by the weight in the corpus. As pointed out by our corpus study, animacy does not seem to play a significant role in ordering French ditransitive complements. This observation remains unexpected since animacy is known to exert a preference on word ordering in several languages (Wasow 2002 ; Branigan *et al*. 2007; Bader and Haussler 2010; Tanaka *et al*. 2011).

In order to get a better idea of the situation for French, we designed an experiment specifically targeted at identifying the status of animacy. Here we go one step further than in the pilot experiment since our test sentences neutralize not only the weight factor but also all other factors known to us: only animacy and word order are allowed to change across sentences. This protocol, inspired by current psycholinguistics methodology, allows us in principle to study the status of animacy independently of other potentially correlated factors that we control in this experimental setting.

The experiment aims to test the following hypothesis: French speakers tend to prefer animate PP before inanimate NP when other variables are neutralized. Since there is a default preference in French for the ditransitive NP-PP order, we expect that an animate PP will be more often used in the PP-NP order than would an inanimate PP if there is an animacy effect.

5.2.1   The design of the experiment

The experiment is designed as a judgment task on a questionnaire. A subject is asked to judge the acceptability of questionnaire examples on a

likert scale ranging from 1 to 5, with 1 being less acceptable to 5 being the most acceptable.

Our two predictor variables are word order (NP-PP/PP-NP) and PP animacy (animate/inanimate) The NP animacy is sytematically set as inanimate. Each test item is made of a context and two possible continuations, one for each possible word order as illustrated below:

33. Il faut que les Israéliens maintenant, dans les prochaines semaines, dans les prochains mois
    ▪ donnent les réponses précises à ces questions.
           1 □□□□□5
    ▪ donnent à ces questions les réponses précises.
           1 □□□□□5

literally 'It is necessary that the Israelis now, in the next few weeks, in the next few month give the precise answers to these questions'

34. Guillaume en était déjà à sa troisième truite, c'est la première fois qu'il en prenait autant en si peu de temps. Fier de sa pêche il se promenait sur la rive
    ▪ et montrait à son copain son précieux butin.
           1 □□□□□5
    ▪ et montrait son précieux butin à son copain.
           1 □□□□□5

literally 'Guillaume was in its third trout, this was the first time that he caught so many of them in such a short time. Proud of his catch, he walked on the bank and showed his loot to his friend'

We manipulate the PP animacy across items in order to have an equal number of items with inanimate (33) and animate (34) PP. Overall a single questionnaire is made up of 16 test items and 12 distractor items.

We submitted the questionnaire to 38 participants, all native speakers of French, selected among Humanities students of Université Paris Diderot. This represents a total of 1216 observations.

 Since the experiment aims at measuring the effect of animacy on word order and nothing else, we tried to control external sources of variation. First, we control for the respective length of each complement. To this end, each item is constructed such that the verbal dependants have an equal number of words. Second, since we know from corpus observation that the verbal lexeme does influence complement ordering, we sampled the verbs involved in test items such that we have an equal proportion of verbs inducing a bias towards NP-PP ordering and PP-NP ordering. Third, we selected contexts that are preferably neither too long, in order to avoid

contextual biases, nor too short in order to keep some naturalness to the task. Fourth, in order to reduce priming effects induced by the questionnaire itself, we also performed a randomized item ordering in each questionnaire, such that no subjects received identical questionnaires.

Despite our efforts to control for external variation, there obviously remain other sources of variations that can be induced by several sources. First, as in any experiment, the subjects themselves may have a natural bias for some ordering stemming from their own linguistic experience. Also, the context and the lexicon used in the items may also introduce some uncontrolled biases.

Apart from external variation, we wanted to ensure that what we believe to measuring is what we are actually measuring. In other words, we want to get subjects to judge the word order variation with respect to animacy and ideally nothing else. In order to maximally ensure that subject's judgments do focus on that specific phenomenon and not on the context, we designed the test items as double continuations focusing on the two possible word orders. The drawback of this choice might be that the participants become aware of the experiment's aims, hence introducing some bias.

Another source of bias may come from the metalinguistic nature of these judgments. We wanted to avoid subjects' judgments being driven by some linguistic norm. To this end, we introduced guidelines inspired by Cowart (1997) on top of each questionnaire, asking the subjects to judge the items as intuitively and quickly as possible.

5.2.2   Analysis of the results

Since the experimental design involves two predictor variables and one response variable, we can measure the response variable independently given each predictor or given the interaction of the two predictors.

First we can observe from Figure XX-3 (left) that independently of animacy considerations, subject' judgments show an overall preference for NP-PP order (mean M=4.06; standard deviation SD=1.07) over PP-NP order (M= 3.68; SD=1.09) which echoes to the corpus frequencies observed above.

As illustrated by Figure XX-3 (right) we can observe that, independently of word order, the animacy of the PP has some influence over the subjects' judgments: subjects do slightly prefer items where the PP is animate (M=3.93; SD=1.09) over inanimate PPs (M=3.80; SD=1.10)
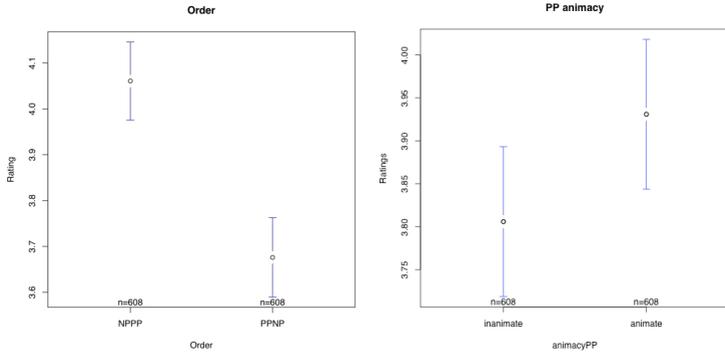
Figure XX-3: Subjects' ratings according to predictor variables

Finally, we can observe the interaction between these two variables (Figure XX-4), which is the important observation here. For the NP-PP case, animacy does not seem to influence the judgments. The judgments do pattern with the slight preference for animate PPs that we just observed before. Now for the PP-NP case, judgments do differ, in that the preference for the PP-NP order seems increased when the PP is indeed animate.
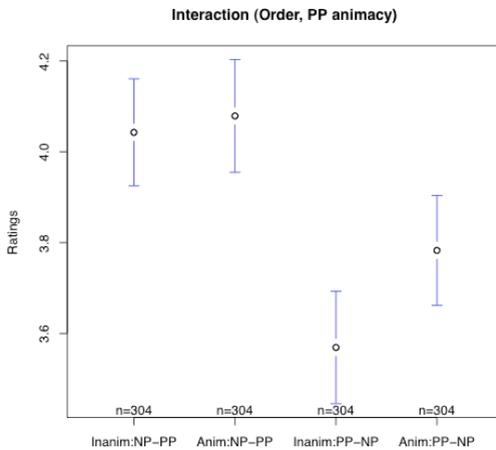
Figure XX-4: Subjects' ratings according to the interaction of the predictor variables.

It remains to assess whether our informal observations are due to mere chance or not. That is, do the observed differences reflect the properties of the specific sample of test items and subjects chosen for the experiment or could these differences reveal some property of French in general?

To provide an answer to this question, we modelled these data with a mixed effect linear regression model (Gelman & Hill 2006), attempting to predict the subject's judgment (**J**) given the order of the ditransitive complements (**O=PP-NP**) and the animacy of the PP (**A=animate**) and their interaction. We included in the model two random effects: one for the subjects (**Subj**) and one for test items[9] (**Item**):

$$J_{i,j} = \alpha + \beta_1 O * \beta_2 A + (1 \mid Subj_i) + (1 \mid Item_j) + \varepsilon$$

By fitting the model to our 1216 observations we observe that the interaction term can be safely removed from the model (χ^2; p= 0.12, the interaction is not significant) whereas the two random effects cannot be eliminated (χ^2; p < 0.001 for each of them). A fit without interaction eventually yields the following coefficient estimates:

	Estimate	Std. Error	t value	p-value
(Intercept)	3.99836	0.09722	41.13	0.0001
A=animate	0.12500	0.10018	1.25	0.2274
O=PP-NP	-0.38487	0.05753	-6.69	0.0001

The model does confirm that neither animacy nor its interaction with word order are significant factors for predicting the judgment made by the subjects. Our experiment focuses on the interaction between word order and animacy. The interaction of word order and animacy is the key factor in this experiment. Since it can be removed from the model, this means that it has absolutely no effect on the judgments made by the subjects.

---

[9] Introducing these two random effects become a common practice in psycholinguistics (Baayen *et al.* 2008). The random intercept set for subjects attempts to model that subjects may have their own bias for some judgment rates (low or high) while the random intercept set for items attempts to model that each item may itself introduce a bias towards a low or high judgment. This bias may be induced for whatever reason such as the lexical items being part of the item, the context used in the item, etc.

To further refine our scope of investigation, we can restrict our data to PP-NP ordering, hence ignoring observations with the other order. In this more specific case, we observe a more important difference in judgments: PP[animate]-NP has mean 3.78 and standard deviation 1.07 while PP[inanimate]-NP has mean 3.57 and standard deviation 1.10. Modelling this second scenario with the same model as before but on this restricted set of observations does lead to similar results:

	Estimate	Std. Error	t value	p value
(Intercept)	3.5691	0.1911	18.677	0.0001
A=animate	0.2138	0.2466	0.867	0.3382

Given this model and the observations considered in our experiment, we cannot conclude that animacy has an effect on subject's judgments. Now, can we definitively conclude that animacy has no effect on the ordering of French postverbal complements? Although our experiments do cast doubt on its significance, we cannot draw here a definitive conclusion.

Indeed, our experimental protocol tried to control for correlations, but several additional reasons can explain the lack of significance. First of all, the experimental protocol itself–the questionnaire–may not be sensitive enough to detect this effect. As pointed out earlier, the design of the questionnaire does carry its own noise, e.g. by setting contexts for introducing test items (even though we tried to keep them minimal in designing the questions). Moreover since the observed difference in means is small, the lack of effect may also come from a lack of statistical power in the design[10]: we could eventually detect such an effect by adding more subjects to our observations. Finally, asking subjects to perform a metalinguistic acceptability judgment is perhaps too indirect. More direct and more sensitive methods like sentence recall tasks used in Tanaka *et al.* (2011) or eye tracking might well turn out to deliver different conclusions.

6   Conclusion

---

[10] Computing the statistical power with a mixed effect model is a non trivial task. Under the simplifying assumption that our current experiment reveals the true estimates, a simulation (Gelman & Hill 2006: chap. 20) suggests that the number of subjects required to get a power of 80% would be more than an hundred for the interaction and close to a thousand for the animacy factor itself.

We have presented the factors, which influence postverbal complement ordering in French. Existing studies (Blinkenberg, 1928-1933 Berrendonner 1987, Schmitt 1987, Abeillé Godard 2004, 2006) have proposed that the weight and the discourse status (given or new) of the complements play a role, as well as the verb meaning, but they are not based on quantitative data. To remedy this lacune, we annotated extracts from 4 corpora (two written and two spoken) for several factors: complement length, animacy, definiteness, pronominality, collocationality of verb-preposition combinations, verb lemma and verb semantic class (following Dubois and Dubois Charlier 1997's classification). We first observe a general preference for direct before indirect complements, in all 4 corpora. We also observe that the following factors may contradict this general tendency: the length of the NP (relative to that of the PP) and the verb lemma (with its semantic class) may reverse the preference and drive an indirect before direct complement ordering. In order to evaluate the respective weights of these factors, and to abstract away from the specificity of each corpus, we have also constructed a multifactorial statistical model, following the methodology of Bresnan *et al.* (2007) and Bresnan and Ford (2010). In this model, the three factors that appeared as statistically significant are the relative length of the complements, the verb combined with its semantic class and the corpus. Contrary to similar studies performed on English or German, pronominality, definiteness and animacy do not play a significant role. In order to complement this result, we conduct two psychological experiments, asking subjects to judge sentences (identical or similar to the ones in the corpora) with equal complement length, and one complement ordering or the other, randomly presented with distractors. Contrary to English and German, again, animacy did not play a significant role, nor did definiteness.

Bibliography

Abeillé, A. & N. Barrier. 2004, "Enriching a French Treebank". In *Proceedings of Language Ressources and Evaluation Conference*, Lisbon.

Abeillé, A., L. Clément, & F. Toussenel. 2003, "Building a treebank for French". In *Treebanks*. Dordrecht : Kluwer.

Abeillé A. & D. Godard, 2004, "De la légèreté en syntaxe". *Bulletin de la Société de Linguistique de Paris*, XCIX(1), 69–106.

Abeillé A. & D. Godard, 2006, "La légèreté en français comme déficience de mobilité". *Lingvisticae Investigationes*, 29(1), 11–24.

Agresti A. 2007. *An Introduction to Categorical Data Analysis*. Wiley.

Arnold J. E., T. Wasow, A. Losongco & R. Ginstrom, 2000, "Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering", *Language*, 76(1), 28–55.

Baayen R. H., D. Davidson & D. Bates, 2008, "Mixed-effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language*, 59(4), 390–412.

Bader M. & J. Häussler, 2010, "Word order in German : A corpus study", *Lingua*, 120, 717–762.

Behaghel O. 1909, "Von deutsher wortstellung", *Insogermanische Forschungen*, 25.

Berrendonner A. 1987, "L'ordre des mots et ses fonctions", *Travaux de linguistique*, 14/15, 9–19.

Branigan H. P. & E. Feleki, 1999, "Conceptual accessibility and serial order in Greek language production", In M. Hahn & S. C. Stoness (eds.), *Proceedings of the 21st Conference of the Cognitive Science Society*, Mahwah: Erlbaum, 96–101.

Branigan, H., M. Pickering & M. Tanaka, 2008, "Contribution of animacy to grammatical function assignment and word order during production", *Lingua* 118.

Bresnan, J., A. Cueni, T. Nikitina, & H. Baayen, 2007, "Predicting the dative alternation", In G. Boume, I. Kraemer, & J. Zwarts (eds.), *Cognitive Foundations of Interpretation*, Amsterdam: Royal Netherlands Academy of Science.

Bresnan J. & J. Hay, 2008, "Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English", *Lingua* 118.

Bresnan, J. & M. Ford, 2010, "Predicting syntax: Processing dative constructions in American and Australian varieties of English" *Language 86* (1).

Cowart W. 1997, *Experimental Syntax: Applying Objective Methods to Sentence Judgments*, Thousand Oaks: Sage Publications.

Cresti E. & M. Moneglia (eds.), 2005, *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam: J. Benjamins.

Dubois J. & F. Dubois-Charlier, 1997, *Les verbes français*, Paris: Larousse-Bordas.

Galliano S., E. Geoffrois, G. Gravier, J-F. Bonastre, D. Mostefa & K. Choukri, 2006, "Corpus description of the ESTER Evaluation Campaign for the Rich Transcription of French Broadcast News", *Proceedings of the 5th international Conference on Language*

Resources and Evaluation (LREC 2006), Genoa: European Language Resources Association (ELRA), 139-142.

Gelman A. & J. Hill, 2006, *Data Analysis Using Regression and Multilevel/ Hierarchical Models*, Cambridge: Cambridge University Press.

Hawkins, J. 1994, *A performance theory of order and constituency*, Cambridge: Cambridge University Press.

Howell, D. C. 2009, Statistical Methods for Psychology, 7th edition, Cengage Lerning.

Kempen G, Harbusch K. 2004. "A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment", In T. Pechmann & C. Habel (eds.), *Multidisciplinary approaches to language production*, Berlin: Mouton de Gruyter, 173–181.

Prat-Sala M. & H. P. Branigan, 2000, "Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish", *Journal of Memory and Language* 42, 168–182.

Schmitt C. 1987a, "À propos de l'impact de la sémantique sur la séquence des compléments d'objets en français moderne", *Travaux de linguistique et de littérature* 25(1), 283–298.

Schmitt C. 1987b, "Sémantique et prédétermination de l'ordre des mots en français contemporain", *Travaux de linguistique* 14/15, 21–31.

Seddah D., M. Candito, B. Crabbé & E. Henestroza Anguiano, 2012, "Ubiquitous usage of a French large corpus: Processing the Est-Republicain corpus", In N. Calzolari, K. Choukri, T. Declerck, M. U. Doǧan, B. Maegaard, J. Mariani, J. Odijk & S. Piperidis (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul: European Language Resources Association (ELRA).

Tanaka, M., H. Branigan, & M. Pickering, 2011, "Conceptual influences on word order and voice in sentence production: Evidence from Japanese", *Journal of Memory and Language* 65(3), 168-182.

Thuilier, J., 2012, *Contraintes préférentielles et ordre des mots en français*, PhD thesis, University of Paris Diderot.

Thuilier, J. & L. Danlos, 2012, "Semantic annotation of French corpora: Animacy and verb semantic classes", In N. Calzolari, K. Choukri, T. Declerck, M. U. Doǧan, B. Maegaard, J. Mariani, J. Odijk & S. Piperidis (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul: European Language Resources Association (ELRA).

Wasow T, 1997, "Remarks on grammatical weight", *Language Variation and Change* 9, 81–105.

Wasow T, 2002, *Postverbal behaviour*, CSLI Publications.

Zaenen A., J. Carletta, G. Garretson, J. Bresnan, A. Koontz-Garboden, T. Nikitina, M. C. O'Connor & T. Wasow, 2004, "Animacy encoding in English: Why and how", In B. Webber & D. K. Byron (eds.), *Proceedings of the ACL 2004 Workshop on Discourse Annotation*, Barcelona : Association for Computational Linguistics, 118–125.