

# Ordre des compléments postverbaux en français : poids et accessibilité discursive

Pegah Faghiri<sup>1</sup>, Juliette Thuilier<sup>2</sup>

<sup>1</sup>Université de Cologne

<sup>2</sup>Université Toulouse Jean Jaurès & CLLE-ERSS

pegah.faghiri@uni-koeln.de, juliette.thuilier@univ-tlse2.fr

**Résumé.** Dans le domaine postverbal, l'ordre des compléments du verbe est relativement libre (Paul donnera un livre à chacun / Paul donnera à chacun un livre). Nous cherchons à déterminer quels sont les facteurs qui influencent le choix d'un ordre par rapport à l'autre. Nous nous intéressons en particulier à deux facteurs : le poids grammatical et l'accessibilité discursive. En ce qui concerne le poids, nous soulevons la question de sa définition : cette notion doit-elle être définie en termes de complexité syntaxique ou de longueur ? L'accessibilité discursive, définie comme le caractère donné ou nouveau des référents, a été mentionnée comme un facteur pertinent dans certains travaux sur le français (Blinkenberg 1928, Berrendonner 1987), mais aucune étude empirique n'a permis de montrer son rôle dans le domaine postverbal. A partir d'une étude expérimentale de complétion de phrases, nous montrons que ces deux facteurs expliquent en partie l'ordre choisi par les locuteurs. Les résultats obtenus permettent d'affirmer que le poids grammatical doit être défini en termes de longueur des constituants et de complexité syntaxique (dans la lignée de Wasow et Arnold 2005, sur l'anglais), et de montrer que l'effet de complexité est plus important que celui de la longueur. De plus, le statut donné ou nouveau des référents a un effet significatif sur le choix des locuteurs, mais cet effet est de petite taille en comparaison à celui du poids grammatical.

**Abstract. Postverbal complement ordering in French: Grammatical weight and discursive accessibility.** Dans le domaine postverbal, l'ordre des compléments du verbe est relativement libre (Paul donnera un livre à chacun / Paul donnera à chacun un livre). Nous cherchons à déterminer quels sont les facteurs qui influencent le choix d'un ordre par rapport à l'autre. Nous nous intéressons en particulier à deux facteurs : le poids grammatical et l'accessibilité discursive. En ce qui concerne le poids, nous soulevons la question de sa définition : cette notion doit-elle être définie en termes de complexité syntaxique ou de longueur ? L'accessibilité discursive, définie comme le caractère donné ou nouveau des référents, a été mentionnée comme un facteur pertinent dans certains travaux sur le français (Blinkenberg 1928, Berrendonner 1987), mais aucune étude empirique n'a permis de montrer son rôle dans le domaine postverbal. A partir d'une étude expérimentale de complétion de phrases, nous montrons que ces deux facteurs expliquent en partie l'ordre choisi par les locuteurs. Les résultats obtenus permettent d'affirmer que le poids grammatical doit être défini en termes de longueur des constituants et de complexité syntaxique (dans la lignée de Wasow et Arnold 2005, sur l'anglais), et de montrer que l'effet de complexité est plus important que celui de la longueur. De plus, le statut donné ou nouveau des référents a un effet significatif sur le choix des locuteurs, mais cet effet est de petite taille en comparaison à celui du poids grammatical.

## 1 Introduction

Le français présente une certaine liberté dans l'ordonnement des constituants de la phrase. Cette liberté relative aboutit à de la variation dans les données. Par exemple, dans le domaine postverbal, la place relative des compléments du verbe peut varier, comme cela est montré en (1) (exemples tirés de Abeillé & Godard, 2006, p. 12).

- (1) a. Paul donnera [un livre] [à chacun]  
b. Paul donnera [à chacun] [un livre]

Après le verbe, l'ordre par défaut est objet direct (OD) suivi d'objet indirect (OI), soit ici SN SP. Blinkenberg (1928, p. 179) l'avait déjà signalé et cela a été confirmé par les travaux sur corpus (Thuilier, 2012a ; Thuilier *et al.*, 2014) et à partir de résultats expérimentaux (Grant *et al.*, 2014 ; Thuilier *et al.*, 2014). Les questions qui guident le travail présenté dans cet article sont les suivantes : quels sont les facteurs qui influencent le choix d'un ordre par rapport à l'autre? Et plus précisément, quelles conditions favorisent le choix de l'ordre SP SN? Nous nous attacherons à discuter les effets de deux facteurs particuliers, le poids grammatical et l'accessibilité discursive, en exploitant les résultats d'une étude expérimentale.

Ce travail s'inscrit dans le cadre d'un projet plus large qui vise à étudier l'ordre des constituants dans deux langues présentant des caractéristiques typologiques différentes : le français et le persan. L'une des propriétés qui distingue les systèmes de ces deux langues est la position du verbe dans la phrase simple : le persan est une langue à verbe final (OV), alors que le français est une langue à verbe initial (VO). Comme les facteurs de poids grammatical et d'accessibilité discursive sont généralement considérés comme ayant une portée universelle, leur étude comparée dans ces deux langues apportera un éclairage nouveau sur la façon dont des systèmes linguistiques différents les mettent en œuvre.

L'objectif de cet article est de présenter les résultats d'une étude menée sur le français, dans le cadre de ce projet. Pour cela, nous commencerons par faire une revue de la littérature concernant les notions de poids grammatical et d'accessibilité. Nous expliquerons ensuite comment nous avons conçu et mis en œuvre notre étude expérimentale, et nous présenterons les résultats obtenus. En guise de conclusion, nous proposerons une discussion autour de ces résultats.

## 2 Etat de l'art

### 2.1 Le poids grammatical

Le poids grammatical est une notion fondamentale pour l'étude des phénomènes d'ordre des mots. En effet, il est généralement admis que lorsque l'ordre n'est pas contraint de façon catégorique par la grammaire, les constituants tendent à s'organiser linéairement en fonction de leur "lourdeur" respective. Le poids peut être défini soit en termes de complexité structurelle des constituants, soit en termes de longueur de la chaîne composant les constituants. Cette tendance a d'abord été identifiée par Behaghel (1909, p. 139), puis documentée par Hawkins (2004) à partir d'études de corpus menées sur une grande variété de langues. L'idée générale qui se dégage est la suivante : dans les langues VO, les constituants tendent à s'organiser du plus court au plus long ; tandis que dans les langues OV, les constituants ont tendance à être placés du plus long au plus court. A titre d'illustration, dans les exemples attestés du français présentés en (2), les compléments du verbe sont ordonnés en suivant l'ordre *court avant long*, respectant la tendance observée dans les langues VO. A l'inverse, en persan, qui est une langue OV, les compléments du verbe ont tendance à

s’organiser selon un principe *long avant court*, comme cela est montré dans les exemples attestés présentés en (3)<sup>1</sup>.

- (2) a. [...] donna [des idées]SN [aux gérants des sicav court terme]SP (*FrenchTreeBank* – FTB, Abeillé *et al.*, 2003 ; Abeillé & Barrier, 2004)  
 b. [...] donne [à ses concitoyens]SP [l’impression d’avoir perdu toute capacité d’initiative]SN. (FTB)
- (3) a. [...] [az in čāh=e margbār=i ke xod ijād karde bud]SP [ānhā=rā]SN  
 de ce puits=EZ funeste=RSTR que soi création faire.PP être.PST.3SG eux=MDO  
 be-rahān-ad  
 SBJV-libérer.PRS-3SG  
 ‘... les libère de ce puits funeste qu’il avait créé lui-même.’ (Corpus de Bijankhan, Bijankhan, 2004)  
 b. [...] [no’=i ehsās=e mahārat va kefāyat]SN [be fard]SP  
 sort=INDF sentiment=EZ compétence et suffisance à individu  
 be-dah-ad  
 SBJV-donner.PRS-3SG  
 ‘... donne un certain sentiment de compétence et de suffisance à l’individu.’ (Corpus de Bijankhan)

Certains auteurs ont émis l’hypothèse selon laquelle le poids grammatical doit être défini en termes de complexité syntaxique et non de longueur (Ross, 1967 ; Chomsky, 1975, parmi d’autres). Des travaux sur corpus menés sur l’anglais (Wasow, 1997 ; Wasow, 2002) ont montré que le poids grammatical agit de manière relative et graduée, et que les mesures de longueur (nombre de mots) et de complexité (nombre de nœuds syntaxiques ou nombre de syntagmes) rendent compte de façon comparable de l’ordre attesté. Les données de corpus ne semblent donc pas permettre de démêler le rôle de la longueur et de la complexité, ce qui n’est pas étonnant étant donné la forte corrélation entre ces deux dimensions. Wasow et Arnold (2005) ont mené une étude sur l’anglais à partir d’un questionnaire d’acceptabilité et ont montré que, dans le domaine postverbal, l’acceptabilité des locuteurs interrogés est affectée par la longueur des constituants, indépendamment de leur complexité. Le travail de Stallings *et al.* (2011) a confirmé les résultats obtenus sur corpus, en montrant que la longueur relative des constituants postverbaux (en nombre de mots) influence significativement l’ordre choisi par les locuteurs dans des tâches expérimentales contrôlées.

Les effets du poids grammatical peuvent être expliqués du point de vue du traitement du langage. L’idée générale est que l’ordre croissant ou décroissant du poids permet de minimiser la distance entre le verbe et ses dépendants et ainsi de faciliter le décodage de la phrase (Hawkins, 1994 ; Gibson, 1998 ; Gibson, 2000 ; Hawkins, 2004). Par exemple, selon Hawkins (2004, principe *Early Immediate Constituent*) dans le syntagme verbal présenté en (2-b), l’ordre attesté permet à l’interlocuteur, qui traite la phrase au fur et à mesure qu’elle est prononcée, de connaître la structure générale du SV (à savoir V SP SN) à partir du cinquième mot (déterminant *l’*). En effet, la présence de l’article défini indique assez clairement que le syntagme prépositionnel est terminé et que commence un syntagme nominal, nouveau constituant immédiat du SV. Si l’ordre était V SN SP, alors l’interlocuteur ne connaîtrait la structure du SV qu’au moment où la préposition *à* est prononcée, c’est-à-dire au onzième mot. Cette idée est résumée dans les schémas présentés en (4).

- (4) a. donne à ses concitoyens l’impression d’avoir perdu toute capacité d’initiative  
 1 2 3 4 5

- b. donne l'impression d'avoir perdu toute capacité d'initiative à ses concitoyens  
1 2 3 4 5 6 7 8 9 10 11

Il est à noter qu'à la différence d'Hawkins qui considère la distance entre la tête et son dépendant comme équivalente au nombre de mots les séparant, Gibson prend en compte le nombre d'expressions référentielles nouvelles (*new discourse referents*), c'est-à-dire, selon l'auteur, les syntagmes nominaux et verbaux, qui séparent la tête de chaque dépendant et le verbe (Gibson, 1998, p. 12). Par exemple, d'après cette théorie, les adjectifs ne participent pas au calcul de la distance entre un verbe et son dépendant, car ils n'introduisent pas de nouveau référent. Ainsi, dans l'exemple (5), la distance entre le verbe *donner* et le SP [à chacun] est du même ordre, car il y a un seul référent qui intervient entre le verbe et son objet indirect (*livre*).

- (5) a. Paul donnera un livre à chacun.  
b. Paul donnera un livre jaune et rouge à chacun.

Blinkenberg (1928) a mentionné le premier les effets de la longueur sur l'ordre des compléments du verbe en français. Beaucoup plus récemment, Abeillé et Godard (2006 ; 2004) ont appliqué la notion de poids pour rendre compte de l'ordonnement des constituants dans le domaine postverbal en français. Ces auteurs ont établi la notion de *légèreté*, qui se définit comme un type de déficience syntaxique. La *légèreté* explique, entre autres choses, le manque de mobilité de certains mots ou constituants. C'est le cas par exemple des noms nus (nom sans déterminant) qui apparaissent obligatoirement en première position après le verbe, en raison de leur caractère *léger*, comme le montrent les exemples (6).

- (6) a. Le pompier a porté secours aux trois victimes.  
b. ??Le pompier a porté aux trois victimes secours.

Le travail à partir de données de corpus de Thuilier (2012a) a montré qu'en français, l'ordre des compléments postverbaux se conforme à l'ordre *court avant long* dans plus de 82% des cas. Mais, les données de corpus ne permettent pas de distinguer un effet de complexité différent de l'effet de longueur en nombre de mots. Cependant, ce travail a pu mettre en évidence que la longueur en termes de nombre de syllabes rend moins bien compte de l'ordre des compléments que la longueur en nombre de mots. Or, le nombre de mots est un indicateur de complexité, car un SN contenant 6 mots est syntaxiquement plus complexe qu'un SN en contenant 3 (dans un SN à 6 mots, soit le nom a plusieurs dépendants, soit un dépendant du nom a lui-même des dépendants). La différence constatée entre nombre de syllabes et nombre de mots suggère donc que l'ordonnement des constituants est tendanciellement plus sensible à une certaine forme de complexité syntaxique en français qu'à la longueur syllabique des constituants.

Le question qui nous intéresse ici est de savoir si le poids peut être réduit à la complexité comme cela est souvent affirmé, ou s'il faut aussi prendre en compte des informations de longueur, en particulier le nombre de syllabes.

## 2.2 L'accessibilité du référent

L'accessibilité linguistique renvoie au statut des référents dans le discours (Chafe, 1976 ; Givón, 1983 ; Gundel *et al.*, 1993). Ce statut évolue à mesure que le discours avance. L'idée est qu'à un moment *t* du discours, certaines informations sont facilement accessibles, alors que d'autres le sont moins. Parmi les paramètres linguistiques qui modulent l'accessibilité d'un référent, on peut citer : le caractère donné (*gi-*

*venness*), la récence et la proéminence syntaxique (Arnold, 2010). Ainsi, un référent *donné*, c'est-à-dire déjà mentionné dans le discours précédent est plus accessible qu'un référent *nouveau*. Dans le même ordre d'idées, un référent qui a été évoqué dans la proposition précédant la proposition produite est plus accessible qu'un référent mentionné deux propositions plus tôt. Enfin, le contexte syntaxique dans lequel le référent a été mentionné joue sur son accessibilité : par exemple, la fonction sujet et les constructions clivées tendent à rendre un référent plus accessible. Le statut du référent a des conséquences sur les formes linguistiques utilisées (l'exemple prototypique est l'utilisation de pronoms pour les référents hautement accessibles), mais également sur les structures syntaxiques utilisées. Par exemple, Prince (1981) présente une analyse détaillée d'une narration orale en anglais, et montre qu'aucun des référents non accessibles (au sens de non mentionnés précédemment) ne remplit la fonction sujet ; alors que 77% des référents accessibles (au sens de mentionnés précédemment ou évoqués par la situation) sont en position sujet. Dans le cadre de cet article, nous restreignons la notion d'accessibilité à la dimension *donnée* ou *nouvelle* des référents. Le principe général peut être résumé de la façon suivante : les informations données tendent à apparaître avant les informations nouvelles (Clark & Clark, 1977 ; Gundel, 1988). Gundel (1988, p. 229) formule les choses de la façon suivante : « *State what is given before what is new in relation to it* ». Arnold *et al.* (2000) ont observé que, dans le domaine postverbal, les données de corpus de l'anglais montraient une tendance à organiser linéairement les dépendants du verbe selon le principe *donné avant nouveau*, quand les dépendants du verbe étaient de même longueur en nombre de mots.

L'accessibilité est également une notion utilisée en psycholinguistique. Elle renvoie à un faisceau de propriétés qui facilite l'accès aux référents dans le traitement du langage, en compréhension et en production. Parmi les propriétés renforçant l'accessibilité d'un référent, il y a des propriétés intrinsèques comme le caractère animé (*animacy*) et des propriétés contextuelles comme son statut discursif ou sa longueur. Ainsi, un référent est particulièrement accessible s'il est animé, court et a été récemment mentionné. Les théories incrémentales de la production (Levelt, 1989 ; Bock & Levelt, 1994 ; De Smedt, 1994) supposent que, dans la parole spontanée, les informations sont traitées au fur et à mesure qu'elles deviennent disponibles. Cela implique que l'ordre dans lequel les éléments apparaissent est lié à l'ordre dans lequel le locuteur accède aux informations. Un constituant dont le référent est animé, court et récemment mentionné est plus accessible qu'un référent ne réunissant pas ces trois propriétés, et par conséquent, est susceptible de le précéder dans le processus de production.

Si l'on s'en tient au caractère donné ou nouveau du référent, les constituants contenant de l'information nouvelle sont plus difficiles à produire car choisir l'expression linguistique appropriée pour renvoyer au référent prend plus de temps. En effet, faire référence à des constituants donnés demande moins de ressources dans la mesure où les représentations conceptuelles et linguistiques sont déjà activées (Arnold *et al.*, 2000, p. 33). Ainsi, respecter le principe *donné avant nouveau* permet d'optimiser le temps de planification et de production du locuteur : en produisant les éléments donnés en premier, il a plus de temps pour résoudre les difficultés liées aux éléments nouveaux.

En ce qui concerne les travaux sur le français, Blinkenberg (1928, p. 181) suggère que si le complément indirect reprend un terme connu, alors ce dernier peut être facilement placé avant le complément direct. Les travaux de Berrendonner (1987) s'appuient sur la notion de *focus*, définie comme « *la partie [du contenu de l'énoncé] qui est présentée [...] comme l'information maximale pertinente au regard de l'état courant du savoir partagé* » (Berrendonner, 1987, p. 10). Selon l'auteur, la dernière position de la phrase est le lieu privilégié du focus. Comme le statut de focus et celui d'information nouvelle sont fortement corrélés, l'auteur s'attend à trouver l'information nouvelle en dernière position. Par exemple, dans une réponse à une question partielle, il est plus naturel, selon l'auteur, de placer le contenu répondant à la question en dernière position comme dans les exemples (7) et (8).

- (7) a. A qui Moscou a-t-il envoyé un message d'appui ?  
b. Moscou a envoyé un message d'appui au mouvement des cent-un.
- (8) a. Qu'est-ce que Moscou a envoyé au mouvement des cent-un ?  
b. Moscou a envoyé au mouvement des cent-un un message d'appui.

Un travail exploratoire sur corpus a été mené par Thuilier (2012a). Dans un échantillon de 166 syntagmes verbaux contenant deux compléments de longueur comparable (différence de longueur n'excédant pas deux mots), l'auteur a annoté les référents des SN et des SP selon deux catégories de Prince (1981): *nouveau* et *évoqué*<sup>2</sup>. Les observations sont les suivantes : lorsque le référent du SN est nouveau, l'ordre SP SN est attesté dans 23% des cas, alors que quand le SN renvoie à un référent donné, cet ordre n'est plus représenté que dans 12% des cas. La tendance observée va dans le sens attendu, car le statut nouveau du SN semble corrélé à une préférence pour la dernière position dans la phrase. Cependant, cette différence n'est pas statistiquement significative. De plus, Thuilier (2012a) a interrogé 28 locuteurs du français *via* un questionnaire afin de tester l'effet de l'ordre des compléments du verbe et du statut donné ou nouveau du référent du SP, sur les jugements d'acceptabilité. Les résultats font clairement apparaître une préférence pour l'ordre par défaut (SN SP), mais n'ont pas mis en évidence une préférence pour l'ordre *donné avant nouveau*.

Enfin, quelques travaux sur le français se sont intéressés à un autre aspect de l'accessibilité : le caractère animé. Globalement, ces travaux ont échoué à montrer qu'il existait une préférence pour l'ordre *animé avant inanimé* dans la domaine postverbal. En effet, dans l'étude sur corpus et le questionnaire d'acceptabilité de Thuilier *et al.* (2014), le caractère animé n'explique pas les ordres attestés ni les jugements des locuteurs. De plus, l'expérience en production (*sentence recall*, rappel de phrases) mise en place par Grant *et al.* (2014) montre clairement que le choix de la voix active ou passive est influencée par le caractère animé du patient d'un verbe transitif. En revanche, le caractère animé du thème d'un verbe ditransitif n'a aucun effet sur l'ordre produit par les sujets de l'expérience. Autrement dit, le fait que le thème soit animé ou inanimé, comme dans l'exemple (9), n'influence pas le choix de l'ordre dans lesquels les deux compléments du verbe sont produits par le locuteur.

- (9) a. Le chef de projet a confié {un nouveau budget / un agent commercial} à un décorateur  
b. Le chef de projet a confié à un décorateur {un nouveau budget / un agent commercial}

En résumé, les travaux existants n'ont pas montré d'effets clairs d'accessibilité sur l'ordre des compléments du verbe en français, que ce soit en termes de caractère donné ou en termes de caractère animé. Cependant, alors que le caractère animé a été étudié par le biais d'une grande variété de données (données attestées, jugement d'acceptabilité et production contrôlée de phrases), le caractère donné ou nouveau du référent ne l'a été que par le biais d'une étude exploratoire sur corpus et d'un questionnaire d'acceptabilité. Or ces outils d'investigation présentent des limites :

- les données attestées présentent des corrélations non contrôlées, ce qui pourrait masquer l'effet d'accessibilité. De plus, la tendance observée par Thuilier (2012a) correspond à une légère préférence pour mettre les référents *nouveaux* en dernière position, mais le faible nombre de données utilisées est peut-être insuffisant pour détecter un effet significatif.
- il est envisageable que les locuteurs ne soient pas sensibles à l'effet d'accessibilité dans une tâche métalinguistique de jugement, mais que ce facteur influence leur production.

La question qui nous intéresse ici est donc de savoir si dans une tâche de production, les locuteurs du français sont sensibles au caractère donné ou nouveau des référents.

## 2.3 Accessibilité et poids

Les travaux sur l'anglais de Ariel (1990) et Arnold (1998) ont montré que le poids et l'accessibilité discursive sont hautement corrélés. Dans une phrase, les constituants lourds et longs ont tendance à être moins accessibles car le locuteur a besoin d'une description linguistique longue et fournie pour référer ; tandis que les constituants légers sont utilisés quand le référent est très accessible. Le cas des pronoms est l'exemple prototypique de la corrélation entre la petite taille d'un constituant et sa haute accessibilité discursive.

La forte corrélation qui existe entre le poids et l'accessibilité discursive amène à se demander si les deux facteurs jouent un rôle, ou si l'un des deux facteurs est réductible à l'autre. Hawkins (1994) estime que l'accessibilité discursive n'est qu'une conséquence de l'effet de poids, autrement dit que, à lui seul, le poids permet d'expliquer les ordres attestés dans les langues. Cependant, à partir d'un corpus de phrases de l'anglais et des résultats d'une expérience d'élicitation de données, Arnold *et al.* (2000) apportent des arguments contredisant le point de vue Hawkins. Ils affirment que « *both grammatical complexity and discourse status influence constituent order* » (Arnold *et al.*, 2000, p. 51).

L'étude que nous présentons dans cet article a également pour but de soulever la question de la relation entre poids et accessibilité en français.

## 3 Étude expérimentale

Notre étude porte sur les effets du poids et de l'accessibilité discursive sur l'ordre des compléments postverbaux en français. Il s'agit d'une expérience en production qui permet de recueillir des phrases construites par des locuteurs natifs dans des conditions contrôlées. L'ordre choisi par les locuteurs dans les différents contextes proposés nous permet d'observer les effets du poids en termes de longueur syllabique et de complexité syntaxique, ainsi que les effets de l'accessibilité discursive, défini comme *référent donné* ou *référent nouveau*. De plus, les résultats obtenus permettent également d'approfondir la discussion sur la relation entre poids et accessibilité.

Nous avons indiqué dans l'introduction que le travail présenté dans cet article s'intègre dans un projet plus large : l'étude de l'ordre des compléments du verbe en français et en persan. Dans cette perspective, l'expérience que nous avons menée sur le français est une réplique de l'expérience Faghiri *et al.* (2014) qui utilisent une tâche de complétion de phrases pour étudier les effets du poids relatif et de l'accessibilité discursive sur l'ordonnement de l'OD et de l'OI dans un paradigme de production en persan. Pour aller plus loin dans l'étude de l'effet du poids grammatical sur l'ordre, nous avons inclus, dans notre expérience sur le français, une manipulation sur la complexité en contrôlant la longueur. Dans l'expérience, nous avons opérationnalisé la distinction entre longueur et complexité en utilisant des constituants de longueur identique en termes de nombre de syllabes, mais se différenciant par la présence ou l'absence d'une subordonnée relative.

Étant donné les études précédentes sur le français, nous nous attendons à avoir une préférence générale pour l'ordre SN-SP. Nous nous attendons également à trouver un effet du poids relatif correspondant à la préférence *court avant long*. Quant à la décorrélation entre la complexité et la longueur, assumant que les constituants complexes sont plus lourds, nous nous attendons à ce qu'à nombre de syllabes égal, le bouleversement de l'ordre soit plus fréquent avec les constituants plus complexes.

En ce qui concerne l'accessibilité discursive, les travaux en linguistique théorique et en psycholinguistique convergent pour admettre la préférence *donné avant nouveau* à vocation universelle. Nous nous attendons donc à trouver une préférence pour cet ordre dans le domaine postverbal du français.

### 3.1 Paradigme expérimental

Le paradigme expérimental que nous utilisons est inspiré des paradigmes de production contrainte, utilisée dans les travaux psycholinguistiques, notamment la tâche du rappel de phrase, utilisée par Stallings *et al.* (1998) pour étudier l'ordre relatif entre des compléments verbaux en anglais, ou par Yamashita et Chang (2001) pour étudier l'ordre des constituants dans les phrases transitives et ditransitives en japonais. Dans ces expériences, les participants doivent construire une phrase à partir des syntagmes affichés sur l'écran, qu'ils auront à formuler (à l'oral) après un laps de temps.

Nous utilisons un questionnaire en ligne et, par conséquent, à la place d'une tâche de rappel, nous employons une tâche de complétion, similaire à une tâche de phrase à trous. Les participants doivent donc compléter une phrase à partir des constituants affichés sur l'écran. Cependant, pour que la tâche ne se résume pas à un choix de l'ordre relatif entre les constituants affichés, la liste comporte un constituant supplémentaire formellement identique à un des autres constituants, mais lexicalement différent. Cela permet non seulement de dissimuler le but de l'exercice aux participants mais aussi de les amener à formuler les phrases à partir d'un contenu sémantique et d'éviter les réponses stratégiques.

### 3.2 Méthode

Nous avons construit les items expérimentaux de façon à tester de façon contrôlée l'effet de deux facteurs : le poids et l'accessibilité. Plus précisément, nous manipulons le poids relatif sur trois niveaux et l'accessibilité discursive sur deux. Nous avons donc mis en place un *design* expérimental 2x3 (un facteur à deux niveaux et un facteur à trois niveaux).

#### Facteurs manipulés :

1. Le poids relatif entre les deux constituants sur trois niveaux (en combinant la complexité et la longueur). La complexité signifie la présence d'une subordonnée relative. La longueur est manipulée par ajout de modifieur ou de coordination.

1. SP > SN : SP long, complexe et SN court
2. SP < SN (Non Comp): SP court et SN long, non complexe
3. SP < SN (Comp) : SP court et SN long, complexe

La différence de longueur entre les deux constituants est d'au moins 5, et, en moyenne, de 7 syllabes. Les deux SN longs, complexe et non complexe, comportent le même nombre de syllabes.

2. l'accessibilité discursive du SP sur deux niveaux. Pour manipuler l'accessibilité, chaque phrase cible est précédée d'un contexte en guise de préambule. Dans le cas des SP donnés, le référent est mentionné dans la phrase de contexte, tandis que le référent d'un SP nouveau est absent du préambule.

1. SP donné
2. SP nouveau

#### Variable observée :

— Ordre relatif entre le SP et le SN dans la phrase complétée par le participant, opérationnalisé par une variable binaire: SN-SP vs. SP-SN.



### 3.2.1 Matériel

Nous avons préparé 6 versions de 24 phrases ditransitives, correspondant à nos 6 conditions expérimentales (cf. tableau 1). Les travaux de Schmitt (1987a ; 1987b) et Thuilier (2012b) ont montré que la sémantique du verbe a une influence sur l'ordre de ses compléments. Nous avons donc choisi 6 verbes sémantiquement comparables dans la mesure où ils assignent des rôles sémantiques similaires : *offrir*, *proposer*, *vendre*, *donner*, *envoyer*, *servir*. Tous les SP sont animés (humains) et les SN inanimés. Conformément à ce que nous avons dit dans la section 3.1, nous avons proposé un constituant supplémentaire. Pour chaque phrase, nous avons donc construit deux SN partageant les mêmes traits formels et la même longueur. Ces deux SN sont des compléments directs possibles du verbe, dans la mesure où aucun ne rend la phrase agrammaticale (ou inacceptable), mais ils n'ont pas forcément la même pertinence dans le contexte.

Un exemple d'item expérimental, dans les six conditions possibles, est présenté dans le tableau 1.

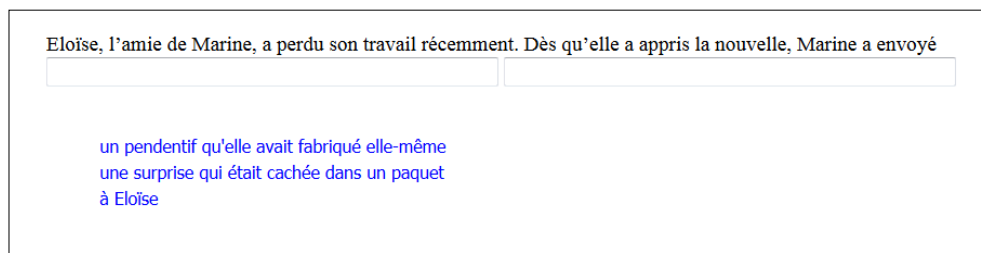
	Contexte <i>SP donné</i>	La voiture s'est arrêtée au carrefour. Le conducteur a ouvert la fenêtre. Un homme habillé en noir s'est appuyé sur la voiture et il a donné
1	SP<SN(Non Comp)	{une petite enveloppe en papier kraft cachetée / une liasse de billets d'environ deux mille euros } {au conducteur}
2	SP<SN(Comp)	{une grande enveloppe qu'il a sortie d'une mallette/ une liasse qui contenait une centaine de billets } {au conducteur}
3	SP>SN	{une enveloppe/ de la monnaie}{au conducteur, qui surveillait les alentours}
	Contexte <i>SP nouveau</i>	La voiture s'est arrêtée au carrefour. Le conducteur a ouvert la fenêtre et, il a donné
4	SP<SN(Non Comp)	{une petite enveloppe en papier kraft cachetée / une liasse de billets d'environ deux mille euros } {au policier}
5	SP<(Comp)	{une grande enveloppe qu'il a sortie d'une mallette/ une liasse qui contenait une centaine de billets } {au policier}
6	SP>SN	une enveloppe/ de la monnaie } {au policier qui surveillait les alentours}

**Tableau 1.** *Un item expérimental complet*

Une liste de 24 items et de 48 distracteurs a été intégrée à IbeX Farm (Drummond, 2013). La présentation suit la procédure du carré latin, ce qui veut dire que chaque sujet juge tous les items et toutes les conditions expérimentales, mais jamais le même item dans plusieurs conditions. L'ordre de présentation des items et des distracteurs est randomisé individuellement pour chaque sujet.

### 3.2.2 Procédure

Chaque item expérimental est présenté par un préambule, une phrase incomplète où il manque l'OD et l'OI d'un verbe ditransitif, représentés par deux cases vides suivant le verbe, une liste de trois syntagmes pour compléter la phrase, incluant deux SN (formellement identiques mais lexicalement différents) et un SP. Pour neutraliser l'impact de l'ordre d'affichage sur l'écran, dans une moitié des items le SP se trouve en haut de la liste des options et dans l'autre moitié, le SP se trouve en bas de la liste. Tous ces éléments sont présentés simultanément sur l'écran (cf. Figure 1). Les participants complètent la phrase en remplissant les cases par les syntagmes de leur choix et dans l'ordre qu'ils souhaitent. Pour éviter de modifier les syntagmes, ils sont priés d'utiliser le copier-coller pour remplir les cases.



**Figure 1.** Copie d'écran pour un item donné

### 3.2.3 Participants

74 locuteurs natifs du français et vivant en France, âgés de 18 à 74 ans (avec des âges moyen de 36,3 et médian de 30 ans), dont 59 femmes et 15 hommes, ont rempli le questionnaire. 22 réponses erronées ont été écartées, ce qui laisse un ensemble de 1778 observations.

### 3.2.4 Résultats

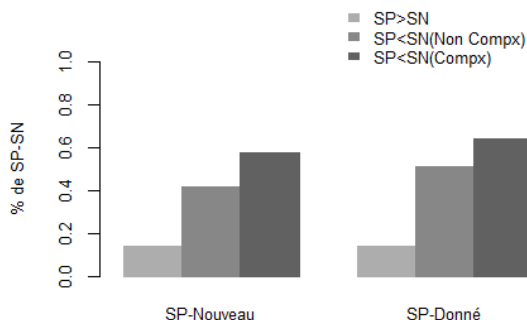
Le taux moyen de l'ordre SN-SP dans nos données, 59,4%, montre une préférence générale pour cet ordre confirmant notre première hypothèse, en accord avec les travaux précédents sur le français. La distribution de l'ordre SN-SP en fonction de nos variables expérimentales ( Figure 2) montre que lorsque le SN est plus léger que le SP, le taux de l'ordre SN-SP est beaucoup plus important (85,9%) que lorsqu'il est plus lourd (46,3%), qu'il soit complexe (39,3%) ou seulement long (53,4%); la complexité favorisant plus le bouleversement de l'ordre SN-SP, conformément aux attentes.

En ce qui concerne le statut discursif du SP, lorsque celui-ci est donné le taux de l'ordre SP-SN est légèrement plus élevé que lors qu'il est nouveau : 43,3% vs. 37,8%. Notons que cette différence disparaît dans la condition où le SN est court, ce qui n'est pas surprenant, car, en effet, un SN court est fortement prédisposé à précéder le SP.

Enfin, il faut noter que la distribution de l'ordre est également fortement sensible à l'ordre d'affichage des syntagmes sur l'écran. A titre d'exemple, lorsque le SP apparaît en haut de la liste (précédant les deux SN), l'ordre SP-SN est plus élevé que lorsqu'il apparaît en dernier (48,8% vs. 32,5%). Nous appellerons cet effet l'*amorçage*. Ainsi, il est important de vérifier que les préférences observées dans les données brutes sont effectivement significatives, une fois la variation liée à l'amorçage captée.

Nous avons analysé les résultats à l'aide d'un modèle de régression logistique<sup>3</sup> mixte (Bates *et al.*, 2012, package "lme4" du logiciel R), en incluant les facteurs "sujet", "item" en tant que facteurs à effet aléatoire. Les facteurs à effet fixe incluent en plus des deux variables expérimentales, une variable binaire représentant l'ordre d'affichage des syntagmes (SN-SP vs. SP-SN), renvoyant à l'effet d'amorçage<sup>4</sup>. Le Tableau 2 présente les résultats du modèle optimal<sup>5</sup>.

Le modèle montre que les variables *poids*, *accessibilité* et *amorçage* ont un effet significatif sur la variable observée, c'est-à-dire l'ordre dans lequel les participants ont choisi de mettre le SN et le SP. Premièrement, l'intercepte positif témoigne d'une préférence générale des participants pour l'ordre SN SP. Deuxièmement, la longueur (SN-LONG) et la complexité (SN-COMP) sont toutes deux significatives, ce qui signifie que les deux dimensions influencent le choix de l'ordre. Les coefficients négatifs montrent que lorsque le SN est long ou complexe, l'ordre SP SN est favorisé. Notons que le coefficient associé à la complexité est supérieur



**Figure 2.** Le taux de l'ordre SP-SN en fonction du poids et de l'accessibilité discursive

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte	0,713	0,204	3,494	<0,0001
POIDS[SN-COMP]	-1,360	0,097	-13,982	<0,0001
POIDS[SN-LONG]	-0,486	0,090	-5,426	<0,0001
ACC[SP-NOUVEAU]	0,205	0,065	3,158	<0,005
LISTORD[SN-SP]	0,577	0,103	5,605	<0,0001

N.B.: Le succès correspond à ORDRE=SN-SP

**Tableau 2.** Résumé des résultats du modèle pour les effets fixes

(en valeur absolue) à celui associé à la longueur : cela indique que les participants ont été plus sensibles à la complexité qu'à la longueur lors du choix de l'ordre. Troisièmement, la présence d'un SP *nouveau* favorise le choix de l'ordre SN SP, comme en témoigne le coefficient positif associé à la variable. Ce coefficient a une valeur peu élevée, en comparaison aux autres coefficients, ce qui signifie que son effet est faible. Enfin, le modèle confirme qu'il y a un effet d'amorçage dans l'expérience : lorsque les syntagmes sont présentés dans un ordre donné, les participants ont tendance à compléter la phrase en utilisant cet ordre. Le point important ici est que le modèle permet de montrer qu'une fois pris en compte cet effet, les variables qui nous intéressent ont toujours un effet significatif. Autrement dit, malgré le biais induit par le dispositif expérimental, les effets de poids et d'accessibilité sont présents.

En résumé, nos données confirment la préférence générale pour l'ordre par défaut SN SP. La complexité mais aussi la longueur ont des effets significatifs, bien que la taille de l'effet soit plus important pour la complexité. Elles montrent également qu'il existe bien un effet de l'accessibilité discursive dans le domaine postverbal du français. L'effet de ce facteur est relativement faible, en comparaison aux autres facteurs pris en compte dans le modèle.

## 4 Discussion et conclusion

Les résultats de notre expérience nous permettent d'apporter des réponses aux questions principales que nous avons définies quant à l'effet de différents facteurs sur l'ordre linéaire des constituants dans le domaine

postverbal du français. Sachant que l'ordre préféré des compléments verbaux est SN SP, notre but était de déterminer quels sont les facteurs susceptibles de favoriser le choix de l'ordre SP SN. Nous avons étudié l'effet de deux facteurs en particulier : le poids grammatical et l'accessibilité discursive.

Les travaux empiriques antérieurs sur le français ont pu montrer un effet du poids correspondant à la préférence *court avant long*. Notre étude a, sans surprise, confirmé cette préférence, mais plus particulièrement, elle a permis de montrer que le poids grammatical ne peut être réduit à la complexité structurale.

Les travaux empiriques précédents n'ont pas réussi à montrer un effet d'accessibilité dans le choix de l'ordre des compléments du verbe. Notre étude semble donc être la première à mettre en évidence une préférence *donné avant nouveau* dans le domaine postverbal du français, confirmant ainsi les intuitions de Blinkenberg (1928) et Berrendonner (1987). Cependant, l'effet du statut discursif (du référent) du SP est relativement faible. En effet, nous constatons que le rôle de l'accessibilité est écrasé par le poids en cas de compétition : par exemple, lorsque le SP est long ou lourd (ce qui favorise l'ordre SN SP) et qu'il est également *donné* (ce qui favorise l'ordre SP SN), l'effet du poids l'emporte largement sur l'effet d'accessibilité.

Cette observation est cohérente avec celle faite par Stallings *et al.* (1998) pour l'anglais. Selon ces auteurs, le domaine postverbal est fortement contraint par la présence du verbe qui a une influence importante sur ce qui suit. Plus précisément, l'ordre des éléments après le verbe serait plus fortement affecté par les facteurs formels comme le poids, que par les facteurs conceptuels tels que l'accessibilité discursive, car le verbe inhiberait en quelque sorte leurs effets. Suite à ces observations, nous faisons l'hypothèse que l'effet d'accessibilité devrait être plus important si le poids du SN et du SP était comparable. De plus, il serait attendu que l'effet d'accessibilité soit plus fort avant le verbe, par exemple dans le cas d'une dislocation.

En ce qui concerne la relation entre poids et accessibilité, nos résultats indiquent que, malgré l'effet largement dominant du poids, la différence de statut du référent du SP (*donné* ou *nouveau*) explique en partie les choix des participants. Rappelons que, dans notre étude, l'accessibilité des référents n'est jamais corrélée à la pronominalité, ce qui signifie que l'effet d'accessibilité mis au jour n'est pas lié à la petite taille des pronoms. Ainsi, en montrant que l'accessibilité discursive du SP peut, indépendamment de sa lourdeur, favoriser son placement avant le SN, nous nous plaçons dans la lignée des travaux qui cherchent à décorréler l'effet de poids et d'accessibilité discursive (Arnold *et al.*, 2000). Notre étude va donc à l'encontre des visions unidimensionnelles telles que celle défendue par Hawkins (1994).

## Références bibliographiques

- Abeillé, A. & Barrier, N. (2004). Enriching a French treebank. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, Lisbonne.
- Abeillé, A., Clément, L. & Toussanel, F. (2003). Building a treebank for French. In *Treebanks*. Dordrecht: Kluwer.
- Abeillé, A. & Godard, D. (2004). De la légèreté en syntaxe. *Bulletin de la Société de Linguistique de Paris*, XCIX(1), 69–106.
- Abeillé, A. & Godard, D. (2006). La légèreté en français comme déficience de mobilité. *Linguisticae Investigationes*, 29(1), 11–24.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Arnold, J. E. (1998). *Reference Form and Discourse Patterns*. PhD thesis, Stanford University, Stanford.

- Arnold, J. E. (2010). How Speakers Refer: The Role of Accessibility. *Language and Linguistics Compass*, 4(4), 187–203.
- Arnold, J. E., Wasow, T., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Bates, D., Maechler, M. & Bolker, B. (2012). lme4: Linear mixed-effects models using eigen and jags.
- Behaghel, O. (1909). Von deutscher Wortstellung [de l'ordre des mots en allemand]. *Indogermanische Forschungen*, 25, 110–142.
- Berrendonner, A. (1987). L'ordre des mots et ses fonctions. *Travaux de linguistique*, 14/15, 9–19.
- Bijankhan, M. (2004). Naqsh-e peykar-hā-ye zabāni dar neveshtan-e dastur-e zabān: mo'areffi-ye yek narmafzār-e rāyāne-i [le rôle de corpus pour écrire une grammaire : présentation d'un logiciel]. *Majalle-ye zabanshenāsi*, 19(2), 48–67.
- Blinkenberg, A. (1928). *L'ordre des mots en français moderne. Première partie*. Copenhagen: Høst & Søn.
- Bock, J. K. & Levelt, W. (1994). Language production: Grammatical encoding. In M. Gernsbacher, Ed., *Handbook of psycholinguistics*, p. 945–984. New York: Academic Press.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. N. Li, Ed., *Subject and Topic*, p. 27–55. New York: Academic Press.
- Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. Cambridge: MIT Press.
- Clark, H. H. & Clark, E. V. (1977). *Psychology and Language: An introduction to psycholinguistics*. New York: Harcourt Press Jovanich.
- De Smedt, K. (1994). Parallelism in incremental sentence generation. In G. Adriaens & U. Hahn, Eds., *Parallel natural language processing*, p. 421–447. Norwood: Ablex.
- Drummond, A. (2013). Ixex farm. Version 0.3.
- Faghiri, P., Samvelian, P. & Hemforth, B. (2014). Accessibility and word order: The case of ditransitive constructions in Persian. In *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*, p. 217–237. S. Müller.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, p. 95–126. Cambridge: MIT Press.
- T. Givón, Ed. (1983). *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam et Philadelphie: John Benjamins.
- Grant, M., Abeillé, A., Crabbé, B. & Thuilier, J. (2014). The role of conceptual accessibility on word order alternations in french: Evidence from sentence recall. Poster présenté à *International Workshop on Language Production 2014*.
- Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. A. Moravcsik & J. R. Wirth, Eds., *Studies in Syntactic Typology*: John Benjamins.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2).
- Hawkins, J. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.
- Levelt, W. J. (1989). *Speaking: From Intention to Articulation*. ACL-MIT Press series in natural-language processing. Bradford Books.

- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole, Ed., *Radical Pragmatics*, p. 223–256. New York: Academic Press.
- Ross, J. R. (1967). *Constraints on variables in syntax*. PhD thesis, Department of Linguistics, MIT.
- 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.
- Stallings, L. M. & MacDonald, M. C. (2011). It's not Just the "Heavy NP": Relative Phrase Length Modulates the Production of Heavy-NP Shift. *Journal of Psycholinguistic Research*, 40(3), 177–187.
- Stallings, L. M., MacDonald, M. C. & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy- NP shift. *Journal of Memory and Language*, 39(3), 392–417.
- Thuilier, J. (2012a). *Contraintes préférentielles et ordre des mots en français*. PhD thesis, Université Paris Diderot.
- Thuilier, J. (2012b). Lemme verbal et classe sémantique dans l'ordonnement des compléments postverbaux. In F. Neveu, V. Muni Toke, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost & S. Teston-Bonnard, Eds., *Actes du 3ème Congrès Mondial de Linguistique Française 2012 (CMLF 2012)*, p. 2451–2469.
- Thuilier, J., Abeillé, A. & Crabbé, B. (2014). Ordering preferences for postverbal complements in french. In Tyne, André, Benzitoun, Boulton & Greub, Eds., *French through Corpora: Ecological and Data-driven Perspectives in French Language Studies*, p. 77–102. Cambridge: Cambridge Scholar Publishing.
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9, 81–105.
- Wasow, T. (2002). *Postverbal Behavior*. CSLI publications.
- Wasow, T. & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, 115(11), 1481–1496.
- Yamashita, H. & Chang, F. (2001). 'Long before short' preferences in the production of a head final language. *Cognition*, 81(2).

---

1. Abréviations utilisées dans la glose des exemples persans : EZ = ezafe ; RSTR = restrictif ; PP = participe passé ; PST = présent ; MDO = marquage différentiel de l'objet ; SBJV = subjonctif ; INDF = indéfini.

2. Les référents ayant le statut d'*inférable* au sens de Prince (1981), ont été écartés de l'analyse

3. Rappelons qu'une variable binaire telle que notre variable ordre relatif ici, peut être assimilée, par le biais d'une transformation logistique, à une fonction de probabilité et modélisée par une courbe sigmoïde (voir par exemple Agresti, 2007). Pour simplifier, le modèle de régression logistique rend la probabilité d'obtenir une des deux valeurs de la variable binaire, désignée comme le scénario du succès, contre l'autre valeur correspondant alors à l'échec: le zéro, correspond à une probabilité de 0,5, signifiant l'impossibilité de trancher en faveur de l'un ou l'autre scénario, les valeurs positives prédisent le succès et les valeurs négatives l'échec. Autrement dit, les coefficients positifs votent pour le succès et les coefficients négatifs pour l'échec.

4. L'interaction entre les deux variables expérimentales était également incluse dans notre modèle maximal. Or, malgré ce que les données brutes (les fréquences) semblaient indiquer celle-ci n'est pas significative ( $p > 0,15$ ). Afin d'obtenir un modèle optimal, nous avons écarté cette interaction en nous basant sur le test du rapport de vraisemblance (*Likelihood ratio test*). Cela s'est fait à l'aide de la fonction ANOVA du logiciel R. Cette fonction permet d'analyser la pertinence d'un facteur (ou d'une interaction de facteurs) en comparant un modèle complet avec un modèle réduit où ce facteur est écarté. La statistique donnée par ce test est le Chi carré :  $\chi^2(2) = 2,6798$ ,  $p = 0,262$ .

5. Notons que tous les facteurs fixes sont codés selon le contraste “somme”. Ainsi, l’intercepte représente la moyenne générale et les coefficients les écarts de chaque groupe par rapport à cette valeur. La présentation des résultats dans le tableau est simplifiée.