



**HAL**  
open science

# La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique

Pegah Faghiri

## ► To cite this version:

Pegah Faghiri. La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique. Linguistique. Université Sorbonne Paris Cité; Université Sorbonne Nouvelle Paris 3, 2016. Français. NNT: . tel-01429416

**HAL Id: tel-01429416**

**<https://theses.hal.science/tel-01429416>**

Submitted on 8 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ÉCOLE DOCTORALE n° 268  
Langage et langues : description, théorisation, transmission

## Thèse de Doctorat en Sciences du Langage

pour obtenir le grade de docteur délivré par

**Université Sorbonne Nouvelle - Paris 3**

*présentée et soutenue publiquement par*

**Pegah Faghiri**

le 09 décembre 2016

### LA VARIATION DE L'ORDRE DES CONSTITUANTS DANS LE DOMAINE PRÉVERBAL EN PERSAN : APPROCHE EMPIRIQUE

Thèse co-dirigée par :

**Pollet Samvelian**

UMR 7528 “Mondes iranien et indien”

et

**Barbara Hemforth**

UMR 7110 “Laboratoire de linguistique formelle”

#### **Jury :**

Philippe BLACHE, Directeur de recherche, CNRS-LPL

Olivier BONAMI, Professeur, Université Paris Diderot

Mohammad DABIR-MOGHADDAM, Professeur, Allameh Tabataba'i University (Rapporteur)

John HALE, Professeur, Cornell University

Barbara HEMFORTH, Directrice de recherche, CNRS-LLF

Pollet SAMVELIAN, Professeur, Université Sorbonne Nouvelle

Shravan VASISHTH, Professeur, University of Potsdam (Rapporteur)

UMR 7528 “Mondes iranien et indien”



# Résumé

## **La variation de l'ordre des constituants dans le domaine pré-verbal en persan : approche empirique**

Cette thèse propose une étude quantitative de la variation de l'ordre des constituants en persan avec un intérêt particulier pour l'ordre relatif entre le COD et le COI étant donné son rôle crucial dans les analyses de la structure du SV.

Afin de remédier à une lacune empirique dont souffre l'étude de la syntaxe du persan, notre premier objectif est d'évaluer, à partir de données empiriques robustes, l'hypothèse largement admise selon laquelle il existe un ordre relatif canonique dichotomique entre les compléments verbaux, dépendant du marquage différentiel de l'objet (MDO). Notre second objectif, relatif à la linguistique générale et à la typologie, est de contribuer aux débats controversés sur les préférences translinguistiques de l'ordre des mots en étudiant, dans une langue SOV à structure mixte, les effets des facteurs tels que le poids (ou la longueur relative).

Les résultats de nos études de corpus et de nos expériences montrent l'inadéquation du critère MDO pour expliquer l'ordre relatif entre le COD et le COI. Cette conclusion nous conduit à réfuter également l'hypothèse de la position syntaxique double de l'objet au profit d'une structure plate pour le SV. De plus, nos données révèlent une préférence "long-avant-court" subordonnée aux facteurs contribuant à la saillance, tels que la définitude, l'animéité et le rôle grammatical. Nous arguons que cette préférence échappe, en partie ou totalement, aux modèles du traitement se fondant sur la distance entre la tête et ses dépendants, alors qu'elle est compatible avec l'hypothèse selon laquelle dans les langues SOV l'accessibilité conceptuelle des constituants longs favorise leur production plus en amont dans la phrase.

**Mots clés :** Syntaxe, Persan, Ordre des mots, Études de corpus, Études expérimentales, Études des préférences, Poids grammatical, Production de la parole, Marquage différentiel de l'objet, Structure syntagmatique.

# Abstract

## **Word order variations in the preverbal domain in Persian : an empirical approach**

This thesis proposes a quantitative study of word order variations in Persian, focusing in particular on the relative order between the direct object (DO) and the indirect object (IO). The latter plays a crucial role in the theoretical analyses of the VP, which in the absence of such studies lack solid empirical underpinning.

My first goal is to contribute to the study of Persian syntax by providing reliable data in order to evaluate the prevailing hypothesis according to which there exists a dual canonical relative order between the two objects triggered by the Differential Object marking (DOM). My second goal, situated in general linguistics and typology, is to contribute to the ongoing debates on word order preferences by bringing in data on an SOV language with mixed head-direction. To this end, I study the effect of factors such as grammatical weight (or relative length), which are claimed to influence the linear order across languages.

First, the results of our corpus and experimental studies show that the DOM account of the relative order between the DO and the IO is flawed. Based on this conclusion, I also reject the two object positions hypothesis and plead for a flat structure view of the VP. Second, our data reveal a long-before-short preference, which is shown to depend on the effect of salience-enhancing factors such as definiteness, animacy and the grammatical role. I argue that while this preference is, either totally or partially, incompatible with the predictions of processing-oriented dependency-based models, it can be accounted for by production models assuming that the greater conceptual accessibility of longer constituents favors their early position in SOV languages.

**Keywords :** Syntax, Persian, Word order, Corpus studies, Experimental studies, Ordering preferences, Length-based effects, Sentence production, Differential object marking, Phrase structure.

## Remerciements

Tout d’abord, je tiens à remercier les membres de mon jury, Philippe Blache, Olivier Bonami, Mohammad Dabir-Moghaddam, John Hale et Shravan Vashishth d’avoir accepté d’évaluer ce travail de recherche.

J’exprime ma vive gratitude envers mes deux directrices de recherches, Barbara Hemforth et Pollet Samvelian, pour avoir encadré mon doctorat et m’avoir formée à la recherche. Je les remercie pour le temps qu’elles m’ont consacré, pour leur disponibilité et leur conseils précieux. Sans leurs encouragements et leur bienveillance, cette thèse n’aurait pas pu aboutir.

Merci Pollet pour m’avoir ouverte à ce domaine de recherche. Merci d’avoir cru en moi et d’avoir insisté dans mes moments de doutes. J’ai appris bien plus que faire de la recherche à tes côtés.

J’adresse ma reconnaissance à l’équipe de recherche de l’UMR “Mondes iranien et indien”, tout particulièrement Christine Rivière, Poupak Rafi-Nejad, Emmanuelle Giraudet, Farzaneh Zarei, Maria Szuppe et Agnès Korn. Je remercie également les membres de l’UMR “Laboratoire de linguistique formelle”, en particulier Anne Abeillé, pour l’intérêt qu’elle a porté à ce travail. Cette thèse a également bénéficié du soutien de l’Axe 2 du Labex “Empirical Foundations of Linguistics” (EFL), dont je tiens à remercier les membres. Je remercie également Anaïd Donnabedian et Benoît Sagot.

Mes remerciements vont également à Benoît Crabbé et Juliette Thuillier pour les discussions fructueuses qui ont contribué au cheminement méthodologique de ma thèse ; à Gwendoline Fox pour m’avoir aidée dans mon auto-apprentissage de Python, mais aussi pour bien plus ; à Ariel Gutman pour son aide avec LateX quand je ne savais pas encore que la compilation peut échouer tout simplement parce que le fichier pdf est ouvert ; à Emilia Ellsiepen pour m’avoir familiarisée avec la plateforme Ibex et m’avoir accompagnée dans la réalisation de mon premier questionnaire.

Merci également à ma famille et à mes amis iraniens pour leurs aides précieuses avec mes questionnaires. Un grand merci à Mahsa et à Shiva, mes fidèles participants cobayes ! Merci à mon père pour la relecture minutieuse de mes phrases et pour beaucoup plus. Merci à Ava, à Farzaneh

à Mazdak et à Pouyan pour la diffusion de mes questionnaires et merci à Leyla pour son aide pour la traduction française.

Merci enfin à mes collègues doctorants et jeunes docteurs Maximilien Guérin, Linh Dao, Céline Pozniak, et Ophélie Gandon. Merci Ophélie pour tes messages de rappel parce que je m’occupe des choses à la dernière minute, et pour bien plus.

Alors que s’achève ce travail, commencé voilà maintenant 6 ans, je suis convaincue qu’une thèse, comme lorsqu’il s’agit de monter jusqu’au sommet, ne peut se réaliser seul. Je n’aurais pas pu “frapper le sommet”, comme on dit en persan, si je n’avais pas eu l’immense privilège de bénéficier du soutien scientifique et logistique au sein de mes deux laboratoires de rattachement, le Labex EFL et l’École doctorale 268 de l’Université Sorbonne nouvelle Paris 3. Mais aussi j’ai été heureuse d’être entourée d’une famille et des amis merveilleux dont je n’aurais pas pu me passer du soutien moral.

Merci à ma mère, à mon père et à mon frère. Merci à Mazdak, Sepideh, Maryam, Vahid, Farzaneh, Yassi, Roja, Carmen et beaucoup d’autres. Merci spécialement à Mahsa, Mahdokht et Shiva, quelle chance que vous soyez à Paris.

Cette thèse doit beaucoup à Pascal, à plus d’un titre, dont le plus important n’est pas son travail de relecture consciencieuse, jusque dans les exemples persans et les gloses. Elle n’aurait pas vu le jour si, voici il y a plus de 10 ans, il ne m’avait pas soutenue dans mon choix de reconversion. Alors que j’étais encore étudiante en 1er cycle de linguistique à Grenoble, il m’a dit un jour en me présentant un article : “Voilà, il faut que tu contactes son auteur, Pollet Samvelian”. Merci Pascal, merci pour tout, merci pour ton soutien, tes encouragements constants et surtout pour ta patience.

J’aimerais finir ces remerciements avec une pensée pour la mémoire du Professeur Rajendra Singh, que j’ai eu l’honneur de connaître en tant que directeur de mon mémoire à l’Université de Montréal. Il m’a donné le goût de la recherche et m’a initiée au regard critique.

# Sommaire

Introduction	1
<b>I Problématique et méthodologie</b>	<b>7</b>
1 Ordre des mots en persan	9
2 Étudier la variation de l'ordre des mots : État de l'art	61
3 Démarche et questions de recherche	105
<b>II Études empiriques</b>	<b>127</b>
4 Variation de l'ordre des mots dans le corpus Bijankhan	129
5 Études expérimentales de la variation de l'ordre des mots	159
<b>III Discussion générale et apports théoriques</b>	<b>209</b>
6 Renouveler l'analyse de la structure syntagmatique du persan	211
7 Considérations sur les préférences de l'ordre des mots	263
Conclusion	273
Bibliographie	278
Annexes	307
Liste des tableaux	368
Table des figures	370

# Abréviations

AUX	Auxiliaire
CLF	Classifieur
COD	Complément d'objet direct
COI	Complément d'objet indirect
COMPR	Comparatif
COP	Copule
CP	Complétive
DEF	Défini
DET	Déterminant
EZ	Ézafé
FUT	Futur
INF	Informel (registre)
IMP	Impératif
IPFV	Imperfectif
LOC	Locatif
NEG	Négation
PL	Pluriel
PP	Participe passé
PRED	Prédicat
PRS	Présent
PST	Passé
RSTR	Restrictive
SG	Singulier
SBJV	Subjonctif
SUJ	Sujet
SPRL	Superlatif
1	Première personne
2	Deuxième personne
3	Troisième personne

# Introduction

La présente thèse porte sur l'ordre des constituants dans le domaine préverbal en persan. Nous y proposons une étude empirique de la variation de l'ordre fondée sur des méthodes quantitatives à partir des données de l'usage, dans la lignée des travaux menés sur des langues telles que l'anglais (par ex. Wasow, 1997; Stallings et al., 1998; Arnold et al., 2000; Bresnan et al., 2007), l'allemand (par ex. Kempen & Harbusch, 2004; Bader & Häussler, 2010) ou le japonais (par ex. Yamashita & Chang, 2001; Yamashita, 2002; Tanaka et al., 2011a). Ces derniers se fondent sur les données de corpus ou les expériences psycholinguistiques et mobilisent des méthodes de la statistique inférentielle pour étudier la variation de l'ordre en termes de préférences dans l'ordonnement linéaire. Ils s'intéressent ainsi aux facteurs cognitivo-fonctionnels susceptibles d'y intervenir.

Nous ferons le choix dans cette étude d'utiliser en tandem ces deux méthodes de collecte de données que nous considérons comme complémentaires. Le corpus offre un volume important de données riches et diverses, produites dans les conditions naturelles, que l'on peut exploiter afin d'observer la distribution effective des différentes alternatives dans les données de l'usage et identifier les facteurs pertinents. Cependant, du fait de leur variation inhérente sur laquelle le chercheur n'a aucune prise, les données extraites de corpus ne peuvent jamais servir pour établir des liens causaux entre les variables d'étude, mais permettent seulement de dévoiler des corrélations entre elles. Dans la méthode expérimentale, en revanche, les données sont construites par le chercheur afin de tester une hypothèse précise de manière contrôlée dans le but d'établir des liens directs entre les variables, c'est-à-dire en contrôlant, dans la mesure du possible, toute source de variation "indésirable" afin d'isoler la variation liée aux variables d'étude. Par ailleurs, étant donné que, dans les deux cas, les données collectées constituent des échantillons dont la représentativité est par nature imparfaite<sup>1</sup>, la combinaison des deux types d'études pourrait consolider les conclusions et faciliter leur généralisation.

---

1. En effet, la population sur laquelle ces échantillons sont prélevés est une population hypothétique et, par définition, insaisissable, car il s'agit pour ainsi dire de l'ensemble des énoncés d'une langue donnée.

Dans notre étude, nous nous concentrerons sur deux phénomènes de la variation de l'ordre dans le domaine préverbal en persan, à savoir l'ordre relatif entre les deux objets, direct et indirect, dans une phrase ditransitive, et l'ordre relatif entre le sujet et l'objet dans une phrase transitive, avec une focalisation sur le premier. L'ordre relatif entre les deux compléments verbaux, direct et indirect, est une question centrale dans l'étude de la structure syntaxique du persan. Or, très peu de travaux ont tenté d'aborder cette question d'un point de vue empirique avec un souci pour la description.

Le persan est une langue SOV à ordre flexible<sup>2</sup>. La question de la variation de l'ordre, bien que toujours présente dans les travaux sur la syntaxe de cette langue, n'a jusqu'à présent pas fait l'objet d'investigations empiriques approfondies<sup>3</sup>. La variation de l'ordre a été essentiellement décrite en termes de fonctions discursives (Dabir-Moghaddam, 1990, 2005; Mahootian, 1997; Karimi, 1994, 2005; Lazard et al., 2006; Windfuhr & Perry, 2009), ce qui s'explique par le lien généralement établi entre les variations de l'ordre des constituants et l'organisation pragmatique de la phrase dans les langues manifestant un ordonnancement relativement libre des mots.

Dans la mesure où il existe par ailleurs un lien entre la variation de l'ordre et le marquage différentiel de l'objet (MDO) en persan, ce dernier phénomène reçoit une attention particulière dans ces travaux. Pour le dire en quelques mots, un complément d'objet direct (COD) défini est obligatoirement marqué par l'enclitique =*rā*. De surcroît, cet enclitique peut également s'adjoindre à d'autres constituants non sujets, tels que circonstants temporels, voire aux COD indéfinis, dans certaines conditions discursives.

La relation entre la variation de l'ordre et le marquage différentiel de l'objet a été au cœur de plusieurs études dans la grammaire générative et a servi à l'élaboration de différentes hypothèses sur la structure syntaxique du syntagme verbal en persan. Ces études s'appuient sur une observation largement répandue dans la littérature, d'après laquelle, en règle générale, le marquage par =*rā* implique l'éloignement du COD du verbe. Cela signifie que, dans un ordre des mots neutre, les COD marqués précèdent le complément d'objet indirect (COI), alors que les COD non marqués le suivent. Ce contraste sert d'argument à l'hypothèse de l'existence de deux positions syntaxiques

---

2. Pour être plus précis, les six ordres de base peuvent (tous) se produire en persan. Toutefois, l'ordre n'est pas totalement libre dans la mesure où les constituants sont soumis à un ordre contraint au sein des structures plus petites que la phrase, tel que le syntagme nominal ou prépositionnel. De plus, au sein du domaine propositionnel, les compléments phrastiques manifestent un ordonnancement rigide.

3. La récente étude de corpus de Rasekhmahand & Ghiasvand (2015), investiguant les facteurs qui favorisent le placement d'un dépendant du verbe dans le domaine postverbal, est une exception notable.

distinctes pour les COD en persan : les COD marqués de  $=r\bar{a}$  apparaissent dans une position syntaxique plus élevée que les COD non marqués (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2003a; Ganjavi, 2007).

Ces analyses, ancrées pour la plupart dans le cadre de la grammaire générative, trouvent leurs fondements empiriques dans des données construites, issues de l'introspection et mobilisées principalement dans le but de contribuer aux débats théoriques sur des questions générales, telles que le *scrambling*, la structure syntagmatique universelle, par exemple. Par conséquent, la question de la variation de l'ordre des mots en tant que telle reste secondaire dans ces études. En effet, l'étude des facteurs qui interviennent pour déterminer le choix d'un ordre donné par rapport un autre, dans les productions des locuteurs, n'est pas pertinente pour ces travaux dont l'objectif est de décrire la compétence linguistique indépendamment de la performance considérée comme contingente.

Notre démarche consiste, au contraire, à étudier la variation de l'ordre à travers les données de la performance que nous considérons au demeurant comme indissociable de la compétence linguistique. Ce choix méthodologique se justifie par la nature des phénomènes sur lesquels porte notre étude.

En règle générale, les phénomènes qui relèvent de la variation n'entraînent pas l'agrammaticalité de la phrase, et par conséquent leur étude ne peut s'appuyer sur des jugements de grammaticalité binaires. Il n'est pas non plus toujours possible de se fier à l'intuition pour identifier l'ordre neutre ou prédominant parmi toutes les possibilités, c'est-à-dire les ordres équivalents sur le plan sémantique et véridictionnel, à l'aide d'une notion aussi vague que la neutralité pragmatique ou discursive. La fréquence d'utilisation, obtenue par la prise en compte d'un nombre relativement élevé de données de manière contrôlée et la plus impartiale possible, s'impose comme le seul critère fiable pour identifier l'ordre prédominant parmi ses concurrents, et le plus communément utilisé à cet effet. Ce critère est effectivement la prémisse sous-entendue dans toutes les études quantitatives existantes sur le sujet.

Il ne s'agit aucunement de nier l'importance de l'intuition, qui peut, dans certains cas de figure, être suffisamment forte et consensuelle pour présenter une base robuste et fiable pour tirer des généralisations solides et indiscutables. Par exemple, le fait que le SVO est l'ordre canonique de la phrase en français ou en anglais n'est sujet à aucune controverse, et la même chose est vrai pour l'ordre SOV en persan. Or, il existe des cas de figure dans lesquels la différence entre les alternatives est beaucoup plus subtile et l'impact des facteurs extra-syntaxiques considérable. Ainsi, les intuitions ne sont pas suffisamment convergentes et seules les méthodes quantitatives peuvent permettre de départager entre les choix non contraints par la

grammaire. L’alternance dative en anglais en constitue un cas d’école (Collins, 1995; Wasow, 2002; Gries, 2003; Bresnan *et al.*, 2007). Cette alternance désigne la possibilité pour l’argument datif dans les constructions ditransitives en anglais d’avoir une réalisation prépositionnelle, ex. (1-a), ou une réalisation directe, ex. (1-b), appelée également construction à double objet.

- (1) a. Kim handed [a toy] [**to the baby**].  
 b. Kim handed [**the baby**] [a toy].

En effet, c’est seulement sur la base d’études quantitatives rigoureuses que les chercheurs ont réussi à obtenir une description adéquate de ce phénomène. À titre d’exemple, il a été démontré que cette alternance ne dépend pas des contraintes catégoriques syntaxiques ou sémantiques, mais relève de l’interaction d’un ensemble de préférences. L’ordre relatif entre les compléments verbaux dans une phrase ditransitive en persan nous semble constituer un cas de cette nature.

Nous situons notre recherche dans la lignée des travaux fondateurs, tels que Wasow (1997), Stallings *et al.* (1998), Arnold *et al.* (2000), Wasow (2002) et Bresnan *et al.* (2007), abordant les préférences de l’ordonnancement linéaire entre les constituants d’une phrase dans une approche cognitivo-fonctionnelle par des méthodes quantitatives, au croisement de différents domaines tels que la syntaxe formelle, la typologie linguistique, la psycholinguistique et les sciences cognitives. Certains de ces travaux, notamment Wasow (2002) et Bresnan *et al.* (2007), ont remis en question la pertinence d’une approche adoptée basée sur la distinction catégorique grammaticale vs. agrammaticale, dans l’étude de la variation de l’ordre. Ils ont proposé des traitements alternatifs des phénomènes tels que l’alternance dative, en prenant appui sur des études de corpus ainsi que des études expérimentales, qui ont mis en lumière la nature probabiliste et/ou préférentielle de ces phénomènes. Ces travaux ont mis en avant l’intervention de facteurs à caractère essentiellement fonctionnel, tels que la longueur relative, la définitude, la structure informationnelle, le caractère animé, le lien collocationnel entre le verbe et les compléments, le lexème verbal, susceptibles, par ailleurs, d’être observés dans les phénomènes de la variation de l’ordre à travers des langues typologiquement différentes.

En cela, ces études se rapprochent des recherches antérieures à caractère typologique et fonctionnel basées sur la performance, notamment celles de Hawkins (1983, 1990, 1994), Siewierska (1993), ainsi que des travaux en sciences cognitives. Comme en témoignent les nombreux travaux en psycholinguistique et en sciences cognitives, qui mobilisent des observations sur l’ordre des constituants, étudier les préférences

dans les phénomènes de variation de l'ordre peut apporter un éclairage sur notre vision de l'architecture du système de la production et/ou de la compréhension de la parole (Bock, 1982; Bock & Warren, 1985; Levelt, 1989; McDonald *et al.*, 1993; Hemforth, 1993; Gibson, 1998; Kempen & Harbusch, 2003; Vasishth, 2003, parmi beaucoup d'autres).

Dans cette optique, la disponibilité des données des langues typologiquement différentes est cruciale. Nous ne pouvons que déplorer la faible diversité des études existantes. C'est ainsi que les données des langues à tête finale, notamment le japonais (Hawkins, 1994; Yamashita & Chang, 2001), ont semé le doute quant à l'universalité de la préférence "court-avant-long", prédite par les modèles les plus répandus de la production, établie sur des données provenant essentiellement de l'anglais et de l'allemand. En tant que langue SOV à structure mixte, le persan présente un intérêt heuristique puisqu'il est à mi-chemin entre, d'une part, une langue SVO comme l'anglais, et, d'autre part, une langue strictement à tête finale comme le japonais.

Il apparaît évident dès lors que l'objectif de cette recherche est double. Il en va de même pour son intérêt. En premier lieu, nous nous efforcerons de remédier à une lacune empirique non négligeable laissant libre cours aux spéculations, dont souffre l'ensemble des travaux en syntaxe formelle qui portent sur la structure du syntagme verbal en persan. Dans un second temps, en apportant les données d'une langue ayant des propriétés en commun avec des langues typologiquement opposées, telles que l'anglais et le japonais, nous souhaitons contribuer aux débats existants sur les préférences d'ordre linéaire à travers les langues du monde.

La thèse s'articule en trois parties organisées de la façon suivante. La première partie sert à expliciter nos problématiques et à présenter notre démarche méthodologique. Cette partie comporte trois chapitres. Dans le premier chapitre, nous aborderons la question de l'ordre des mots en persan dans tous les aspects évoqués dans la littérature et présenterons les analyses proposées pour rendre compte de la variation de l'ordre dans cette langue. Pour ce faire, nous commencerons par une description des propriétés syntaxiques du persan qui interviennent d'une manière ou d'une autre dans l'étude de l'ordre. Le deuxième chapitre est consacré à l'étude de la littérature sur la variation de l'ordre des mots. Nous commencerons tout d'abord par une définition de ce que l'on entend par variation et présenterons ensuite l'étude de la variation de l'ordre au sein de différents courants théoriques portant un intérêt pour cette question. Nous énumérerons enfin les différentes problématiques sur lesquelles les études existantes se sont penchées. Dans le troisième chapitre, nous justifierons la méthodologie adoptée dans cette étude et circonscriurons les questions

de recherche.

La deuxième partie de la thèse est entièrement consacrée à la présentation de notre étude qui comporte deux volets, le premier basé sur le corpus et le second sur la méthode expérimentale. Cette partie comporte deux chapitres consacrés à la présentation détaillée de nos études respectives pour les deux volets.

Enfin, dans la troisième partie de la thèse nous proposerons une discussion des conclusions de notre étude quantitative et de ses enjeux théoriques que nous voulons mettre en perspective, en fonction de nos questions de recherche, avec les hypothèses et analyses existantes dans la littérature. Cette partie comporte deux chapitres correspondant aux deux objectifs que nous avons poursuivis dans ce travail. Le premier chapitre est consacré à l'étude de la structure syntagmatique du persan. Nous y effectuerons une étude fine des analyses existantes, notamment l'hypothèse de la position syntaxique double de l'objet (PSDO), que nous remettrons en question à la lumière des conclusions de notre étude. Cette remise en cause sera confirmée par des études complémentaires pour examiner l'ensemble des arguments évoqués en faveur de cette hypothèse. Dans le dernier chapitre, nous aborderons les résultats de notre étude en ce qui concerne les préférences de l'ordonnancement linéaire et passerons en revue les effets des différents facteurs étudiés. Plus particulièrement, nous nous focaliserons sur l'enjeu théorique et empirique de la préférence "long-avant-court" observée pour le persan qui, étant donné les spécificités de cette langue, pourrait livrer un éclairage précieux quant à la façon dont l'on peut expliquer l'effet de la longueur dans une perspective translinguistique.

# Première partie

## Problématique et méthodologie



# Chapitre 1

## Ordre des mots en persan : Aperçu général

Ce premier chapitre présente un aperçu général de l'ordre des mots en persan ayant pour objectif de mettre en évidence les phénomènes empiriques qui feront l'objet d'une investigation approfondie dans les chapitres suivants. Cette présentation inclura également la description de certaines spécificités syntaxiques du persan qui ont une pertinence dans l'étude de l'ordre des constituants, notamment la direction de la tête (§1.1.1), la détermination nominale (§1.1.2), le marquage différentiel de l'objet (MDO) (§1.1.3), les prédicats complexes (§1.1.4) et la réalisation des arguments (§1.1.5).

Le persan est considéré de façon consensuelle comme une langue SOV (section 1.2). Elle manifeste néanmoins un ordre (relativement) libre permettant à tous les dépendants verbaux d'apparaître avant ou après le verbe (section 1.3). L'ordre canonique ainsi que la liberté de l'ordonnancement s'observent aussi bien dans la phrase subordonnée (§1.2.3) que dans la phrase matrice (§1.2.1), quelle que soit, par ailleurs, sa modalité, par exemple interrogative, impérative, etc. (§1.2.2).

Dans la littérature existante sur la variation de l'ordre en persan, les ordres non canoniques sont décrits essentiellement en termes de considérations discursives générales. En effet, comme c'est souvent le cas dans des langues à ordre flexible, l'ordre linéaire de la phrase en persan est en lien étroit avec sa configuration discursive (§1.3.1). En outre, comme cela a été évoqué dans plusieurs études, l'accent phrastique joue un rôle significatif pour déterminer la configuration discursive et, de ce fait, intervient dans la variation de l'ordre (§1.3.2). Notons dès à présent que cette thèse se focalise sur la variation de l'ordre dans le domaine préverbal, néanmoins, dans ce chapitre (§1.3.3), nous présenterons les ordres non canoniques tels que décrits dans la littérature, sans écarter le domaine postverbal.

Au-delà des travaux à caractère descriptif, l'ordre des constituants en persan a fait l'objet d'études en syntaxe formelle (section 1.4), notamment dans le cadre de la grammaire générative. Ces études adoptent une vision configurationnelle du persan et se rejoignent pour poser l'ordre SOV comme ordre sous-jacent de la phrase en persan. La plupart de ces travaux se concentrent sur l'ordre relatif entre les deux compléments verbaux, à savoir le complément d'objet direct (COD) et le complément d'objet indirect (COI). Cet intérêt s'explique par l'existence du marquage différentiel de l'objet, qui intervient dans la position du COD en favorisant son éloignement du verbe, alors que les objets non marqués occupent, de préférence, une position adjacente au verbe. Les travaux dans le cadre génératif se sont saisis de ce phénomène afin de postuler deux positions syntaxiques distinctes pour les COD en fonction de leur marquage. Nous présenterons très brièvement les hypothèses les plus répandues dans ce chapitre. Toutefois, l'examen détaillé de ces dernières se fera dans la troisième partie de cette thèse (ch. 6).

En ce qui concerne la variation de l'ordre, les travaux en syntaxe formelle se limitent essentiellement à ceux de Karimi (1994, 2003a, 2005). Dans le cadre de la grammaire générative, et plus précisément du projet minimaliste, l'auteur rend compte des ordres non canoniques par le biais de mouvements déclenchés par des fonctions discursives, sujets, selon l'auteur, aux contraintes catégoriques de nature essentiellement sémantique. Ces mouvements concernent avant tout le domaine préverbal (§1.4.1), domaine qui constitue également le focus du présent travail. Néanmoins, nous rencontrons aussi dans cette littérature des contraintes formulées à l'égard du positionnement des dépendants verbaux dans le domaine postverbal (§1.4.2).

## 1.1 Quelques éléments de la syntaxe du persan

Dans cette section, nous présentons des spécificités de la syntaxe du persan qui sont essentielles en vue du traitement de la question de l'ordre des constituants. Après une présentation sommaire sur la direction de la tête dans divers constituants (§1.1.1), nous nous concentrerons sur deux phénomènes en particulier, le marquage différentiel de l'objet (MDO) (§1.1.3) et les prédicats complexes (§1.1.4).

Le premier phénomène intervient pour déterminer la position du complément d'objet direct (COD) et, à ce titre, se retrouve au cœur des travaux portant principalement ou non sur l'ordre des constituants en persan.

Les prédicats complexes (locutions verbales formées d'un verbe et d'un élément non verbal), quant à eux, introduisent de la complexité dans l'ordre des constituants

en persan, dans la mesure où la composante non verbale fonctionne comme un dépendant du verbe et, à ce titre, doit être ordonnée par rapport aux autres dépendants. Nous aborderons préalablement la détermination nominale (§1.1.2), fondamentale pour saisir le fonctionnement du marquage différentiel de l'objet, ainsi que les problématiques liées à la formation des prédicats complexes. Enfin, nous évoquerons la réalisation des arguments, notamment la réalisation nulle (§1.1.5.1) et enclitique des compléments (§1.1.5.2), les constructions à double objet (§1.1.5.3), qui ont un impact sur le nombre de constituants à ordonner par rapport au verbe.

### 1.1.1 La direction de la tête

À la différence par exemple du japonais, langue uniformément à tête finale, le persan est une langue SOV à structure mixte (Karimi, 1994) et manifeste certaines caractéristiques typiques des langues à tête initiale<sup>1</sup>. Contrairement à ce qui est attendu pour une langue à tête finale, en persan : a) le syntagme nominal est fortement à tête initiale (Samvelian, 2007) ; b) il y a des prépositions et pas de postpositions<sup>2</sup> ; c) le complémenteur précède la proposition subordonnée :

- (1) [SP dar [SN in ketāb=e jāleb [CP ke diruz xānd-am]]]  
 dans ce livre=EZ intéressant que hier lire.PST-1SG  
 ‘dans ce livre intéressant que j’ai lu hier’

En outre, même dans le domaine verbal, l'ordre n'est pas strictement tête finale, dans la mesure où les compléments phrastiques se positionnent obligatoirement dans le domaine postverbal, et que, dans le registre informel, les arguments directionnels apparaissent après le verbe sans pour autant modifier la prosodie (cf. §1.2.1 pour plus de détails)<sup>3</sup>.

1. Dabir-Moghaddam (2001, 2006) suggère que la langue persane est entrée dans un processus de transition depuis le moyen perse. De langue à ordre libre, *free word order* (non configurationnelle), à dominance SOV en ancien perse, elle tend vers de plus en plus de configurationnalité et vers le type VO (dans la typologie greenbergienne, Greenberg (1963, 1966); Dryer (1992)), étant donné le développement des prépositions et l'établissement des ordres Nom > Génitif, Nom > Adjectif et Auxiliaire (*xāstan* et *dāštan*) > Verbe, et Verbe > Compléments phrastiques.

2. Il faut noter que l'enclitique =*rā*, utilisé pour le marquage différentiel de l'objet (cf. §1.1.3), se comporte comme un mot graphique et, du fait de sa postposition au nom tête, est souvent traité comme une postposition (Lazard et al., 2006; Windfuhr, 1979; Dabir-Moghaddam, 1992), et plus précisément comme une adposition dans la littérature persanophone, puisque le terme utilisé généralement en persan, *harf-e ezāfe*, est neutre vis-à-vis de la position.

3. N'omettons pas de noter que ces observations ont conduit certains linguistes générativistes à postuler un ordre sous-jacent SVO pour le persan dans tous les cas de figure, tout en admettant que l'ordre SOV est l'ordre de surface dominant (Marashi, 1970; Moinzadeh, 2001). Cette analyse reste, toutefois, marginale dans la littérature.

- (2) a. Maryam be Sepide fahmānd ke Ali doruq  
 Maryam à Sepideh faire.comprendre.PST.3SG que Ali mensonge  
 mi-guy-ad  
 IPFV-dire.PRS-3SG  
 ‘Maryam a fait comprendre à Sepideh que Ali ment.’
- b. Maryam raft birun  
 Maryam aller.PST.3SG dehors  
 ‘Maryam est allée dehors.’

Au sein du syntagme nominal, les déterminants (non clitiques et non affixaux), les quantifieurs (y compris les numéraux cardinaux) et les classifieurs précèdent toujours le nom tête, ex. (3-a) et (3-b).

- (3) a. in se(=tā) ketāb  
 ce trois(=CLF) livre  
 ‘ces trois livres’
- b. har ketāb=i<sup>4</sup>  
 chaque livre=INDF  
 ‘chaque livre’

En revanche, tous les modifieurs (adjectif, syntagme prépositionnel, nom épithète), les compléments adnominaux, les compléments prépositionnels et les relatives suivent le nom tête. À l’instar de la plupart des langues iraniennes occidentales, le nom en persan est lié à ses dépendants par l’enclitique *ezāfé*, réalisé =*(y)e*<sup>5</sup>. La structure du syntagme nominal est illustrée schématiquement en (4), inspiré de Samvelian (2007). Notons que le déterminant n’est pas obligatoire en persan.

- (4) a. (Dét) N=EZ SAdj=EZ SPrép=EZ SN(Possesseur) Relative  
 b. (in) lebās=e sefid=e bi āstin=e Maryam ke (...)  
 (ce) vêtement=EZ blanc=EZ sans manche=EZ Maryam que (...)  
 ‘La robe blanche sans manche de Maryam, qui/que...’

Contrairement aux langues à tête finale qui ont tendance à avoir des postpositions, le persan ne dispose que de prépositions<sup>6</sup>. Ces dernières peuvent être regroupées en trois classes en fonction de leur combinaison avec l’*ezāfé* (Samiian, 1994)<sup>7</sup> :

1. les prépositions qui ne prennent jamais l’*ezāfé*, *az* ‘de’, *bā* ‘avec’, *bar/be* ‘à’

4. L’enclitique =*i* est un déterminant indéfini (voir §1.1.2.2).

5. Réalisé *e* devant une consonne, *sag=e* ‘chien=EZ’, et *ye* devant les voyelles, *gorbe=ye* ‘chat=EZ’.

6. Voir la note de bas de page 2.

7. Samiian (1994) propose deux catégories de prépositions : 1) celles qui ne prennent pas l’*ezāfé* et 2) celles qui le prennent ; et sépare ensuite, à l’intérieur de la seconde catégorie, les prépositions avec l’*ezāfé* obligatoire de celles avec l’*ezāfé* facultatif.

et *dar* ‘dans’;

2. celles qui peuvent prendre l’*ezāfē* de manière facultative, ex. *tu* ‘dans’, *ru* ‘sur’<sup>8</sup>;
3. celles qui le prennent obligatoirement, ex. *zīr* ‘sous’, *qabl* ‘avant’, *pošt* ‘derrière’, etc.

Notons, par ailleurs, que les prépositions de la classe 1. peuvent prendre comme compléments des syntagmes prépositionnels dont la tête est une préposition des classes 2. et 3., ex. *be ru=ye* ‘sur (Lit. à sur)’, *dar tu=ye* ‘dans (Lit. dans intérieur)’, etc.

Enfin, les subordonnées introduites par l’unique complémenteur *ke* (cf. §1.2.3) sont également à tête initiale :

- (5) Maryam [ke dānešju=ye adābiyāt ast] mi-guy-ad  
 Maryam que étudiant=EZ littérature être.PRS.3SG IPFV-dire.PRS-3SG  
 [(ke) in ketāb jāleb ast]  
 (que) ce livre intéressant être.PRS.3SG  
 ‘Maryam, qui est étudiante en littérature, dit que ce livre est intéressant.’

### 1.1.2 La détermination nominale

Le persan connaît ce que Corbett (2000) appelle un “nombre général” (*general number*) exprimé par le singulier. En d’autres termes, la forme singulière non déterminée est neutre vis-à-vis du nombre et peut être employée pour désigner une quantité indéterminée, cela non seulement dans le cas des noms massifs, mais également pour les noms comptables.

Comme le montrent les exemples suivants, un nom comptable (ex. *medād* ‘crayon’, *printer* ‘imprimante’) peut avoir une lecture singulière ou plurielle en fonction du contexte, du sémantisme du nom ou des propriétés aspectuelles du verbe, etc.

- (6) a. medād xarid-am  
 crayon acheter.PST-1SG  
 ‘J’ai acheté un/des crayon(s).’  
 b. printer xarid-am  
 imprimante acheter.PST-1SG  
 ‘J’ai acheté une imprimante.’  
 c. printer ham mi-foruš-im  
 imprimante aussi IPFV-vendre.PST-1PL  
 ‘Nous vendons aussi des imprimantes.’

---

8. L’absence de l’*ezāfē* a une connotation familière.

- (7) a. medād (ru=ye miz) hast  
 crayon sur=EZ table exister.PRS.3SG  
 ‘Il y a un/des crayon(s) (sur la table).’  
 b. printer (tu=ye daftar) hast  
 imprimante dans=EZ bureau exister.PRS.3SG  
 ‘Il y a une imprimante (dans le bureau).’

Par ailleurs, le persan standard ne dispose pas de déterminant défini<sup>9</sup>. Seule l’indéfinitude est marquée formellement dans ce registre. Par conséquent, un nom qui ne porte aucune marque de détermination, *šišē*, ex. (8), est susceptible d’avoir deux lectures différentes : une lecture définie-anaphorique, ‘la vitre’ (8-a), et une lecture non déterminée non spécifiée pour le nombre, ‘une/des vitre(s)’ (8-b). Par “nom nu”, *bare noun*, nous faisons seulement référence à cette deuxième lecture. Notons que ces deux lectures diffèrent sur le plan prosodique. Nous verrons toutefois que cette ambiguïté disparaît dans la position du complément de l’objet direct.

- (8) šišē šekast  
 vitre casser.PST.3SG  
 a. ‘La vitre s’est cassée.’  
 b. ‘Un/des vitre(s) se sont cassées.’

En fonction de leur niveau de détermination, les syntagmes nominaux peuvent être regroupés en trois classes : noms nus, syntagmes nominaux indéfinis et syntagmes nominaux définis.

### 1.1.2.1 Les noms nus

Les noms nus sont neutres vis-à-vis du nombre et reçoivent une lecture non référentielle<sup>10</sup>. Cette lecture peut être générique (non existentielle) dénotant un type (*kind-level reading*), ex. (9-a) et ex. (9-b), ou existentielle dénotant une entité (sous-spécifiée pour le nombre), ex.(9-c) et ex.(9-d), en fonction du contexte.

- (9) a. šišē be-rāhati mi-šekan-ad  
 vitre facilement IPFV-casser.PRS-3SG  
 ‘Une vitre, ça se casse facilement.’  
 b. na-bāyad dar sāxt=e in asar šišē be kār  
 NEG-faut dans construction=EZ ce œuvre vitre à emploi

9. Dans le registre familier, en revanche, il existe un suffixe, *-(h)e*, qui marque un nom singulier comme défini-anaphorique, cf. §1.1.2.3.

10. Faut-il rappeler que la notion de référentialité ne se réduit pas simplement à la présence ou non d’un référent particulier, potentiellement identifiable par l’énonciateur (Fodor & Sag, 1982).

be-bar-id

SBJV-porter-2PL

‘Il ne faut pas que vous utilisez de vitre dans la construction de cette œuvre.’

- c. morāqeb bāš injā šīše šekaste  
attention faire.IMP ici vitre casser.PP  
‘Attention! Il s’est cassé une/des vitres ici.’
- d. barāye panjare šīše xarid-am  
pour fenêtre vitre acheter.PST-1SG  
‘J’ai acheté une vitre/des vitres pour la fenêtre.’

Un nom nu peut prendre un ou plusieurs modifieurs, ex. (10). Seules les relatives sont exclues. Ces noms nus, que nous appellerons “noms nus modifiés” par la suite, reçoivent les mêmes lectures que les noms nus non modifiés.

- (10) a. film=e xāreji  
film=EZ étranger  
‘un/des film(s) étranger(s)’
- b. film=e bā zirnevis  
film=EZ avec sous-titres  
‘un/des film(s) sous-titré(s)’
- c. film=e xāreji bā zirnevis  
film=EZ étranger avec sous-titres  
‘un/des film(s) étranger(s) sous-titré(s)’

### 1.1.2.2 Les syntagmes nominaux indéfinis

L’indéfini est marqué soit par le déterminant enclitique =*i*, ex. (11-a), soit par le cardinal *yek*, prononcé *ye* dans le registre informel, ex. (11-b), soit par la combinaison des deux, ex. (11-c)<sup>11</sup>. Bien que ces différentes réalisations partagent les valeurs de l’indéfini, elles ne sont pas strictement équivalentes et ne peuvent pas toujours commuter, notamment dans le cas de la négation (voir Ghomeshi, 2003, pour plus de détails). À la différence des noms nus, les indéfinis ont nécessairement une lecture existentielle, qui, en revanche, peut être, *de dicto*, non spécifique (un livre quelconque), ex. (12-a), ou, *de re*, spécifique (un certain livre), ex. (12-b).

- (11) a. ketāb=i  
livre=INDF
- b. ye(k) ketāb  
un livre
- c. ye(k) ketāb=i  
un livre=INDF

11. L’emploi de l’enclitique seul est rare dans le registre familier.

- (12) a. (yek) ketāb=i gom šod  
 (un) livre=INDF perdu devenir.PST.3SG  
 ‘Un livre (quelconque) a été perdu.’
- b. (yek) ketāb=i gom šod va hargez peydā  
 (un) livre=INDF perdu devenir.PST.3SG et jamais trouvé  
 na-šod  
 NEG-devenir.PST.3SG  
 ‘Un (certain) livre a été perdu et n’a jamais été retrouvé.’

Les syntagmes indéfinis sont également formés par les numéraux, ex. (13-a), ou les autres quantifieurs, ex. (13-b) et (13-c). Avec les numéraux et les quantifieurs, le nom reste au singulier, même lorsqu’il réfère à une entité multiple. Les numéraux ne sont pas compatibles avec l’enclitique =*i*, mais certains quantifieurs requièrent la présence de celui-ci.

- (13) a. se(=tā) ketāb  
 trois(=CLF) livre  
 ‘trois livres’
- b. čand(=tā) ketāb  
 quelques(=CLF) livre  
 ‘quelques livres’
- c. har/hič ketāb=i  
 chaque/aucun livre=INDF  
 ‘chaque livre/aucun livre’

### 1.1.2.3 Les syntagmes nominaux définis

En persan standard, il n’y a pas d’article défini à proprement parler. Les syntagmes nominaux définis ne contiennent aucun déterminant, ou sont introduits par des déterminants, comme les démonstratifs, qui impliquent la définitude, ex. (14-a). Le registre informel, en revanche, dispose d’un suffixe, *-(h)e*, qui marque un nom singulier comme défini-anaphorique, ex. (14-b). Ce suffixe est en distribution complémentaire avec le suffixe *-(h)ā* du pluriel<sup>12</sup>, ex. (14-c). En effet, ces pluriels impliquent généralement une lecture définie, ex. (15).

- (14) a. (in) ketāb  
 (ce) livre
- b. (in) ketāb-e  
 (ce) livre-DEF  
 ‘ce/le livre’
- c. (in) ketāb-hā  
 (ce) livre-PL  
 ‘ces/les livres’

12. Le persan dispose de plusieurs suffixes pour marquer le pluriel (*-(h)ā*, *-ān* et *-āt*), mais connaît aussi les pluriels non concaténatifs (*fekr<sub>sg</sub>* / *afkār<sub>pl</sub>*). La marque du pluriel par défaut est *-(h)ā*. Cette marque s’adjoint à n’importe quel nom pour former le pluriel (pour une présentation détaillée, voir Lazard et al. (2006) et Faghiri (2010)).

- (15) a. ketāb(-e) gom šod  
livre(-DEF) perdu devenir.PST.3SG  
'Le livre a été perdu.'
- b. ketāb-hā gom šod-and  
livre-PL perdu devenir.PST-3PL  
'Les (\*des) livres ont été perdus.'

Par ailleurs, la combinaison des numéraux avec le pluriel déclenche une lecture définie ou partitive, ex. (16-a) et (16-b) respectivement. Notons toutefois que le pluriel n'est pas incompatible avec la détermination indéfinie, =*i* ou *yek*, ex. (17). Une présentation exhaustive de la syntaxe et de la sémantique des pluriels nominaux dépasse le cadre du présent travail<sup>13</sup>.

- (16) a. se=tā ketāb-hā gom šod-and  
trois=CLF livre-PL perdu devenir.PST-3PL  
'Les trois livres ont été perdus.'
- b. čand=tā/se=tā az ketāb-hā gom šod-and  
quelques=CLF/trois=CLF de livre-PL perdu devenir.PST-3PL  
'Quelques/trois des livres ont été perdus.'
- (17) (yek) ketāb-hā=i heyn=e asbābkeši gom šod-and  
un livre-PL=INDF pendant=EZ déménagement perdu deveni.PST-3PL  
'Quelques-uns des livres ont été perdus pendant le déménagement.'

### 1.1.3 Le marquage différentiel de l'objet

La plupart des travaux traitant de l'ordre des constituants en persan se sont concentrés sur l'ordre relatif entre les deux compléments verbaux, à savoir le complément d'objet direct (COD) et le complément d'objet indirect (COI). Cet intérêt s'explique par l'existence du marquage différentiel de l'objet (MDO)<sup>14</sup>, qui intervient dans la position du COD en favorisant son éloignement du verbe, alors que les objets non marqués occupent, de préférence, une position adjacente au verbe. Les travaux dans le cadre génératif se sont saisis de ce phénomène afin de postuler deux positions syntaxiques distinctes pour les COD en fonction de leur marquage.

Le marquage différentiel de l'objet est une des spécificités du persan qui a fait couler le plus d'encre (Windfuhr, 1979) et qui continue de faire l'objet de débats controversés encore aujourd'hui. Dans ce qui suit, nous tâcherons de présenter les

13. Voir Ghomeshi (2003) pour une étude détaillée du lien entre la définitude et le marquage du pluriel.

14. Cette appellation, proposée par Bossong (1985), désigne la propriété de certaines langues qui marquent formellement une partie des objets directs en fonction de leurs traits sémantiques et/ou pragmatiques (voir par ex. Hopper & Thompson, 1980; Aissen, 2003; Lazard, 1982, 1994).

différents caractéristiques de son fonctionnement et à en résumer les analyses les plus répandues, sans chercher à trancher en faveur de l'une entre elles.

En persan, le marquage différentiel se réalise par l'enclitique  $=r\bar{a}$ , dans le registre formel, et  $=ro$ ,  $=o$ , dans le registre informel, après une finale vocalique et une finale consonantique respectivement, ex. *maqāla=rō* 'article= $R\bar{A}$ ' et *ketāb=o* 'livre= $R\bar{A}$ '<sup>15</sup>. Nous utiliserons la variante formelle tout au long de cette thèse, sauf dans les cas où le registre est expressément familier.

L'enclitique  $=r\bar{a}$  est traditionnellement analysé comme un marqueur de l'objet direct défini (Sadeghi, 1970; Khanlari, 1972; Gharib et al., 1994; Mahootian, 1997). La spécificité et/ou la topicalité ont été considérées ultérieurement comme des facteurs déterminants au détriment de la définitude (Browne, 1970; Peterson, 1974; Browning & Karimi, 1994; Karimi, 1990; Dabir-Moghaddam, 1990).

Dans le cadre de la syntaxe formelle,  $=r\bar{a}$  a été également traité comme un marqueur de Cas, dont la présence est expliquée par le filtre du Cas et/ou le critère thématique (Browning & Karimi, 1994; Ghomeshi, 1997; Ganjavi, 2007).

Un objet direct défini est toujours marqué de  $=r\bar{a}$ , quel que soit son statut discursif, ex. (18-a). Par conséquent, un objet non marqué reçoit obligatoirement une lecture indéfinie, ex. (18-b).

- (18) a. Maryam ketāb= $r\bar{a}$  xarid  
 Maryam livre= $R\bar{A}$  acheter.PST.3SG  
 'Maryam a acheté le livre.'
- b. Maryam ketāb xarid  
 Maryam livre acheter.PST.3SG  
 'Maryam a acheté un/des livre(s).'

Bien que le lien entre la définitude de l'objet direct et le marquage différentiel en persan soit incontestable, la définitude à elle seule ne suffit pas pour rendre compte de tous les emplois de  $=r\bar{a}$ . En effet, ce dernier peut être co-occurent avec le déterminant indéfini,  $=i$ , ex. (19), ce qui implique en général une lecture indéfinie spécifique de l'objet, qui peut être paraphrasée par 'un certain' en français.

- (19) a. Maryam (yek) ketāb= $i=r\bar{a}$  xarid  
 Maryam (un) livre= $INDF=R\bar{A}$  acheter.PST.3SG  
 'Maryam a acheté un (certain) livre.'
- b. Maryam ketāb= $i$  xarid  
 Maryam livre= $INDF$  acheter.PST.3SG

---

15. Il faut noter que  $=r\bar{a}$  se comporte comme un mot graphique et, du fait de sa postposition au nom tête, est traité, par les grammaires traditionnelles, ainsi que par certaines grammaires de référence (Lazard et al., 2006; Windfuhr, 1979), comme une postposition.

‘Maryam a acheté un livre (quelconque).’

Ce fait a conduit certains linguistes (Browne, 1970; Karimi, 1990; Browning & Karimi, 1994) à suggérer que le marquage par  $=r\bar{a}$  est déclenché par la spécificité plutôt que par la définitude, sachant qu’un constituant défini est *a fortiori* spécifique. D’autres auteurs (Peterson, 1974; Windfuhr, 1979; Dabir-Moghaddam, 1992) ont privilégié le lien entre la spécificité et la structure informationnelle et ont proposé d’analyser  $=r\bar{a}$  comme un marqueur de topicalisation, plus précisément comme un marqueur du topic secondaire (*secondary topic*), pour exclure ainsi le sujet qui n’est jamais marqué de  $=r\bar{a}$ <sup>16</sup>. Cette analyse s’appuie sur le fait que  $=r\bar{a}$  peut servir à topicaliser, particulièrement à l’oral, d’autres constituants dans la phrase, sujet excepté, ex. (20) à (22).

- (20) ru=ye yax(= $r\bar{a}$ ) āb rixt-and  
sur=EZ glace= $R\bar{A}$  eau verser.PST-3PL  
‘(Ils) ont versé de l’eau sur la glace.’
- (21) emruz(=o) dars mi-xun-am  
aujourd’hui(= $R\bar{A}$ ) leçon IPFV-lire.PRS.INF-1SG  
‘Aujourd’hui, j’étudie.’ ou ‘Toute la journée d’aujourd’hui, j’étudie.’<sup>17</sup>
- (22) Amir=o barā=š ketāb xarid-am (vs. barāye Amir)  
Amir= $R\bar{A}$  pour=3SG livre acheter.PST-1SG  
‘Pour Amir, je lui ai acheté un/des livres.’

Notons que lorsqu’un constituant enchâssé dans un autre syntagme (prépositionnel ou nominal) est ainsi topicalisé, un pronom clitique résomptif doit être réalisé dans sa position canonique dans la phrase, ex. (22).

En outre, l’analyse de  $=r\bar{a}$  comme “topicalisateur” permet de rendre compte des cas où le COD marqué de  $=r\bar{a}$  ne peut recevoir une lecture spécifique et/ou définie<sup>18</sup>, mais une lecture générique, ex. (23) à (25).

- (23) serke šir= $r\bar{a}$  mi-bor-ad  
vinaigre lait= $R\bar{A}$  IPFV-couper.PRS-3SG  
‘Le vinaigre fait tourner le lait.’

16. Pour cette même raison, on a qualifié  $=r\bar{a}$  de marqueur oblique, cf. Windfuhr (1979); Karimi (1990).

17. Lazard (1982) évoque la notion de la *totalité* pour analyser le rôle de  $=r\bar{a}$  dans les cas illustrés par cet exemple et celui qui précède.

18. Pour rendre compte de ces observations, certains auteurs (Shokouhi & Kipka, 2003; Roberts et al., 2009) ont proposé d’analyser  $=r\bar{a}$  comme un marqueur d’“identifiabilité” (*identifiability*, cf. Lambrecht (1994)). Ces analyses s’appuient, par ailleurs, sur des études quantitatives de corpus (oraux pour la première référence, et écrits pour la seconde) qui ont démontré que  $=r\bar{a}$  peut être joint à un constituant non spécifique représentant une information nouvelle sur le plan discursif.

Dabir-Moghaddam (1992, p. 557)

- (24) jāni=rā mi-koš-and  
 criminel=RĀ IPFV-tuer.PRS-3PL  
 ‘Les criminels, on les tue.’

Lazard (1994, p. 170)

- (25) xodā ensān=rā dust dār-ad  
 dieu humain=RĀ ami avoir.PRS-3SG  
 ‘Dieu aime l’homme.’

Meunier &amp; Samvelian (1997, p. 210)

Plusieurs études ont avancé que le fonctionnement de =rā ne peut être saisi de manière exhaustive par une analyse unifiée basée sur un trait binaire (voir Lazard, 1982, 1994; Meunier & Samvelian, 1997; Ghomeshi, 1997). En effet, un ensemble de facteurs, tels que la définitude, la topicalité, l’individuation, la totalité, l’humanité, ou encore le sémantisme du COD et le verbe, qui contribuent tous à la transitivité (Hopper & Thompson, 1980), semble intervenir dans le marquage différentiel par =rā.

Néanmoins, Dabir-Moghaddam (1990, 1992, 2005) plaide en faveur d’une analyse unifiée de =rā, à l’intersection de la syntaxe et de la pragmatique, en tant que marqueur du topic secondaire (*secondary topic*). Il postule que ce dernier n’est pas une notion absolue (catégorique), mais se définit sur un *continuum* où les facteurs mentionnés dans le précédent paragraphe interviennent. En outre, cet auteur adopte une conception du topic qui n’exclut pas l’information nouvelle (cf. §1.3.1 pour plus de détails)<sup>19</sup>.

Présentons enfin un dernier cas de figure qui, bien que très marginalement évoqué la littérature, n’est pas sans intérêt dans le cadre du présent travail<sup>20</sup>. Lazard (1982)

---

19. Dalrymple & Nikolaeva (2011), dans leur étude du marquage différentiel de l’objet et de la structure informationnelle à travers une grande variété de langues, s’inscrivent dans la démarche proposée par Dabir-Moghaddam (1990, 1992, 2005), tout en émettant deux réserves : 1) un COD marqué n’est pas forcément le topic secondaire, mais peut également correspondre au topic principal de la phrase, notamment lorsque le sujet est en focus, ex. (i). Nous ne sommes pas tout à fait d’accord avec le jugement de grammaticalité proposé par les auteurs dans cet exemple ; 2) dans certains cas le marquage s’explique uniquement par des propriétés sémantiques, notamment la définitude, sans que le statut discursif intervienne. En effet, nous l’avons également noté plus haut dans cette section, un COD défini est toujours marqué quelque soit son statut discursif.

(i) ki ketāb=i=rā / ?\*ketāb=i xarid  
 qui livre=INDF=RĀ / livre=INDF acheter.PST.3SG  
 ‘Qui a acheté un livre?’

Dalrymple &amp; Nikolaeva (2011, p. 112)

20. Un autre cas de figure marginalement évoqué qui mérite d’être mentionné est la présence obligatoire de =rā, quel que soit son statut discursif, dans les constructions à l’attribut de l’objet (Sadeghi, 1970, pp. 14-15) – les exemples de Sadeghi sont tirés d’un livre publié en 1964 – :

note que “le poids relatif du groupe objet et de ce qui suit jusqu’au verbe inclus” favorise le marquage par  $=r\bar{a}$ , comme l’illustre la paire d’exemples ci-dessous<sup>21</sup> :

- (26) a. bošqāb=e sard=i=rā jolow=ye boxār=e samāvar  
 assiette=EZ froid=INDF=RĀ devant=EZ vapeur=EZ samovar  
 be-gir-id  
 SBJV-tenir.PRS-2PL  
 ‘Exposez une assiette froide à la vapeur d’un samovar.’  
 b. jolow=ye boxār=e samāvar bošqāb=e sard=i be-gir-id

Lazard (1982, p. 338)

Il faut noter que le marquage différentiel ne peut affecter qu’un seul COD dans la même phrase, ce qui est attendu si l’on compare  $=r\bar{a}$  à une sorte de marquage casuel accusatif. On peut cependant observer plus d’un constituant marqué de  $=r\bar{a}$  dans une même phrase, étant donné l’usage de  $=r\bar{a}$  pour topicaliser tout constituant extrait d’un syntagme non sujet, cf. ex. (21) et (22), qui peut cumuler ou non avec un COD marqué de  $=r\bar{a}$ , ex. (27) à (30). Pour plus de détails, voir Lazard (1982); Dabir-Moghaddam (1992).

- (27) a. Amir=o barā=š in ketāb=o xarid-am  
 Amir=RĀ pour=3SG ce livre=RĀ acheter.PST-1SG  
 ‘Quant à Amir, je lui ai acheté un/des livres.’  
 b. barā=ye Amir in ketāb=o xarid-am
- (28) a. Amir=o in bār=o barā=š ketāb xarid-am  
 Amir=RĀ ce fois=RĀ pour=3SG livre acheter.PST-1SG  
 ‘Quant à Amir, cette fois, je lui ai acheté un/des livres.’  
 b. in bār barā=ye Amir ketāb xarid-am
- (29) a. māšin=o dar=eš=o be-band  
 voiture=RĀ portière=3SG=RĀ IMP-fermer.PRS.2SG  
 ‘Ferme la porte de la voiture!’  
 b. [dar=e māšin]=o be-band
- (30) a. havij=o bāyad pust=eš=o kand  
 carotte=RĀ il-faut peau=3SG=RĀ arracher.PST  
 ‘La carotte, il faut l’éplucher.’

- 
- (i) tassavor ne-mi-kard-am inčenin moqe’=i=rā ... šāyeste  
 imagination NEG-IPFV-faire.PST-1SG tel moment=INDF=RĀ approprié  
 be-dān-i  
 SBJV-considérer.PRS-2SG  
 ‘Je n’imaginai pas que tu considères ... un tel moment approprié.’

21. L’auteur a emprunté cet exemple à J. A. Boyle, *Grammaire of Modern Persian* (Wiesbaden, 1966, p. 58).

- b. [pust=e havij]=o bāyad kand

Morphologiquement, =*rā* est un affixe syntagmatique qui n’affecte pas l’accent lexical de l’élément auquel il s’adjoint et qui se place à la frontière droite, c’est-à-dire sur le dernier élément du syntagme nominal :

- (31) [ketāb=e dastur=e zabān=e fārsi=ye čāp=e jadid]=**rā**  
 livre=EZ grammaire=EZ langue=EZ persan=EZ édition=EZ nouveau=RĀ  
 xarid-am  
 acheter.PST-1SG  
 ‘J’ai acheté la dernière édition du livre de la grammaire (de la langue) per-  
 sane.’

Néanmoins, dans le cas où le nom tête est modifié par une subordonnée relative, =*rā* se place soit juste avant la relative (placement prescrit par la grammaire normative), ex. (32-a), soit à la frontière droite du syntagme, ex. (32-b). Notons que le marquage double, bien que proscrit par le bon usage, n’est pas pour autant exclu, ex. (32-c).

- (32) a. [ketāb=i=**rā** ke ru=ye miz bud] xānd-am  
 livre=RSTR<sup>22</sup>=RĀ que sur=EZ table être.PST.3SG lire.PST-1SG  
 b. [ketāb=i ke ru=ye miz bud]=**rā** xānd-am  
 livre=RSTR que sur=EZ table être.PST.3SG=RĀ lire.PST-1SG  
 c. [ketāb=i=**rā** ke ru=ye miz bud]=**rā** xānd-am  
 livre=RSTR=RĀ que sur=EZ table être.PST.3SG=RĀ lire.PST-1SG  
 ‘J’ai lu le livre qui était sur la table.’

### 1.1.4 Les prédicats complexes

Le persan dispose d’un nombre restreint de verbes simples, moins de 250 verbes, dont seulement la moitié est employée dans le langage courant<sup>23</sup>. Le lexique verbal du persan est essentiellement constitué de combinaisons syntaxiques, appelées “prédicats complexes”<sup>24</sup> et s’enrichit régulièrement par la création de nouvelles combinaisons<sup>25</sup>. Ces combinaisons incluent un verbe simple et un élément non verbal,

22. Il existe deux déterminants enclitiques =*i* distincts en persan. Le premier est le déterminant indéfini que nous avons présenté en §1.1.2.2. Le second apparaît exclusivement devant une relative restrictive (voir Lazard, 1966; Lazard et al., 2006; Samvelian, 2006b).

23. Khanlari (1986) fournit une liste de 279 verbes simples au total en persan. Sadeghi (1993) évalue à 252 le nombre de verbes simples, dont 115 communément employés.

24. D’autres appellations courantes sont : verbes complexes, *Compound Verbs*, et constructions à verbe support, *Light Verb Constructions (LVCs)*.

25. Bien qu’elle soit disponible, la conversion d’un nom et d’un adjectif en verbe, ex. *raqs<sub>N</sub>* ‘danse’ : *raqsidan<sub>V</sub>* ‘danser’, *ārām<sub>Adj</sub>* ‘calme’ : *ārāmidan<sub>V</sub>* ‘se calmer, se reposer’, est rarement effective.

tel qu'un nom, un adjectif, un syntagme prépositionnel, ou une particule (cf. ex. Goldberg, 1996; Karimi, 1997; Karimi-Doostan, 1997; Megerdooian, 2002) :

- N + V : *negāh kardan* 'regarder (Lit. regard faire)'
- Adj + V : *derāz kešidan* 's'allonger (Lit. allongé tirer)'
- SP + V : *az dast dādan* 'perdre (Lit. de main donner)'
- Particule + V : *bar dāštan* 'prendre (Lit. PARTICULE avoir)'

Les prédicats complexes sont des combinaisons syntaxiques comparables à des lexèmes (Bonami & Samvelian, 2010), en ceci qu'ils sont souvent lexicalisés et ont une interprétation conventionnelle (c'est-à-dire idiomatique). Ainsi, sur le plan sémantique, ils se comportent comme des verbes simples et peuvent prendre divers compléments. Par exemple, *negāh kardan* 'regarder', de même que *az dast dādan* 'perdre', s'assimilent à des verbes simples transitifs et prennent un COD, ex. (33), *derāz kešidan* peut prendre un complément prépositionnel de lieu, ex. (34), *nešān dādan* s'assimile à un verbe ditransitif et s'accompagne d'un COD et d'un COI, ex. (35), enfin *harf zadan* peut prendre plusieurs COI, ex. (36).

- (33) Sepide [Ali=rā] **negāh kard**  
 Sepideh Ali=RĀ regard faire.PST.3SG  
 'Sepdideh regarda Ali.'
- (34) Sepide [ru=ye taxt] **derāz kešid**  
 Sepideh sur=EZ lit allongé tirer.PST.3SG  
 'Sepdideh s'allongea sur le lit.'
- (35) Sepide [Ali=rā] [be Puyān] **nešān dād**  
 Sepideh Ali=RĀ à Puyan signe donner.PST.3SG  
 'Sepdideh montra Ali à Puyan.'
- (36) Sepide [bā Ali] [darbāre=ye Puyān] **harf zad**  
 Sepideh avec Ali de/sur=EZ Puyan parole frapper.PST.3SG  
 'Sepdideh parla de Puyan à Ali.'

Tout en fonctionnant comme des unités lexicales sur le plan sémantique, les prédicats complexes – hormis les combinaisons particule-verbe – présentent toutes les propriétés syntaxiques d'une combinaison complément-verbe ordinaire, comme cela a été démontré par Samvelian (2001, 2006a, 2012). Citons notamment l'élément non verbal du prédicat complexe, qui peut être séparé du verbe comme peut l'être un complément verbal ordinaire. Comparez, dans les exemples (37) et (38) ci-dessous, les phrases en (a) illustrant la combinaison du verbe *xaridan* 'acheter' avec son COD *ketāb* 'livre' et les phrases en (b) avec le prédicat complexe *qowl dādan* 'promettre (Lit. promesse donner)'

- (37) a. Maryam ketāb [hičvaqt] ne-mi-xar-ad  
 Maryam livre jamais NEG-IPFV-acheter.PRS-3SG  
 ‘Maryam n’achète jamais de livres.’  
 b. Maryam qowl [hičvaqt] ne-mi-dah-ad  
 Maryam promesse jamais NEG-IPFV-donner.PRS-3SG  
 ‘Maryam ne promet jamais.’
- (38) a. Maryam ketāb [be bačče-hā] xāh-ad dād  
 Maryam livre à enfant-PL AUX.FUT-3SG donner.PST  
 ‘Maryam donnera un/des livre(s) aux enfants.’  
 b. Maryam qowl [be bačče-hā] xāh-ad dād  
 Maryam promesse à enfant-PL AUX.FUT-3SG donner.PST  
 ‘Maryam fera une promesse/promettra aux enfants.’

Citons également l’élément nominal du prédicat complexe, qui, au même titre qu’un complément verbal, peut être déterminé/quantifié, modifié, extrait et passivisé, ex. *harf* ‘parole’ dans le prédicat complexe *harf zadan* ‘parler (Lit. parole frapper)’, en (39).

- (39) a. Ali in harf=rā / yek harf=e jāleb zād  
 Ali ce parole=RĀ / un parole=EZ intéressant frapper.PST.3SG  
 ‘Ali dit ceci/une chose intéressante.’  
 b. harf=i az Ali momken ast na-zan-i ?  
 parole=INDF de Ali possible être.PST.3SG NEG-frapper.PRS-2SG  
 ‘Est-ce possible que tu ne dises rien d’Ali?’  
 c. in harf zade šod  
 ce parole frapper.PP devenir.PST.3SG  
 ‘Ceci fut dit.’

Samvelian (2001, 2006a, 2012) en conclut que le parallélisme syntaxique entre les séquences complément-verbe ordinaires et celles formant un prédicat complexe est incontestable, bien que les généralisations ci-dessus connaissent des exceptions<sup>26</sup> au demeurant marginales. En somme, les différences que l’on peut relever entre les deux types de combinaison demeurent d’ordre sémantique; par conséquent, il est impossible d’établir une distinction catégorique ou formelle entre ceux-ci<sup>27</sup>.

26. Certains prédicats complexes non-verbe manifestent un degré de fusion comparable aux combinaisons particule-verbe, ex. *qarār dādan* ‘installer (Lit. installation donner)’, *qarār gereftan* ‘s’installer (Lit. installation prendre)’. Dans d’autres, l’élément nominal s’apprête difficilement à être déterminé, quantifié ou modifié, ex. *dust dāstan* ‘aimer (Lit. ami avoir)’, *negah dāstan* ‘maintenir (Lit. regard avoir)’, *ersāl kardan* ‘envoyer (Lit. envoyer faire)’, etc.

27. Notons que cette analyse des prédicats complexes ne fait pas consensus dans la mesure où la majorité des travaux, admettant les propriétés syntagmatiques des prédicats complexes (Karimi, 1997; Karimi-Doostan, 1997; Megerdooian, 2002), accorde un statut syntaxique à part aux séquences nom-verbe des prédicats complexes. Toutefois, ces travaux, développés pour la plupart dans la visée d’une théorie formelle précise, s’appuient sur un ensemble restreint de données, gé-

Ces faits ne restent pas sans conséquences dans l'étude de l'ordre des compléments verbaux, dans la mesure où cela implique que le constituant non verbal d'un prédicat complexe soit traité comme un complément du verbe, ce dernier s'inscrivant dans la liste des compléments sélectionnés du verbe au même titre que les autres arguments du prédicat complexe (Samvelian, 2012, p. 86).

À titre d'exemple, comme dans le cas d'un verbe simple comme *dādan* 'donner', où il y a trois dépendants à positionner par rapport au verbe, dans le cas d'un prédicat complexe comme *qowl dādan* 'promettre (Lit. promesse donner)', ex. (40), il y a également trois arguments à ordonner. En effet, le verbe simple *dādan* prend le sujet et le COI du prédicat complexe ainsi que l'élément nominal, *qowl*, analysable comme le COD du verbe *dādan*.

- (40) a. Sepide [COD in qowl=rā] [COI be Maryam] dād  
 Sepideh ce promesse=RĀ à Maryam donner.PST.3SG  
 'Sepideh donna cette promesse à Maryam.'
- b. Sepide [COI be Maryam] [COD qowl] dād  
 Sepideh à Maryam promesse donner.PST.3SG  
 'Sepideh donna cette promesse à Maryam.'

Ainsi, le nombre de dépendants verbaux à réalisation directe (c'est-à-dire comme syntagme nominal) peut augmenter : ex. (41), formé avec le prédicat complexe *hediye dādan* 'offrir (un cadeau)'. Nous reviendrons sur ce point en §1.1.5.3 où nous aborderons la possibilité d'avoir dans une même phrase plusieurs compléments à réalisation directe.

- (41) a. Sepide [SP be Maryam] [SN (yek) hediye(=i)] dād  
 Sepideh à Maryam (un) cadeau(=INDF) donner.PST.3SG  
 'Sepideh offrit un cadeau à Maryam.'
- b. Sepide [SN in ketāb=rā] [SP be Maryam] [SN hediye] dād  
 Sepideh ce livre=RĀ à Maryam cadeau donner.PST.3SG  
 'Sepideh offrit ce livre à Maryam (en cadeau).'
- c. Sepide [SP be Maryam] [SN ketāb(=i)] [SN hediye] dād  
 Sepideh à Maryam livre(=INDF) cadeau donner.PST.3SG  
 'Sepideh offrit un livre/des livres à Maryam (en cadeau).'

---

néralement issues uniquement de l'introspection, et écartent les cas de figure problématiques. Le traitement proposé par Samvelian (2001, 2006a, 2012) et Samvelian & Faghiri (2013, 2014, 2016) se distingue par une couverture empirique plus étendue.

## 1.1.5 La réalisation des arguments

### 1.1.5.1 Les arguments implicites

Le persan est une langue à sujet nul, ou *pro-drop* dans la terminologie générative, ce qui signifie que la réalisation syntaxique du sujet grammatical n'est pas obligatoire même dans une proposition finie ayant un verbe fléchi/conjugué.

- (42) vali ba'dan tānk-hā=rā be xiāban-hā ferestād-and  
 mais après tank-PL=RĀ à rue-PL envoyer.PST-3PL  
 'Mais plus tard ils ont envoyé les tanks dans les rues.'

Corpus Bijankhan<sup>28</sup>

Il faut noter qu'il n'existe pas de syncrétisme dans la conjugaison verbale du persan. Chaque désinence réalise sans ambiguïté aucune la personne et le nombre du sujet grammatical<sup>29</sup>. Le tableau 1.1 présente les deux paradigmes des désinences personnelles verbales utilisées respectivement pour les temps présent et passé<sup>30</sup>.

<b>Passé</b>		<b>Présent</b>		
RAD.PST + suffixes personnels		IPFV + RAD.PRS + suffixes personnels		
SG	PL	SG	PL	
1	-am xord-am	-im xord-im	-am mi-xor-am	-im mi-xor-im
2	-i xord-i	-id (-in) xord-id (xord-in)	-i mi-xor-i	-id (-in) mi-xor-id (mi-xor-in)
3	∅ xord	-and (-an) xord-and (xord-an)	-ad (-e) mi-xor-ad (mi-xor-e)	-and (an) mi-xor-and (mi-xor-an)

TABLE 1.1 – Suffixes personnels verbaux (pour *xordan* 'manger')

La réalisation anaphorique (par un pronom) des COD ou COI n'est pas obligatoire, tel que précisé dans les deux exemples suivants :

- (43) aks-hā=rā did-i? did-am  
 photo-PL=RĀ voir.PST-2SG voir.PST-1SG  
 'As-tu vu les photos? Je \*(les) ai vues.'

- (44) be Maryam goft-i? goft-am  
 à Maryam dire.PST-2SG dire.PST-1SG

28. Se reporter à la section 4.1 pour une présentation de ce corpus.

29. Notons que le genre grammatical n'existe pas en persan. Par commodité, nous glosons la troisième personne par le masculin en français.

30. Certaines désinences ont deux formes différentes en fonction du registre; les formes notées entre parenthèses renvoient aux variantes informelles.

‘As-tu dit à Maryam ? Je \*(lui) ai dit.’

### 1.1.5.2 La réalisation pronominale enclitique de l’objet

Outre les pronoms personnels toniques, le persan dispose d’un paradigme de pronoms personnels clitiques, présentés dans le tableau 1.2<sup>31</sup>. Ces pronoms réalisent diverses fonctions, telles que le complément adnominal (génitif), ex. (46), le complément d’une préposition, ex. (45), le COD, ex. (47), et marginalement le COI, ex. (49).

	SG	PL
1	-am	-emān (-emun)
2	-at (-et)	-etān (-etun)
3	-aš (-eš)	-ešān (-ešun)

TABLE 1.2 – Pronoms personnels clitiques

Lorsqu’ils réalisent la fonction du complément d’une préposition, ex. (45), ou celui d’un nom, ex. (46), ces pronoms sont *in situ*. En assumant la fonction du COD (ou du COI), ils se réalisent comme un enclitique sur le verbe lexical, ex. (47) ou sur l’auxiliaire, ex. (48). Avec les prédicats complexes, le pronom clitique peut également se réaliser sur l’élément non verbal d’un prédicat complexe, ex. (50) (voir Samvelian & Tseng, 2010, pour le placement des clitiques et leurs propriétés).

- (45) az Maryam / az=aš  
de Maryam de=3SG  
‘de Maryam’ / ‘d’elle’
- (46) ketāb=e Maryam / ketāb=aš  
livre=EZ Maryam livre=3SG  
‘le livre de Maryam’ / ‘son livre’
- (47) ketāb=rā xarid-am / xarid-am=aš  
livre=RĀ acheter.PST-1SG acheter.PRS-1SG=3SG  
‘j’ai acheté le livre’ / ‘je l’ai acheté.’
- (48) ketāb=rā xaride bud-am / xaride bud-am=aš  
livre=RĀ acheter.PP être.PST-1SG acheter.PP être.PST-1SG=3SG  
‘j’ai acheté le livre’ / ‘je l’ai acheté.’
- (49) be Ali goft-am / goft-am=aš  
à Ali dire.PST-1SG dire.PST-1SG=3SG

31. Ces clitiques connaissent deux formes différentes en fonction du registre ; les formes entre parenthèses correspondent à la variante informelle

‘j’ai dit à Ali’ / ‘je lui ai dit.’

- (50) Ali=rā negāh kard-am / negāh(=aš) kard-am(=aš)  
 Ali=RĀ regard faire.PST-1SG regard(=3SG) faire.PST-1SG(=3SG)  
 ‘j’ai regardé Ali’ / ‘je l’ai regardé.’

Dans le registre familier, le clitique peut être doublé d’un objet syntaxique clivé.

- (51) ketāb=o xarid-am=eš  
 livre=RĀ acheter.PST-1SG=3SG  
 ‘Le livre, je l’ai acheté.’

### 1.1.5.3 Les constructions à double objet

Il est possible en persan qu’un verbe prenne plusieurs compléments directs. C’est le cas des verbes, comme *nāmīdan* ‘nommer’, *xāndan* ‘appeler’ ou *dānestan* ‘considérer’, prenant un complément attributif qui peut se réaliser comme un syntagme nominal, comme illustré dans ex. (52) et (53). C’est également le cas des constructions appelées “à double objet” comparables à l’alternance dative de l’anglais, ex. (54) et (55), et celles qui rappellent l’alternance dite *spray-load* de l’anglais ou “standard-croisé” du français, ex. (56).

- (52) (man) [SN in ketāb=rā] [SN behtarīn ketāb=e Maryam]  
 (moi) ce livre=RĀ meilleur livre=EZ Maryam  
 mi-dān-am  
 IPFV-considérer.PRS-1SG  
 ‘Je considère ce livre (comme) le meilleur livre de Maryam.’

- (53) man ... [SN majles=e šešom=rā] [SN majles=e eslāhāt]  
 moi parlement=EZ sixième=RĀ parlement=EZ réformes  
 nāmīd-am  
 nommer.PST-1SG  
 ‘... j’ai appelé le sixième parlement le parlement des réformes.’

Corpus Bijankhan

- (54) a. Sepide [SP be bačče-hā] [SN qazā] dād  
 Sepideh à enfant-PL repas donner.PST.3SG  
 ‘Sepideh donna à manger aux enfants.’  
 b. Sepide [SN bačče-hā=rā] [SN qazā] dād  
 Sepideh enfant-PL=RĀ repas donner.PST.3SG  
 ‘Sepideh fit manger les enfants.’
- (55) a. ... agar [SP be nān-hā] [SN rowqan] mi-zad-and  
 si à pain-PL huile IPFV-frapper.PST-3PL  
 ‘... s’ils mettaient de l’huile sur les pains’

- b. ... ke Hasan va Hoseyn [SN nān-hā=rā] [SN rowqan]  
 que Hassan et Hosseyn pain-PL=RĀ huile  
 be-zan-and  
 SBJV-frapper.PRS-3PL  
 ‘... que Hassan et Hosseyn huilent les pains’

Corpus Bijankhan

- (56) a. Omid [SP dar in zamin] [SN gandom] kāšt  
 Omid dans ce terre blé planter.PST.3SG  
 ‘Omid planta du blé dans cette terre.’  
 b. Omid [SN in zamin=rā] [SN gandom] kāšt  
 Omid ce terre=RĀ blé planter.PST.3SG  
 ‘Omid planta cette terre de blé.’

Samvelian (2012, p. 51)

Par ailleurs, comme nous l’avons vu précédemment (§1.1.4), cette construction se manifeste avec les prédicats complexes nom-verbe prenant un COD, comme c’est le cas de *vazn kardan* ‘peser (Lit. poids faire)’ en ex. (57). Notons que, dans ce cas de figure, le nombre de compléments à réalisation nominale n’est pas limité à deux (Lazard, 1982), comme illustré en ex. (58).

- (57) Maryam [SN baste-hā=rā] [SN vazn] kard  
 Maryam paquet-PL=RĀ poids faire.PST.3SG  
 ‘Maryam pesa les paquets.’

- (58) [SN Nowruz=rā] be šomā [SN tabrik] [SN arz]  
 Nowruz=RĀ à vous félicitation présentation  
 mi-kon-am  
 IPFV-faire.PRS-1SG  
 ‘Je vous présente mes vœux de bonne année.’

Lazard (1982, p. 342)

Il faut noter que, dans les cas de figure mentionnés ci-dessus, les différents compléments nominaux n’ont pas le même statut syntaxique. Rappelons que, quelle qu’en soit la configuration, une phrase simple n’admet qu’un seul COD marqué de =rā (cf. §1.1.3). Le complément, qui peut être marqué de =rā, dispose d’une mobilité nettement plus importante que le second complément nominal. Ce dernier se place adjacent au verbe dans l’ordre canonique<sup>32</sup>.

32. Samvelian (2001) rend compte de cette différence en se basant sur la hiérarchie d’accessibilité ou d’obliquité (*obliqueness*) développée par Keenan & Comrie (1977). Le complément susceptible d’être marqué de =rā, qui est également le complément occupant la fonction sujet lors de la passivisation, est considéré moins oblique (ou plus accessible) que le second complément nominal

- (59) a. Maryam [<sub>SN</sub> baste-hā=rā] [<sub>SN</sub> vazn(\*=rā)] kard  
 Maryam paquet-PL-RĀ poids faire.PST.3SG  
 ‘Maryam pesa les paquets.’
- b. ??Maryam [<sub>SN</sub> vazn] [<sub>SN</sub> baste-hā=rā] kard  
 Maryam poids paquet-PL=RĀ faire.PST.3SG

## 1.2 L’ordre canonique des constituants dans la phrase

Le persan est considéré comme une langue SOV (voir Lazard *et al.*, 2006; Mahootian, 1997; Windfuhr, 1979; Dabir-Moghaddam, 2001; Karimi, 1994). L’ordre canonique<sup>33</sup> exige une organisation verbe final indépendamment de la modalité de la phrase : Sujet > Prédicat > Copule ou Sujet > Compléments > Verbe (cf. §1.2.2).

- (60) a. [<sub>Suj</sub> in film] [<sub>Prédicat</sub> behtarin film=e Farhādi] [<sub>Cop</sub> ast]  
 ce film meilleur film=EZ Farhadi être.PRS.3SG
- b. [<sub>Suj</sub> film=e Farhādi] [<sub>COD</sub> Oskār=e behtarin film=e xāreji=rā]  
 film=EZ Farhadi Oskar=EZ meilleur film=EZ étranger=RĀ  
 [<sub>V</sub> bord]  
 gagner.PST.3SG  
 ‘Le film de Farhadi a gagné l’Oscar du meilleur film étranger.’
- c. [<sub>Suj</sub> film=e Farhādi] [<sub>COI</sub> be Oskār] [<sub>V</sub> raft]  
 film=EZ Farhadi à Oskar=EZ aller.PST.3SG  
 ‘Le film de Farhadi est allé aux Oscars.’

L’ordre SOV est incontestablement l’ordre non marqué<sup>34</sup> ou neutre du point de vue discursif et prosodique et l’ordre prédominant en persan – au moins en ce qui concerne la langue écrite –, qui manifeste un ordre SOV (presque) rigide<sup>35</sup> (Karimi, 1994; Dabir-Moghaddam, 2001). Néanmoins, tous les autres ordres, ex. SVO, OVS, VSO, etc., peuvent également se produire<sup>36</sup>, faisant du persan une langue à ordre (relativement) libre.

---

(voir également Lazard, 1994, p. 95).

33. Également appelé ordre préféré, ordre normal, ordre de base, ordre non marqué, ordre neutre ou encore ordre naturel dans la littérature.

34. Pour une définition de cette notion, voir §3.2.2.2 au chapitre 3.

35. Il est à noter que, historiquement, le persan moderne, comme en témoigne la littérature de l’époque, jouissait d’une liberté d’ordonnement beaucoup plus prononcée aux débuts de son développement, c’est-à-dire au 13<sup>ème</sup> siècle, et qu’il n’est devenu “conservateur” vis-à-vis de l’ordre que dans son évolution tardive : l’ordre SOV s’est imposé selon un processus graduel (et inégal) qui s’est stabilisé vers la fin du 19<sup>ème</sup> (Khanlari, 1994, pp. 18–21 et 266–267).

36. Notons que les dérogations à l’ordre canonique se produisent avec une fréquence inégale (Lazard *et al.*, 2006, p. 194). Notamment, certains ordres sont propres à l’oral et présentent une prosodie fortement marquée (cf. la section 1.3).

Dans cette section, nous présentons l'ordre canonique des constituants tout d'abord dans les phrases déclaratives (§1.2.1), ensuite dans les phrases non déclaratives (§1.2.2), et enfin dans les propositions subordonnées (§1.2.3). Nous aborderons les ordres non canoniques dans la section 1.3.

### 1.2.1 Phrase déclarative

Les grammaires traditionnelles et scolaires donnent des indications générales sur l'ordre canonique dans la phrase en persan, présenté comme l'ordre “naturel” ou “préférable” ou encore la “position propre” des constituants (voir par ex. Khanlari, 1972; Givi Ahmadi & Anvari, 1995; Gharib et al., 1994).

Ces indications peuvent être synthétisées comme suit :

- le sujet se met “généralement” en position initiale,
- le verbe est “toujours” en position finale,
- l'objet ou les objets se met(tent) entre le sujet et le verbe,
- les circonstanciels peuvent se mettre à tout endroit dans le domaine préverbal.

Les mêmes observations se rencontrent dans les travaux plus récents sur la syntaxe du persan. Il faut ajouter à cela que les compléments phrastiques (complétives) se placent régulièrement et sans exception après le verbe.

L'ordre canonique dans la phrase déclarative peut par conséquent être schématisé comme en (61)<sup>37</sup> :

(61) (S) (OI) (OD) /(OD) (OI) V (Complétive)

Comme ce schéma le montre, l'ordre relatif entre les deux compléments direct et indirect peut varier. Le COD peut précéder ou suivre le COI. L'ordre relatif entre les deux objets est généralement associé au marquage différentiel de l'objet.

(62) a. (S) OD= $r\bar{a}$  OI V  
 b. (S) OI OD V

Comme le montre le schéma (62), il est admis que le COD marqué de  $=r\bar{a}$ <sup>38</sup> précède le COI, ex. (63), tandis que le COD non marqué suit celui-ci, ex. (64) (voir Karimi, 1994; Browning & Karimi, 1994; Mahootian, 1997; Rasekhmahand, 2004; Ganjavi, 2007; Windfuhr & Perry, 2009; Roberts et al., 2009) :

37. Rappelons qu'en persan les compléments verbaux peuvent être implicites ou avoir une réalisation enclitique sur le verbe (cf. §1.1.5.1 et §1.1.5.2).

38. Notons que les généralisations que l'on retrouve dans la littérature ne sont pas nécessairement formulées en termes de marquage différentiel de l'objet, mais en faisant appel aux propriétés qui déclenchent ce dernier, notamment la définitude (Mahootian, 1997; Roberts et al., 2009; Windfuhr & Perry, 2009) ou la spécificité (Karimi, 1994; Browning & Karimi, 1994).

- (63) dāvar-ān [jāyeze=ye behtarin film=rā] [be film=e Farhādi]  
 arbitre-PL [prix=EZ meilleur film=RĀ] [à film=EZ Farhādi]  
 dād-and  
 donner.PST-3PL  
 ‘Les membres du jury ont attribué le Prix du meilleur film au film de Farhādi.’
- (64) irān [be kešvar-hā=ye orupāi] [naft] ne-mi-foruš-ad  
 Iran [à pays-PL=EZ européen] [pétrole] NEG-IPFV-vendre.PRS-3SG  
 ‘L’Iran ne vend pas de pétrole aux pays européens.’

Le même constat est partagé par les grammairiens, même si certains comme Givi Ahmadi & Anvari (1995) établissent trois cas de figure en tenant compte de la présence de =i de l’indéfini (cf. §1.1.2.2) : 1) le COD marqué de =rā précède le COI ; 2) le COD non marqué de =rā suit le COI ; 3) Le COD indéfini non marqué de =rā peut suivre et précéder le COI. Les exemples de Givi Ahmadi & Anvari (1995, p. 305) pour illustrer ces cas sont donnés en (65).

- (65) a. Yusef ketāb=rā be ketābxāne dād  
 Yusef livre=RĀ à librairie donner.PST.3SG  
 ‘Yusef donna le livre à la librairie.’
- b. Yusef az ketābxāne ketāb gereft  
 Yusef de librairie livre prendre.PST.3SG  
 ‘Yusef pris un livre/des livres de la librairie.’
- c. (i) Yusef ketāb=i az ketābxāne gereft  
 Yusef livre=INDF de librairie prendre.PST.3SG
- (ii) Yusef az ketābxāne ketāb=i gereft  
 Yusef de librairie livre=INDF prendre.PST.3SG  
 ‘Yusef pris un livre de la librairie.’

Comme il a été dit précédemment, les compléments phrastiques se mettent obligatoirement après le verbe, ex. (66). Il faut néanmoins noter qu’un complément phrastique peut être nominalisé – à l’aide du déterminant démonstratif *in* ‘ce’ – et, dans ce cas, il se comporte comme un complément nominal et peut se trouver dans le domaine préverbal.

- (66) a. mi-dān-am (ke) bārān āmade ast  
 IPFV-savoir.PRS-1SG (que) pluie venu.PP être.PRS.3SG  
 ‘Je sais qu’il a plu.’
- b. \*[(ke) bārān āmade ast] mi-dān-am  
 que pluie venu.PP être.PRS IPFV-savoir.PRS-1SG
- (67) a. [in ke bārān āmade ast]=rā mi-dān-am  
 ce que pluie venu.pp être.PRS=RĀ IPFV-savoir.PRS-1SG

- b. in=rā ke bārān āmade ast mi-dān-am  
 ce=RĀ que pluie venu.pp être.PRS IPFV-savoir.PRS-1SG  
 ‘Qu’il a plu, je le sais.’

Les circonstanciels prépositionnels et adverbiaux peuvent se placer à tout endroit de la phrase (Mahootian, 1997, p. 50). Certaines préférences d’ordre ont été signalées en fonction du sémantisme de ces circonstanciels dans divers travaux. Nous en donnons deux exemples ici.

- (68) a. Sujet < CC. de temps < COD < CC. de source < CC. locatif < CC. bénéfactif < CC. instrumental  
 Mahootian (1997, p. 50)
- b. Temps < Manière < Locatif-Directionnel/Lieu  
 Windfuhr & Perry (2009, p. 480) et Mace (2003, p. 144)

Il est important de noter que, dans le registre informel (exclusivement), les compléments directionnels (“allatif” dans la terminologie de Creissels, 2006) et/ou locatifs, ex. (69) à (72), se placent systématiquement après le verbe. En d’autres termes, dans le cas de ces objets, l’ordre linéaire non marqué sur le plan discursif et prosodique est SVO<sup>39</sup>. De plus, la préposition *be* ‘à’, qui, dans l’ordre SOV, introduit les compléments directionnels et certains compléments locatifs, est omise dans le domaine postverbal. Dans les exemples ci-dessous, la variante informelle est donnée en (a) et la variante formelle en (b).

- (69) a. Sepide rafte bud [Yazd]  
 Sepideh aller.PP être.PST.3SG Yazd
- b. Sepide [be yazd] rafte bud  
 Sepideh à Yazd aller.PP être.PST.3SG  
 ‘Sepideh était allée à Yazd.’
- (70) a. Sepide bačče-hā=ro bord [pārk]  
 Sepideh enfant-PL=RĀ emmener.PST.3SG parc
- b. Sepide bačče-hā=rā [be pārk] bord  
 Sepideh enfant-PL=RĀ à parc emmener.PST.3SG  
 ‘Sepideh a emmené les enfants au parc.’

39. Lazard *et al.* (2006, p. 196) notent, en effet, que “c’est l’ordre inverse qui marque une nuance particulière” :

- a. raft-an bāq  
 aller.PST-3PL jardin  
 ‘Ils sont allés au jardin.’
- b. bāq raft-an  
 jardin aller.PST-3PL  
 ‘C’est au jardin qu’ils sont allés.’

- (71) a. resid-im [bālā=ye čěšme]  
arriver.PST-1PL haut=EZ fontaine
- b. [(be) bālā=ye čěšme] resid-im  
à haut=EZ fontaine arriver.PST-1PL  
'Nous sommes arrivés en haut de la fontaine.'
- (72) a. kelid=o gozāš-am [zir=e farš]  
clé=RĀ mettre.PST-1SG [sous=EZ tapis]
- b. kelid=rā [zir=e farš] gozāš-am  
clé=RĀ [sous=EZ tapis] mettre.PST-1SG  
'J'ai mis la clé sous le tapis.'

Les grammaires traditionnelles, qui n'abordent pas généralement les ordres non verbes finaux, signalent la possibilité d'omettre la préposition *be* et, de surcroît, le positionnement postverbal des compléments directionnels à l'oral (par ex. Givi Ahmadi & Anvari, 1995; Mace, 2003).

- (73) asr mi-rav-im [gardeš] -> asr [be gardeš] mi-rav-im  
soir IPFV-aller.PRS-1PL promenade  
'Nous allons nous promener ce soir.'

Givi Ahmadi & Anvari (1995, p. 212)

De même, dans le cadre génératif, Karimi (2003a, p. 124), en observant que les "éléments directionnels" (*directional elements*) peuvent apparaître dans le domaine postverbal sans modifier l'intonation neutre, avance, dans une note de bas de page, que la position canonique de ce type d'éléments est après le verbe. Elle donne l'exemple suivant pour illustrer son propos :

- (74) Kimea raft madrese  
Kimea aller.PST.3SG école  
'Kimea alla à l'école.'

Karimi (2003a, p. 124)

Notons que les compléments datifs, ex. (75), se placent également régulièrement dans le domaine postverbal où la préposition *be* est fréquemment (mais pas systématiquement) omise.

- (75) a. kelid=o dād-am [(be) Maryam]  
clé=RĀ donner.PST-1SG [à Maryam]
- b. kelid=rā [be Maryam] dād-am  
clé=RĀ [à Maryam] donner.PST-1SG  
'J'ai donné la clé à Maryam.'

Certains travaux ne distinguent pas entre les deux cas de figure. Les exemples en (76) sont de Mahootian (1997, p. 49) afin d'illustrer les compléments d'objets indirects introduits par la préposition *be* et le fait que cette dernière est presque toujours omise à l'oral.

- (76) a. bilit-a=ro dād be man  
 billet-PL=RĀ donner.PST.3SG à moi  
 'Il m'a donné les billets'
- b. fardā mi-r-an (be) Hamedan  
 demain IPFV-aller.PRS.INF-3PL (à) Hamedan  
 'Ils iront à Hamedan demain.'

Mahootian (1997, p. 49)

Rappelons que la langue écrite est réticente aux ordonnancements non verbes finaux. Les deux extraits ci-dessous, tirés d'un article de nécrologie<sup>40</sup>, illustrent bien l'écart entre les deux registres. Dans le premier extrait, ex. (77), l'auteur fait une citation rapportée dans son registre familier propre à l'oral, tandis que le deuxième extrait, ex. (78), relève de l'écrit dans sa forme standard.

- (77) yek ruz ye āqā=i bā kolāh šāpo **umad**  
 un jour un monsieur=INDF avec chapeau "chapeau" venir.PST.3SG  
**xune=ye mā.** pedar=am goft (...) man dār-am  
 maison=EZ nous. père=1SG dire.PST.3SG (...) moi avoir.PRS-1SG  
**mi-r-am matab.** doroške xabar kon  
 IPFV-aller.PRS-1SG cabinet (de médecine) calèche appel faire.IMP.2SG  
**išun=rā be-bar qahve-xune=ye domil (...)**  
 lui=RĀ IMP-emmener.PRS.2SG café=EZ Domil (...)  
 'Un jour un monsieur avec un chapeau est venu chez nous. Mon père a dit  
 (...) je vais au cabinet. Appelle une calèche et emmène le monsieur au café  
 Domil ...'

- (78) čand sāl=i ke gardun=rā montašer mi-kard-am  
 quelques années=RSTR que Gardun=RĀ publier IPFV-faire.PST-1SG  
 aqlab bā u sobhāne mi-xord-am, va ba'd **be daftar**  
 souvent avec elle petit-déjeuner IPFV-manger.PST-1SG et après à bureau  
**mi-raft-am.**  
 IPFV-aller.PST-1SG  
 'Les années où je publiais (le magazine) Gardun, je prenais souvent le petit-déjeuner avec elle avant d'aller au bureau.'

40. Paru sur le site de *BBC Persian* le 09 mars 2012, consulté à l'adresse suivante : [http://www.bbc.co.uk/persian/arts/2012/03/120309\\_141\\_book\\_daneshvar\\_memorial\\_marooofi.shtml](http://www.bbc.co.uk/persian/arts/2012/03/120309_141_book_daneshvar_memorial_marooofi.shtml)

## 1.2.2 Phrase non déclarative

L'ordre des constituants dans les phrases non déclaratives, c'est-à-dire interrogative, exclamative et impérative, est identique à l'ordre des constituants dans une phrase déclarative. Tous les agencement linéaires attestés dans les phrases déclaratives se rencontrent également dans les phrases non déclaratives. L'interrogation est indiquée par la prosodie et éventuellement par les mots interrogatifs, mais jamais par une modification de l'ordre (Lazard et al., 2006; Mahootian, 1997; Windfuhr & Perry, 2009). C'est également le cas des phrases exclamatives. L'impératif est indiqué par l'emploi du mode verbal impératif (ou prohibitif), sans qu'un changement dans l'ordre intervienne.

- (79) a. Sepide=rā baxšid-i. (prosodie assertive)  
 Sepideh=RĀ pardonner.PST-2SG  
 'Tu as pardonné à Sepideh.'
- b. Sepide=rā baxšid-i? (prosodie interrogative)  
 'As-tu pardonné à Sepideh?'
- c. Sepide=rā baxšid-i! (prosodie exclamative)  
 'Tu as pardonné à Sepideh!'
- d. Sepide=rā be-baxš / na-baxš!  
 Sepideh=RĀ IMP-pardonner.2SG / NEG-pardonner.2SG  
 'Pardonne à Sepideh! / Ne pardonne pas à Sepideh!'

### 1.2.2.1 L'interrogatif total

L'interrogatif total peut être indiqué par des mots (particules) interrogatifs *āyā* 'est-ce que' ou *magar*<sup>41</sup> 'n'est-ce pas', qui se mettent généralement, mais pas obligatoirement, en tête de la phrase. Rappelons qu'une interrogation totale peut être marquée par la prosodie seulement. Les exemples suivants sont tirés du corpus de Bijankhan :

- (80) šomā mi-gu-id moxātab na-dār-ad ?  
 vous IPFV-dire.PRS-2PL public NEG-avoir.PRS-3SG ?  
 'Vous dites qu'il n'y a pas de public?'
- (81) āyā hamegi=ye šomā ham-nazar hast-id ?  
 est-ce que tout=EZ vous même-avis être.PRS-2PL ?  
 'Est-ce que vous êtes tous de même avis?'

---

41. Prononcé *mage* dans le registre informel.

- (82) dar yek orupā=ye 40-ozvi āyā in kār sāde-tar  
 dans un Europe=EZ de 40 membre est-ce que ce tâche facile-COMPR  
 xāh-ad bud ?  
 AUX.FUT-3SG être.PST ?  
 ‘Cette tâche sera-t-elle plus simple dans une Europe composée de 40 membres ?’
- (83) šik-tar az in eǰdehā-hā magar peydā mi-šav-ad ?  
 chic-COMPR de ce dragon-PL n'est-ce pas trouver IPFV-devenir.PRS-3SG ?  
 ‘Est-ce possible de trouver des dragons plus chics que ceux-là ?’

### 1.2.2.2 L'interrogatif partiel

Les interrogatifs partiels sont formés par les pronoms ou déterminants interrogatifs, ex. *če* ‘que ou quel’, *či* ‘quoi’, *čeqadr* ‘combien’, *četon* ‘comment’, *ki* ‘qui’, *kojā* ‘où’, etc. Les mots *Qu-* ou *Wh-* du persan sont *in situ*, ex. (84) et (85)<sup>42</sup> mais, ils peuvent également apparaître en position initiale, en particulier lorsqu'ils sont “en focus” (Mahootian 1997, p. 18, ex. (86), et Windfuhr & Perry 2009, p. 482).

- (84) in rāh **be kojā** xatm xāh-ad šod ?  
 ce route à où fin AUX.FUT-3SG devenir.PST ?  
 ‘Où est-ce que cette route se termine ?’
- (85) ... hattāki=ye ānān **če** asar=i dār-ad ?  
 ... insulte=EZ eux quel effet=INDF avoir.PRS-3SG ?  
 ‘... quels effets vont avoir leurs insultes ?’
- (86) **či**=ro rezā be man dad ?  
 quoi=RĀ Reza à moi donner.PST.3SG ?  
 ‘C'était quoi que Reza m'a donné ?’

Notons que les travaux générativistes affirment que le persan ne manifeste pas le mouvement *Qu-* (*wh-movement*) (Kahnemuyipour, 2001; Karimi, 2005)<sup>43</sup>.

### 1.2.2.3 L'exclamatif

L'exclamatif est indiqué par un mot *Qu-* exclamatif, ex. (87) et (88), ou par des particules spécialisées telles que *kāš*, ou encore par l'intonation seulement, ex. (90) (Lazard *et al.*, 2006, p. 199).

42. Exemples tirés du corpus de Bijankhan.

43. Plus précisément, Kahnemuyipour (2003, p. 197) analyse l'interrogative partielle du persan comme un cas de *focus fronting*. En effet, il admet trois types d'interrogative partielle (*wh-question*) : 1. *wh-movement* ; 2. *wh-in situ* ; et 3. *focus fronting*. En écartant le mouvement *Qu-* obligatoire pour le persan, il considère les interrogatives partielles de cette langue comme *focus fronting*.

- (87) **čeqadr** xaste=am !  
 combien fatigué=COP.1SG  
 ‘Que je suis fatigué!’
- (88) **če** bāq=e qašang=i !  
 quel jardin=EZ beau=INDF  
 ‘Quel beau jardin!’
- (89) **kāš** zud xub š-e !  
 PART vite bien devenir.SBJV-3SG  
 ‘Qu’il guérisse vite!’
- (90) tamum šod !  
 fini devenir.PST.3SG  
 ‘C’est fini!’

Notons que ces mots *Qu-* dans les phrases exclamatives ne se mettent pas obligatoirement en tête de phrase.

- (91) (**čeqadr**) in bāq (**čeqadr**) qašang=e (**čeqadr**)!  
 (combien) ce jardin (combien) beau=COP.3SG (combien)  
 ‘Que ce jardin est beau!’

#### 1.2.2.4 L’impératif

L’impératif (ou prohibitif) est marqué par l’emploi du mode impératif, réalisé par la flexion verbale sans ambiguïté. Rappelons que l’ordre des constituants dans une phrase impérative ne diffère pas de celui d’une phrase déclarative.

- (92) a. in ketāb=rā mi-xān-am  
 ce livre=RĀ IPFV-lire.PRS-1SG  
 ‘Je lis ce livre!’
- b. in ketāb=rā be-xān  
 ce livre=RĀ IMP-lire.2SG  
 ‘Lis ce livre!’

### 1.2.3 Proposition subordonnée

Le persan dispose d’un complémenteur unique et invariable, *ke*, qui introduit indifféremment les complétives et les relatives. Celui-ci précède toujours la proposition subordonnée et peut être omis uniquement dans les subordonnées complétives.

- (93) mi-dān-am (ke) ejāze=ye parvāz ne-mi-dah-and  
 IPFV-savoir.PRS-1SG (que) permis=EZ vol NEG-IPFV-donner.PRS-3PL  
 ‘Je sais qu’ils ne donnent pas l’autorisation de vol.’

Dans les subordonnées relatives, le nom tête précède la subordonnée et, dans le cas d'une relative restrictive, celui-ci sera marqué par un enclitique, =*i*, distinct de celui de l'indéfini (Samvelian, 2006b), ex. (94). La reprise d'élément relativisé par un pronom résomptif – clitique ou tonique – dans la subordonnée est obligatoire quand cet élément occupe la fonction du complément prépositionnel au sein de la relative, ex. (95). Lorsque l'élément relativisé occupe la fonction du COD, la reprise par un pronom résomptif clitique est possible, ex. (96), bien qu'elle soit proscrite par les grammaires normatives. Il faut noter que la subordonnée relative ne se positionne pas obligatoirement immédiatement après le nom tête et peut être extraposée dans le domaine postverbal de la phrase matrice, ex. (97).

- (94) [mard=i ke sāheb-e maqāze bud] ketāb=rā be  
 homme=RSTR que propriétaire=EZ magasin être.PST.3SG livre=RĀ à  
 man dād  
 moi donner.PST.3SG  
 'L'homme qui était le propriétaire du magasin m'a donné un livre.'
- (95) [mard=i ke **az=aš** ketāb=rā gereft-am] sāheb=e  
 homme=RSTR que de=3SG livre=RĀ prendre.PST-1SG propriétaire=EZ  
 maqāze bud  
 magasin être.PST.3SG  
 'L'homme dont j'ai pris le livre était le propriétaire du magasin.'
- (96) [ketāb=i ke ān mard be man dād]=rā  
 livre=RSTR que ce homme à moi donner.PST.3SG=RĀ  
 xarid-am=aš  
 acheter.PST-1SG=3SG  
 'Je n'ai pas acheté le livre que cet homme m'a donné.'
- (97) ketāb=i=rā xarid-am ke ān mard be man dād  
 livre=RSTR=RĀ acheter.PST-1SG que ce homme à moi donner.PST.3SG  
 'Je n'ai pas acheté le livre que cet homme m'a donné.'

L'ordre des constituants dans une proposition subordonnée ne diffère pas de celui d'une proposition matrice. Autrement dit, la construction enchâssée n'implique pas de changement d'ordre. De plus, les variations de l'ordre attestées dans la phrase matrice sont également possibles dans les subordonnées. Mahootian (1997, p. 29) mentionne une exception à cette généralisation en se référant aux travaux de Karimi : l'antéposition du verbe (\*Comp V S PP). En effet, Karimi (1994, p. 56) exclut cet ordre, mais seulement dans le cas des syntagmes nominaux complexes, comme *le fait que...* . Elle l'admet cependant pour les compléments verbaux phrastiques<sup>44</sup>.

44. Karimi (1994, p. 56) considère toutefois que les syntagmes nominaux complexes sont des subordonnées typiques et que les compléments verbaux phrastiques, étant comparables au discours

### 1.3 La variation de l'ordre dans la phrase déclarative

Tout en étant incontestablement SOV, le persan est une langue à ordre flexible, ou une langue *scrambling* d'après la terminologie générative (Karimi, 1994; Dabir-Moghaddam, 1995; Mahootian, 1997; Windfuhr, 1979; Dabir-Moghaddam, 2001; Lazard *et al.*, 2006).

L'ordre canonique SOV peut être bouleversé, produisant ainsi tous les cas de figure possibles<sup>45</sup>, c'est-à-dire OSV, SVO, OVS, VSO et VOS :

- (98) a. Puyān Sepide=rā did  
Puyan Sepideh=RĀ voir.PST.3SG  
'Puyan a vu Sepideh.'
- b. Sepide=rā Puyān did (OSV)
- c. Puyān did Sepide=rā (SVO)
- d. Sepide=rā did Puyān (OVS)
- e. did Puyān Sepide=rā (VSO)
- f. did Sepide=rā Puyān (VOS)

De même, dans les phrases ditransitives ayant l'ordre canonique S-OD-OI-V, les possibilités d'ordonnement sont multipliées par quatre :

- (99) a. Puyān ketāb=rā be Maryam dād  
Puyan livre=RĀ à Maryam donner.PST.3SG  
'Puyan a donné le livre à Maryam.'
- b. be Maryam Puyān ketāb=rā dād (OI-S-OD-V)
- c. Puyān be Maryam ketāb=rā dād (S-OI-OD-V)
- d. be Maryam ketāb=rā Puyān dād (OI-OD-S-V)
- e. ketāb=rā be Maryam Puyān dād (OD-OI-S-V)
- f. ketāb=rā Puyān be Maryam dād (OD-S-OI-V)
- g. ketāb=rā Puyān dād be Maryam (OD-S-V-OI)
- h. Puyān ketāb=rā dād be Maryam (S-OD-V-OI)
- etc.

Outre les réagencements dans la proposition simple, le persan manifeste, particulièrement dans le registre familier, la possibilité pour les constituants d'une

---

rapporté, se comportent comme une proposition principale de manière universelle.

45. Notons que les dérogations à l'ordre canonique se produisent avec une fréquence inégale (Lazard *et al.*, 2006, p. 194). Notamment, comme nous le verrons, certains ordres sont propres à l'oral et présentent une prosodie fortement marquée.

proposition complétive de se mêler avec les constituants de la phrase matrice, à condition que le verbe régi se trouve dans le domaine postverbal du verbe recteur<sup>46</sup>. Les exemples (100) à (101) ci-dessous illustrent cette possibilité, qui est désignée “*long distance scrambling*” (*scrambling* à distance) dans la littérature générative (Browning & Karimi, 1994; Karimi, 2005). Mentionnons le fait qu’un élément extrait de la proposition matrice peut être marqué de =*rā* (cf. §1.1.3), ex. (101-b) et (101-c), quelle que soit sa fonction dans la phrase enchâssée (Dabir-Moghaddam, 1990, 1992).

- (100) a. Maryam mi-xāst (ke) Ali=rā/ro be-bin-ad/e  
Maryam IPFV-vouloir.PST.3SG (que) Ali=RĀ SBJV-voir.PRS-3SG  
‘Maryam voulait voir Ali.’  
b. Ali=ro Maryam mi-xāst (ke) be-bin-e  
c. Maryam Ali=ro mi-xāst (ke) be-bin-e  
d. mi-xāst Maryam (ke) Ali=ro be-bin-e  
etc.
- (101) a. Ali fekr mi-kard (ke) Maryam xābide  
Ali pensée IPFV-faire.PST.3SG (que) Maryam dormir.PP  
‘Ali pensait que Maryam dormait.’  
b. Maryam=o Ali fekr mi-kard (ke) xābide  
c. Ali Maryam=o fekr mi-kard (ke) xābide  
etc.

De plus, il existe en persan une forme verbale périphrastique, le progressif, formée avec le semi-auxiliaire *dāštan* ‘avoir’, dans laquelle, contrairement aux autres formes périphrastiques<sup>47</sup>, le verbe principal est conjugué tout comme l’auxiliaire (Lazard *et al.*, 2006; Bonami & Samvelian, 2009). Ce dernier peut se placer assez librement dans la phrase dès lors qu’il se trouve avant le verbe principal, ex. (102).

- (102) a. Ali dāšt bā Maryam harf mi-zad  
Ali AUX.PST.3SG avec Maryam parole IPFV-frapper.PST.3SG  
‘Ali était en train de parler avec Maryam.’  
b. Ali bā Maryam dāšt harf mi-zad

46. Rappelons qu’en persan les constructions dites “à montée” (*raising*) ou “à contrôle” (*control*) sont normalement finies.

47. Notamment le futur formé par l’auxiliaire *xāstan* ‘vouloir’, ex. *xāh-am raft* (AUX.FUT-1SG aller.PST) ‘je partirai’, le parfait aux temps passé et présent formés par l’auxiliaire *budan* ‘être’ – avec réalisation lexicale ou clitique respectivement – ex. *rafte=am* (aller.PP=COP.1SG) ‘je suis (déjà) allé’ et *rafte bud-am* (aller.PP être.PST-1SG) ‘j’étais allé’, et le passif formé avec l’auxiliaire *šodan* ‘devenir’, ex. *xorde šod* (manger.PP devenir.PST.3SG) ‘il a été mangé’ (cf. Bonami & Samvelian, 2009, pour une présentation détaillée des formes périphrastiques du persan).

- c. bā Maryam Ali dāšt harf mi-zad
  - d. bā Maryam dāšt Ali harf mi-zad
  - e. bā Maryam harf dāšt Ali mi-zad
  - f. \*Ali bā Maryam harf mi-zad dāšt
- etc.

Bien que les dérogations à l'ordre SOV soient particulièrement fréquentes à l'oral (plus précisément dans le discours spontané), elles se produisent également à l'écrit. En effet, même les grammaires traditionnelles et scolaires admettent la possibilité de variation de l'ordre, non seulement à l'oral, mais aussi à l'écrit pour des considérations discursives (ex. : la mise en relief, l'emphase, la spécification ou le contraste)<sup>48</sup> ou stylistiques (Gharib *et al.*, 1994; Givi Ahmadi & Anvari, 1995).

Ci-dessous des exemples attestés, extraits du corpus Bijnakhan :

- (103) [COD ān ruz-hā=rā] [Suj man] be xāter dār-am  
 ce jour-PL=RĀ moi à mémoire avoir.PRS-1SG  
 'Je me rappelle de ces jours-là.'
- (104) kānton=e stānvud dar avāel sāl=e jāri [COI be zanān]  
 canton=EZ Stanwood en début année=EZ en cours à femmes  
 [COD=rā haqq=e entexāb=rā] dād  
 droit=EZ choisir=RĀ donner.PST.3SG  
 'Au début de l'année en cours, le canton de Stanwood a donné aux femmes le droit de choisir.'
- (105) vaqti be mā resid [COD bolandgu=rā] [V časbānd]  
 quand à nous arriver.PST.3SG mégaphone=RĀ coller.PST.3SG  
 [COI be guš=e barādar=e Asqar Molavi]  
 à oreille=EZ frère=EZ Asqar Molavi  
 'Quand il est arrivé près de nous, il a collé le mégaphone sur l'oreille du frère de Asqar Molavi.'

Les grammaires descriptives ainsi que les travaux relevant de cadres théoriques plus récents rendent compte des ordres divergents principalement par des considérations discursives générales accompagnées de changement prosodique, plus pré-

48. Un exemple donné pour illustrer cette observation est la différence entre les deux phrases suivantes (Gharib *et al.*, 1994, p. 283) :

- a. Sa'di šāer ast  
 Sa'di poète être.PRS.3SG  
 'Sa'di est un poète.'
- b. šāer Sa'di ast  
 poète Sa'di être.PRS.3SG  
 'C'est Sa'di qui est un poète (et pas les autres).'

cisément le déplacement de l'accent phrastique et/ou la pause (Mahootian, 1997; Karimi, 1994; Windfuhr & Perry, 2009; Lazard *et al.*, 2006, entre autres).

Dans la suite de cette section, nous allons tout d'abord présenter brièvement la configuration discursive de la phrase canonique ainsi que le placement de l'accent phrastique (appelé plus couramment l'accent nucléaire). Nous passerons ensuite à la présentation des ordres dérogatoires et des analyses syntaxiques existantes.

### 1.3.1 L'ordre linéaire et la structure informationnelle

Il est généralement admis que l'ordre linéaire (non marqué) SOV en persan correspond, sur le plan discursif, à l'articulation thème - rhème (*topic - comment*)<sup>49</sup>. À titre d'exemple, Dabir-Moghaddam (1990, 1992, 2005) suggère l'organisation discursive schématisée en (106) pour la phrase canonique en persan.

- (106) a. Primary Topic Comment  
 Subjet Objet direct Verb
- b. Primary Topic Secondary Topic Comment  
 Subjet Objet direct Verb

Comme ce schéma le montre, Dabir-Moghaddam admet différents niveaux de topicalisation, notamment le topic principal et le topic secondaire, et, outre l'articulation binaire, ex. (106-a), suppose une articulation ternaire, ex. (106-b), comparable à celle proposée ultérieurement par Nikolaeva (2001), c'est-à-dire *topic-secondary topic-focus*. Les deux exemples (107-b) et (107-a) illustrent cette organisation en réponse, respectivement, aux questions *qu'est-ce que tu as fait, hier?*, et *qu'est-ce que tu as fait concernant le livre?*

- (107) a. man ketāb=i xarid-am  
 moi livre=INDF acheter.PST-1SG  
 'J'ai acheté un livre (quelconque).'
- b. man ketāb=rā xarid-am  
 moi livre=RĀ acheter.PST-1SG  
 'J'ai acheté le livre.'

Notons que Dabir-Moghaddam, en se joignant à Halliday (1985), opère une distinction entre, d'un côté, le statut informationnel donné/nouveau (orientée vers l'interlocuteur) et, de l'autre, l'articulation topic/focus (orientée vers le locuteur), et ainsi adopte une définition du topic qui est indépendante de la nouveauté ou non de son référent dans le discours<sup>50</sup>.

49. Pour une définition de la structure informationnelle, voir §3.2.2.2 au chapitre 3.

50. Voir §2.3.1.2 au chapitre 2 pour plus de détail sur la séparation de ces deux dimensions.

Dans le cadre de la grammaire générative, il a été suggéré que le persan est une langue *topic prominent* et non *subject prominent* (Karimi, 2005, p. 111 et pp. 116-117), signifiant que l'argument externe du  $v''$  n'est pas (forcément) le sujet de la phrase. Karimi (2005) postule qu'il est possible, notamment dans une phrase *all focus*, que tous les dépendants du verbe (y compris le sujet) restent *in situ*, c'est-à-dire sous le  $v''$ , et que n'importe quel dépendant du verbe peut sortir du  $v''$  afin de recevoir une fonction discursive telle que le topic (Top'' ou T'') ou le focus contrastif (Foc'') (Karimi, 2005, pp. 111-113).

L'auteur postule la configuration schématisée en (108) pour la phrase en persan. Nous reviendrons sur cette configuration en §1.4.1.

$$(108) \quad [CP [TopP [FocP [TP [T' [ [vP [v' \underbrace{[XP [X']]_{PredP}} v ]]]]]]]]]]$$

Adapté de Karimi (2005, p. 123)

En s'inspirant de la définition du focus élaboré par Kiss (1998)<sup>51</sup>, Karimi (2005) définit le focus contrastif, en opposition au focus inhérent ou primaire, ex. (109), de la façon suivante :

Contrastive focus operates on a closed set of entities whose members are known to the participants in the discourse.

Karimi (2005, p. 132)

(109) hālā            barā Kīmea pro nāme mi-nevis-an  
 maintenant pour Kīmea      lettre IPFV-écrire.PRS-3PL  
 'Maintenant, c'est à Kīmea qu'ils écrivent une lettre (en opposition à quel-  
 qu'un d'autre).'

De plus, Karimi (2005) postule deux types de topic en persan : "*shifted topic*" (TopP) et "*background topic*" (TP). Dans l'exemple (110) ci-dessous, *Kīmea* correspond au *shifted topic*, car "[il] indique un changement de l'attention dans le discours". En revanche, *unā=ro*, référant à un élément déjà introduit dans le discours (*ketāb-hā* 'les livres'), correspond au *background topic*. Ces deux types de *topic* nous rappellent les deux niveaux de topicalisation (primaire et secondaire) définis par Dabir-Moghaddam (1992).

51. Kiss (1998) distingue deux types de focus : "*informational focus*", une information non pré-supposée dans le discours que l'on peut appeler focus inhérent ou focus primaire, et "*identificational focus*", défini comme suit : « *An identificational focus represents a subset of the set of contextually or situationally given elements for which the predicate phrase can potentially hold; it is identified as the exhaustive subset of this set for which the predicate phrase actually holds.* » (1998, p. 245).

- (110) a. ketāb-ā čī šod  
 livre-PL quoi devenir.PST.3SG  
 'Qu'est-ce qui est arrivé aux livres?'
- b. Kimea unā=ro diruz be man dād  
 Kimea ce.PL=RĀ hier à moi donner.PST.3SG  
 'Kimea me les a donnés hier.'

Karimi (2005, p. 129)

Notons que Karimi (2005) souligne par ailleurs que le focus inhérent ou primaire (PredP<sup>52</sup> dans le schéma (108)), présentant toujours une information nouvelle (contrairement au focus contrastif), est confiné à la position préverbale<sup>53</sup>.

### 1.3.2 Le placement de l'accent nucléaire

Kahnemuyipour (2004), qui a étudié le placement de l'accent phrastique (appelé également accent nucléaire) dans un cadre génératif, en postulant que l'accent phrastique tombe sur le focus, suppose qu'en persan l'accent phrastique tombe sur l'élément qui se trouve à la frontière gauche de celui-ci, en l'occurrence sur l'objet non marqué, ex. (111-a), ou le verbe, ex. (111-b), dans une phrase SOV.

- (111) a. Maryam (yek) ketāb xarid  
 Maryam un livre acheter.PST.3SG  
 'Maryam a acheté un livre.'
- b. Maryam ketāb=rā xarid  
 Maryam livre=RĀ acheter.PST.3SG  
 'Maryam a acheté le livre.'

Plus généralement, il est admis que l'accent phrastique en persan tombe sur l'information nouvelle, l'élément mis en relief ou en focus, et que celui-ci est progressif en persan, contrairement à l'accent lexical qui lui est régressif (Windfuhr, 1990; Dabir-Moghaddam, 1990; Mahootian, 1997; Kahnemuyipour, 2003; Lazard *et al.*, 2006). Ceci permet de lever l'ambiguïté dans une séquence telle que *bāzkon* entre un nom-composé 'ouvre-boîte', *bāz-kon*, et une phrase 'ouvre! (Lit. ouvert fais!)', *bāz kon* (Windfuhr, 1990).

La modification de l'accent saillant de la phrase est un moyen pour provoquer le changement de la structure informationnelle, sans entraîner de modifications de l'ordre des constituants, comme l'illustrent les exemples ci-dessous.

52. Cette projection, représentant le prédicat exprimé par le verbe, est la projection maximale qui contient les arguments internes du verbe.

53. Rappelons que Karimi (2005) exclut les compléments directionnels de cette généralisation et admet la position postverbale comme la position sous-jacente pour ces éléments (cf. §1.2.1).

- (112) a. Ali Maryam=o busid  
 Ali Maryam=RĀ embrasser.PST.3SG  
 ‘Ali a embrassé Maryam.’  
 b. Ali Maryam=o busid  
 Ali Maryam=RĀ embrasser.PST.3SG  
 ‘C’est Maryam que Ali a embrassée.’  
 Kahnemuyipour (2004, p. 170)
- (113) a. Ali sib xord  
 Ali pomme manger.PST.3SG  
 ‘Ali a mangé une/des pommes.’  
 b. Ali sib xord  
 Ali pomme manger.PST.3SG  
 ‘Des pommes, Ali en a mangé.’  
 Kahnemuyipour (2004, p. 206)
- (114) a. **man** hast-am  
 moi être.PRS-1SG  
 ‘C’est moi. (en réponse à la question : Qui est-ce?)’  
 b. man **hast-am**  
 moi être.PRS-1SG  
 ‘Je suis là. (en réaction à la phrase : Je ne veux pas laisser l’enfant seul.)’  
 Mahootian (1997, p. 322)
- (115) a. **Reza** bud  
 Reza être.PST.3SG  
 ‘C’était Reza. (en réponse à la question : Qui a sonné?)’  
 b. Reza **āmad**  
 Reza venir.PST.3SG  
 ‘Reza est venu. (sachant qu’on parlait de Reza)’  
 Dabir-Moghaddam (1990)

En outre, les modifications prosodiques accompagnent les ordres non canoniques pour réaliser différentes configurations discursives, ex. (116).

- (116) a. Ali xeyli sib xord  
 Ali très pomme manger.PST.3SG  
 ‘Ali a beaucoup mangé de pommes.’  
 b. Ali sib xeyli xord  
 Ali pomme très manger.PST.3SG  
 ‘Des pommes, Ali en a beaucoup mangé.’  
 Kahnemuyipour (2004, p. 206)

Par ailleurs, Karimi (2005, p. 16) considère qu'un élément déplacé (*scrambled element*) peut être interprété comme topic ou focus contrastif en fonction de l'accent qu'il porte. Un accent fort (*heavy stress*) correspond selon elle au focus contrastif, ex. (117).

- (117) a. **pirhan=o** Parviz barā Kimea xarid  
 robe=RĀ Parviz pour Kimea acheter.PST.3SG  
 'Quant à la robe, Parviz l'a achetée pour Kimea.'  
 ou 'C'était la robe que Parviz a achetée pour Kimea.'
- b. **be Hasan** hame mi-xand-an  
 à Hassan tout-le-monde IPFV-rire.PRS-3PL  
 'Quant à Hassan, tout le monde rit de lui.'  
 ou 'C'est de Hassan que tout le monde rit.'

Karimi (2005, p. 16)

### 1.3.3 Les ordres non canoniques

Les variations de l'ordre sont décrites dans la littérature principalement en référence à la configuration discursive. Les fonctions discursives générales, telles que la topicalisation, l'emphase, le focus et le contraste sont évoquées pour rendre compte des ordres non canoniques. À titre d'exemple, Windfuhr (1990) avance que les dérogations de l'ordre SOV, l'inversion, selon ses propres termes, sont dues à la topicalisation, et que les deux positions initiale, ex. (118-a), et (immédiatement) préverbale, ex. (118-b), accueillent généralement le topic "*are topical*" :

- (118) a. be u **javāb** dād-am  
 à lui réponse donner.PST-1SG
- b. **javāb** be u dād-am  
 réponse à lui donner.PST-1SG  
 'Je lui ai donné une réponse.'

Windfuhr (1990)

Il semble cependant que l'auteur entend par topicalisation plus largement la mise en relief (l'emphase) d'un constituant. En effet, dans un autre ouvrage (Windfuhr & Perry, 2009)<sup>54</sup>, dans lequel les dérogations (à l'ordre SOV) les plus fréquentes sont présentées de manière plus détaillée, il emploie le terme "emphase" (*emphasis*). Les deux dérogations mentionnées dans cette référence sont les suivantes : 1) OSV (*object raising*) qui permet de mettre le sujet en emphase en le positionnant immédiatement

54. Intitulé *Persian and Tajik*, celui-ci est rédigé conjointement par G. L. Windfuhr et J. Perry. Ce dernier, comme précisé dans l'introduction (Windfuhr, 2009, p. 2), s'est chargé de la partie sur le tadjik.

avant le verbe, ex. (119-a), et 2) OVS (*concomitant subject lowering*) pour mettre l'emphase sur l'objet direct, ex. (119-b).

- (119) a. sag=rā **man** košt-am  
 chien=RĀ moi tuer.PST-1SG  
 'C'est moi qui ai tué le chien.'
- b. **sib=rā** ne-mi-xar-id šomā ?  
 pomme=RĀ NEG-IPFV-acheter.PRS-2PL vous  
 'Vous n'allez pas acheter de pommes? (*Aren't you going to buy any apples*)'<sup>55</sup>

Windfuhr & Perry (2009, p. 480)

Comme en témoignent ces deux derniers exemples, l'élément mis en emphase peut être le topic, comme dans (119-b), aussi bien que le focus, comme dans (119-a). Rappelons qu'en persan la position immédiatement avant le verbe est en général focale, tandis que la position initiale est réservée au topic (cf. §1.3.1).

Par ailleurs, il est généralement admis que la topicalisation d'un constituant, quelle que soit sa fonction s'effectue communément, par sa mise en tête de la phrase, ex. (120) (Dabir-Moghaddam, 1990; Karimi, 1994; Mahootian, 1997; Lazard *et al.*, 2006). Cela est également le cas dans les phrases à complément phrastique, où un élément peut être extrait de la phrase enchâssée pour être mis en tête de la phrase matrice, selon Dabir-Moghaddam (1990, 1992), avec pour condition suffisante d'être marqué de =rā, ex. (121).

- (120) a. be Mahin bilit=o ( )<sup>56</sup> dād-am  
 à Mahin billet=RĀ ( ) donner.PST-1SG  
 'À Mahin j'ai donné le billet.'
- b. bā Aqdas man ( ) raft-am tā'ātr  
 avec Aqdas moi ( ) aller.PST-1SG théâtre  
 'Avec Aqdas, je suis allé au théâtre.'
- c. bā otobus Parviz ( ) raft Širāz  
 avec bus Parviz ( ) aller.PST.3SG Shiraz  
 'En bus, Parviz est allé à Shiraz.'
- d. māhi behtar=e ( ) na-xar-i  
 poisson mieux=COP.3SG ( ) NEG-acheter.PRS-2SG  
 'Quant au poisson, il vaut mieux que tu n'en achètes pas.'

Mahootian (1997, p. 123)

55. Il est précisé que l'objet direct, étant donné son marquage par =rā, est employé de manière contrastive, *grosso modo* '... quant aux pommes (en plus du yaourt, etc.), ...'

56. Les parenthèses vides représentent, dans la notation de Mahootian (1997), la position canonique des constituants déplacés.

- (121) a. man bā'es šod-am [ke Ali xāne=rā be Mina  
moi cause devenir.PST-1SG que Ali maison=RĀ à Mina  
na-foruš-ad]  
NEG-vendre.PRS-3SG  
'Par mon fait, Ali n'a pas vendu la maison à Mina.'
- b. Ali=ro man bā'es šod-am [ke  $\emptyset_i$  xāne=rā be Mina na-foruš-ad]
- c. xāne=rā<sub>i</sub> man bā'es šod-am [ke Ali  $\emptyset_i$  be Mina na-foruš-ad]
- d. ?Mina=ro<sub>i</sub> man bā'es šod-am [ke Ali xāne=rā beh=eš<sub>i</sub> na-foruš-ad]  
Dabir-Moghaddam (1990, p. 129)

D'autres ordonnancements dérogatoires liés à l'articulation discursive de la phrase sont également évoqués dans la littérature (Karimi, 1994; Mahootian, 1997). Nous pouvons mentionner l'antéposition du verbe "*verb preposing*", ex. (122), qui, comme les traductions proposées en témoignent, permettrait de mettre le verbe en relief, ou, selon Mahootian (1997), en emphase contrastive. Mentionnons également la postposition d'un constituant (au verbe), ex. (123), qui exprimerait une emphase non contrastive (Mahootian, 1997, p. 117).

- (122) a. **gereft-am** man ketāb=o  
prendre.PST-1SG moi livre=RĀ  
'J'ai bien pris le livre. (*I did get the book.*)'
- b. **xarid** Sepide ketāb=o barā=t  
acheter.PST.3SG Sepideh livre=RĀ pour=2SG  
'Est-ce que Sepideh a acheté le livre pour toi? (*Did Sepideh buy the book for you?*)'
- Karimi (1994, p. 55)
- c. xord-am šām=am=o man  
manger.PST-1SG dîner=1SG=RĀ moi  
'J'ai déjà dîné. (*I already ate dinner.*)'
- Mahootian (1997, p. 127)
- (123) a. mardom t<sub>i</sub> komak kard-and **be dānešju-hā**  
gens aide faire.PST-3PL à étudiant-PL  
'Les gens ont bien aidé les étudiants.'
- Karimi (1994, p. 54)
- b. šām=am=o xord-am man  
dîner=1SG=RĀ manger.PST-1SG moi  
'Moi, j'ai déjà dîné. (*Me, I already ate dinner.*)'
- Mahootian (1997, p. 117)

Force est de constater que le seul cas de figure absolument exclu est le positionnement d'un élément focal, c'est-à-dire une information nouvelle qui n'a pas été déjà

mentionnée dans le discours, dans le domaine postverbal<sup>57</sup>. Selon Karimi (2003a, p. 115), l'impossibilité pour les mots Qu-, focals par définition, d'apparaître dans ce domaine démontrerait l'existence de cette contrainte en persan :

- (124) \*Ramin barā Kimea xarid                      čī / čī=ro  
Ramin pour Kimea acheter.PST.3SG quoi / quoi=RĀ

Outre les ordonnancements non canoniques associés à l'articulation discursive, nous rencontrons dans la littérature de manière relativement marginale quelques cas de réagencement expliqués par la contrainte du poids ou de la longueur. Mahootian (1997, pp. 127-128) mentionne la postposition des relatives (cf. §1.2.3) et l'antéposition ou la postposition des COI complexes (c'est-à-dire portant une relative), qu'elle qualifie de "*heavy shift*", ex. (125). Lazard et al. (2006, p. 194) mentionnent, quant à eux, des ordres dérogatoires liés à la longueur au sein des propositions prédicatives (à copule), dans lesquelles "lorsque [le prédicat nominal] est un groupe de mots relativement long, une partie en est souvent rejetée après la copule verbale", ex. (126).

- (125) a. pul=o      be un mard=i                      ( ) dād-im                      [ke  
argent=RĀ à ce homme=RSTR ( ) donner.PST-1PL que  
kotšalvār=e arzun=i                      pušide      bud]  
costume=EZ bon-marché=INDF habiller.PP être.PST.3SG  
'Nous donnâmes l'argent à cet homme en costume bon marché.'
- b. pul=o      ( ) dād-im                      [be un mard=i                      ke  
argent=RĀ ( ) donner.PST-1PL à ce homme=RSTR que  
kotšalvār=e arzun=i                      pušide      bud]  
costume=EZ bon-marché=INDF habiller.PP être.PST.3SG
- c. [be un mard=i                      ke kotšalvār=e arzun=i  
à ce homme=RSTR que costume=EZ bon-marché=INDF  
pušide      bud]                      pul=o      ( ) dād-im  
habiller.PP être.PST.3SG argent=RĀ ( ) donner.PST-1PL  
Mahootian (1997, p. 128)
- (126) miz=e      bozorg ... ke mamlov bud                      [az ruznāme majalle va  
table=EZ grand ... que plein être.PST.3SG de journal revue et  
ālbom=e aks]<sup>58</sup>  
album=EZ photo  
'La grande table ... qui était pleine de journaux, de revues et d'albums de  
photographies.'

Lazard et al. (2006, p. 194)

57. À l'exception bien évidemment des compléments phrastiques ainsi que des compléments directionnels et/ou datifs dans le registre informel, qui se placent par défaut dans ce domaine comme nous l'avons évoqué plus haut (cf. §1.2.1).

## 1.4 Les traitements syntaxiques existants

Outre les travaux à caractère descriptif, la variation de l'ordre des constituants en persan a reçu une analyse syntaxique formelle dans le cadre notamment de la grammaire générative<sup>59</sup>. Dans ce cadre théorique, le persan est décrit comme une langue configurationnelle<sup>60</sup>, c'est-à-dire manifestant une structure syntagmatique hiérarchisée de la phrase avec une asymétrie entre l'argument interne et l'argument externe du verbe (Moyne, 1974; Karimi, 1994; Browning & Karimi, 1994; Ghomeshi, 1996; Karimi-Doostan, 1997; Darzi, 1996; Lotfi, 2003; Kahnemuyipour, 2004; Karimi, 2005; Darzi, 2008; Toosarvandani, 2008; Ganjavi, 2007; Adli, 2010; Ilkhanipour, 2014; Modarresi, 2014, entre autres). L'ordre sous-jacent (excepté les compléments phrastiques) étant considéré SOV, les variations de l'ordre sont analysées comme des instances de différents types de mouvement ; ce qui requiert généralement l'existence de contraintes formelles et, de surcroît, l'exclusion de certaines possibilités. Les divergences résident notamment dans les types de mouvements postulés et la structure syntagmatique proposée.

Ces travaux se focalisent en majorité sur la position syntaxique de l'objet direct du fait du marquage différentiel de celui-ci, et, au-delà de leurs divergences, se rejoignent pour postuler que le COD marqué de  $=r\bar{a}$  occupe une position plus élevée et se caractérise par une indépendance plus forte vis-à-vis du verbe que sa contrepartie non marquée (Karimi, 1990; Browning & Karimi, 1994; Ghomeshi, 1997; Karimi, 2003a; Ganjavi, 2007). Ci-après, nous ferons référence à cette hypothèse comme celle de la "position syntaxique double de l'objet" (hypothèse de la PSDO).

Cette hypothèse repose sur le fait que, comme nous l'avons vu, le marquage différentiel intervient dans la détermination de l'ordre linéaire en favorisant l'éloignement du COD marqué du verbe. Les objets non marqués nus, quant à eux, occupent de préférence une position adjacente au verbe. Ce phénomène se manifeste notamment dans l'ordonnancement relatif des deux objets direct et indirect, qui a suscité, à ce titre, une attention particulière dans la littérature sur l'ordre des constituants en persan. De même, ce dernier se trouve au cœur des problématiques traitées dans notre recherche et, par conséquent, fera l'objet d'un examen approfondi dans les chapitres dédiés (ch. 4 et ch. 5) de la deuxième partie de cette thèse. Nous verrons que les résultats de ces études ne sont pas solidaires des prédictions attendues.

De ce fait, nous reviendrons sur le statut syntaxique du COD dans la troisième

---

58. L'auteur illustre, dans cet exemple, la postposition du complément prépositionnel de l'adjectif *mamlov*.

59. Pour une présentation du programme de grammaire générative et le traitement de la question de l'ordre des mots au sein de ce cadre théorique, se reporter à §2.2.1.3 au chapitre suivant.

60. Pour une définition, se reporter à §2.2.1.3.

partie (ch. 6) et ferons une présentation détaillée des analyses syntaxiques existantes (section 6.1). Par la même occasion, nous examinerons le fondement de l'hypothèse PSDO (section 6.2), au-delà de l'ordre relatif entre le COD et le COI, et démontrons que celui-ci est empiriquement contestable.

Une fois la position syntaxique du COD établie, certains auteurs s'intéressent au *scrambling*. Parmi les différents ordres non canoniques attestés, c'est l'extraction à partir de la phrase matrice, ou *long distance scrambling*, qui est abordée par la majorité des auteurs. Cela se fait généralement dans la continuité de l'analyse de  $=r\bar{a}$ , plus précisément en lien avec son fonctionnement discursif. De même, nous aborderons ce phénomène succinctement dans la continuité de l'analyse proposée par ces auteurs pour rendre compte du fonctionnement de  $=r\bar{a}$ , sachant que celui-ci dépasse le cadre de ce travail qui s'en tient à l'étude à la phrase simple.

Il faut noter qu'au-delà de la position syntaxique du COD et du fonctionnement de  $=r\bar{a}$ , la littérature générative traitant de la variation de l'ordre en persan se résume aux travaux de Karimi (1994, 2003a, 2005), dans lesquels elle propose avant tout une analyse syntaxique des ordres non canoniques, dans le domaine préverbal comme postverbal, en termes de *scrambling* (cf. §2.2.1.3) déclenchés par des traits discursifs. Par ailleurs, Karimi est, à notre connaissance, le seul auteur qui propose une argumentation en faveur de l'analyse configurationnelle du persan. En effet, dans les autres ouvrages susmentionnés, la vision configurationnelle du persan est considérée comme acquise. Dans la troisième partie (section 6.3), nous reviendrons également sur la vision configurationnelle du persan et présenterons une revue critique des arguments avancés dans ce sens et envisagerons une vision alternative, c'est-à-dire une structure syntagmatique plate de la phrase.

Par la suite, nous présentons l'analyse de Karimi pour rendre compte du *scrambling* dans le domaine préverbal (§1.4.1), ainsi que les contraintes formulées par l'auteur quant au domaine postverbal (§1.4.2).

### 1.4.1 *Scrambling* dans le domaine perverbal

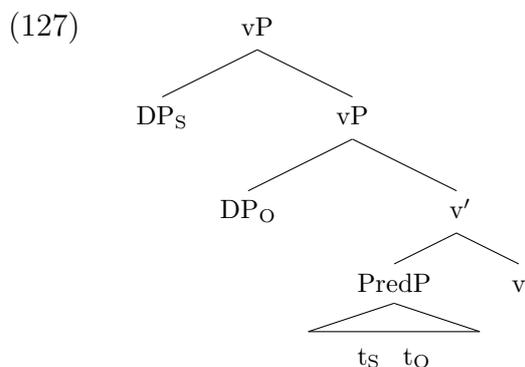
Pour rendre compte des données du persan et plus généralement des langues similaires, Karimi postule deux types de mouvement.

1/ En premier lieu, le déplacement des dépendants nominaux du verbe afin de recevoir l'interprétation, en d'autres termes les mouvements plus au moins standards du modèle génératif<sup>61</sup>, sans lien avec la configuration discursive, qui ne produisent

---

61. En réalité, Karimi (2005) s'écarte quelque peu des traitements classiques disponibles dans ce modèle et propose de nouvelles analyses pour le persan du fait de ces spécificités syntaxiques, notamment l'absence de sujet grammatical dans les constructions impersonnelles, l'absence d'accord systématique du verbe en nombre, etc. Néanmoins, la présentation détaillée de ces analyses

pas d'ordres non canoniques. Il s'agit notamment du déplacement obligatoire du COD spécifique, "object shift", ainsi que du déplacement obligatoire du sujet spécifique défini de manière analogue par Karimi (2005). Le schéma (127) illustre les positions vers lesquelles le COD et le sujet sont déplacés,  $DP_O$  et  $DP_S$  respectivement.



Adapté de Karimi (2005, p. 109)

Pour résumer, un  $COD_{[-spécifique]}$  demeure sous  $v''$ , plus précisément sous  $Pred''$ , c'est-à-dire à l'intérieur du prédicat exprimé par le verbe dont, selon l'auteur, il fait partie. Un  $COD_{[-spécifique]}$ , en revanche, doit se déplacer vers la position de spécifieur du  $v''$  pour recevoir son interprétation, autrement dit pour échapper au domaine de la clôture existentielle (cf. §6.1.1 pour une présentation détaillée).

En ce qui concerne le sujet, Karimi (2005) observant quelques spécificités du sujet en persan<sup>62</sup>, notamment l'absence d'accord (obligatoire)<sup>63</sup> du verbe en nombre avec le sujet non spécifique, ex. (128), postule qu'en persan le filtre du Cas et le critère thématique sont vérifiés non par le mouvement de celui-ci vers la position du

dépasse largement le cadre de ce travail (voir Karimi, 2005, ch. 3, pour plus de détails).

62. Ces observations sont abordées avec un fort ancrage théorique qui dépasse le cadre de ce travail. Par conséquent, nous n'entreprendrons pas ici la présentation intégrale et détaillée de ces dernières.

63. Nous n'adhérons pas à l'analyse de Karimi (2005), qui propose que le verbe s'accorde en nombre seulement avec son sujet spécifique. Cette analyse catégorique est sujette à caution. Bien qu'il soit incontestable qu'en persan l'accord du verbe en nombre avec son sujet inanimé n'est pas systématique, les facteurs qui y interviennent restent à déterminer. Quoi qu'il en soit, une analyse en termes de traits binaires est discutable. D'autres facteurs, comme l'agentivité et l'individualité du sujet, les propriétés sémantico-aspectuelles du verbe, etc., ex. (i), sont, en effet, évoqués pour rendre compte de ce phénomène (voir Lazard *et al.*, 2006, pp. 164-165) :

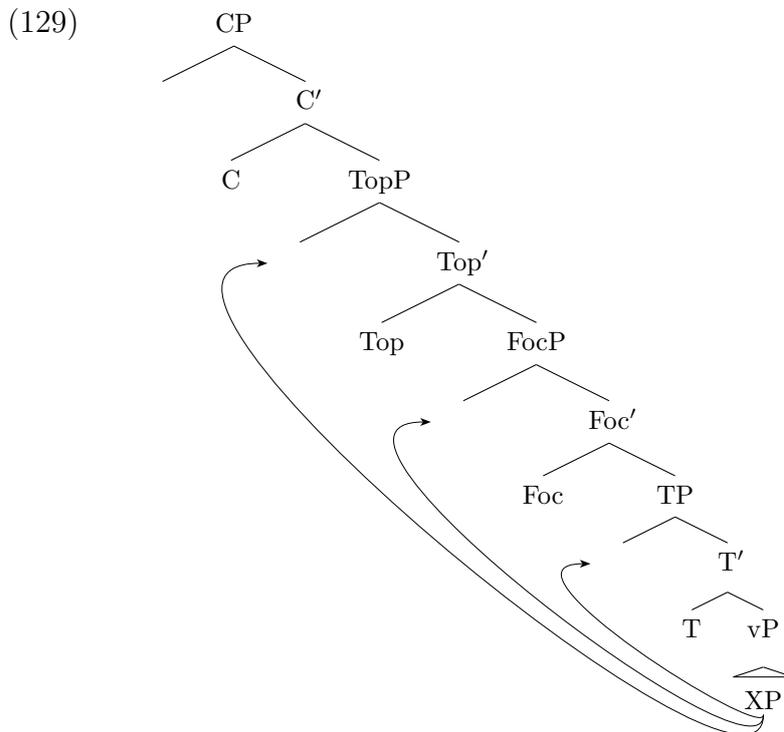
- (i) a. divār-hā=ye otāq bu=ye nam mi-dād  
 mur-PL=EZ chambre odeur=EZ humidité IPFV-donner.PST.3SG  
 'Les murs de la chambre sentaient l'humidité.'
- b. čerāq-hā (yeki-yeki) xāmuš šod(-and)  
 lampe-PL un-par-un éteint devenir.PST.3SG(-3PL)  
 'Les lampes se sont éteintes (une par une/toutes ensemble).'

spécifieur de I'' (ou T''), ni par le mouvement du v comme c'est généralement le cas dans les langues *pro drop*, mais par la morphologie verbale riche de cette langue.

- (128) ketāb-hā ru=ye miz-e/an  
 livre-PL sur=EZ table-COP.3SG/COP.3PL  
 'Les livres sont sur la table.'

Karimi (2005, p. 97)

Il en va de même pour la satisfaction de l'EPP (*Extended Projection Principle*, Principe de la Projection Étendue), selon lequel une phrase finie doit obligatoirement avoir un sujet<sup>64</sup>. Il s'ensuit que le sujet grammatical peut rester *in situ*, c'est-à-dire sous Pred''. Karimi (2005, p. 95) avance que les SN se trouvant dans ce domaine, faisant partie du prédicat exprimé par le verbe, sont neutres vis-à-vis du filtre du Cas. Le sujet spécifique, au contraire, ne fait pas partie du prédicat et doit sortir du domaine de la clôture existentielle, c'est-à-dire du Pred'', pour recevoir son interprétation. De plus, ce dernier requérant l'accord du verbe, il est donc légitime de postuler qu'il se déplace vers la position de spécifieur du v'' où le filtre du Cas et le critère thématique sont vérifiés par la morphologie verbale (Karimi, 2005, p. 95-96).



64. Rappelons que Karimi considère le persan comme une langue *topic prominent* et non *subject prominent*. Ce qui signifie que cette langue ne requiert pas la présence d'un dépendant nominal dans une position externe au v''. Elle affirme en revanche qu'en persan, un dépendant du verbe, qui n'est pas forcément le sujet, se déplace vers la position du spécifieur de T'' lorsqu'il porte le trait [+topic] (Karimi, 2005, p. 111 et pp. 116-117).

Adapté de Karimi (2005, p. 108)

2/ En second lieu viennent des mouvements relevant du *scrambling*, c'est-à-dire ceux qui aboutissent en ordres non canoniques, déclenché par des traits discursifs. Il s'agit du déplacement d'un constituant (XP), satisfaisant les conditions sémantiques spécifiées par l'auteur, vers la position de spécifieur d'une tête fonctionnelle, comme cela est schématisé en (129).

Nous avons présenté brièvement la configuration discursive de la phrase proposée par Karimi (2005) dans la section précédente (cf. §1.3.1). Comme nous l'avons vu, elle définit trois fonctions discursives – *shifted topic* (TopP), *background topic* (TP) et *contrastive focus* (FocP) – en mesure de provoquer ce déplacement. Rappelons que *background topic*, se trouvant dans une position plus basse, correspond au topic et *shifted topic* correspond au changement de topic dans les phrases successives. L'auteur précise qu'en l'absence de *shifted topic*, le *background topic* peut être placé dans la position initiale de la phrase.

Rappelons que, d'après Karimi (2005), dans une phrase de type *all focus* tous les dépendants verbaux restent *in situ*<sup>65</sup>, c'est-à-dire sous  $v''$  ou plus précisément sous  $\text{PredP}''$ . En d'autres termes, le focus inhérent, contrairement au focus contrastif, est confiné à la position préverbale. Par ailleurs, elle affirme que le déplacement d'un élément contrasté est optionnel.

- (130) a. man **divān=e Hāfez=o** bārā Kimea xarid-am  
 moi recueil=EZ Hafez=RĀ pour Kimea acheter.PST-1SG  
 b. [<sub>FocP</sub> *divān=e Hāfez=o* [man  $t_i$  bārā Kimea xarid-am]]  
 'C'était le recueil des poèmes de Hafez que j'ai acheté pour Kimea (et pas autre chose).'

Karimi (2005, p. 135)

- (131) a. man be **Kimea** ye ketāb dād-am  
 moi à Kimea un livre donner.PST-1SG  
 b. [<sub>FocP</sub> be **Kimea** [man  $t_i$  ye ketāb dād-am]]  
 'C'est à Kimea que j'ai donné le livre (non à quelqu'un d'autre).'

Karimi (2005, p. 135)

Karimi (2005) suppose que le *scrambling* est soumis à certaines restrictions basées notamment sur les conditions sémantiques requises par les différentes fonctions discursives postulées. Elle affirme que les dépendants non spécifiques, faisant partie du prédicat exprimé par le verbe, correspondent toujours au focus primaire (inhérent) et jamais à la fonction du topic, qu'il soit de type *shifted* ou *background*. De ce

65. Voir la note de bas de page 64.

fait, ils ne peuvent satisfaire que la condition sémantique requise pour la fonction de focus contrastif. Leurs contreparties spécifiques, en revanche, satisfont la condition sémantique requise pour les trois fonctions.

Ces conditions entraîneraient des restrictions sur les positions linéaires disponibles pour les constituants non spécifiques, s'appliquant également à l'élément non verbal d'un prédicat complexe, et expliqueraient la liberté de positionnement plus importante des constituants spécifiques. Ainsi, par exemple, un COD<sub>[-spécifique]</sub> ne peut précéder le COI ou prendre la position initiale que s'il correspond au focus contrastif, ce qui requiert l'accent fort, ex. (132). Il en va de même pour l'élément non verbal d'un prédicat complexe, ex. (133), qui, selon Karimi, doit contenir un élément quantificationnel (*če* en l'occurrence) et porter l'accent fort pour être séparé du verbe.

- (132) a. Parviz **pirhan** barā Kimea xarid  
 Parviz chemise pour Kimea acheter.PST.3SG  
 b. **pirhan** Parviz barā Kimea xarid  
 'C'est **des chemises** que Parviz acheta pour Kimea (non des robes).'  
 c. \*Parviz pirhan barā Kimea xarid  
 Pour dire 'Parviz acheta des chemises pour Kimea.'

Adapté de Karimi (2005, p. 18)

- (133) a. Kimea [če zamin=e saxt=i]<sub>i</sub> diruz t<sub>i</sub> xord  
 Kimea quel terre=EZ dur=INDF hier manger.PST.3SG  
 'Quelle mauvaise chute Kimea a fait hier.'  
 b. \*Kimea zamin=e saxt=i diruz xord

Adapté de Karimi (2005, p. 19)

Ces assertions sont contestables. En ce qui concerne les prédicats complexes, comme nous en avons discuté précédemment (cf. §1.1.4), Samvelian (2001, 2012) démontre, exemples attestés à l'appui, que les composantes d'un prédicat complexe peuvent être séparées par divers constituants syntaxiques, et cela sans qu'une lecture quantificationnelle soit requise, comme l'illustre ex. (134).

Quant aux COD non spécifiques, nous rencontrons également des exemples attestés dans lesquels un COD non marqué précède le sujet et/ou le COI, sans pour autant qu'une lecture contrastive de celui-ci soit requise, ex. (135).

- (134) ān zan bargāšt va sili be surat=am zad  
 ce femme retourner.PST.3SG et gifle à visage=1SG frapper.PST.3SG  
 'Cette femme se retourna et me donna une gifle.'

Extrait de *Suvašun* de Simin Daneshvar (*in* Samvelian, 2012, p. 40)

- (135) harčand [COD mavāred=i] niz [S<sub>uj</sub> komisyon] [COI be lāyehe=ye  
 bien que cas.PL=INDF aussi commission à projet de loi  
 dowlat] afzude ast  
 gouvernement ajouter.PP être.PRS.3SG  
 ‘Bien que la commission ait aussi ajouté quelques cas au projet de loi du  
 gouvernement.’

Corpus Bijankhan

## 1.4.2 Le positionnement postverbal des constituants

Les phénomènes de la variation de l’ordre dans le domaine préverbal au sein de la phrase simple feront l’objet d’une étude empirique approfondie dans la deuxième partie de cette thèse. Les phénomènes qui relèvent du domaine postverbal ne seront, par contre, pas traités. Ceux-ci n’ont pas reçu l’attention attendue dans la littérature. Ils mériteraient un examen empirique de même ampleur, ce qui dépasse le cadre de la présente thèse.

Par souci d’exhaustivité, avant de finir cette section, nous aborderons brièvement les contraintes formulées dans la littérature quant au *scrambling* des dépendants verbaux dans le domaine postverbal, appelé généralement l’antéposition du verbe (*verb preposing*). Rappelons tout d’abord que la position postverbale est considérée unanimement comme la position canonique (ou sous-jacente) des compléments phrastiques, ainsi que la position par défaut des compléments directionnels/allatifs dans le registre informel (cf. §1.2.1).

Karimi (1994) relève les contraintes suivantes quant à la postposition des constituants au verbe. Tout d’abord, elle avance que le *gapping* – matérialisé dans les exemples par un cadratin – n’est pas possible dans le domaine postverbal en persan, ex. (136). Cet argument confirmerait que l’ordre sous-jacent est SOV en persan, notamment par opposition à l’hypothèse de Kayne (1994), qui postule l’ordre SVO comme ordre sous-jacent universel de la phrase.

- (136) a. man ketāb=o mi-xun-am to ruznāma=ro —  
 moi livre=RĀ IPFV-lire.PRS-1SG to journal=RĀ  
 ‘Je lis le livre (et) toi le journal.’  
 b. \*man mi-xun-am ketāb=o to — ruznāma=ro
- (137) a. man bā Sāsān raqsid-am Rāmin bā Sepide —  
 moi avec Sasan danser.PRS-1SG Ramin avec Sepideh  
 ‘J’ai dansé avec Sasan (et) Ramin avec Sepideh.’  
 b. \*man raqsid-am bā Sāsān Rāmin — bā Sepide

Karimi (1994, p. 50)

Ces phrases nous paraissent également très difficilement acceptables. Nous expliquons cette difficulté par le fait qu'en règle générale en persan, le domaine postverbal n'est pas favorable aux constituants focaux.

De plus, Karimi (1994) affirme que les COD non spécifiques (non marqués), ex. (138), ainsi que les sujets non spécifiques, ex. (139), sont exclus de cette position :

- (138) a. \*Sepide emruz jāru kard se=tā otāq  
 Sepideh aujourd'hui balai faire.PST.3SG trois=CLF chambre  
 Pour dire 'Sepideh a balayé trois chambres aujourd'hui.'  
 Adapté de Karimi (1994, p. 61)

- b. \*gozāšt-i ru miz gol  
 mettre.PST-2SG sur table fleur  
 Pour dire 'As-tu mis des fleurs sur la table?'  
 Adapté de Karimi (1994, p. 64)

- (139) \*xābid-an do=tā pesar  
 dormir.PST-INF.3PL deux=CLF garçon  
 Pour dire 'Deux garçons dorment.'  
 Adapté de Karimi (1994, p. 65)

Or, ces ordres ne nous semblent pas pour autant impossibles, comme l'illustrent les exemples ci-dessous. Le nom nu *mive* 'fruit', occupant la fonction du sujet et du COD dans ex. (140) et ex. (141) respectivement, bien que non spécifique, correspond à une information donnée dans le discours dans les deux phrases. Par conséquent, il n'est pas focal et peut apparaître dans le domaine postverbal. Par ailleurs, nous acceptons parfaitement la phrase (138-b) ci-dessus, dans un contexte similaire où, par exemple, l'interlocuteur est supposé avoir mis des fleurs sur la table. Rappelons que l'antéposition du verbe, permettant la mise en relief de celui-ci (cf. §1.3.3), est considérée comme un ordre peu fréquent, ayant une connotation familière forte et nécessitant une prosodie marquée (Lazard *et al.*, 2006, p. 197).

- (140) a. mive bo-xor  
 fruit IMP-manger.2SG  
 'Mange des fruits!'  
 b. xord-am mive  
 manger.PST-1SG fruit  
 'J'ai bien mangé des fruits.'

- (141) a. mive na-dār-im  
 fruit NEG-avoir.PRS-1PL  
 'Il n'y a plus de fruits?'  
 b. hast tu yaxčāl mive  
 exister.PRS.3SG dans frigo fruit  
 'Il y en a, dans le frigo, des fruits.'

- (142) a. bilit xarid-in  
 billet acheter.PRS-2PL  
 ‘Avez vous acheté des billets?’
- b. na-resid bilit be=hemun  
 NEG-arriver.PST.3SG billet à=3PL  
 ‘Il n’y avait plus de billets.’

En outre, cet ordre serait également exclu dans les subordonnées occupant la fonction du complément nominal, ex. (143). Nous sommes encore en désaccord avec ces jugements de grammaticalité et acceptons des phrases avec cet ordre, ex. (144).

- (143) \*in mozu ke nevešte bud Rāmin un maqāla=ro barā ruznāme  
 ce sujet que écrire.PP être.PST.3SG Ramin ce article=RĀ pour journal  
 sehhat na-dār-e  
 validité NEG-avoir.PRS-3SG  
 Pour dire ‘Il n’est pas vrai que Ramin avait écrit cet article pour le journal’  
 Karimi (1994, p. 56)
- (144) in harf ke na-did Rāmin Sarā=ro tu sinema  
 ce parole que NEG-voir.PST.3SG Ramin Sarah=RĀ dans cinéma  
 doruq=e  
 mensonge=COP.3SG  
 ‘Dire que Ramin n’a pas vu Sarah au cinéma est un mensonge.’

## 1.5 Conclusion

Dans ce chapitre, nous avons abordé les phénomènes relevant de la variation de l’ordre des constituants en persan, une langue avec ordre canonique SOV, dans laquelle tous les autres cas de figure (SVO, OSV, etc.) peuvent également se produire. Cette présentation a permis notamment de souligner le lien étroit entre l’ordre linéaire et l’articulation discursive de la phrase en persan ; celle-ci constitue, en effet, la base commune des analyses proposées pour rendre compte de la variation de l’ordre dans cette langue. Pour simplifier, l’on peut dire que la position initiale est généralement réservée aux constituants topicalisés, alors que la position immédiatement préverbale est focale. De plus, le domaine postverbal n’est pas favorable aux constituants focaux – à l’exception des compléments phrastiques et des compléments directionnels. Nous avons également mis en évidence le rôle singulier du marquage différentiel de l’objet, qui a retenu l’attention des chercheurs en syntaxe formelle. Les travaux de ces auteurs feront l’objet d’une étude détaillée dans le premier chapitre de notre troisième partie (cf. section 6).

En outre, nous avons passé brièvement en revue les analyses syntaxiques exis-

tantes quant à la variation de l'ordre des constituants en persan, qui relèvent toutes du cadre de la grammaire générative. Les auteurs des travaux que nous avons examinés n'abordent pas les phénomènes de la variation de l'ordre à proprement parler. En effet, la variation en tant que telle ne rentre pas dans les objectifs de ce cadre théorique.

Dans le chapitre suivant, nous effectuerons une présentation de ce cadre en vue de mettre en évidence les limites qu'il présente pour l'étude de la variation de l'ordre et présenterons une démarche alternative, développée dans les années 2000 au croisement des disciplines de la linguistique théorique et de la psycholinguistique. Ces études se fondent sur les données d'usage et privilégient des méthodes quantitatives, rendues possibles par la disponibilité de corpus annotés de grande taille, ainsi que des protocoles expérimentaux inspirés des méthodes ayant cours en psychologie. En nous inscrivant dans cette démarche, nous avons conduit une étude empirique basée sur l'usage des phénomènes de la variation de l'ordre dans le domaine préverbal au sein de la phrase simple en persan, dont nous rendrons compte dans la deuxième partie de cette thèse.

## Chapitre 2

# Étudier la variation de l'ordre des mots : État de l'art

Dans ce chapitre, nous allons passer en revue les questions relatives à la variation de l'ordre des mots. Dans la première section, nous en proposerons une définition et présenterons les différents phénomènes qui s'étudient sous cet angle. Cela nous sera utile pour définir notre méthodologie.

Étudier la variation de l'ordre revient à poser la question de la préférence entre deux (ou plusieurs) options autorisées par la grammaire. Ainsi, les contraintes grammaticales ne rentrent pas en ligne de compte. Ce sont, en revanche, des facteurs extra-grammaticaux qui peuvent influencer le choix, qui reflètent des comportements tendanciels et n'agissent pas de manière binaire. Par conséquent, les méthodes empiriques en vigueur en syntaxe, c'est-à-dire le jugement de grammaticalité par introspection, se révèlent inadéquates.

Néanmoins, dans d'autres domaines des sciences du langage, notamment la typologie linguistique et la psycholinguistique, ces facteurs continuent de susciter un intérêt. Cela conduit finalement à engendrer, dans le domaine de la syntaxe, un changement de paradigme à l'initiative des syntacticiens qui considèrent l'étude de la performance linguistique comme indispensable et au moins aussi légitime que celle de la compétence. Ces auteurs s'approprient les outils et méthodes utilisés dans d'autres disciplines pour l'analyse quantitative des comportements langagiers afin de proposer une étude appropriée des préférences.

Pour saisir ce changement de paradigme, nous proposons dans la deuxième section (2.2) un aperçu de l'étude de la variation de l'ordre à travers la littérature. Dans la troisième section (2.3), nous nous concentrerons sur les problématiques évoquées dans cette littérature.

## 2.1 Qu'est-ce que la variation de l'ordre des mots ?

Nous pouvons définir la variation de l'ordre des constituants, appelée généralement et par commodité variation de l'ordre des mots (*word order variations*), comme l'existence de différentes possibilités d'ordonnement des constituants pour un même contenu propositionnel, autrement dit des phrases ayant la même valeur véridictionnelle, mais différentes quant à l'ordre. Des raisons d'ordre cognitif, pragmatique ou stylistique peuvent être à l'origine de ces variations<sup>1</sup>.

Deux types de langues se distinguent dans la littérature : les langues à ordre rigide (appelé également – par simplification – ordre fixe) vs. les langues à ordre flexible (appelé également ordre relativement libre ou encore – par simplification – ordre libre)<sup>2</sup>.

### 2.1.1 Phénomènes relevant de la variation de l'ordre

Les phénomènes de la variation de l'ordre ne se limitent pas à ceux rencontrés dans les langues à ordre libre. À titre d'exemple, les phrases ci-dessous présentent trois phénomènes d'alternance de l'ordre des mots en français, langue SVO à ordre rigide, à savoir l'ordre relatif entre les compléments verbaux, ex. (1), le placement de l'adjectif épithète, ex. (2), et l'inversion dite stylistique du sujet, ex. (3).

- (1) a. J'ai donné une lettre au responsable administratif.  
 b. J'ai donné au responsable administratif une lettre.
- (2) a. C'était une soirée agréable.  
 b. C'était une agréable soirée.
- (3) a. Un ermite vivait dans la forêt.  
 b. Dans la forêt vivait un ermite.

Pareillement, dans une langue à ordre flexible, comme le persan, le japonais ou le turc, il existe des relations de préférence imposées par la grammaire dont la violation rendrait la séquence agrammaticale. Ainsi, deux cas de figure sont possibles en ce qui concerne l'ordre : l'ordre peut être contraint par la grammaire, ex. l'ordre Dét < Nom (\*Nom < Det) dans les langues à tête initiale (ou encore l'ordre entre

---

1. Il va de soi que la variation de l'ordre n'implique pas que les deux (ou plusieurs) variantes possibles soient équivalentes ou qu'elles se produisent de manière égalitaire. Effectivement, dans une majorité de cas de figure un ordre par défaut (ou dominant) peut être observé.

2. Cette dichotomie rappelle celle posée en termes de configurationnalité que nous allons évoquer plus loin (cf. §2.2.1.3) et sur laquelle nous reviendrons dans la troisième partie de la thèse (section 6.3). Rappelons que, dans les deux cas, il ne s'agit pas d'une classification absolue, mais du placement des langues sur un continuum.

une préposition/postposition et son complément) ou ne pas l'être, ex. l'ordre dans une coordination<sup>3</sup>. La formulation de ce type de généralisation remonte assez loin dans l'histoire des sciences du langage, notamment à Behaghel (1909, p. 39), qui postule, entre autres, un principe connu sous la désignation *End weight* pour rendre compte de la tendance, observée en grec, en latin et en allemand, à ordonner les éléments plus courts avant les plus longs. Nous reviendrons sur de telles hypothèses dans la section 2.3 de ce chapitre.

Il faut mentionner que des phénomènes variés se regroupent sous la même appellation. Tout d'abord, ils peuvent concerner l'ordre relatif entre une tête et son dépendant, comme c'est le cas du placement de l'adjectif épithète en français. Ils peuvent également concerner l'ordre relatif entre différents dépendants d'une même tête, comme c'est le cas de l'ordre relatif des compléments postverbaux en français. Il peut s'agir des constituants de nature différente, mais aussi des constituants de même nature, partageant le même rôle grammatical, comme c'est le cas dans une coordination, ou encore de l'ordre relatif entre les compléments verbaux prépositionnels, ou entre ajouts adverbiaux, prépositionnels ou adjectivaux.

Par ailleurs, la variation de l'ordre au sens strict n'entraîne aucune autre modification dans la phrase comme c'est le cas dans l'ensemble des exemples mentionnés jusqu'à présent. Mais celle-ci peut avoir une portée large et englober les phénomènes qui ne sont pas de simples alternances de l'ordre et qui entraînent d'autres modifications dans la structure de la phrase. Nous pouvons citer la dislocation comme exemple de ce type de variation de l'ordre, ex. (4).

- (4) a. Marie a cassé le vase.  
 b. Le vase, Marie l'a cassé.  
 c. Marie, le vase, elle l'a cassé.  
 etc.

Le français connaît un ordre canonique SVO relativement rigide. Ceci étant, il est possible, par exemple, de placer l'argument réalisé comme le COD avant le verbe. Dans ce cas, celui-ci devrait être repris par un pronom résomptif enclitique au verbe<sup>4</sup>. Il en va de même pour les structures dites clivées, ex. (5). En outre, l'alternance de la voix (ou diathèse), c'est-à-dire voix active, ex. (6-a), voix passive, ex. (6-b), ou voix moyenne, ex. (6-c), s'étudie sous le même angle, mais, dans ces cas, il est plus

---

3. Notons que l'ordre linéaire des éléments coordonnés n'est pas pour autant nécessairement libre. Par exemple, lorsque celui-ci correspond à un ordre chronologique ou qu'il est déterminé par d'autres éléments – extra-linguistiques – relevant du contexte ou lorsqu'il est figé comme dans *noir et blanc*. Ce nonobstant, l'ordre inverse n'est jamais agrammatical, strictement parlant.

4. L'argument déplacé est traité par certains auteurs comme "détaché".

exact de parler de la variation de l'ordre des arguments, dans l'acception sémantique du terme.

- (5) a. Marie a cassé le vase.  
 b. C'est le vase que Marie a cassé.  
 etc.
- (6) a. Marie a cassé le vase.  
 b. Le vase a été cassé par Marie.  
 c. Le vase s'est cassé.

Dans le même ordre d'idées, nous pouvons évoquer l'alternance dative (*dative alternation* ou *dative shift*), ex. (7), largement étudiée dans la littérature sur la langue anglaise (Collins, 1995; Wasow, 2002; Gries, 2003; Bresnan *et al.*, 2007), de même que l'alternance génitive dans cette langue, ex. (8) (Rosenbach, 2005).

- (7) a. Kim handed a toy to the baby.  
 b. Kim handed the baby a toy.
- (8) a. The king's palace.  
 b. The palace of the king.

### 2.1.2 La variation de l'ordre comme objet d'étude en syntaxe

La question de la variation de l'ordre est une des questions de la syntaxe qui a permis la (re)mise en question de la définition de l'objet d'étude et le choix méthodologique en syntaxe. Depuis Chomsky (1957), l'objet d'étude de la linguistique se définit en référence à l'opposition compétence/performance – ou son équivalent langue interne (*I-Language*) vs. langue externe (*E-language*). Dans cette optique, la question de la variation intra-langue (et *a fortiori* celle de l'ordre) relève de la sphère de la performance et, à ce titre, n'est pas pertinente pour la théorie de la grammaire.

Pour le dire autrement, la variation de l'ordre ne relève pas de contraintes grammaticales à proprement parler, dites “dures” (*hard*)<sup>5</sup>, autrement dit des contraintes qui déterminent la grammaticalité d'une séquence et, de ce fait, composent (le noyau de) la connaissance linguistique. Les contraintes de nature graduelle et/ou préférentielle, dites “molles” (*soft*), qui interviennent pour déterminer les préférences, font

---

5. Terminologie répandue notamment grâce à Keller (2000); Sorace & Keller (2005).

en revanche partie de la performance, et sont, par conséquent, “non pertinentes” (*irrelevant*) selon les termes de Chomsky, pour la théorie linguistique (1965, p. 3).

Or, plus récemment, des linguistes travaillant en particulier sur l'alternance dative en anglais (Wasow, 2002; Bresnan *et al.*, 2007; Bresnan, 2007; Bresnan & Ford, 2010) plaident pour la reconnaissance de ce type de contraintes comme faisant partie intégrante de la connaissance linguistique. Ils rejoignent ainsi les auteurs adhérant aux théories basées sur l'usage, qui considèrent l'étude de la performance indissociable de l'étude de la compétence, et suggèrent, de surcroît, que cette dernière comporte une part probabiliste. Ces auteurs affirment notamment qu'il est possible d'extraire des tendances stables et convergentes, chez les locuteurs d'une langue donnée, qu'il n'est pas théoriquement justifié d'exclure de cette entité abstraite qu'est la connaissance linguistique commune.

Ce plaidoyer pour la redéfinition de l'objet d'étude en syntaxe s'accompagne bien entendu d'un changement méthodologique, puisque la méthodologie traditionnelle, basée essentiellement sur l'introspection, se révèle absolument incompatible avec l'étude des tendances qui appelle une approche quantitative. Ce changement de paradigme est rendu possible grâce aux avancées et pratiques effectives des autres domaines connexes. D'une part, la linguistique informatique a mis à disposition des chercheurs des corpus annotés de grande taille. Ne négligeons pas le fait que des spécialistes en linguistique informatique ont milité pour ce changement de paradigme (Schutze, 1996; Manning, 2003; Meurers, 2005). D'autre part, la méthodologie et les protocoles expérimentaux sont en vigueur dans les domaines tels que la phonétique, l'acquisition, la psycholinguistique et la sociolinguistique qui font par ailleurs usage de diverses méthodes d'analyse statistique, notamment en mobilisant des modèles de régression.

## 2.2 Évolution de l'étude de la variation de l'ordre

Dans cette section, nous proposons un panorama des études de la variation de l'ordre dans la littérature<sup>6</sup>. Différents domaines des sciences du langage, allant de la syntaxe générale à la psycholinguistique en passant par la typologie linguistique, se sont intéressés à cette question, chacun en lien avec ses propres préoccupations. Notons que les études pluridisciplinaires sur le sujet sont de plus en plus fréquentes.

En syntaxe, dont la définition traditionnelle est l'étude de la combinaison des mots pour former des séquences grammaticales, la question de l'ordre des mots (*word order*) est omniprésente. Ceci est dû au fait qu'elle va de pair avec la question

---

6. Pour une revue de la littérature de la question de l'ordre des mots, voir Song (2012).

de la structure et celle de la constituance, et se trouve directement ou indirectement impliquée dans la détermination des relations grammaticales. La variation de l'ordre en tant que telle n'éveille, en revanche, qu'un intérêt ambivalent dans ce domaine.

L'ordre des mots est également une des questions fondamentales de la typologie linguistique, qui s'intéresse aux tendances générales observables à travers les langues du monde et se trouve au cœur des travaux fondateurs de ce domaine (Greenberg, 1963, 1966; Dryer, 1992). La variation de l'ordre, à proprement parler, y a également fait l'objet de plusieurs études (Hawkins, 1983, 1990, 1994; Siewierska, 1988, 1993; Givón, 1983, 1988).

Un intérêt notable sur la question de la variation de l'ordre est porté par les chercheurs en sciences cognitives qui s'intéressent activement aux mécanismes de la production et du traitement de la parole en temps réel, comme en témoignent les nombreux travaux dans cette perspective consacrés à ce sujet (voir Garrett, 1980; Bock, 1982; Bock & Warren, 1985; Hemforth, 1993; McDonald *et al.*, 1993; Gibson, 1998; Konieczny, 2000; Kempen & Harbusch, 2003; Vasishth, 2003; Christianson & Ferreira, 2005; Jaeger & Norcliffe, 2009; Chang, 2009; Tily, 2010; MacDonald, 2013, parmi beaucoup d'autres). Cela s'explique par le fait que l'étude des facteurs intervenant dans les phénomènes qui relèvent des contraintes préférentielles peut apporter un éclairage sur le fonctionnement du système de la production et/ou du traitement de la parole. La variation de l'ordre est, en effet, un terrain propice à cet égard. À la différence de la typologie linguistique, les sciences cognitives, du fait de leur méthodologie plus contraignante, se concentrent généralement sur une langue donnée et se basent, somme toute, sur un ensemble restreint de langues.

Dès les années 1990, la question de la variation de l'ordre en tant que telle revient sur le devant de la scène en syntaxe grâce aux auteurs qui, du fait de leur sensibilité aux faits de l'usage réel, appellent à un changement de perspective en syntaxe et s'approprient des approches quantitatives, y compris des méthodes expérimentales. Ainsi, une nouvelle approche voit le jour, que l'on peut qualifier de cognitivo-fonctionnelle et que nous avons adoptée dans cette thèse.

Dans cette section, nous tâcherons de tracer le cheminement de l'étude de la variation de l'ordre vers cette approche, tout en ayant à l'esprit que faire un état de la littérature des différents domaines et approches théoriques concernés dépasse largement le cadre de ce travail. Nous parlerons, tout d'abord, du courant génératif, qui a eu un impact indéniable sur les pratiques actuelles de la recherche en linguistique générale (§2.2.1), et ensuite brièvement de la typologie linguistique (§2.2.2). Nous présenterons enfin les travaux adoptant une approche cognitivo-fonctionnelle (§2.2.3). Dans la section suivante (section 2.3), nous exposerons les hypothèses émer-

gentes de l'ensemble de ces travaux pour rendre compte des préférences observées.

### 2.2.1 La grammaire générative

Les pratiques actuelles de la linguistique théorique sont toutes plus au moins marquées par les travaux de Noam Chomsky, le fondateur de la linguistique générative née d'une rupture avec la linguistique structurale.

La description des langues non décrites était un des objectifs majeurs de la linguistique structurale (représentée par les travaux de Ferdinand de Saussure). La méthodologie dominante dans ce cadre théorique, mise en exercice notamment par le courant distributionnaliste (développé par Leonard Bloomfield dans les années 1920), consiste à étudier les langues à travers des corpus d'élicitation collectés par des linguistes épaulés par des informateurs bilingues, souvent au cours de séjours sur le terrain. La démarche de la linguistique distributionnelle peut être résumée ainsi : inventorier des unités de la langue, sous forme de couplages de formes et de sens, et établir des règles qui rendent compte de leurs relations, autrement dit qui explicitent les régularités observées.

Cette démarche, dominante jusque dans les années 1960, est vivement critiquée par Chomsky (1957, 1965, 1975, 1986) qui plaide pour une réflexion épistémologique et la redéfinition de l'objet d'étude en linguistique. Celle-ci consiste à déplacer l'attention des études linguistiques de la "langue externe" *E-language* vers la "langue interne" *I-language* : "*from the study of language regarded as an externalized object to the study of knowledge of language attained and internally represented in mind/brain*" (Chomsky, 1986, p. 24)<sup>7</sup>.

Le changement du paradigme suscité par les travaux de cet auteur a ainsi affecté la recherche en linguistique théorique sur deux plans essentiels : son objet d'étude et sa méthodologie. L'étude de la variation de l'ordre des mots est directement concernée par ce changement sur les deux plans.

Dans ce qui suit, nous nous livrons tout d'abord à une brève présentation accompagnée de remarques critiques succinctes du programme génératif (§2.2.1.1 et §2.2.1.2), en vue de mettre en évidence l'impact de ce cadre théorique sur l'étude de la variation de l'ordre<sup>8</sup>. Ensuite, nous aborderons l'ordre des mots en syntaxe généra-

---

7. Dans la vision bloomfieldienne, la langue est définie comme "*the totality of utterances that can be made in a speech community*" (Bloomfield, 1926, p. 155) et la grammaire est envisagée comme "*the meaningful arrangement of forms*" (Bloomfield, 1984, p. 163). Pour Chomsky, cette démarche sous-entend la conception de la langue comme un "artéfact", un "épiphénomène", c'est-à-dire comme quelque chose d'externe (Chomsky, 1986, p. 19). Or, le langage est une propriété de la pensée humaine, une connaissance intériorisée par les locuteurs (Chomsky, 1986, p. 24).

8. Pour une discussion critique relativement exhaustive tout en étant synthétique de ce programme, voir Wasow (2002, pp. 131-148). L'auteur y argumente notamment contre : 1) la vision

tive transformationnelle et la place fondamentale qu'il y occupe (§2.2.1.3) et, enfin, nous évoquerons brièvement la vision des courants génératifs non transformationnels (§2.2.1.4). Notons que la syntaxe générative chomskyenne a connu plusieurs développements depuis son émergence<sup>9</sup>, dont nous ferons abstraction en faisant usage de l'appellation générale de syntaxe générative transformationnelle.

### 2.2.1.1 Définir les objectifs de la théorie

Premièrement, la langue interne, objet que la théorie générative se propose d'étudier, est une conception abstraite : le savoir linguistique tacite d'un locuteur idéal (la compétence), indépendant et indemne vis-à-vis de son usage effectif (la performance).

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.

Chomsky (1965, p. 3)

La compétence représente la connaissance linguistique stable et identique pour tous les locuteurs partageant la même langue, c'est-à-dire appartenant à la même *speech community*, tandis que la performance est l'utilisation réelle de ce savoir et ainsi variable d'un locuteur à l'autre et d'une situation à une autre, en fonction de nombreuses considérations extralinguistiques portant sur le locuteur – des paramètres plus au moins fixes dépendant de la biographie du locuteur (expérience linguistique, âge, éducation, situation socio-culturelle, situation géographique, etc.) ou passagers comme le niveau d'attention au moment de l'acte de parole – ou sur la situation de l'énonciation. Dans le cadre de la théorie générative, une distinction entre les deux sphères s'impose en vue de concevoir l'objet d'étude de la linguistique<sup>10</sup>.

---

binaire des contraintes et le rejet total d'information quantitative, 2) la priorité de l'étude de la compétence linguistique sur celle de la performance et 3) la perspective "internaliste" et la mise à l'écart de la langue externe (pour une critique méthodologique des pratiques de recherche dans ce cadre, voir ex. Wasow, 1986; Schutze, 1996; Cowart, 1997; Dąbrowska, 2010; Gibson & Fedorenko, 2010, 2013).

9. Il s'agit dans l'ordre chronologique de la Théorie Standard, la Théorie Standard Étendue, la Théorie du Gouvernement et du Liage et le Programme Minimaliste.

10. « *We must make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations)* » Chomsky (1965, p. 4).

Il faut noter que cette distinction reste encore à l'état d'idée et n'est pas dotée d'une définition explicite indépendante. La compétence n'est pas dissociable de la performance, dans la mesure où nous n'avons qu'un accès indirect à la connaissance linguistique et cela par le biais de la performance. Toute décision quant à la ligne de démarcation entre les deux sphères sera circulaire ou relèvera de la spéculation. Comment décider qu'un fait ou une observation langagière relève de la compétence et non de la performance et *vice versa*<sup>11</sup> ?

En outre, d'autres recherches, relevant des théories dites "basées sur l'usage" (*usage-based* : Kaplan & Bresnan, 1982; Langacker, 1987; Givón, 1995; Bybee & Hopper, 2001; Croft, 2001; Tomasello, 2003; Bybee, 2006; Bresnan, 2007), avancent que la connaissance linguistique d'un locuteur entretient un lien étroit avec son expérience linguistique, c'est-à-dire l'ensemble des instances de sa production et de sa compréhension de la parole, et qu'il y a des raisons de croire que l'usage façonne la connaissance linguistique. La fréquence de l'usage, à titre d'exemple, peut avoir un impact sur la représentation linguistique abstraite (Bybee, 2006). Cela signifie que la compétence est susceptible d'être affectée par la performance et que, par conséquent, ces deux sphères ne sont pas étanches.

Deuxièmement, la grammaire générative suppose l'existence d'une grammaire universelle et innée (*Universal Grammar*, UG) qui concrétise l'hypothèse de la faculté de langage (Chomsky & Lasnik, 1993). Cette grammaire innée permet à l'enfant d'apprendre parfaitement une langue naturelle, sa langue native en l'occurrence, à l'aide de son expérience linguistique propre, autrement dit les données linguistiques fournies par son environnement. Le résultat de cet apprentissage est une spécification de la grammaire universelle, la grammaire de la langue lambda, qui est intériorisée et partagée par tous les locuteurs de cette langue.

Cette grammaire universelle doit, d'un côté, être assez générale pour comprendre les traits universaux communs à toutes les langues du monde, et, de l'autre, assez flexible pour pouvoir saisir les variations existantes entre elles. Selon la théorie *principes et paramètres* (Chomsky & Lasnik, 1993), la répartition des tâches entre les deux grammaires se fait de la manière suivante : la grammaire universelle fournit les *principes* invariants, communs à toutes les langues et intrinsèques à la nature de la faculté de langage, et la grammaire spécifique de chaque langue détermine les valeurs des *paramètres* qui peuvent, à un certain degré, varier d'une langue à l'autre. Autrement dit, la grammaire spécifique résulte du croisement des principes

---

11. Voir par exemple le travail de Hofmeister & Sag (2010) sur les effets d'îlots qui met en évidence l'impact des facteurs relevant de la performance, c'est-à-dire les facteurs dont la manipulation influe le traitement (*processing*) d'une construction sans en modifier la structure, dans les jugements portés sur la grammaticalité de ces constructions.

de la grammaire universelle avec les valeurs des paramètres précisées localement<sup>12</sup>.

Il va de soi que les principes et les paramètres ne jouissent pas du même statut dans ce cadre théorique. Pour simplifier, l'on peut dire qu'il n'est pas intéressant pour la théorie d'expliquer ou de rendre compte des phénomènes locaux observés dans une langue donnée, si cela ne peut pas se faire par des principes généraux ou des paramètres applicables ou appliqués par ailleurs. Une hiérarchisation s'exerce en effet entre les phénomènes qui appartiennent au cœur de la compétence linguistique et les autres qui sont par conséquent considérés comme périphériques. Or, de même que pour la distinction compétence/performance, il y a lieu de s'interroger sur les conditions de cette hiérarchie, dans la mesure où celle-ci s'est initialement établie sur la base des données d'un nombre restreint de langues.

### 2.2.1.2 La place des données dans la démarche générativiste

Ce qui précède nous conduit à évoquer la méthodologie et le statut de données empiriques dans ce cadre théorique. La grammaire générative revendique une démarche scientifique conforme à la vision popperienne de la science fondée sur le critère de la falsifiabilité (Popper, 1963), c'est-à-dire une démarche hypothético-déductive dans laquelle la théorie précède les données. Ainsi, les données empiriques ne sont envisagées qu'en fonction de leur apport théorique<sup>13</sup>. Or, l'essentiel des hypothèses sont émises à partir d'observations faites sur quelques langues dotées de surcroît de caractéristiques similaires, tandis que les faits linguistiques sont innombrables et en partie encore méconnus. Nous ne mettons naturellement pas en cause cette démarche scientifique que nous partageons également, mais nous notons que, même dans une démarche déductive, la théorie initiale émerge d'une observation : plus cette observation est exhaustive, plus cette théorie est susceptible d'être validée par ailleurs – c'est-à-dire par les données collectées à cet effet.

Pour aborder la méthode de la collecte de données, penchons-nous sur la façon dont la grammaire est concrètement envisagée dans la théorie générative. La gram-

---

12. Pour donner un exemple concret, selon le Principe de la Projection Étendue (*Extended Projection Principle*, EPP) de la grammaire universelle, la position sujet est une position obligatoire de la proposition finie. Autrement dit, chaque verbe fléchi nécessite un sujet manifeste. Cette règle est satisfaite dans les langues à flexion verbale (relativement) pauvre telles que l'anglais et le français. Toutefois, dans beaucoup de langues (souvent à flexion verbale riche comme l'italien ou le persan) la présence d'un sujet grammatical explicite n'est pas obligatoire. Pour permettre à la grammaire d'appréhender ce fait, un paramètre, celui du sujet nul, *Null Subject Parameter* (Rizzi, 1989), a été introduit. Ainsi, les langues à sujet nul permettent la non réalisation du sujet. Notons cependant que, pour que le principe universel de l'EPP ne soit pas violé, il est assumé qu'un sujet nul, le *pro*, occupe la position du sujet grammatical.

13. « *Data were of interest only to the extent that they provided support for or evidence against some theoretical hypothesis.* » Wasow (2002, p. xi).

maire doit formaliser les règles qui permettent d'obtenir l'ensemble des énoncés bien formés d'une langue donnée, et rien d'autre. Cette définition se fonde sur une distinction catégorique, qui sépare les énoncés bien formés, dits grammaticaux, des énoncés mal formés ou agrammaticaux.

L'objectif fondamental de l'analyse linguistique d'une langue L est de séparer les suites *grammaticales* qui sont des phrases de L, des suites *agrammaticales* qui ne sont pas des phrases de L, et d'étudier la structure des suites grammaticales. [...] Un moyen de tester l'adéquation d'une grammaire proposée pour L est de déterminer si les suites qu'elle engendre sont réellement grammaticales ou non, c'est-à-dire acceptables par un locuteur indigène.

Chomsky (1957, p. 15)

Chomsky (1957, pp. 17-19) prend le soin de préciser que la notion de “grammatical” ne doit pas être confondue avec celle “doué de sens” – sa fameuse phrase *D'incolores idées vertes dorment furieusement* est devenue un classique au-delà même de la communauté des linguistes. La confusion avec des notions d’“approximation statistique” telles que “possible”, dans le sens de “peuvent se rencontrer”, doit être également soigneusement évitée, car le fait qu'une phrase n'est pas attestée – c'est-à-dire réellement produite – ne prouve en rien que celle-ci est agrammaticale (comme en témoignent, entre autres, les phrases grammaticales dépourvues de sens). Ce qui revient à dire qu'une preuve négative (dans le sens de l'agrammaticalité) ne peut se trouver dans un ensemble de phrases attestées, autrement dit dans un corpus, sachant que “l'ensemble des phrases grammaticales ne peut être assimilé à aucun corpus particulier d'énoncés, recueilli par le linguiste” (Chomsky, 1957, p. 17).

Ainsi, la seule méthode envisageable pour juger si une phrase est grammaticale ou agrammaticale est le recours à l'intuition du locuteur natif, autrement dit, le jugement de grammaticalité, opéré sur une base binaire absolue<sup>14</sup>.

Or, cette méthode peut être mise en question, car, d'un côté, les conditions de la collecte des jugements ne sont soumises à aucun protocole commun, ce qui peut compromettre leurs fiabilité (Schutze, 1996, pp. 52-53), et, de l'autre, les jugements portent sur des phrases construites et décontextualisées. L'absence de contextes appropriés pourrait favoriser le rejet excessif des phrases<sup>15</sup>. En effet, il semblerait

---

14. À ce propos, notons que le caractère binaire et catégorique de la notion de grammaticalité a été depuis remis en question par de nombreux chercheurs qui plaident pour une vision graduée de celle-ci (Keller, 2000; Manning, 2003; Sorace & Keller, 2005; Blache *et al.*, 2006).

15. En effet, certains travaux portant sur des thématiques variées signalent la divergence entre les jugements reportés dans les travaux linguistiques et les phrases attestées en corpus ou les jugements à partir des phrases contextualisés (ex. Miller & Hemforth, 2014). Cf. également nos contre-exemples dans la section 1.4 du précédent chapitre.

que la variabilité des jugements de grammaticalité, d'un locuteur à l'autre, d'une situation à l'autre ou d'une phrase à l'autre, ne soit pas prise en considération par les générativistes (Wasow, 2002, pp. 157-158).

De plus, dans la démarche générativiste, la source primaire (pour ne pas dire exclusive) de données empiriques utilisée par les linguistes est l'introspection, sachant que ceux-ci travaillent majoritairement sur leur(s) langue(s) native(s). Or, ces derniers ne sont pas des locuteurs comme les autres et il n'est pas acquis que la manipulation constante de leur objet d'étude ne compromette pas la naïveté de leurs jugements. À cet égard, la comparaison de Schutze (1996, p. 15) entre la linguistique et la parfumerie est éloquente<sup>16</sup>. De plus, il a été démontré que les jugements des linguistes ne coïncident pas toujours avec ceux des locuteurs naïfs (cf. Dąbrowska, 2010).

Par conséquent, l'on peut légitimement s'interroger sur la fiabilité des jugements de grammaticalité ainsi obtenue et les analyses qui en découlent<sup>17</sup>.

Pour conclure, la vision idéaliste de la langue entraîne la simplification des faits répertoriés par rapport à ce qui se produit réellement. En d'autres termes, les productions langagières qui divergent de ce qui est attendu par la grammaire – par exemple les phrases qui ne suivent pas l'ordre “canonique” – sont souvent mises à l'écart au profit des variantes les plus normées.

### 2.2.1.3 L'étude de la variation de l'ordre des mots dans le courant générativiste

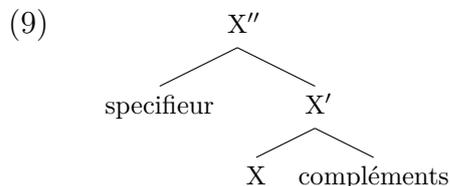
Au sein du courant générativiste, et en syntaxe post-structuraliste plus généralement, l'étude de la variation de l'ordre des mots remonte à la thèse de doctorat de Ross (1967), qui forge le terme de *scrambling* pour désigner une règle de transformation “stylistique”, autrement dit un “mouvement”.

La structure syntagmatique, une notion fondamentale dans la syntaxe générative, à partir desquelles les relations grammaticales sont définies, est étroitement liée avec l'ordre des mots dans une phrase canonique. La structure syntagmatique est une hiérarchie supposée universelle, communément représentée sous forme d'un schéma arborescent binaire. Tout syntagme est représenté par une tête ( $X^0$  ou  $X$  tout simplement), sa projection maximale ( $X''$ ) et les projections intermédiaires ( $X'$ ) –

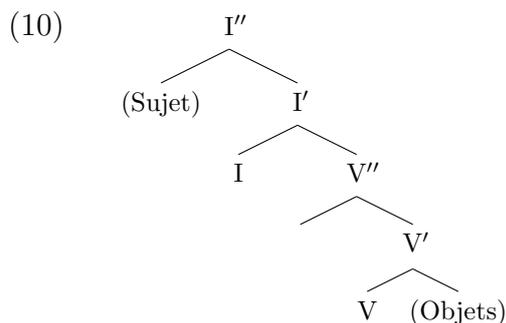
16. « *These experts [perfume smellers] might differ from naïve perfume smellers in the same ways that linguists differ from naïve sentence judges.* »

17. Notons néanmoins que certaines études comparatives montrent que les jugements de grammaticalités informels présentés dans une sélection de travaux publiés en syntaxe, convergent souvent avec ceux obtenus suivant une méthodologie quantitative (par ex. Sprouse & Almeida, 2012; Sprouse et al., 2013).

ce qui permet la récursivité. Cette projection comprend, pour tout syntagme, une position unique de “spécifieur” et une position récursive de “compléments”, voir en (9)<sup>18</sup>.



Les relations grammaticales entre les constituants d'une phrase se définissent selon les positions qu'ils occupent au sein de la projection maximale de la phrase, représentée généralement par la tête fonctionnelle ( $I^0$ , pour *Inflectional phrase*), car celle-ci est supposée porter les marques de la flexion verbale), que le schéma (10) illustre sous une forme simplifiée<sup>19</sup>. Comme le montre ce schéma, cette configuration suppose que le verbe et ses compléments d'objet forment un constituant, le syntagme verbal ( $V''$ ), d'où le sujet est exclu, impliquant une asymétrie (structurale) entre le sujet, qualifié d'argument externe, et le(s) objet(s), qualifiés d'arguments internes<sup>20</sup>.



Ainsi, cette configuration reflète un ordre linéaire précis, identique pour toutes les phrases, appelé ordre sous-jacent ou structure profonde. Les ordres linéaires effectifs, dits de surface, sont obtenus par “dérivation”, qui consiste en le “mouvement” d'un constituant de sa position de base vers une autre position en laissant une copie (la trace) dans la position initiale. Différents types de mouvements sont définis dans ce cadre par un ensemble de contraintes (ou règles) catégoriques à vocation universelle.

Cette théorie peut rendre compte des données des langues telles que l'anglais

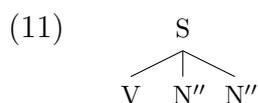
18. Notons que la direction du branchement, c'est-à-dire la position respective de la tête, est un paramètre, qui, en fonction des langues, peut se faire à droite (dans les langues à tête finale) ou à gauche (dans les langues à tête initiale).

19. Rappelons que cette théorie a connu de nombreux développements ultérieurs et que nous nous limitons ici à son socle fondamental.

20. En effet, le(s) complément(s) d'objet se mettent en position de “complément” du  $V''$ , qui, à son tour, occupe cette même position en  $I''$ , tandis que le sujet occupe la position de “spécifieur” de ce dernier. En termes techniques, le sujet c-commande le(s) objet(s).

ou le français, manifestant un ordre des mots rigide, dans lesquelles l'on peut dire que c'est l'ordre linéaire qui détermine les relations grammaticales et, par là, le sens vériconditionnel de la phrase (ex.  $AXB \neq BXA$ ). Or, dans les langues où l'ordre des constituants n'est pas contraint par la grammaire et ne détermine pas les relations grammaticales (ex.  $AXB \equiv BXA$ ), l'hypothèse d'une syntaxe des positions est problématique. Au début des années 1980, avec la prise en compte des données de telles langues, dites "à ordre libre", comme le japonais (Farmer, 1980) et le walpiri (Hale, 1981), la notion typologique de la "(non) configurationnalité" est alors forgée.

Les langues seront ainsi classées en deux catégories : configurationnelle lorsque les relations grammaticales sont définies par une syntaxe des positions, et non configurationnelle, ou à structure plate (*flat structure*), dans le cas contraire (Hale, 1981; Chomsky, 1981; Hale, 1983)<sup>21</sup>. Plus précisément, cette dernière désignation s'applique aux langues pour lesquelles l'on ne dispose pas de données empiriques pour postuler une hiérarchie entre les constituants de la phrase, notamment entre le sujet et l'objet. La phrase, dans ce cas de figure, est présentée par une structure syntaxique non hiérarchisée ou plate, ex. (11).



Il est usuel depuis de traiter les variations de l'ordre des mots (ou le *scrambling*<sup>22</sup>) dans des langues non configurationnelles comme différents ordres sous-jacents (*base-generated word orders*), c'est-à-dire sans faire intervenir un quelconque mouvement. Les variations de l'ordre dans des langues configurationnelles sont, en revanche, traitées comme dérivées d'un même ordre sous-jacent.

La question se pose alors de déterminer dans quel cas (ou dans quelle langue) les variations de l'ordre correspondent telles quelles à plusieurs ordres de base disponibles (approche par *base generation* ou non configurationnelle), et dans quel autre cas ces variations relèvent d'un mouvement à partir d'un ordre sous-jacent (approche par mouvement ou configurationnelle). Dans ce dernier cas, le mouvement peut être considéré comme optionnel et vide du point de vue sémantique ou comme

21. L'étiquette typologique "non configurationnel" inclut selon Hale (1983, p. 5) entre autres : 1) l'ordonnement libre des mots, 2) l'emploi de séquences syntaxiques discontinues et 3) l'emploi extensif de pronoms nuls. L'auteur précise dans une note de bas de page qu'il considère l'anglais et le walpiri comme des exemples canoniques de langues respectivement configurationnelle et non configurationnelle, tout en reconnaissant que beaucoup de langues peuvent se situer entre ces deux pôles.

22. Notons que, dans la littérature, le terme *scrambling* est utilisé parfois comme synonyme d'ordonnement libre des mots.

non optionnel ayant un impact sur l'interprétation sémantique de la phrase<sup>23</sup>.

Ainsi, l'étude de la variation de l'ordre au sein de ce courant se limite essentiellement à un débat théorique entre les tenants de ces différents points de vue. Cette perspective est manifeste dans les travaux sur la variation de l'ordre en persan, notamment ceux de Karimi (1994, 2003a, 2005), que nous avons présentés dans le chapitre précédent (section 1.4) et sur lesquels nous reviendrons dans la troisième partie (sections 6.2 et 6.3).

#### 2.2.1.4 Les théories génératives non transformationnelles

Les modèles génératifs en syntaxe ne se limitent pas aux modèles transformationnels (dits également dérivationnels ou chomskyens). Ceux-ci comprennent également des modèles non transformationnels, telles que LFG, *Lexical Functional Grammar* (Kaplan & Bresnan, 1982), GPSG, *Generalized Phrase Structure Grammar* (Gazdar, 1985) abandonné par la suite au profit de sa version plus récente HPSG, *Head-driven Phrase Structure Grammar* (Pollard & Sag, 1994)<sup>24</sup>. Contrairement aux modèles transformationnels, ces derniers ne postulent qu'un seul niveau de représentation syntaxique (représentation monostratale) et n'admettent aucune dérivation intermédiaire.

Pour rendre compte de la variation de l'ordre, ces modèles postulent une séparation entre les relations de précedence linéaire (ou l'ordre des mots) et les relations de dominance immédiate (ou la structure syntagmatique) (Gazdar & Pullum, 1981; Pullum, 1982; Falk, 1983; Reape, 1993, 1996). Ainsi, les phénomènes qui relèvent de la variation de l'ordre sont étudiés sous le thème général de la "linéarisation", englobant l'ensemble des contraintes qui détermine l'ordre linéaire des éléments linguistiques (sans distinguer entre mots et constituants). La linéarisation comprend donc les phénomènes tels que l'inversion (*inversion construction*), le *scrambling*, la fusion des syntagmes (*clause union*), l'extraction, etc.

Cette approche a une conséquence directe sur l'étude des préférences dans l'ordre linéaire, car, contrairement aux modèles génératifs dérivationnels, elle permet de concevoir les contraintes préférentielles susceptibles d'intervenir dans le choix de l'ordre linéaire indépendamment des règles catégoriques qui agissent au niveau de la constituance (autrement dit, les relations de dominance). Notons que les travaux précurseurs sur les contraintes préférentielles en jeu dans les phénomènes d'alternance de l'ordre dans le domaine postverbal en anglais, adoptant une approche

---

23. Pour une revue de la littérature, voir par ex. Grewendorf & Sternefeld (1990); Corver & van Riemsdijk (1994); Karimi (2003b); Sabel & Saito (2005).

24. Appelées également grammaires d'unification (pour une présentation de ces modèles on pourra se reporter à Abeillé, 2007).

cognitivo-fonctionnelle, nous sont livrés par des syntacticiens proches des théories HPSG et LFG, à savoir Thomas Wasow et Joan Bresnan et leurs collègues (Wasow, 1997, 2002; Wasow & Arnold, 2003; Arnold *et al.*, 2000, 2004; Bresnan *et al.*, 2001, 2007; Bresnan, 2007; Bresnan & Ford, 2010).

Avant de finir cette sous-section, il importe de mentionner le courant de la Théorie de l'Optimalité (Prince & Smolensky, 2008), qui émerge dans les années 1990, comme alternative à la vision de la Grammaire Universelle du modèle chomskyen (Chomsky, 1981) tout en demeurant dérivationnelle. L'idée fondatrice de cette théorie réside dans l'ordonnement local des contraintes universelles de bonne formation au détriment de la dichotomie *principes* et *paramètres*. Cet ordonnancement peut être envisagé de manière absolue, tel que pratiqué dans la version standard de la théorie, ou de manière probabiliste, d'où l'émergence de la théorie de l'optimalité stochastique (Boersma, 1997; Boersma & Hayes, 2001). Ce courant adopte également la dissociation des contraintes qui régissent la hiérarchie structurale de la phrase de celles qui interviennent sur l'ordre linéaire – y compris éventuellement les contraintes liées à la performance notamment au sein de la version probabiliste –, et, par là, offre également un contexte propice pour l'étude de la variation de l'ordre, dont la thèse de Choi (1996) en constitue un exemple.

## 2.2.2 La typologie linguistique

Nous pouvons caractériser le domaine de la typologie linguistique par l'intérêt porté aux tendances générales observées dans les données recueillies dans des langues variées, à la vision fonctionnelle et à la suffisance descriptive. Dans ce domaine, l'ordre des mots tient une place centrale dont témoigne l'ouvrage fondateur de Greenberg (1966), dans lequel 34 universaux de langage sont formulés en référence explicite à l'ordre des mots dans la phrase, ou en prenant les mots de l'auteur, à "l'ordre des éléments doués de sens" (*the order of meaningful elements*).

Dans ce cadre, la question primaire porte sur "l'ordre de base" (*basic word order*, Derbyshire & Pullum, 1981; Hawkins, 1983; Brody, 1984; Tomlin, 1986; Siewierska, 1988; Comrie, 1981; Whaley, 1997; Dryer, 2007) des constituants "majeurs", autrement dit l'ordre "neutre" dans lequel apparaissent le sujet, l'objet (direct) et le verbe, ayant un sémantisme – de préférence – prototypique des rôles grammaticaux attribués (agent, patient, action) dans une phrase simple et indicative<sup>25</sup>.

---

25. Voici pour en donner un exemple la définition avancée par Siewierska (1988, p. 8), largement reprise dans la littérature : « [...] *the order that occurs in stylistically neutral, independent, indicative clauses with full noun phrase (NP) participants, where the subject is definite, agentive and human, the object is a definite semantic patient, and the verb represents an action, not a state or an event.* » Notons que cette notion est quelque peu problématique et sa définition ne fait pas toujours

En effet, les travaux typologiques sur l'ordre sont consacrés en premier lieu à la distribution des types structuraux, c'est-à-dire les six classes de langues proposées initialement par Greenberg (1963) – ce classement évoluera dans les travaux ultérieurs (Lehmann, 1973; Vennemann, 1974; Hawkins, 1983; Tomlin, 1986; Dryer, 1992) –, et leurs universaux dits implicationnels (ou corrélationnels)<sup>26</sup>.

La question de la variation de l'ordre (à l'intérieur d'une même langue) et celle de l'ordre relatif entre des constituants non "majeurs", par exemple entre le COD et le COI, sera abordée plus tard et reste encore relativement marginale. Des typologues, tels que Talmy Givón, John A. Hawkins, Anne Siewierska, ont consacré une partie de leur travail à ces questions. Des théories unidimensionnelles, basées uniquement sur le poids grammatical (ex. Hawkins, 1990) ou sur la structure informationnelle (ex. Givón, 1983, 1988), ainsi que des théories multidimensionnelles (ex. Tomlin, 1986; Siewierska, 1993; Hawkins, 2004, 2008), sont proposées pour rendre compte de ces variations.

Givón (1983, 1988) propose des principes d'ordonnement en fonction seulement du statut informationnel des éléments, notamment "l'ordre de topicalité décroissante" qui suggère l'ordre thème - rhème (*topic - comment*) ou le principe selon lequel les locuteurs s'occupent en premier lieu des tâches les plus urgentes "*attend first to the most urgent task*" (Givón, 1988, p. 252). Un principe similaire connu sous la formulation "*given-before-new*" est également proposé dans la littérature (Gundel, 1988).

Hawkins (1990, 1994), quant à lui, propose une théorie des préférences de l'ordre linéaire motivée par l'analyse syntaxique (*parsing*), plus précisément de son efficacité face à la complexité syntaxique. Il propose, notamment, un principe, *Early Immediate Constituent* (EIC), basé sur la complexité relative des éléments, permettant de comparer, toute chose égale par ailleurs, deux ordres linéaires alternatifs. Nous reviendrons sur ce principe en §2.3.2.1. De plus, l'auteur compare son principe avec ceux basés sur les considérations pragmatiques à partir d'une étude de fréquence sur plusieurs langues typologiquement différentes (Hawkins, 1994, pp. 214-242) et conclut que le principe EIC est un meilleur prédicteur de l'ordre linéaire que les analyses concurrentes.

Siewierska (1993), comparant ces deux propositions, en conclut qu'aucun principe unidimensionnel — qu'il soit basé sur les facteurs structurels ou formels, tels

---

consensus (Mithun, 1992). L'ordre de base est également défini en association avec la fréquence d'usage et la notion de marque. Voir une discussion sur la neutralité pragmatique et l'ordre non marqué par Lambrecht (1994) en §3.2.2.2 au chapitre 3.

26. La revue de ces travaux dépasse le cadre de ce travail, voir Song (2012, ch. 2) et Song (2014, ch. 2) pour une synthèse.

que la longueur ou la complexité, ou sur les facteurs conceptuels ou informationnels – ne pourrait rendre compte de manière adéquate des préférences de l'ordonnement relatif des constituants. Elle propose différentes hiérarchies, regroupées en trois classes, susceptibles d'avoir un impact sur la linéarisation des dépendants verbaux, telles que la longueur, la complexité, la personne, le rôle sémantique, la familiarité discursive, etc.

Plus récemment, Hawkins (2004) a proposé une version améliorée de sa théorie qui n'est plus exclusivement basée sur la dépendance ou la distance structurale, mais qui inclut, par exemple, la dépendance sémantique comme une autre force susceptible de provoquer l'adjacence linéaire des constituants.

Les travaux de ces auteurs ont contribué à l'émergence des approches cognitivo-fonctionnelles. Dans les sous-sections suivantes, nous allons revenir sur une partie d'entre eux..

### 2.2.3 Les approches cognitivo-fonctionnelles

Le phénomène connu sous la désignation *Heavy NP Shift* (HNPS), et plus généralement les alternances de l'ordre dans le domaine postverbal en anglais, se trouvent à l'origine de la tournure méthodologique qui s'est produite en syntaxe théorique, dont la monographie de Wasow (2002) en est l'ouvrage emblématique.

Par une série d'études de corpus et d'expériences sur ces différents phénomènes (Wasow, 1997; Arnold *et al.*, 2000; Wasow & Arnold, 2003), l'auteur met en évidence le fait qu'une variété de facteurs intervient pour déterminer pourquoi "*English speakers produce postverbal phrases in one order, rather than another, when both are grammatical*" (Wasow, 2002, p. 164). Il argumente que la dichotomie compétence / performance conduit à la mise à l'écart de ce type d'études au détriment des généralisations convergentes qu'elles peuvent contribuer à formuler. De plus, notre accès à la compétence linguistique restant indirect et intimement lié à la performance, la décision, qui consiste à écarter certains phénomènes de l'objet d'étude de la linguistique théorique, serait arbitraire. Ainsi, sur la base de la littérature croissante des modèles basés sur l'usage, il conclut, sans adhérer à la vision entièrement probabiliste de la connaissance linguistique, que, jusqu'à preuve du contraire, il n'est pas judicieux d'exclure de cette dernière toutes considérations probabilistes.

Par la même occasion, Wasow (2002) revient sur les objectifs et la méthodologie du projet génératif, notamment en ce qui concerne le lien entre la théorie et les données, et attire l'attention sur le bien-fondé empirique des données dans une discipline qui se veut scientifique. Plus précisément, il met en cause les jugements de grammaticalité informels, réhabilite l'étude quantitative basée sur corpus et plaide

pour une approche méthodologique plurielle. Pour cet auteur, cette méthodologie consiste en : 1) le recours à l'introspection comme étape pilote ou lorsque les jugements sont suffisamment robustes et consensuels, 2) l'étude multifactorielle basée sur corpus et 3) la mise en œuvre d'expériences psycholinguistiques.

Une autre contribution notable dans cette perspective est celle de Bresnan et al. (2007) portant sur l'alternance dative en anglais, notamment sur le plan méthodologique. Les auteurs proposent une modélisation poussée en fonction d'un ensemble extensif de facteurs et de leur poids respectifs pour prédire l'alternance dative à travers des données de corpus. Ils évaluent ensuite leur modèle de prédiction par un questionnaire psycholinguistique. Les auteurs répondent ainsi aux critiques remettant en question la validité des généralisations faites à partir d'études de corpus.

Cette nouvelle approche méthodologique est rendue possible par les progrès accomplis dans les domaines externes. Premièrement, les domaines de la linguistique appliquée, notamment le traitement automatique des langues (TAL), progressent rapidement grâce aux avancées informatiques. Bien que le paradigme génératif ait stigmatisé la linguistique de corpus, ce dernier est resté une source importante pour la linguistique appliquée. La génération automatique des langues (naturelles) vise la simulation de la langue effective, autrement dit la performance, et pour cela elle doit s'attacher aux productions (réelles) de la langue et ne peut se limiter à la connaissance linguistique tacite. Linguistes et informaticiens collaborent de plus en plus et une nouvelle discipline émerge, la linguistique informatique. La linguistique de corpus a prospéré à partir de cette collaboration et s'est créée une place au sein de la linguistique théorique<sup>27</sup>. Des corpus de grande taille, avec des annotations riches, voient le jour non seulement pour le registre écrit mais aussi pour la parole spontanée, par exemple le corpus *Switchboard* (Godfrey et al., 1992). Ceci promeut les études quantitatives de plus en plus sophistiquées ayant recours aux méthodes statistiques poussées et déjà pratiquées dans les domaines tels que la phonétique, la sociolinguistique, la psycholinguistique, etc.

Deuxièmement, malgré la mise en écart de l'étude des facteurs extralinguistiques du projet de la recherche en linguistique théorique, celle-ci poursuit son cours dans les domaines externes en s'intéressant à l'architecture du système de la production et/ou l'analyse de la parole, telles que la psycholinguistique et l'acquisition du langage, ainsi que des domaines connexes tels que les sciences cognitives (Garrett, 1980; Bock, 1982; Bock & Warren, 1985; Frazier & Fodor, 1978; Hemforth et al., 1993; Bock & Levelt, 1994; Gibson, 2000; Konieczny, 2000; Kempen & Harbusch, 2003; Vasishth & Lewis, 2006; Vasishth, 2011; Tanaka et al., 2011b; MacDonald, 2013; Hemforth

---

27. Cf. la parution de la revue *Corpus Linguistics and Linguistics Theory* depuis 2005.

et al., 2014, entre autres). En effet, Chomsky, lui-même, a encouragé de tels projets de recherches :

Mais il est clair que l'utilisation du langage telle qu'on l'observe – la performance effective – ne reflète pas seulement les relations intrinsèques entre le son et le sens établi par le système des règles linguistiques. La performance implique aussi bien de nombreux autres facteurs [...]. Des considérations extralinguistiques, portant sur le locuteur et sur la situation, jouent un rôle fondamental lorsqu'il s'agit de déterminer comment le discours est produit, identifié et compris. La performance linguistique, en outre, est régie par des principes de la structure cognitive (par exemple, par les limitations de la mémoire) qui ne sont pas, à proprement parler, des aspects du langage.

Chomsky (1969, p. 125-126)

Dans cette perspective, de nombreux auteurs se sont intéressés à la variation de l'ordre (Osgood & Bock, 1977; Bock & Warren, 1985; McDonald et al., 1993; Gibson, 1998; Stallings et al., 1998; Prat-Sala & Branigan, 2000; Yamashita & Chang, 2001; Ferreira & Yoshita, 2003; Kempen & Harbusch, 2003, 2004; Vasishth, 2004; Kempen & Harbusch, 2005; Christianson & Ferreira, 2005; Branigan et al., 2008; Branigan & Feleki, 1999; Chang, 2009; Tily, 2010; Tanaka et al., 2011a; Stallings & MacDonald, 2011; Vasishth et al., 2012; Ros et al., 2015), un terrain propice qui peut apporter un éclairage sur le fonctionnement du système langagier, car, en évacuant les contraintes grammaticales spécifiques à chaque langue, il permet d'étudier les facteurs qui ont une portée universelle.

Deux hypothèses sont largement admises dans cette littérature pour rendre compte de la production de la parole<sup>28</sup>. Selon la première hypothèse, le processus de la production de la parole se déroule en plusieurs niveaux<sup>29</sup> (Garrett, 1980; Levelt, 1989; Bock & Levelt, 1994).

- le niveau du message (appelé également niveau conceptuel) fournit la matière première au niveau de l'encodage grammatical (*grammatical encoding*) – niveau qui nous intéresse ici – où le message est formulé verbalement, c'est-à-dire sous forme d'expression linguistique. Celui-ci fournit à son tour la matière première au niveau phonologique où l'articulation du message est program-

28. Les travaux susmentionnés concernent aussi bien le domaine de la production que celui du traitement et de la compréhension. Étant donné que les études empiriques menées dans le cadre de cette thèse se situent essentiellement dans le paradigme de la production, nous n'approfondirons pas ici les modèles portant sur la compréhension.

29. Cette hypothèse est notamment motivée par l'analyse des erreurs de la production spontanée de la parole.

mée.

- l'encodage grammatical se déroule sur deux plans. Premièrement, au niveau fonctionnel, la sélection lexicale et l'assignation des fonctions grammaticales ont lieu. Ensuite, le niveau positionnel est responsable de l'ordre linéaire – c'est-à-dire l'assemblage des constituants – et de la flexion.

Selon la seconde hypothèse, la production de la parole se fait de manière incrémentale, c'est-à-dire segment par segment (Kempen & Hoenkamp, 1987; De Smedt & Kempen, 1987; Levelt, 1989; Bock & Levelt, 1994; De Smedt, 1994), ce qui signifie que le passage par les différents niveaux n'est pas sériel, dans le sens où un niveau doit avoir terminé son traitement avant que le niveau suivant ne puisse débuter. La production incrémentale suppose, au contraire, que les traitements sur différents plans peuvent se dérouler simultanément, c'est-à-dire que les informations sont traitées par chaque niveau au fur et à mesure qu'elles deviennent disponibles. Ceci implique que : 1) les éléments plus facilement accessibles seront traités avant ceux qui le sont moins et 2) l'ordre dans lequel les informations sont livrées d'une composante à l'autre peut avoir un impact sur l'ordre dans lequel les éléments apparaissent dans la parole (Bock & Levelt, 1994, p. 949).

Dans cette approche, il est judicieux de supposer que, dans les cas où l'ordre des mots n'est pas contraint par la grammaire, la précedence linéaire observée dans la production reflète l'accessibilité (relative) des constituants au niveau notamment de l'encodage grammatical (Bock, 1982; Bock & Levelt, 1994; Kempen & Harbusch, 2003)<sup>30</sup>. Le *Heavy NP Shift* est un exemple par excellence de cette observation.

Tout les travaux convergent pour confirmer qu'en anglais, il y a une tendance à reporter les SN "lourds" vers la position finale, c'est-à-dire après le SP. Le terme "*Heavy NP Shift*" est forgé initialement dans la littérature générative (Kimball, 1973) pour désigner ce type de "déplacement", dont la première mention remonte à Ross (1967)<sup>31</sup>.

De prime abord, les auteurs rendent compte de ce phénomène en termes de traitement (*processing*) (Bever, 1970; Kimball, 1973; Frazier & Fodor, 1978; Hawkins, 1990; Gibson, 1998; Hawkins, 2004), dans la mesure où le HNPS facilite le traitement

---

30. « *The incremental nature of human grammatical encoding is an important source of word order variation even in languages with relatively strict word order. [...] constituents whose shape is determined at an earlier point in time will tend to precede constituents that 'arrive' later. And the correlation between the arrival time of a constituent (the moment its shape is fixated) and its linear position will tend to be stronger the greater the amount of flexibility. Arrival time may depend not only on syntactic factors such as the complexity of the syntactic assembly process but also on lexical factors (e.g., different retrieval times for low- and high-frequency words) and on semantic factors (e.g., salient fragments of the to-be-expressed meaning being conceptualized prior to less salient ones; see Yamashita & Chang (2001)).* » Kempen & Harbusch (2003, p. 141).

31. L'emploi du terme HNPS s'est ensuite généralisé, notamment grâce aux travaux de Wasow.

de la phrase et, de ce fait, il est judicieux de supposer que, dans la limite des règles de la grammaire, l'analyseur humain privilégie cet ordre.

Parallèlement, cette préférence s'explique de manière simple et concise par un modèle de production incrémentale sur la base de la hiérarchie d'accessibilité formelle. Les constituants plus longs ou plus complexes requièrent un traitement plus "coûteux" – en temps et/ou en mémoire – que les constituants plus "légers". Cette différence se traduit par un délai de traitement qui peut être compensé par le "report" des constituants "coûteux" dans le processus de l'encodage grammatical. Ceci se reflète dans l'ordre linéaire de la phrase produite (Wasow, 1997; Stallings et al., 1998; Arnold et al., 2000; Wasow, 2002; Kempen & Harbusch, 2003; Stallings & MacDonald, 2011).

Il convient de noter que ces deux approches ne sont pas contradictoires. À l'inverse, elles sont convergentes, dans la mesure où l'idée fondamentale des deux approches est l'optimisation des ressources mobilisées, et manifestent une certaine complémentarité<sup>32</sup>. Les difficultés pour la production – à condition que l'on adopte une vision incrémentale de celle-ci – ne sont pas différentes de celles rencontrées en compréhension. Elles proviennent dans les deux cas des éléments "coûteux" ou "complexes" (Gibson, 1998, p. 52).

Cet exposé rend compte du rapprochement entre les chercheurs qui poursuivent cette ligne de recherche au sein de la linguistique théorique et ceux qui s'intéressent à l'architecture du système de la production et la compréhension de la parole. Ce rapprochement a abouti à la prolifération des projets de recherche portant sur la variation de l'ordre des mots. Nous consacrerons la section suivante à la présentation des problématiques traitées dans ces travaux.

---

32. À ce propos, il faut noter que Wasow et ses collègues ont avancé un certain nombre d'arguments pour montrer qu'une analyse en termes purement de traitement ne suffit pas pour rendre compte du HNPS en anglais. Premièrement, à l'instar de Clark (1994, 1996), Arnold et al. (2000, pp. 46-48) utilisent comme indice les erreurs de la production spontanée de la parole, notamment les "disfluences" ou *fillers*. Observant, à partir d'une expérience de production, que le taux de HNPS est corrélé avec celui des disfluences, ils avancent que le HNPS facilite la planification et la production de la parole. Un deuxième argument porte sur le rapport collocationnel entre les verbes et ses compléments sur lequel nous viendrons plus loin (cf. §2.3.1.5). En bref, ils concluent que le HNPS sert l'intérêt du locuteur – allant dans le sens du point de vue de la production – plus que celui de l'interlocuteur – allant dans le sens du point de vue de l'analyse et compréhension (Wasow, 1997, pp. 95-97).

## 2.3 Les problématiques principales liées à la variation de l'ordre des mots

Différentes questions de recherche quant à la variation de l'ordre linéaire se trouvent au cœur de travaux qui s'inscrivent dans la lignée des approches cognitivo-fonctionnelles. Nous les présentons ici pour exposer les principales hypothèses proposées dans la littérature pour prédire les préférences dans l'ordre linéaire, avec comme ligne directrice l'ordre relatif entre les dépendants verbaux. Ces hypothèses concernent, en premier lieu, les différents facteurs identifiés et leurs interactions (§2.3.1). Comme la vision incrémentale de la production de la parole le suggère, ces facteurs relèvent principalement, mais non exclusivement, de l'accessibilité des constituants. En second lieu, il s'agit de comparer les langues SVO et SOV (§2.3.2). Ce questionnement est dû au fait que des tendances contredisant les hypothèses répandues, notamment quant à l'effet du poids, sont observées dans les langues SOV (Hawkins, 2008).

### 2.3.1 Les facteurs influant le choix de l'ordre linéaire

Tout d'abord, notons que l'accessibilité dans le processus de la production est à envisager sur deux dimensions. D'un côté, il y a les facteurs liés à la forme, généralement regroupés dans la littérature linguistique sous l'appellation "poids grammatical" et largement débattu dans cette littérature<sup>33</sup>. De l'autre côté, il y a les facteurs conceptuels.

Différentes propriétés sémantiques ou pragmatiques des constituants sont susceptibles de contribuer à leur accessibilité conceptuelle<sup>34</sup> : le caractère humain ou animé (Bock *et al.*, 1992; Ferreira, 1994; Kempen & Harbusch, 2004), la définitude (Grieve & Wales, 1973; Siewierska, 1993; Bader & Häussler, 2010), le statut informationnel (donné < nouveau ou thème < rhème) (Bock & Irwin, 1980; Givón, 1983, 1988; Gundel, 1988; Arnold *et al.*, 2000; Ferreira & Yoshita, 2003), et la saillance du référent (Osgood & Bock, 1977; Bock & Warren, 1985; Prat-Sala & Branigan, 2000). Plusieurs questions se posent sur la façon dont ces facteurs interviennent

---

33. Il est à noter que dans la littérature psycholinguistique, ces facteurs sont également désignés sous l'appellation facteur lexical. Ces derniers, parmi lesquels l'on peut mentionner la fréquence de l'usage et la complexité phonologique, concernent en particulier le mot, notamment dans la question de l'ordre dans une coordination (Bock, 1982, 1986; McDonald *et al.*, 1993).

34. Les deux dimensions, sémantique et pragmatique, de l'accessibilité conceptuelle rejoignent respectivement la classification accessibilité inhérente (ou fixe) vs. accessibilité dérivée (ou temporaire) (cf. Prat-Sala & Branigan, 2000; Branigan *et al.*, 2008). Le caractère animé, à titre d'exemple, est une propriété sémantico-lexicale inhérente d'une entité, tandis que le caractère donné est lié à la configuration discursive et est par conséquent temporaire.

dans le choix de l'ordonnement, ainsi que sur leurs interactions avec le poids grammatical.

Dans ce qui suit, nous présenterons brièvement tout d'abord le débat sur la notion de poids grammatical, notamment sa définition théorique et opératoire, ainsi que celui sur le statut informationnel, notamment en vue de clarifier certaines confusions terminologiques. Ensuite, nous nous attarderons sur les questions soulevées à propos des facteurs conceptuels notamment en psycholinguistique. Premièrement, nous discuterons de la question du rapport entre l'ordre linéaire et les fonctions grammaticales et, en second lieu, des convergences entre différents facteurs conceptuels ainsi qu'entre les facteurs conceptuels et le poids grammatical. Enfin, nous aborderons d'autres facteurs tels que biais lexical et le lien collocationnel ou idiomatique.

### 2.3.1.1 Le poids grammatical

La définition opératoire et théorique du poids grammatical est une question emblématique de la littérature. La mention d'un principe de la grammaire relative au poids remonte à Behaghel (1909), notamment sa fameuse "loi des constituants croissants" (*Gesetz der wachsenden Glieder*)<sup>35</sup>, fondée sur des observations portant sur le latin, le grec et l'allemand. Ce principe est généralement connu sous l'appellation *End weight* :

(12) Les syntagmes se présentent dans l'ordre croissant du poids<sup>36</sup>.

Wasow (2002, p. 3), à partir de Quirk et al. (1972)

Encore faudrait-il s'accorder sur la définition de la notion de poids. Elle peut être comprise de deux façons différentes : 1) la complexité structurelle (simple < complexe) ou 2) la longueur de la chaîne composant le constituant (court < long). La définition structurelle est sans surprise privilégiée dans les travaux relevant de la syntaxe générative<sup>37</sup>, où la complexité se définit en termes d'enchâssements. Un SN est ainsi considéré lourd lorsqu'il contient une proposition. À titre d'exemple, Chomsky (1975, p. 477) compare les deux phrases *they brought all the leaders of the riot in* et *they brought the man I saw in* et, jugeant la première plus naturelle que la seconde, souligne que ce n'est pas la longueur en nombre de mots qui est déter-

---

35. *Law of increasing constituents* en anglais, tiré de Bader & Häussler (2010). « *So bildet sich unbewußt in den Sprachen ein eigenartiges rhythmisches Gefühl, die Neigung, vom kürzeren zum längeren Glied überzugehen [...] was ich [...] als das Gesetz der wachsenden Glieder bezeichnen möchte.* » cité par Wasow (1997, p. 103).

36. « *Phrases are presented in order of increasing weight.* »

37. Rappelons que dans ce cadre un "mouvement" stylistique est prévu pour rendre compte du phénomène de HNPS (Ross, 1967; Kimball, 1973).

minante dans HNPS, mais la complexité. Dans le même ordre d'idée, des définitions en termes de nombre de syntagmes ou de nœuds compris dans le SN ont également été proposées.

Wasow (1997) présente une étude de corpus afin de comparer les différentes définitions disponibles afin d'arriver à une définition adéquate et opératoire. Il compare alors deux types de mesure du poids. D'un côté, trois mesures graduées, en nombre de mots, de nœuds ou de syntagmes, de l'autre côté, des mesures binaires, telles que la présence de propositions enchâssées, etc. Tout d'abord, il rejette les définitions basées sur les mesures binaires montrant qu'aucune d'entre elles n'a de capacité de prédiction adéquate, car aucune ne permet une approximation des conditions à la fois suffisantes et nécessaires du HNPS. La comparaison des trois mesures graduées révèle, en revanche, que celles-ci assurent toutes une prédiction satisfaisante et qu'il n'est pas possible de trancher en faveur de l'une ou d'une autre. Ceci est attendu du fait que ces mesures sont fortement corrélées. La conclusion essentielle de l'étude de Wasow est que le poids grammatical agit de manière relative et graduée et, par conséquent, la variable à prendre en compte dans les études est le poids relatif en tant que variable numérique et non binaire. Cette conclusion sera confirmée ultérieurement par une étude expérimentale qui adopte un paradigme de production de phrases (Stallings & MacDonald, 2011).

S'il n'est pas possible de prendre une décision empiriquement motivée quant à la définition théorique du poids, cela l'est d'un point de vue pratique. Une annotation en nombre de mots est nettement plus aisée à mettre en œuvre qu'une annotation avec les autres mesures évoquées, étant donné que celle-là se révèle au moins aussi efficace que les autres. Il est judicieux d'opter pour une définition opératoire du poids, en termes de nombre de mots, généralement désignée par la longueur<sup>38</sup>. Il convient d'insister sur le fait que la complexité et la longueur sont deux facteurs convergents bien que différents, qu'il n'est pas aisé de décorrélérer dans une étude de corpus, du fait même de cette convergence<sup>39</sup>.

---

38. Une mesure de la longueur en termes phonologiques, nombre de syllabes par exemple, n'a pas été envisagée. Il faut noter que celle-ci ne semble pas nécessairement pertinente, dans la mesure où, comme le font remarquer Stallings *et al.* (1998, p. 411), dans la vision modulaire et incrémentale de la production présentée plus haut (cf. §2.2.3), l'ordre linéaire est déterminé au niveau de l'encodage grammatical et non à celui de l'articulation. Ainsi, la longueur doit être comprise dans sa complexité syntaxique, conceptuelle ou prosodique, mais non dans sa complexité phonologique (voir également Hemforth *et al.*, 2013). Il faut néanmoins noter que, dans son étude basée sur corpus (*French Tree Bank*, Abeillé *et al.*, 2003), Thuilier identifie la longueur (absolue) en termes de nombre de syllabes comme un des facteurs dont dépend la position de l'adjectif épithète en français, dans la mesure où les adjectifs composés de plus de deux syllabes ont tendance à être postposés au nom, tandis que les adjectifs plus courts préfèrent l'antéposition (Thuilier, 2012, pp. 143–144).

39. Les travaux expérimentaux ultérieurs (Wasow & Arnold, 2003) ont montré que les deux facteurs sont susceptibles d'intervenir de manière indépendante dans le choix de l'ordre linéaire,

Le lien entre le poids grammatical et les facteurs tels que le statut informationnel, le caractère animé et la pronominalité du référent, constitue également une question de recherche, qui a attiré l'attention d'un bon nombre de chercheurs, sur laquelle nous reviendrons plus bas (§2.3.1.3)

### 2.3.1.2 Le statut informationnel

Un autre facteur largement débattu dans la littérature, au même titre que le poids grammatical, est le statut informationnel des constituants. Les travaux en linguistique théorique et en psycholinguistique convergent pour admettre la préférence “donné avant nouveau” à vocation universelle (Clark & Haviland, 1977; Bock & Irwin, 1980; Givón, 1983; Gundel, 1988). Cette préférence, comme celle relative au poids grammatical, est corroborée par la théorie de la production incrémentale. Un constituant, dont le référent a été précédemment mentionné dans le discours (appelé “donné” par commodité), est effectivement plus accessible qu'un constituant dont le référent n'a pas été mentionné auparavant (appelé “nouveau”) et, par conséquent, est susceptible de le précéder dans le processus de la production.

Cependant, comme un bon nombre d'auteurs le font remarquer, deux dimensions discursives distinctes semblent se confondre dans la littérature (ex. Halliday, 1985; Gundel, 1988; Vallduví, 1992). Il s'agit, d'un côté, de la dimension qui se rapporte à la nouveauté du référent dans le discours, *Referential Givenness* dans les termes de Gundel (1988), et, de l'autre, de la dimension qui relève de la configuration discursive de la phrase et qui se rapporte à l'importance relative de l'information apportée par l'élément<sup>40</sup>, *Relational Givenness* (Gundel, 1988). L'articulation thème (*topic*) vs. rhème (*comment*) a trait à cette dimension. Cette confusion s'explique, en effet, par la subtilité de la différence entre les deux dimensions. Le thème correspond généralement à une information donnée, tandis que le rhème correspond à une information nouvelle. Cette vision est toutefois quelque peu réductrice, puisqu'un bon nombre de cas de figure y échappe<sup>41</sup>.

Dans la littérature sur la variation de l'ordre, l'étude du statut informationnel des constituants se fait généralement sur la base de la dimension référentielle<sup>42</sup> sur une

---

bien que leurs effets convergent.

40. Appelée occasionnellement *newsworthiness* (par ex. Mithun, 1992).

41. Par ailleurs le problème de l'hétérogénéité terminologique présente dans ce domaine, notamment relative aux notions telles que thème, rhème ou focus, rend difficile l'appréciation des théories avancées par les uns et les autres (ex. Gundel, 1988; Lambrecht, 1994; Arnold, 1998; Ward & Prince, 1991). Comme le font remarquer Ward & Prince (1991), des analyses contradictoires rencontrées dans la littérature sont en partie dues à une terminologie malheureuse.

42. Notons qu'une annotation en termes de nouveauté ou de mention précédente dans le discours est, pratiquement parlant, plus facile à mettre en œuvre qu'une annotation en termes d'articulation discursive, beaucoup plus subtile.

échelle binaire (donné vs. nouveau), bien que des échelles graduées ont également été proposées. À titre d'exemple, Prince (1981) propose une échelle à trois degrés<sup>43</sup> en intercalant, entre les deux, le statut "inférable"<sup>44</sup>.

Notons avant de finir que, malgré un consensus dans la littérature portant sur donné < nouveau (et thème < rhème), l'ordre inverse est également suggéré par certains principes. À titre d'exemple, *First Things First Principle* (Gundel, 1988) ou *Attend first to the more urgent task* (Givón, 1988) suggèrent que le rhème (ou l'information nouvelle) soit positionné avant le thème (ou information ancienne). Cet ordre est également documenté pour certaines langues (ex. Payne, 1987; Mithun, 1992). Il s'agit notamment des langues à ordre libre où, selon les auteurs, l'ordre linéaire des constituants dépend purement des considérations pragmatiques sur la base de l'importance informative (*newsworthiness*) relative des constituants. Néanmoins, il convient de noter que les études sur la variation de l'ordre abordées dans ce chapitre se concentrent sur des langues à ordre rigide et examinent les préférences dans quelques cas de figure précis où ces langues autorisent la variation. En effet, il apparaît judicieux de considérer que les préférences de l'ordre quant au statut informationnel ne se manifestent pas de la même manière dans tous les cas de figure.

### 2.3.1.3 Le lien entre l'ordre linéaire et les fonctions grammaticales

Une problématique importante, largement évoquée dans la littérature en psycholinguistique, est la dissociation entre l'effet de l'accessibilité conceptuelle des référents sur l'assignation des fonctions grammaticales et sur l'ordre linéaire, en particulier dans une langue à ordre rigide comme l'anglais. Car les différentes hiérarchies d'accessibilité conceptuelle, que nous énumérons ci-dessous, sont corrélées. Pour en donner un exemple, le sujet est prototypiquement apparié avec le rôle sémantique d'agent qui est, à son tour, prototypiquement assigné à un référent placé haut dans la hiérarchie de personne. En effet, ces trois hiérarchies convergent pour favoriser la précédence linéaire d'un tel élément dans la phrase :

— la hiérarchie des rôles grammaticaux (Keenan & Comrie, 1977) :

(13) sujet < COD < COI

43. L'auteur distingue, par ailleurs, deux façons par lesquelles un référent peut être donné ou nouveau. La première se réfère à l'interlocuteur (*hearer-old* vs. *hearer-new*) – plus précisément la supposition du locuteur en ce qui concerne ce dernier – et la seconde au discours (*discourse-old* vs. *discourse-new*).

44. Dans leur étude basée sur corpus, Arnold et al. (2000) ont initialement utilisé cette échelle. Toutefois, ils ont été amenés à regrouper les deux catégories inférable et donné, faute de représentativité de la catégorie inférable dans leurs données.

— la hiérarchie des rôles sémantiques<sup>45</sup> (Hopper & Thompson, 1980) :

(14) agent < patient < destinataire < bénéfactif < instrumental...

— la hiérarchie de personne – appelée également hiérarchie référentielle (*referentiality hierarchy*), d'animéité<sup>46</sup> (*animacy hierarchy*), ou hiérarchie d'agentivité (*agentivity hierarchy*) (Silverstein, 1976; Dixon, 1979; Comrie, 1981; Croft, 2002)<sup>47</sup> :

(15) 1ère personne < 2ème personne < 3ème personne humaine < 3ème personne non humaine < nom propre < nom commun humain < nom commun non humain < non commun non humain animé < non commun non humain inanimé (matière non organique < abstrait)

Deux hypothèses sont envisagées quant à l'effet du caractère animé sur l'ordre linéaire : soit celui-ci est susceptible d'intervenir directement sur l'ordre linéaire, indépendamment de son impact sur l'assignation des rôles grammaticaux (Kempen & Harbusch, 2004), soit son impact sur l'ordre linéaire est indirect par le biais de son effet sur les rôles grammaticaux (McDonald et al., 1993; Stallings et al., 1998). À cet égard, l'étude des langues à ordre flexible, qui présentent de meilleurs terrains pour distinguer entre l'influence des différents facteurs, est à privilégier (Branigan & Feleki, 1999; Prat-Sala & Branigan, 2000; Kempen & Harbusch, 2004; Branigan et al., 2008; Tanaka et al., 2011a). L'ordre relatif dans une coordination ou l'alternance de voix sont néanmoins de possibles objets d'étude dans les langues à ordre fixe où l'ordre des constituants et les fonctions grammaticales sont interdépendants.

#### 2.3.1.4 La corrélation entre différents facteurs

Une autre convergence, traitée plus souvent dans la littérature linguistique qu'en psycholinguistique, porte sur le lien entre la hiérarchie de personne et les hiérarchies liées au statut informationnel du référent dans le discours. Ce dernier est en relation étroite avec la réalisation syntaxique du référent, notamment quant à sa référentialité, cf. (15), et/ou sa définitude dont l'échelle est donnée en (16).

45. Il faut noter que cette hiérarchie est loin d'être consensuelle (voir par ex. Levin & Rappaport Hovav, 2005).

46. La terminologie française n'est pas homogène. Beaucoup d'auteurs utilisent le "caractère animé" ou encore "animacité" comme équivalent de *animacy* de l'anglais. Nous optons pour cette terminologie à l'instar des travaux en linguistique générale (par ex. Lemaréchal, 1993).

47. La première mention d'une telle hiérarchie remonte à Silverstein (1976), Dixon (1979). Comme le fait remarquer Croft (2002, p. 130), trois dimensions y sont confondues : hiérarchie de personne, hiérarchie référentielle et hiérarchie d'animéité.

- (16) (référentiel) défini < (référentiel) indéfini (spécifique < non spécifique) < non référentiel (< générique)

Un référent donné, à titre d'exemple, requiert une détermination définie. Il peut être réalisé par un pronom personnel, un nom propre ou des syntagmes nominaux portant une réalisation définie (ex. déterminants démonstratif, possessif, défini, etc.). Un SN indéfini ne peut, en revanche, correspondre à un référent mentionné précédemment dans le discours.

Notons à ce propos que différentes hiérarchies relatives à la fois au statut discursif et à la forme du référent – qui se rejoignent souvent (Arnold, 1998) – sont proposées dans la littérature : *Topicality continuum* (Givón, 1983), *Familiarity scale* (Prince, 1981), *Accessibility hierarchy* (Ariel, 1988), *Givenness hierarchy* (Gundel et al., 1993), *Activation states* (Chafe, 1994).

Le cas des pronoms personnels, placés haut sur les deux types de hiérarchie (c'est-à-dire la hiérarchie relative au statut informationnel et la hiérarchie de personne), est particulièrement éloquent. Les pronoms personnels sont par définition référentiels et définis, donc donnés – dans la mesure où leur référent est déjà introduit dans le discours (pronoms anaphoriques) ou qu'il est connu par le contexte (pronoms déictiques). Par ailleurs, sur le plan formel, ces éléments sont prototypiquement légers et donc hautement accessibles. Ainsi, différents facteurs contribuent à leur procurer une accessibilité élevée.

En outre, il est plus généralement suggéré que la définitude et l'animéité sont tous deux fortement corrélées avec la complexité formelle. Les référents donnés / définis / humains ont tendance à être plus légers que les référents nouveaux / indéfinis / inanimés (Hawkins, 1994; Arnold, 1998). Il s'agira alors de savoir si ces propriétés convergentes sont réductibles ou non à un seul facteur, notamment au poids grammatical comme cela a été suggéré par Hawkins (1994).

Cette question oppose les partisans des théories unidimensionnelles aux tenants des approches multidimensionnelles. Il faut noter qu'aujourd'hui, grâce aux études quantitatives poussées basées sur corpus et/ou expérimentales, la dernière approche prend le dessus. À titre d'exemples, Arnold et al. (2000) montrent que le poids grammatical et le statut informationnel, en dépit de leur forte corrélation, jouent chacun un rôle (indépendant) dans le choix de l'ordre entre les compléments. Gries (2003) et Bresnan et al. (2007) montrent, à leur tour, que, contrairement à ce qui est suggéré dans la littérature, le caractère pronominal joue un rôle indépendant dans l'alternance dative en anglais qui ne peut être réduit au statut informationnel ou au poids. Quant au caractère animé vis-à-vis du poids, Rosenbach (2005) montre que les deux jouent un rôle significatif dans l'alternance génitive en anglais, de même

que Bresnan et al. (2007) pour l'alternance dative.

De nombreux auteurs s'inscrivant dans cette lignée poursuivent des recherches similaires. À titre d'exemple, Bader & Häussler (2010) proposent une étude de corpus, ayant trait à l'ordre relatif entre le sujet et l'objet dans différentes types de propositions en allemand, afin d'examiner les hypothèses existantes à partir d'une annotation riche d'un ensemble extensif de facteurs à l'instar de Siewierska (1993). Thuilier (2012) étudie l'ordre relatif entre les compléments verbaux du français en adoptant la méthodologie utilisée par Bresnan et al. (2007) pour l'étude de l'alternance dative de l'anglais. Son étude a montré un impact significatif du poids relatif tel qu'il est attendu dans une langue SVO (c'est-à-dire court-avant-long), mais pas du caractère animé, ni du caractère défini.

### 2.3.1.5 Autres facteurs

Les facteurs évoqués dans la littérature pour rendre compte des préférences de l'ordre linéaire ne se limitent pas à ceux qui traitent de l'accessibilité. Des facteurs supplémentaires susceptibles d'intervenir, tels que le biais lexical, la dépendance sémantique, l'amorçage syntaxique (*syntactic priming*) ou encore l'ambiguïté structurelle, sont également étudiés.

**Biais lexical.** Le fait que le choix du verbe intervient dans l'ordonnement linéaire des compléments est une observation largement documenté dans les phénomènes d'alternance pour l'anglais, notamment quant à l'alternance dative (Ferreira, 1994, 1996; Wasow, 1997; Stallings et al., 1998; Bresnan et al., 2007)<sup>48</sup>. Ceci est également confirmé par des travaux sur d'autres langues comme le français (Thuilier, 2012).

L'impact du lemme verbal peut être envisagé sous deux angles différents. D'un côté, il peut être expliqué par des propriétés syntaxiques du verbe et de ses arguments, autrement dit son cadre de sous-catégorisation, et, de l'autre, en termes de fréquence de l'usage. La première approche est soutenue par Wasow (1997) et la seconde par Stallings et al. (1998) qui se situe dans la littérature psycholinguistique. Dans les deux approches, l'idée de base repose sur le fait que des informations statistiques, émanant de l'usage effectif de chaque verbe, sont intégrées dans le savoir linguistique des locuteurs et interviennent lors de la production de phrases sous forme d'un biais envers une structure particulière (Wasow, 2002, p. 88). Ceci n'est

---

48. À titre d'exemple, dans un échantillon extrait du *New York Times*, les phrases avec le verbe *tell* et *give* manifestent l'alternance dative dans 100% et 84% des cas respectivement, tandis que cette fréquence s'élève seulement à 40% pour le verbe *bring* ou encore 20% pour *send* (Wasow, 2002, p. 87). Des observations similaires ont été faites pour le HNPS.

pas une idée isolée en littérature psycholinguistique, notamment en ce qui concerne la compréhension (MacDonald, 1994; Tanenhaus & Trueswell, 1995; Seidenberg & MacDonald, 1999).

Wasow (1997) avance que les verbes disposent de cadres de sous-catégorisation spécifiques qui déterminent les structures syntaxiques dans lesquelles ils peuvent apparaître. Par exemple, un verbe comme *bring* requiert un complément nominal dans tous ces cadres de sous-catégorisation, tandis que pour un verbe comme *write* un complément nominal est en option. Les locuteurs de l'anglais ayant intégré ces informations ont, pour chaque verbe, des attentes différentes quant à la suite de la phrase en fonction des éléments disponibles<sup>49</sup>. Celles-ci interviennent directement dans le choix de l'ordre linéaire, ce qui se traduit par le biais lexical du verbe envers un ordre particulier<sup>50</sup>.

Stallings et al. (1998) expliquent au contraire le biais lexical du verbe par la fréquence de l'usage. Plus précisément, les auteurs avancent que les verbes portent en eux une information sur l'historique des structures dans lesquelles ils apparaissent. Cet historique influence leurs comportements futurs. À titre d'exemple, les auteurs soulignent que les verbes, qui sélectionnent un complément phrastique, manifestent un taux plus élevé de HNPS. Ce constat peut s'expliquer par le fait que le complément phrastique étant obligatoirement postposé, ces verbes se trouvent fréquemment dans les structures où ils sont suivis d'un complément prépositionnel. Cette fréquence est enregistrée dans leur historique et se traduit par un biais général, en faveur de l'ordre SP-SN, se manifestant même dans leur emploi avec des compléments non phrastiques.

N'omettons pas de mentionner l'impact du lemme verbal sur le plan sémantique, notamment sur le sémantisme de ces compléments – l'appariement entre les rôles sémantiques et les fonctions grammaticales. En effet, la façon dont chaque verbe sous-catégorise ses arguments a un impact sur la position respective de ces compléments sur la hiérarchie d'accessibilité. Celle-ci joue à son tour un rôle dans le choix de l'ordre linéaire (voir ex. Ferreira, 1994, 1996). Or, déterminer l'aspect purement sémantique de l'impact du biais lexical reste une question ouverte. Notons néanmoins que Bresnan et al. (2007) ont montré, à l'aide de modélisations statistiques, que l'effet du lemme verbal perdure au-delà des effets relatifs à l'accessibilité, notamment le caractère animé et le statut informationnel des référents.

---

49. Par exemple, un syntagme prépositionnel est toujours attendu après un verbe comme *bring*, contrairement à un verbe comme *write*.

50. L'auteur envisage l'impact de ces différences sur le HNPS en termes d'engagement précoce vs. engagement tardif, qui peut être expliqué du point de vue de la production ou celui de la compréhension. Or, les données recueillies pour tester les deux possibilités soutiennent seulement les prédictions allant dans le sens de la production.

**Dépendance sémantique.** L'idée selon laquelle le lien sémantique ou collocationnel entre les constituants, en l'occurrence entre le verbe et ses dépendants, favorise leur adjacence remonte encore une fois à Behaghel (1932) et demeure consensuelle dans la littérature. Empruntant les mots de Hawkins, nous désignons ce lien par "dépendance sémantique"<sup>51</sup>. Effectivement, il est judicieux de postuler que cette dépendance entre les constituants entraîne leur proximité linéaire. Cette intuition est formulée en termes simples par Behaghel (1932, p. 4) : "ce qui est mentalement lié et est également positionné à proximité" (adapté de Hawkins, 2001, p. 1)<sup>52</sup>. De nombreux travaux empiriques basés sur corpus ont depuis confirmé l'intuition de Behaghel, notamment dans différents phénomènes d'alternances en anglais (Wasow, 1997; Hawkins, 2000, 2004; Lohse et al., 2004).

Wasow (1997) propose une annotation (intuitive) en trois catégories pour les séquences V-SP. Tout d'abord, il sépare les séquences où le SP entretient un rapport collocationnel avec le verbe ; les autres sont considérées alors "non collocationnelles". Ensuite, il oppose les séquences collocationnelles "transparentes", telles que *bring to an end*, aux séquences collocationnelles "opaques", telles que *take into account*. Les taux respectifs de HNPS dans ces trois cas de figure sont : 26%, 47% et 60%. La différence entre ces taux démontre clairement un effet croissant du rapport collocationnel. Plus le lien sémantique est étroit et plus le SP a tendance à apparaître adjacent au verbe.

Pour définir la dépendance sémantique entre un verbe et son complément prépositionnel, Hawkins (2000) propose un test basé sur la relation d'implication. Plus précisément, il postule qu'un verbe entretient un lien de dépendance sémantique avec un SP si l'omission de ce dernier de la phrase initiale résulte dans une phrase non entraînée par la phrase initiale.

Par exemple, la phrase *John waited for Mary at noon* entraîne *John waited*, mais la phrase *John accounted for this fact in his book* n'entraîne ni *John accounted in his book*, ni *John accounted*, mais seulement *John accounted for this fact*. Ainsi, dans le premier cas, le verbe et les deux compléments sont indépendants, tandis

---

51. Hawkins définit la relation de "dépendance" comme suit (2004, p. 226) : « *Two categories are in a relation of dependency iff syntactic or semantic properties of B require access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified or ambiguously or polysemously specified.* » Dans le modèle proposé par l'auteur, les relations de dépendance et/ou combinatoires s'appliquent dans deux domaines distincts : le domaine du syntagme (*Phrase Combination Domain*) s'occupe de la reconnaissance des constituants immédiats, c'est-à-dire les dépendances syntaxiques, et le domaine lexical (*Lexical Domain*) rend compte des liens sémantiques. Un principe commun (*Minimize Domains*) rend compte du fait que, dans chacun de ces domaines, l'ordre, qui minimise les distances où les liens de dépendance s'exercent, optimise le traitement et sera, par conséquent, préféré.

52. « [...] *das geistig eng Zusammengehörige auch eng zusammengestellt wird.* »

que, dans le deuxième cas, le verbe dépend du SP *for this fact*. Les données de corpus étudiées par l'auteur permettent de mettre en évidence que ce lien sémantique favorise l'adjacence.

Lohse *et al.* (2004) appliquent cette définition à la relation verbe-particule<sup>53</sup> pour étudier l'effet de la dépendance sémantique sur l'ordre linéaire dans ces constructions et parviennent à un résultat similaire à celui de Hawkins (2000).

Plus particulièrement, ces travaux montrent, par le biais des analyses statistiques, que l'effet de la dépendance sémantique sur l'ordre linéaire, bien que moins important, perdure au-delà de l'effet du poids grammatical.

**Amorçage syntaxique.** L'amorçage syntaxique ou structurel se définit par la réutilisation de la structure syntaxique d'une phrase amorce, c'est-à-dire la phrase qu'un locuteur vient de percevoir ou de produire (Bock, 1986; Pickering & Branigan, 1999; Gries, 2005)<sup>54</sup>. Différentes études expérimentales ou basées sur corpus signalent que cet effet se manifeste également par rapport à l'ordre linéaire (Hartsuiker *et al.*, 1999; Hartsuiker & Westenberg, 2000). Ainsi, il est possible que le choix de l'ordre dans une phrase étudiée soit influencée par la phrase précédente. Dès lors, il importe de prendre cet effet en compte dans les études sur l'ordre linéaire, car celui-ci pourrait biaiser les résultats ou fournir une explication quand l'ordre observé n'est pas l'ordre prédit par le modèle.

**Ambiguïté de rattachement.** L'ambiguïté est un des sujets largement étudié dans la littérature en psycholinguistique. Les phrases dites *garden path* ont notamment fait couler beaucoup d'encre (Frazier, 1987; Frazier & Rayner, 1982; Trueswell & Tanenhaus, 1994). Il a été avancé que les phrases ambiguës peuvent présenter une charge supplémentaire pour le traitement, notamment dans le cas où elles doivent faire l'objet de réanalyse.

L'ordre linéaire est une source d'ambiguïté structurelle. La phrase *La femme a frappé l'homme avec un parapluie* est un classique des cours de syntaxe pour illustrer l'ambiguïté liée au rattachement. Cette ambiguïté disparaît, en revanche, dans l'ordre SP - SN des compléments, ex. (17).

(17) a. La femme a frappé [<sub>SN</sub> l'homme qui se promenait] [<sub>SP</sub> avec un parapluie].

53. Pour donner un exemple illustrant cette définition par rapport à cette relation, *to lift up* entraîne *to lift* et *up*, mais *to look up* n'entraîne ni *to look*, ni *up*.

54. Notons que l'amorçage syntaxique est un des aspects de l'amorçage, qui consiste plus généralement en la tendance des locuteurs à reprendre dans leur production des éléments de ce qui a été produit immédiatement avant et qui peut se manifester au niveau sémantique, lexical, morphologique et syntaxique.

- b. La femme a frappé [<sub>SP</sub> avec un parapluie] [<sub>SN</sub> l'homme qui se promenait].

Il a été suggéré que le fait d'éviter l'ambiguïté structurelle peut constituer une motivation dans le choix de l'ordre linéaire. Ce dernier peut effectivement être choisi de façon à faciliter la compréhension de l'interlocuteur. Or, un tel effet n'a pas été corroboré par des études empiriques (Wasow, 2002; Wasow & Arnold, 2003), dont les résultats ne sont pas compatibles avec la prise en compte de l'intérêt de l'interlocuteur par le locuteur lors de la production.

### 2.3.2 Langues SVO *vs.* Langues SOV

Les divergences et les similitudes observées entre les langues SVO et SOV constituent une question de recherche centrale, notamment quant à l'universalité des préférences. Celles-ci, telles que formulées dans la majorité des travaux, sont censées s'appliquer de manière universelle, car elles découleraient des principes cognitifs universaux qui sous-tendent le système de la production et/ou du traitement de la parole. Or, les travaux initiaux sur ces préférences portaient essentiellement sur l'anglais et l'allemand.

En effet, certaines préférences, telles que “animé-avant-inanimé” ou “donné-avant-nouveau”, s'observent également dans des langues SOV comme le japonais ou le hindi. Or, en ce qui concerne l'effet du poids grammatical, les données du japonais (Dryer, 1980; Hawkins, 1994; Yamashita & Chang, 2001) et du coréen (Choi, 2007) ont remis en question l'hypothèse d'une préférence universelle “court-avant-long”.

La préférence inverse “long-avant-court”, “l'image miroir” selon les termes de Hawkins (1994), est relevée dans les données de corpus pour le japonais et pour le coréen (Hawkins, 1994; Choi, 2007, respectivement), ainsi que, pour le japonais, par des expériences psycholinguistiques en paradigmes de production de phrase ou de jugement d'acceptabilité (Yamashita & Chang, 2001; Yamashita, 2002; Long, 2010). Plus récemment, Ros *et al.* (2015) ont rapporté cette préférence pour la langue basque sur la base des expériences menées dans le même paradigme que celles de Yamashita & Chang (2001).

Deux types d'analyses sont disponibles pour en rendre compte. Le premier adopte le point de vue du traitement (*human language processing*) et de la compréhension (Hawkins, 1990, 2004, 2008; Gibson, 1998, 2000; Wasow, 2013). Le second privilégie le point de vue de la production (Yamashita & Chang, 2001, 2006; Chang, 2009).

Hawkins (1994, 2004) et Gibson (1998, 2000) avancent que les préférences entre différentes alternatives peuvent être expliquées du point de vue du traitement sur la base de la distance entre une tête et ses dépendants. Pour simplifier, les auteurs

proposent différentes métriques pour déterminer, toutes choses égales par ailleurs, quelle alternative requiert le traitement le moins “coûteux”. Étant donné que les modèles proposés tiennent compte des relations de la dépendance et non des distances en tant que telles, la position de la tête y acquiert une place déterminante. Ainsi, ces modèles prédisent correctement la préférence “long-avant-court” observée dans les langues à tête finale. Mentionnons aussi l’hypothèse de Wasow (2013), qui tend à expliquer la préférence “long-avant-court” dans les langues SOV, notamment lorsque le constituant long comporte une ou des enchâssées, par une propension à éviter les constituants complexes au milieu (c’est-à-dire *center-embedded structures*), difficiles et “coûteux” à analyser.

Quant à Yamashita & Chang (2001), ils proposent de rendre compte de la préférence “long-avant-court” dans les langues SOV par un même modèle de production prédisant la préférence inverse dans les langues SVO, sur la base de l’hypothèse de l’accessibilité conceptuelle des constituants longs.

Par la suite, nous nous attardons tout d’abord sur le principe *Early Immediate Constituent* de Hawkins (1994, 2004), ensuite sur le modèle *Dependency Locality Theory* de Gibson (1998, 2000), et enfin sur l’hypothèse de l’accessibilité conceptuelle de Yamashita & Chang (2001).

### 2.3.2.1 Le principe *Early Immediate Constituent* de Hawkins

Sur la base d’une étude extensive des langues typologiquement différentes, Hawkins (1990, 1994) propose une théorie de la performance visant l’efficacité d’analyse face à la complexité des structures. Selon cette théorie, l’analyseur humain (*human parser*) préfère, toutes choses égales par ailleurs, un ordre linéaire qui optimise le traitement, ce qui se traduit par la reconnaissance la plus rapide des constituants immédiats.

L’auteur prévoit un mécanisme de calcul permettant une comparaison chiffrée entre deux alternatives. À cet effet, il postule notamment le principe de *Early Immediate Constituent* (EIC) visant la minimisation des distances entre une tête (le verbe en l’occurrence) et ses dépendants. Ce principe opère de la façon suivante. Tout d’abord, il faut définir le domaine de la reconnaissance des constituants, *Constituent Recognition Domain* (CRD), du syntagme<sup>55</sup>. Celui-ci correspond à l’ensemble des

---

55. Dans des travaux ultérieurs, l’auteur apporte quelques modifications à sa théorie initiale (Hawkins, 2004). Le principe *EIC* est remplacé par le principe *Minimize Domains* (MiD), qui résume *EIC* et *CRD*. Ce principe, à la différence de la théorie originale, comprend en plus du domaine syntagmatique, *Phrase Combination Domain* (PCD), un domaine lexical afin de rendre compte de l’impact de la dépendance sémantique (Hawkins, 2001). Cette modification a été opérée, par ailleurs, dans le but de rendre le modèle compatible avec la production. Toutefois, excepté la prise en compte de la dépendance sémantique, ce modèle demeure essentiellement identique au

nœuds terminaux et non terminaux qu'il faut traiter afin de reconnaître le syntagme et ses constituants immédiats (*Immediate Constituents*, IC). Cette séquence se déroule, de gauche à droite, depuis le nœud terminal qui permet d'identifier le premier constituant immédiat jusqu'au nœud terminal du dernier constituant. Notons que celle-ci est la séquence minimale qui permet d'obtenir la construction d'un syntagme – sa tête et ces constituants immédiats – sans pour autant traiter tous les mots qui les construisent.

(18) I [<sub>VP</sub> introduced [<sub>NP</sub> some friends that John had brought] [<sub>PP</sub> to Mary]]

À titre d'exemple, pour un syntagme verbal ditransitif de l'anglais, ex. (18), ce domaine s'étend du verbe *introduced* jusqu'à la préposition *to*. Le nombre de nœuds à traiter, voir en (19), est un indice de la rapidité par laquelle la phrase peut être traitée et permet d'ores et déjà la comparaison chiffrée des deux alternatives. Le traitement optimal se produit lorsqu'un maximum d'IC, en l'occurrence 3 IC, est reconnu avec un minimum de nœuds terminaux, en d'autres termes, lorsque le rapport entre le nombre de mots parcourus et le nombre des IC reconnus, *IC-to-Word*, est maximal.

Dans les exemples ci-dessous, c'est la phrase (b) qui obtient le score le plus élevé (3/4 vs. 3/11). Ainsi, le modèle prédit correctement le HNPS ou la préférence "court-avant-long".

(19) a. I [<sub>VP</sub> introduced [<sub>NP</sub> some friends that John had brought to the  
           1                  2      3          4      5      6      7          8  9  
           party] [<sub>PP</sub> to Mary]]  
           10          11

b. I [<sub>VP</sub> introduced [<sub>PP</sub> to Mary] [<sub>NP</sub> some friends that John had  
           1                  2  3          4  
           brought to the party]]

Hawkins (1990, p. 233)

L'auteur propose un second calcul plus complexe pour simuler l'efficacité du traitement en temps réel (*on-line*) de gauche à droite, appelé ultérieurement *Maximize On-line Processing* (MaOP) (Hawkins, 2004). Pour ce faire, le rapport *IC-to-Word* est calculé pour chaque mot de la séquence CRD. La moyenne de toutes ses fractions, exprimée sous forme d'un taux, indique le rapport *IC-to-Word* pour le syntagme et constitue un second indice pour comparer les alternatives, en particulier lorsque le principe *EIC* ne permet pas de trancher en faveur d'une entre elles. Les scores

---

modèle initial, notamment en ce qui concerne l'effet du poids grammatical. Dans la littérature, la formulation originale, le principe *EIC*, continue à prévaloir.

pour les phrases en (20), sont respectivement 46,6%<sup>56</sup> et 85,4%<sup>57</sup>. Cette différence soutient encore une fois la préférence “court-avant-long” pour l'anglais.

- (20) a. I [VP introduced [NP some friends that John had brought to the  
           1/1                   2/2 2/3    2/4 2/5 2/6 2/7    2/8 2/9  
           party] [PP to Mary]]  
           2/10       3/11
- b. I [VP introduced [PP to Mary] [NP some friends that John had  
           1/1                   2/2 2/3           3/4  
           brought to the party]]

Hawkins (1990, p. 235)

De la même manière, dans une langue (strictement) tête finale comme le japonais, l'ordre préconisé par ce principe est celui qui réduit la distance entre la tête et ses dépendants. À titre d'exemple, ce principe prédit correctement la préférence pour la préposition des compléments phrastiques en japonais, documentée par Dryer (1980, p. 134), comme illustré par la paire de phrases en (22), 3/6 vs. 3/3. L'auteur étudie le placement des compléments phrastiques dans un certain nombre de langues SOV, y compris le persan dans lequel, comme nous l'avons vu dans le chapitre précédent, ceux-ci se positionnent obligatoirement après le verbe, ce qui est compatible avec les prédictions du principe *EIC* (Hawkins, 1990, p. 248).

- (21) [IP [NP Mary-ga] [VP [CP kinoo John-ga kekkonsi-ta to]] it-ta]  
       Mary                   hier John s'est marié que       dit  
       'Mary a dit que John s'est marié hier.'

- (22) a. [IP [NP Mary-ga] [VP [CP kinoo John-ga kekkonsi-ta to]] it-ta]  
           1                   2    3           4           5    6
- b. [IP [CP kinoo John-ga kekkonsi-ta to] [NP Mary-ga] [VP it-ta]]  
   1           2           3

Adapté de Hawkins (1990, p. 249)

Quant à l'ordre relatif entre les compléments, le modèle, comme illustré en (24), 3/5 vs. 3/3, prédit correctement la préférence “long-avant-court”, confirmée pour le japonais à partir des données du corps (Hawkins, 1994), ainsi que dans des expériences psycholinguistiques (Yamashita & Chang, 2001; Yamashita, 2002; Long, 2010).

- (23) watashi-wa [VP [PP Mary-ni] [NP Jon-ga tsureteki-ta tomodachi-o]]  
       Moi                   Mary           John emmena ami

56.  $(\frac{1}{1} + \frac{2}{2} + \frac{2}{3} + \frac{2}{4} + \frac{2}{5} + \frac{2}{6} + \frac{2}{7} + \frac{2}{8} + \frac{2}{9} + \frac{2}{10} + \frac{3}{11})/11$

57.  $(\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + \frac{4}{4})/4$



- (25) a. The landlord talked [PP to the visitor who had just arrived] [PP about the problem]
- b. The landlord talked [PP about the problem] [PP to the visitor who had just arrived]

Adapté de Gibson (1998, p. 51)

La phrase (25-a) ci-dessous est difficile à analyser et à comprendre à cause de la longue distance qui sépare la préposition *about*, la tête du second complément prépositionnel, et le verbe *talked*, c'est-à-dire, le nombre élevé d'expressions référentielles qui doivent être analysées avant *about*. Inverser l'ordre entre les deux compléments prépositionnel, comme l'illustre ex. (25-b), minimise cette distance<sup>60</sup>, et permet ainsi un traitement moins "coûteux" (en mémoire) et facilite ainsi la compréhension. Ainsi, déplacer le complément long à droite permet d'optimiser le traitement et cela explique la préférence "court-avant-long" dans les langues SVO.

- (26) a. ... [PP Mary-ni] [NP Jon-ga tsureteki-ta tomodachi-o] shoukaishi-ta
- b. ... [NP Jon-ga tsureteki-ta tomodachi-o] [PP Mary-ni] shoukaishi-ta

Dans une langue SOV, pour minimiser la distance lorsque l'un des compléments verbaux est plus long, ex. (26) (cf. (24) ci-dessus), il faudrait le déplacer vers la gauche. En effet, le traitement de la phrase (26-a) est plus coûteux que la phrase<sup>61</sup>. En d'autres termes, l'ordre "long-avant-court" est à préférer, car il permet d'optimiser le traitement et faciliter la compréhension.

### 2.3.2.3 L'approche par production incrémentale de Yamashita & Chang

Yamashita & Chang (2001) se donnent comme objectif de proposer une analyse de la préférence "long-avant-court" de la langue japonaise, située dans la théorie de l'encodage grammatical de Bock & Levelt (cf. 1994) et compatible avec les modèles incrémentaux de production de la parole que nous avons présentés plus haut (§2.2.3).

En effet, les données de la langue japonaise viennent compromettre l'universalité de la préférence "court-avant-long" prédite par ces modèles<sup>62</sup>. Or, ces derniers

60. En (25-b) les distances entre le verbe et ces deux compléments prépositionnels sont de 0 et 1 unité, tandis qu'en (25-a), elles sont de 0 et 4 unités.

61. En (26-a) les distances entre le verbe et ces compléments sont de 3 et 2 (ou 0 si l'on compte à partir de la tête) unités, tandis qu'en (26-b), elles sont de 3 (ou 1 si l'on compte à partir de la tête) et 0 unités.

62. Rappelons que ces modèles expliquent les préférences de l'ordre linéaire, c'est-à-dire lorsque

expliquent cette préférence par des considérations purement cognitives, qui se doit d'être valide pour toutes les langues. Dès lors, Yamashita & Chang (2001) cherchent à proposer une analyse unifiée des deux préférences.

Les auteurs affirment qu'il est possible de rendre compte de ces préférences opposées, et de prime abord contradictoires, par une analyse unifiée tout en restant dans le cadre de ladite théorie, à condition de tenir compte des paramètres spécifiques de chaque langue. Dans ce qui suit, nous présentons l'argumentation mise en avant par les auteurs quant à la façon dont ces spécificités peuvent intervenir dans ce modèle de la production (Yamashita & Chang, 2001, pp. 52-54)<sup>63</sup>.

S'appuyant sur Bock (1982), les auteurs avancent que, dans le cadre de la théorie de l'encodage grammatical, les décisions quant à l'ordre linéaire dépendent des spécificités des différents niveaux du système de production<sup>64</sup>. Une partie de ces décisions relève de considérations conceptuelles, tandis que d'autres relèvent de considérations formelles. Dans la sphère conceptuelle, d'après Osgood & Bock (1977), il y a une tendance à placer les éléments saillants plus tôt dans la phrase, tandis que, dans la sphère formelle, d'après Ferreira (1996), le système choisit des éléments de manière incrémentale.

Dans cette vision, les constituants longs ont des propriétés concurrentes. D'un côté, ceux-ci sont plus saillants que leurs contreparties courtes, dans la mesure où les éléments lexicaux supplémentaires améliorent le sens attribué au nom tête. Cette richesse sémantique contribue à l'accessibilité conceptuelle des constituants longs. De l'autre côté, ils contiennent plus de mots et sont par conséquent formellement plus complexes, ce qui pourrait retarder leur disponibilité dans la chaîne du traitement au profit de constituants plus simples.

Les langues anglaise et japonaise ont des différences grammaticales qui peuvent être déterminantes dans la façon dont ces deux biais s'équilibrent entre elles.

Since accessibility of meaning and form can have different influences in production (Bock, 1982), it could be that Japanese speakers are focusing more on conveying meaning (putting enriched material earlier), while English speakers are focusing on sequencing forms (putting easily accessed word earlier).

---

la grammaire autorise toutes les variations étudiées, par le biais de l'accessibilité relative des constituants. Celle-ci prédit la préférence "court-avant-long" et non la préférence inverse que manifeste le japonais contrairement à l'anglais.

63. Voir également Yamashita & Chang (2006) et Chang (2009).

64. Rappelons que cette théorie postule que le processus de la production se déroule, de manière simultanée et non sérielle, sur différents niveaux, notamment le niveau de message, le niveau de l'encodage grammatical regroupant une composante fonctionnelle et une composante positionnelle, et le niveau phonologique.

Yamashita & Chang (2006, p. 8)

Tout d'abord, le japonais est une langue *pro drop* à ordre flexible dans laquelle les rôles grammaticaux ne sont pas déterminés par l'ordre mais par marquage, notamment les marquages casuels pour les cas nominatif (le sujet) et accusatif (l'objet) et les postpositions. L'anglais, au contraire, est une langue à ordre rigide qui requiert la réalisation grammaticale de tous les dépendants verbaux<sup>65</sup>. De plus, la linéarisation des compléments se fait dans le domaine préverbal en japonais, tandis qu'elle a lieu après le verbe en anglais. Comme cela a été démontré, en anglais le verbe exerce une influence importante sur les constituants qui le suivent (Stallings et al., 1998), tandis qu'en japonais il exerce une influence moindre sur les constituants le précédant (Yamashita, 1997, 2000).

Pour résumer, les locuteurs japonais disposent d'une liberté syntaxique plus importante et leur linéarisation est, par conséquent, davantage guidée par le contenu sémantique du message, ce qui leur permettrait d'utiliser l'ordre linéaire afin de marquer la saillance des différents constituants. En anglais, en revanche, la production des phrases se fait essentiellement selon les contraintes relatives aux structures syntaxiques, ce qui pourrait réduire l'influence des facteurs conceptuels au profit des facteurs liés à la forme lors de la linéarisation des constituants. Le fait qu'en anglais l'animéité n'est pas identifiée comme significative dans le HNPS, contrairement au poids grammatical et au biais lexical du verbe (Stallings et al., 1998), constitue un argument dans ce sens.

Ainsi, cette analyse permet de rendre compte des préférences de l'ordre linéaire dans les deux langues sur la base de la concurrence entre les différents facteurs intervenant pendant la production incrémentale. Cette concurrence trouve son équilibre en fonction des propriétés grammaticales spécifiques de chacune des langues.

If the shifts in human languages are motivated by both conceptual and lexical factors, then the language-specific differences in sensitivity to these factors can provide a uniform account for both “short before long” and “long before short” preferences.

Yamashita & Chang (2001, p. 54)

L'anglais est plus sensible aux facteurs formels. Dans cette langue, la production incrémentale favorise alors la postposition des constituants plus longs, c'est-à-dire

---

65. Cet argument nous semble essentiel, car il peut potentiellement expliquer par exemple l'absence d'une préférence “long-avant-court” dans une langue comme l'allemand qui manifeste un certain degré de la variabilité de l'ordre et notamment des structures à verbe final. Effectivement, malgré ces dernières propriétés, l'allemand est plus proche de l'anglais que du japonais, en ce qui concerne la rigidité, telle qu'entendue par Yamashita & Chang (2001), étant donné qu'il n'est pas *pro drop* et qu'il requiert la réalisation grammaticale de tous les dépendants verbaux.

plus complexes et moins accessibles formellement. Le japonais, à l'inverse, est plus sensible aux facteurs conceptuels et, dans cette langue, toutes choses égales par ailleurs, la préposition des constituants longs, c'est-à-dire plus saillants et plus accessibles conceptuellement, est favorisée.

#### 2.3.2.4 Bilan

Les deux approches, c'est-à-dire celles orientées vers le traitement et la compréhension et celles orientées vers la production, insistent, d'une manière ou d'une autre, sur l'importance de la position du verbe. Dans la première approche, le rôle joué par la position du verbe est directement pris en compte. Quant à la seconde approche, il est suggéré que celui-ci a un impact sur la façon dont l'accessibilité s'exerce à travers différents facteurs. Dans le domaine préverbal, c'est l'aspect conceptuel de l'accessibilité qui domine, tandis que, dans le domaine postverbal, c'est son aspect formel qui l'emporte.

Les auteurs s'accordent pour dire que départager les approches requiert la prise en compte de davantage de données intra-langue et inter-langue. Le présent travail s'inscrit dans cette démarche. Comme nous le verrons dans le chapitre suivant, le persan présente à cet égard un terrain propice, car il manifeste des propriétés communes à la fois avec l'anglais comme avec le japonais.

Il importe de noter qu'au-delà de l'effet du poids, vérifier si les différentes hypothèses – à vocation universelle –, formulées à partir d'observations sur un nombre restreint de langues, s'appliquent effectivement à des langues typologiquement différentes est un objectif que poursuivent nombre de travaux.

## 2.4 Conclusion

Ce chapitre nous a permis de faire le point sur la littérature portant sur la question de la variation de l'ordre des mots. Pour ce faire, nous avons, en premier lieu, mis en discussion la grammaire générative et son impact sur la définition de l'objet d'étude et la méthodologie en linguistique théorique, ainsi que la place de la question de l'ordre en syntaxe générative. En deuxième lieu, nous avons abordé la typologie linguistique pour laquelle l'ordre des mots est une notion fondamentale.

Finalement, nous nous sommes attardée sur les approches que nous avons nommées cognitivo-fonctionnelles, qui s'intéressent à la question de la variation de l'ordre à proprement parler. Ces travaux traitent de différentes problématiques, en rapport notamment avec l'architecture du système de la production et la compréhension de la parole, et tentent de rendre compte des phénomènes de la variation de l'ordre par

des principes à vocation universelle.

Le présent travail s'inscrit dans la lignée de ces approches et un volet des objectifs que nous nous assignons est la mise en perspective d'une partie de ces principes avec les données du persan. Cette mise en perspective pourrait permettre de départager entre les différentes théories concurrentes. À cet égard, le persan se montre particulièrement intéressant.

Nous détaillerons nos questions de recherche dans le chapitre suivant. Nous y expliciterons notamment ce que ces principes prédisent sur la base des propriétés du persan. La présentation de nos méthodes et outils se fera dans la deuxième partie de la thèse.



## Chapitre 3

# Démarche et questions de recherche

Le premier chapitre nous a permis d'explorer les phénomènes empiriques que le persan offre comme objet d'étude sur la variation de l'ordre des constituants, et le deuxième d'apprécier les différentes approches développées dans les travaux précédents. Nous allons pouvoir présenter dans ce dernier chapitre de la première partie les questions de recherche que nous proposons de traiter dans cette thèse, ainsi que la méthodologie que nous adoptons pour ce faire.

Les questions de recherche investies dans cette thèse comprennent deux volets. Le premier volet regroupe les questions relatives à la syntaxe du persan, soulevées pour certaines et ignorées pour d'autres dans les débats théoriques en cours. Il s'agit, tout d'abord, d'examiner les hypothèses existantes sur l'ordre non marqué entre le COD et le COI, et, ensuite, d'étudier les facteurs qui interviennent dans les préférences dans l'ordonnement linéaire des dépendants verbaux dans le domaine préverbal. Dans le second volet, il est question de l'apport théorique des données du persan pour l'étude de la variation de l'ordre des mots. Dans le chapitre précédent, outre des questions méthodologiques, nous avons évoqué les hypothèses, à vocation universelle, proposées pour rendre compte des préférences attestées dans les langues du monde. Il s'agit de mettre ces hypothèses à l'épreuve des données de la langue persane, langue SOV à ordre flexible. Il va sans dire qu'il se produit un va-et-vient constant entre ces deux volets, car étudier les données du persan à la lumière des hypothèses émises dans la littérature revient à mettre ces dernières à l'épreuve des données du persan. Ainsi, dans la section dédiée au second volet, nous nous concentrerons sur les analyses concurrentes proposées, notamment pour rendre compte de l'effet du poids grammatical.

Ce chapitre est organisé de la façon suivante. Dans la première section, nous jus-

tifierons notre divergence méthodologique par rapport aux travaux précédents (sur le persan) et préciserons l'approche adoptée pour cette étude. Dans la seconde section, nous détaillerons nos questions de recherche après avoir circonscrit les phénomènes étudiés.

## 3.1 Méthodologie et approche adoptées

Dans cette section, tout d'abord, en nous basant sur la revue critique des approches et méthodologies disponibles pour étudier la variation de l'ordre des mots faite dans le chapitre précédent (cf. section 2.2), nous exposons le choix de l'approche et la méthodologie. Ensuite, nous introduisons brièvement les méthodes quantitatives auxquels l'on peut avoir recours dans le cadre d'une étude linguistique dans le domaine de la syntaxe.

### 3.1.1 Le choix méthodologique

Dans le premier chapitre, nous avons vu que les phénomènes relevant de la variation de l'ordre des constituants en persan sont exposés dans des travaux de sensibilités différentes. Les travaux relevant de la grammaire générale et descriptive énumèrent, de manière plus ou moins exhaustive, les différentes possibilités d'ordonnement des constituants. Les ordres non canoniques sont généralement décrits dans ces travaux en rapport avec l'articulation discursive de la phrase. Ceux-ci prennent souvent appui sur des exemples attestés, dont ils se servent de manière ponctuelle pour illustrer leurs observations, tandis qu'aucune étude quantitative n'y est proposée.

Les travaux en syntaxe formelle reprennent une partie de ces observations et en rendent compte avec un appareil formel. Ces derniers se situant dans le cadre de la grammaire générative se soucient en premier lieu de l'apport théorique de leur analyse et de sa capacité de généralisation au-delà des données de la langue étudiée.

Plus précisément, dans un modèle théorique où différents ordres linéaires (ou de surface) sont dérivés d'un même ordre sous-jacent par des règles spécifiques de la grammaire, l'objectif étant d'explicitier les règles à partir desquelles les ordres non canoniques s'obtiennent. Le but ultime consiste à départager entre les analyses concurrentes telles que les approches par ordres de base distincts (*base generation*) et celles par mouvements (*scrambling*), voir §2.2.1.

À titre d'exemple, Karimi (2005), adoptant une vision configurationnelle du persan, avance que les différents ordres non canoniques observés dans cette langue sont régis par des contraintes sémantiques spécifiques – qui agissent de manière binaire à

l’instar des règles traitant de la grammaticalité – et résultent de mouvements comparables à ceux rencontrés dans les langues à ordre rigide (cf. section 1.4). Puis, prenant appui sur les données du persan, elle revisite les analyses existantes du *scrambling* et en propose de nouvelles définitions. De ce fait, nous pouvons dire que l’auteur n’étudie pas la variation de l’ordre des constituants à proprement parler, c’est-à-dire telle que nous l’avons définie dans le deuxième chapitre (cf. section 2.1), comme une alternance “libre” eu égard des contraintes catégoriques.

Dans cette thèse, nous nous éloignons de cette approche et adoptons une méthodologie différente. La question de la validité empirique des données est au centre de cette divergence. À l’inverse de ces travaux théoriques se fondant sur une sélection partielle des faits empiriques, nous proposons une étude basée sur l’usage dans la lignée des travaux adoptant une approche cognitivo-fonctionnelle (cf. §2.2.3).

En l’absence d’études empiriques robustes démontrant le contraire, nous considérons que la variation de l’ordre des constituants en persan n’est pas régie par des contraintes catégoriques de la grammaire, mais, à l’instar des phénomènes exposés dans le chapitre précédent, est susceptible de relever des préférences générales établies à travers des études sur des langues variées (cf. section 2.3).

### 3.1.2 Les méthodes quantitatives en linguistique

Le recours aux méthodes quantitatives basées sur des outils statistiques est habituel dans d’autres domaines des sciences humaines (sociologie, psychologie, etc.), ainsi que dans certains domaines de la linguistique tels que phonétique, sociolinguistique, psycholinguistique, et commencent également à se faire une place au sein de domaines comme la phonologie, la morphologie, la syntaxe, la sémantique et la pragmatique.

Il s’agit de méthodes qui permettent de faire, à partir d’un échantillon (représentatif), c’est-à-dire une sélection qui reflèterait la population cible<sup>1</sup> de la manière la plus exacte possible telle qu’un échantillon tiré par échantillonnage arbitraire ou probabiliste (*random sampling*)<sup>2</sup>, des inférences sur l’ensemble de la population<sup>3</sup>.

Les statistiques inférentielles sont basées sur un principe simple : si l’échantillon est suffisamment grand et prélevé de manière parfaitement aléatoire, les proportions observées dans celui-ci ont une certaine probabilité d’être relativement proches de

---

1. Le terme “population” est à comprendre dans son sens technique : ensemble d’individus, d’évènements ou d’unités de même espèce dont on souhaite étudier les caractéristiques.

2. C’est-à-dire de manière à garantir que chaque individu de la population a une chance égale pour se trouver au sein de l’échantillon.

3. Pour en donner un exemple tangible, les sondages d’opinion, tels que l’estimation des résultats d’une élection, se font sur la base de telles méthodes.

celles observées dans la population. Ainsi, un comportement (ou une caractéristique) observé dans un tel échantillon peut, dans une certaine mesure, être généralisé à l'ensemble de la population. Plus techniquement, et dans la terminologie en vigueur, une observation peut être généralisée, avec une marge d'erreur donnée, à l'ensemble de la population à condition d'être significative sous le seuil de cette marge. Les tests statistiques sont utilisés pour en décider.

Par ailleurs, la modélisation statistique, notamment les modèles de régression, permet d'effectuer des études multifactorielles, c'est-à-dire la prise en compte d'un ensemble de variables pour étudier leurs effets respectifs ainsi que leurs interactions. Il est possible d'utiliser ces modèles pour "prédire" le comportement d'une unité d'étude en fonction de ses caractéristiques<sup>4</sup>.

L'enjeu essentiel dans ces méthodes est la constitution de l'échantillon, dont dépend notamment la validité dite externe de la recherche, c'est-à-dire sa force de généralisation. En fonction de la population étudiée, différentes méthodes de collecte ou de tirage sont possibles. Tout d'abord, il faut circonscrire la population et choisir l'unité d'étude. En syntaxe, l'unité d'étude (maximale) est la phrase, mais, en fonction de l'objet de recherche, l'unité peut être le syntagme, voire le mot. En revanche, la population étudiée est un ensemble infini impossible à saisir, car il s'agit pour ainsi dire de l'ensemble des énoncés d'une langue donnée. L'échantillonnage se fait sur la base d'une population hypothétique et la représentativité de l'échantillon, le corpus d'étude en l'occurrence, quelle que soit sa taille, ne peut être que foncièrement approximative<sup>5</sup>. Les travaux en linguistique ayant recours aux méthodes quantitatives basées sur corpus, le font en dépit de ce problème théorique, au demeurant insurmontable (Evert, 2006; Baroni & Evert, 2008)<sup>6</sup>.

---

4. Par exemple, les sondages d'opinion, au-delà de l'estimation des résultats, tentent de prédire l'effet des caractéristiques, tel que l'âge, le sexe, etc., sur le vote des électeurs.

5. En réalité, les échantillons utilisés, c'est-à-dire les corpus, sont généralement partiels, dans la mesure où ils sont compilés à partir des données spécifiques, tels que des extraits d'un journal, des communications téléphoniques, etc. Cependant, le fait d'être constitué d'un ensemble cohérent permet, au détriment de la représentativité, somme toute imparfaite, d'apporter une certaine homogénéité qui permet des généralisations relativement solides sur une population circonscrite, en l'occurrence une variété spécifique de la langue étudiée.

6. Il importe de mentionner qu'en sciences humaines et sociales, en général, l'échantillon aléatoire reste souvent un idéal inatteignable. Cela n'interdit pas la recherche quantitative dans ces domaines, mais exige de la précaution lors de l'interprétation des résultats, notamment quant à leur généralisation à l'ensemble de la population (cf. entre autres Howell *et al.*, 2008, p. 2). De plus, bien que ce défaut affaiblit la validité externe de la recherche, il est tout à fait réalisable d'en garantir la validité interne, c'est-à-dire la fiabilité des conclusions internes de la recherche, qui dans la pratique se montre généralement bien plus importante. Il suffit pour cela que les groupes, que l'on cherche à comparer, partagent exactement les mêmes conditions d'échantillonnage. En effet, quand le but de la recherche est de comparer différentes catégories entre elles, ce qui importe, en premier lieu, est que l'échantillon, quelque peu représentatif qu'il soit, ne présente aucun biais envers l'une ou l'autre catégorie.

C'est la raison pour laquelle la prolifération des études similaires sur un même objet d'étude, mais avec des données différentes, est la bienvenue. Si les généralisations faites à partir de données variées convergent, il y aura lieu dès lors de les considérer comme valides. Par ailleurs, pour pallier cette imperfection, deux méthodes de collecte de données, la méthode de corpus et la méthode expérimentale, peuvent être utilisées en tandem. Dans le cas de cette dernière, le matériel expérimental (les stimuli) consiste généralement en phrases fabriquées artificiellement "en laboratoire" par le chercheur, c'est-à-dire en dehors des situations ordinaires de production, et l'échantillonnage se fait sur la population des locuteurs de la langue<sup>7</sup>. En effet, la linguistique expérimentale, inspirée des protocoles en vigueur en psychologie, met le locuteur au centre et s'intéresse avant tout à la performance langagière de celui-ci – tant en production qu'en compréhension.

En fonction de leurs différences, la démarche et l'interprétation des résultats ne sont pas identiques pour ces deux méthodes de collecte de données langagières. La méthode expérimentale consiste en l'observation d'une variable d'étude (appelée variable dépendante) par la variation systématique d'un ou plusieurs facteurs (appelés variables indépendantes) à travers un échantillon, tout en contrôlant, dans la mesure du possible, toute autre source de variation indésirable considérée comme bruit ou biais, appelée variation externe. De ce fait, en l'absence de variables parasites, une différence statistiquement significative, observée au niveau de la variable dépendante, peut être directement attribuée aux variables manipulées dans l'expérience. Les données extraites de corpus, en revanche, sont produites dans les conditions dites "écologiques" et présentent une variation inhérente sur laquelle le chercheur n'a aucune possibilité d'intervention. Par conséquent, du fait de la présence de sources incontrôlées de variation, les méthodes statistiques ne peuvent que détecter des corrélations entre les variables sans jamais pouvoir établir de lien causal entre elles.

Les données de corpus ont l'avantage de permettre l'exploration d'un volume important de données riches et diverses et, par là, la prise en compte d'un éventail plus large de facteurs. Tandis que, dans les études expérimentales, du fait d'un rapport exponentiel entre le nombre de facteurs et le nombre d'items nécessaires pour les stimuli, les variables manipulées sont limitées, souvent à deux, maximum à trois. De plus, dans une étude de corpus, les groupes sont constitués *a posteriori*. Il est ainsi possible d'étudier différents facteurs avec les mêmes données. Dans la méthode expérimentale, en revanche, les variables sont définies avant la collecte des données, et un changement dans la conception de l'expérience nécessite normalement

---

7. Précisons que les données expérimentales peuvent également être considérées comme un échantillon d'énoncés prélevé sur la population de la construction étudiée, par exemple les phrases à double objets de l'anglais, les phrases ditransitives du persan, etc.

le renouvellement de celle-ci, avec en principe de nouveaux participants.

Pour toutes ces raisons, il convient d’explorer les données du corpus afin d’observer la distribution effective des différentes alternatives dans les données naturelles et d’identifier les facteurs pertinents. Il est ensuite possible d’étudier ces facteurs de manière ciblée avec des expériences contrôlées. Cette démarche en tandem est celle que nous avons entreprise pour notre étude quantitative présentée dans la partie suivante. Nous nous appuyerons, tout d’abord, sur des données de corpus, ch. 4, pour ensuite effectuer des expériences psycholinguistiques pointues, ch. 5.

Nonobstant leurs problèmes inhérents, ces méthodes, à l’inverse des méthodes traditionnellement employées en syntaxe, ont l’avantage d’être reproductibles, ce qui contribue à assurer leur objectivité relative.

Quelle que soit la source, l’analyse des données quantitatives repose sur les méthodes de la statistique inférentielle. Dans la sous-section suivante, nous énumérons les principaux outils statistiques auxquels nous avons eu recours pour l’analyse de nos données.

### 3.1.3 Les outils statistiques utilisés dans cette étude

Les méthodes que nous utilisons varient dans les différents étapes d’analyse des données. En somme, nous avons utilisé deux méthodes principales : les tests tels que le test  $\chi^2$  de Pearson pour les variables catégorielles, et le test  $t$  pour les variables numériques, ainsi que les modèles de régression mixtes (appelés en simplifiant les modèles mixtes)<sup>8</sup>. L’ensemble de nos analyses statistiques et de nos graphiques est effectué avec le logiciel en accès libre R (R Core Team, 2013)<sup>9</sup>. Nous nous sommes servi en particulier du package “lme4” (Bates et al., 2012), spécialisé dans les modèles mixtes.

Notons que le test  $\chi^2$  est traditionnellement utilisé en linguistique, mais cette pratique a graduellement laissé la place à la méthode des modèles de régression. Tout d’abord, la régression multiple permet une analyse multifactorielle des données par la prise en compte simultanée de plusieurs variables explicatives et, le cas échéant, de leurs interactions (voir par ex. Gelman & Hill, 2006). Par ailleurs, une variable catégorielle, en l’occurrence une variable binaire telle que l’ordre relatif entre deux constituants, peut être assimilée, par le biais d’une transformation logistique, à une fonction de probabilité et modélisée au lieu d’une droite de régression par une courbe sigmoïde (voir par ex. Agresti, 2007). Plus précisément, à l’aide de cette fonction, il

---

8. Ces modèles sont également connus sous l’appellation “modèle de régression hiérarchique” (*hierarchical regression*) ou encore “modèle de régression multi-niveau” (*multilevel regression*).

9. Disponible à l’adresse suivante : <http://cran.r-project.org>.

est possible de calculer la probabilité d'obtenir une des deux valeurs possibles de la variable désignée comme le scénario du succès contre l'autre valeur correspondant alors à l'échec. Cette méthode permet ainsi de modéliser une variable catégorielle par une combinaison de variables numériques et catégorielles.

En outre, les modèles (de régression) mixtes – utilisés notamment, mais pas exclusivement, pour l'analyse des données expérimentales –, qu'ils soient linéaires ou logistiques ont l'avantage de permettre en plus la prise en compte des sources de variation externe systématique et, de ce fait, d'assurer une meilleure estimation du rapport entre les variables de l'étude<sup>10</sup>.

Enfin, il est possible d'évaluer la force explicative de chaque variable ou interaction en comparant des modèles de régression calculés sur le même jeu de données, en incluant de manière progressive l'ensemble des différentes variables et interactions disponibles<sup>11</sup>. Cette méthode est en outre utilisée pour réduire le modèle de régression maximal afin d'obtenir un modèle optimal (voir Bates et al., 2015).

## 3.2 Questions de recherche

Dans le projet de recherche que nous menons à travers l'étude de deux phénomènes de variation de l'ordre des constituants en persan, nous poursuivons deux objectifs généraux. Premièrement, nous souhaitons apporter un point de vue basé sur l'usage aux débats théoriques portant sur différents problèmes de la syntaxe du persan. Le présent travail est, à notre connaissance, le premier à offrir un tel point de vue avec une méthodologie quantitative qui correspond à l'état de l'art. Deuxièmement, nous souhaitons contribuer aux débats théoriques sur les préférences d'ordonnement linéaire des mots à travers les langues du monde. Ces débats souffrent d'un déséquilibre dû à la sur-représentativité des langues telles que l'anglais.

Dans la suite de cette section, nous présentons, dans l'ordre, les questions de recherche relatives à chacun de ces objectifs. Mais, avant cela, il convient de circonscrire les phénomènes que nous allons étudier dans le cadre du présent travail.

---

10. Les modèles mixtes, particulièrement pertinents pour l'analyse des données expérimentales, sont des modèles de régression dans lesquels l'on inclut, en plus des variables indépendantes de l'étude, appelées variables à effet fixe, des sources de variation externe systématique, telles que la variation due aux items et aux participants, appelées variables à effet aléatoire. Cette méthode permet de minimiser la part de la variation non expliquée par le modèle afin de faciliter la généralisation des résultats, obtenus à partir d'un échantillon, à l'ensemble de la population (voir par ex. Baayen et al., 2008; Jaeger, 2008). Le biais lexical est une autre source de variation qui, dans une étude de corpus aussi bien que dans une étude expérimentale, peut être intégrée au modèle en tant que variable à effet aléatoire. Précisons que celui-ci peut en fonction des objectifs de l'étude également être intégré au modèle en tant que facteur à effet fixe.

11. Nous nous sommes servi pour cela de la fonction *anova* du logiciel R, basée sur le test du rapport de vraisemblance (*Likelihood ratio test*).

### 3.2.1 Le choix des phénomènes étudiés

Le panorama des phénomènes de la variation de l'ordre des constituants en persan, exposé dans le chapitre 1, nous a permis de poser le constat que ces phénomènes sont multiples et complexes. Ceux-ci comprennent l'ordonnement linéaire des dépendants sélectionnés par rapport au verbe et leur ordre relatif, ainsi que le placement des ajouts et des modificateurs. Ces phénomènes se manifestent, d'un côté, dans le domaine préverbal et, de l'autre, dans le domaine postverbal, et concernent aussi bien la phrase simple que la phrase complexe.

Le cadre de ce travail ne permet pas d'aborder l'ensemble de ces phénomènes et nous conduit à opérer des choix. Nous nous focaliserons donc sur la variation de l'ordre au sein de la proposition simple entre les dépendants sélectionnés du verbe dans le domaine préverbal. Celle-ci n'a jamais fait, à notre connaissance, l'objet d'une étude quantitative méthodique, comparable à celle que nous proposons de mener dans cette thèse dans la lignée des travaux mentionnés dans le chapitre 2 (cf. §2.2.3). Nous nous attacherons ainsi à deux phénomènes, à savoir l'ordre relatif entre les deux objets dans une construction ditransitive et l'ordre relatif entre le sujet et l'objet dans une construction transitive. Le premier, comme nous l'avons vu dans le chapitre 1, du fait du marquage différentiel de l'objet (§1.1.3), constitue un problème central de la syntaxe du persan et, à ce titre, a reçu une attention particulière dans la littérature sur le persan. Le second, au contraire, n'a pas suscité de débats significatifs et demeure relativement peu abordé.

La question de l'ordre relatif entre les compléments verbaux, notamment le COD et le COI, est, par ailleurs, un sujet bien étudié et relativement bien documenté à travers les langues du monde, comme nous l'avons vu dans le chapitre 2. Cet objet d'étude, comme celui de l'ordre relatif dans une coordination, est propice à l'étude des préférences dans la production de la parole et, de ce fait, offre un terrain favorable à l'étude du fonctionnement du système langagier.

Il faut rappeler que cet intérêt a une signification particulière pour l'anglais, langue SVO à ordre rigide, dans laquelle le verbe s'intercale entre le sujet et les autres dépendants verbaux. De nombreux travaux en syntaxe générale ont été consacrés à l'étude des différents phénomènes liés à la linéarisation des compléments verbaux dans cette langue, tels que le HNPS ou l'alternance dative, et ces derniers se trouvent au cœur des travaux précurseurs adoptant une approche cognitivo-fonctionnelle. Il semblerait que cela s'est traduit par un intérêt général plus prononcé porté à l'ordre relatif entre les compléments, hormis le sujet. Par ailleurs, que l'on adhère ou non au point de vue génératif qui distingue entre l'argument externe (le sujet) et les arguments internes (les compléments d'objet), le fait que le sujet et le(s)

objet(s) n'ont pas le même statut quant à la hiérarchie des rôles grammaticaux est unanimement admis en linguistique générale. Cette asymétrie se manifeste, par exemple, au niveau des relations de l'antécédence pour les anaphores, autrement dit les relations du liage dans la terminologie générative<sup>12</sup>.

Les langues SOV manifestent, par contraste, un cas de figure dans lequel le verbe ne s'intercale pas entre les dépendants et le sujet entre directement en ligne de compte dans la question de la linéarisation des dépendants verbaux. Les travaux similaires sur de telles langues, notamment le japonais (Yamashita & Chang, 2001), incluent effectivement l'ordre relatif entre le sujet et le(s) objet(s)<sup>13</sup>. En effet, dans une langue SOV, saisir les préférences de l'ordonnancement linéaire dans la globalité ne peut se faire sans la prise en compte du sujet.

### 3.2.2 Questions relatives à la syntaxe du persan

Une question essentielle de la syntaxe du persan, à laquelle nous nous proposons d'apporter une réponse empirique robuste, porte sur l'ordre non marqué entre le COD et le COI qui, du fait du marquage différentiel de l'objet, s'est trouvé au centre d'une bonne partie des travaux. À travers cet examen, nous souhaitons tout particulièrement relancer le débat théorique sur le statut syntaxique du COD (cf. section 6.1); débat qui a fait couler beaucoup d'encre, mais qui nous semble avoir été bâti, jusqu'à présent, sur des hypothèses de départ qui ne peuvent être que contestées. Pour ce faire, nous passerons en revue, tout d'abord, les hypothèses existantes quant à l'ordre non marqué entre le COD et le COI, et, ensuite, nous nous pencherons sur la définition de l'ordre non marqué, que nous adopterons en prémisses de notre étude.

Le second sujet que nous envisageons d'approfondir, tout aussi essentiel mais relativement peu investi dans la littérature, est la variation de l'ordre dans le domaine préverbal à proprement parler. Dès lors, nous élargirons notre champ d'étude pour inclure le sujet.

---

12. Nous reviendrons sur cette asymétrie dans la troisième partie (cf. section 6.3).

13. L'allemand présente un cas intéressant à noter à cet égard, car il s'agit d'une langue dite V2, qui connaît un ordre relativement plus flexible que l'anglais, pour laquelle la question de l'ordre est largement documentée. Contrairement à l'anglais, l'allemand permet la permutation du sujet et de l'objet (SO vs. OS). Ainsi, l'ordre relatif entre le sujet et l'objet (direct/accusatif ou indirect/datif) dans cette langue est également débattu dans la littérature. En outre, l'ordre SOV se manifeste aussi dans certaines constructions en allemand, notamment dans les subordonnées (voir Bader & Häussler, 2010).

### 3.2.2.1 Ordre canonique entre le COD et le COI

L'ordre relatif entre le COD et le COI dans une phrase non marquée sert généralement d'argument pivot pour corroborer l'hypothèse de la position syntaxique double du COD (PSDO) en persan<sup>14</sup>. En effet, Browning & Karimi (1994) et Karimi (2003a) partent de l'observation selon laquelle les deux COD, spécifique et non spécifique<sup>15</sup>, ne se trouvent pas dans la même position linéaire dans une phrase canonique, ex. (1) et (2)<sup>16</sup>.

- (1) a. Hassan=*rā* be Ali mo'arrefi kard-am  
Hassan=*RĀ* à Ali présentation faire.PST-1SG  
'J'ai présenté Hassan à Ali.'
- b. be Ali yek ādam=*e* xub moa'rrefi kard-am  
à Ali un personne=*EZ* bien présentation faire.PST-1SG  
'J'ai présenté une personne bien à Ali.'
- Browning & Karimi (1994, p. 68)
- (2) a. Kimea aqlab barā mā še'r mi-xun-e  
Kimea souvent pour nous poésie IPFV-lire-3SG  
'C'est souvent le cas que Kimea nous lit de la poésie.'
- b. Kimea aqlab barā mā ye še'r az Hāfez mi-xun-e  
Kimea souvent pour nous un poème de Hafez IPFV-lire-3SG  
'C'est souvent le cas que Kimea nous lit un poème de Hafez.'
- c. Kimea aqlab hame=*ye* še'r-ā=*ye* tāza=*š=ro* barā mā  
Kimea souvent tous=*EZ* poème-PL=*EZ* nouveau=*3SG=RĀ* pour nous  
mi-xun-e  
IPFV-lire-3SG  
'C'est souvent le cas que Kimea nous lit un de ses nouveaux poèmes.'
- d. Kimea aqlab ye še'r az Hāfez=*ro* barā mā mi-xun-e  
Kimea souvent un poème de Hafez=*RĀ* pour nous IPFV-lire-3SG  
'C'est souvent le cas que Kimea nous lit un (certain) poème de Hafez.'
- Karimi (2003a, pp. 91–92)

Browning & Karimi (1994) avancent que les ordres inverses, ex. (3), bien que grammaticaux, sont jugés "marqués" par les locuteurs. Plus précisément, les constituants en gras sont perçus comme "déplacés" par ces derniers<sup>17</sup>.

14. Pour une présentation détaillée, se reporter à la section 6.1 dans la troisième partie.

15. Rappelons que ces auteurs admettent le trait [ $\pm$ spécifique] comme le seul trait binaire dont dépend le marquage par *=rā* (cf. §1.1.3).

16. Nous avons effectivement vu, dans le chapitre 1, que l'ordre relatif entre le COD et le COI est considéré à la quasi-unanimité comme dépendant du marquage différentiel de l'objet (cf. §1.2.1).

17. « *Although these sentences are grammatical, native speakers judge them to be marked. In particular, underlined constituents are regarded as having been scrambled.* » (Browning & Karimi, 1994, p. 69)

- (3) a. **be Ali** Hassan=*rā* moa'rrefi kard-am  
 b. **yek adam=e xub** be Ali moa'rrefi kard-am

Browning & Karimi (1994, p. 68)

Karimi (2003a, p. 92) affirme, quant à elle, que seules les conditions discursives précises pourraient provoquer un bouleversement de cet ordre neutre. Un COD<sub>[-spécifique]</sub>, à titre d'exemple, ne peut être séparé du verbe, selon l'auteur, que s'il correspond à un "focus contrastif"<sup>18</sup>, ex. (4)<sup>19</sup>.

- (4) Kimea aqlab (ye) ketāb=e dāstān barā bačče-hā mi-xun-e  
 Kimea souvent (un) livre=EZ histoire pour enfant-PL IPFV-lire.PRS-3SG  
 'Kimea lit souvent un livre d'**histoires** (plutôt qu'un livre de poésie) pour les enfants'

Karimi (2003a, p. 92)

Tout d'abord, il faut noter que l'assertion selon laquelle les COD marqués, d'un côté, et les COD non marqués, de l'autre, constituent des classes homogènes est discutable, dans la mesure où les COD non marqués varient en fonction de leur niveau de détermination. Ces COD peuvent avoir une réalisation nue, c'est-à-dire sans aucune détermination, ni quantification, tout aussi bien qu'une réalisation déterminée ou quantifiée (cf. §1.1.2).

Le COD nu est considéré de manière consensuelle comme ayant une relation étroite avec le verbe (cf. par ex. Lazard, 1982, 1994; Ghomeshi & Massam, 1994; Samvelian, 2001, 2012). En règle générale, la combinaison d'un nom et d'un verbe, en persan, se prête à une interprétation (sémantique) fusionnelle, dans le sens où l'ensemble désigne un événement dans lequel le COD agit comme un "modificateur restrictif du verbe", selon les termes de Creissels (2006, p. 128), ex. *ketāb xāndan* 'lire un ou des livre(s) (Lit. livre lire)', désigne un type particulier de l'activité de lire dont l'objet de lecture est un livre.

Ganjavi (2007) met également cette affirmation en question, dans sa thèse intitulée "*Direct Objects in Persian*", et suggère trois classes de COD en fonction de sa structure syntagmatique : DP, NumP et NP, respectivement, COD marqué de =*rā*, COD non marqué portant une détermination quantitative, COD nu<sup>20</sup>.

18. Pour la définition du focus contrastif adoptée par l'auteur, voir §1.3.1 au chapitre 1.

19. « *This element [nonspecific object] can be separated from the verb only in a very limited fashion, by representing contrastive focus.* » (Karimi, 2003a, p. 92)

20. Précisons, toutefois, que Ganjavi (2007, pp. 27-29) souscrit à l'hypothèse évoquée dans les travaux susmentionnés en ce qui concerne l'ordre linéaire, affirmant que ce dernier est un des points sur lequel les COD non marqués se rejoignent pour s'opposer aux COD marqués.

De plus, nous remarquons que dans le peu d'études à caractère quantitatif, qui cherchent à démontrer la véracité empirique de cette hypothèse, le COD nu semble être pris comme le prototype du COD non marqué. Dans son article, qui se situe dans la lignée des travaux de Karimi, Rasekhmahand (2004, p. 59) avance que le résultat de l'étude de la distribution des fréquences de l'apparition des deux ordres en fonction du marquage du COD par  $=r\bar{a}$ , dans un corpus de 300 phrases extraites du *Dictionnaire du persan familier*, confirme la distribution binaire de l'ordre. Or, l'auteur ne fournit pas les détails méthodologiques de son étude, notamment quant à l'extraction et à l'échantillonnage. Faute de connaître la distribution des différents types de COD dans l'échantillon étudié par l'auteur, nous ne pouvons pas nous prononcer sur la représentativité de l'échantillon, et, sachant que les exemples cités dans le texte de l'article sont exclusivement des objets nus (Rasekhmahand, 2004, p. 61), nous émettons des réserves quant à sa force de généralisation. En effet, la sur-représentation des COD nus pourrait expliquer les résultats probants dans le sens de l'hypothèse défendue, dans la mesure où, bien qu'une forte préférence pour la position adjacente au verbe soit tout à fait plausible pour les objets nus, la formation des prédicats complexes (cf. §1.1.4) en témoigne par ailleurs, il n'en va pas de même pour les COD indéfinis ou quantifiés.

En outre, comme nous l'avons vu dans le chapitre 1, nous rencontrons dans les travaux des grammairiens une généralisation quant à l'ordre canonique des constituant, qui admet trois cas de figure (cf. ex. (65), p. 32 en §1.2.1). Selon ces auteurs, les COD non marqués qui portent la marque de la détermination indéfini  $=i$ , contrairement aux COD (non marqués) nus, peuvent indifféremment précéder ou suivre le COI.

La première question de recherche que nous entreprenons dans cette thèse consiste à examiner cette éventualité, c'est-à-dire à vérifier si l'hypothèse de la dichotomie marqué vs. non marqué quant à l'ordre linéaire est empiriquement fondée. Rappelons que les travaux précédents formulent leurs hypothèses en termes de traits sémantiques supposées déclencher le marquage par  $=r\bar{a}$ , ex.  $[\pm\text{spécifique}]$ ,  $[\pm\text{défini}]$ . Nous remplacerons ces traits, au demeurant discutables, par le trait  $[\pm r\bar{a}]$  qui permet d'éviter les biais d'une annotation subjective. Pour ce faire, nous opérons une annotation qui oppose, d'un côté, les COD $_{[+r\bar{a}]}$  aux COD $_{[-r\bar{a}]}$ , et, de l'autre, les COD $_{[+nu]}$  aux COD $_{[-nu]}$ . Cette annotation permet de distinguer trois catégories du COD : 1) COD $_{[+r\bar{a}]}$  ; 2) COD $_{[+nu,-r\bar{a}]}$  et 3) COD $_{[-nu,-r\bar{a}]}$ . À l'issue des premiers résultats, nous affinerons davantage notre annotation, le cas échéant.

La question est plus précisément de savoir, pour chacune de ces trois catégories, quel est l'ordre relatif non marqué entre le COD et le COI. L'hypothèse courante

prédit l'ordre COI-COD pour les COD<sub>[+nu,-rā]</sub> et les COD<sub>[-nu,-rā]</sub>, et COI-COD pour les COD<sub>[+rā]</sub>.

À ce stade, il est nécessaire de formuler notre prémisse quant à la définition de l'ordre non marqué. Nous nous appuyerons sur celle explicitée par Lambrecht (1994), largement admise dans la littérature, que nous présentons par la suite.

### 3.2.2.2 Qu'est-ce que l'ordre non marqué ?

La notion de l'ordre non marqué<sup>21</sup> va de pair avec la notion de la structure informationnelle<sup>22</sup>. Ce terme désigne une organisation des constituants de la phrase qui permet de rendre compte du fait que plusieurs énoncés construits avec les mêmes constituants, mais distincts au niveau de l'ordre des constituants, de la structure prosodique (le placement de l'accent saillant), de la construction ou du marquage des constituants, peuvent avoir le même contenu sémantique (c'est-à-dire les mêmes valeurs véri-conditionnelles), appelées allophrases (*allo-sentences* en anglais) par Lambrecht (1994). À titre d'exemple, les phrases suivantes ont toutes le même contenu sémantique, mais servent différentes fonctions pragmatiques, du fait de leur prosodie (l'accent saillant est indiqué par la mise en gras du constituant), ou de leur ordre ou d'une construction particulière (Marandin, 2006)<sup>23</sup>.

- (5)
- a. Marie a rencontré Paul hier.
  - b. Mairie a rencontré **Paul** hier.
  - c. **Marie** a rencontré Paul hier.
  - d. Marie a rencontré Paul **hier**.
  - e. Hier, Marie a rencontré Paul.
  - f. Marie, Paul l'a rencontrée hier.
  - g. Paul, il a rencontré Marie hier.
  - h. C'est hier que Marie a rencontré Paul.

21. Notons que l'opposition marqué vs. non marqué, utilisée ici, fait référence à la notion de la marque (ou du marquage), *markedness*, qui remonte aux travaux de Jakobson et de Troubetzkoy au sein du Cercle linguistique de Prague (Lambrecht, 1994, p. 18). Pour simplifier à l'extrême, et sans entrer dans les débats théoriques controversés, nous pouvons dire que le marquage (ou la forme marquée) est utilisé, en règle générale, pour caractériser, de manière binaire, une forme plus complexe ou un usage plus restreint ou plus spécifique (cf. ex. Waugh, 1982; Andrews, 1990; Battistella, 1990).

22. Le terme *information structure* a été forgé par Halliday (1967), mais l'hypothèse de la structure informationnelle remonte à Paul (1880), qui a postulé l'articulation binaire sujet – prédicat indépendamment de la relation sujet/verbe (cf. Marandin, 2006).

23. À ce propos, il est important de rappeler le lien étroit entre l'ordre linéaire et la configuration discursive de la phrase noté pour les données du persan. En effet, nous avons pu observer dans le chapitre 1 (§1.3.1 et §1.4.1) que l'ordre linéaire en persan a reçu une analyse en termes d'articulation discursive de la phrase et que les ordres non canoniques sont principalement décrits dans la littérature en fonction de celle-ci.

etc.

D'après Lambrecht (1994, pp.15-17), strictement parlant, qualifier de neutre ou de non marqué, du point de vue de la structure informationnelle, un ordre linéaire (ou un schéma prosodique) ne fait pas sens, dans la mesure où cela implique l'hypothèse difficilement défendable selon laquelle il existe des phrases neutres sur le plan pragmatique. Or, toute phrase a une structure informationnelle et chaque structure informationnelle a une fonction pragmatique, ce qui revient à dire qu'il ne peut exister de fonction pragmatique neutre. De ce fait, il est seulement possible de postuler qu'un ordre est non marqué lorsqu'il peut servir plusieurs fonctions pragmatiques, autrement dit lorsqu'il est susceptible de correspondre à différentes articulations de la structure informationnelle. Un ordre marqué, en revanche, ne peut servir qu'une seule fonction pragmatique.

À titre d'exemple, une phrase comme *Marie aime le thé* peut avoir une structure informationnelle à "focus large", *all focus* dans la terminologie de Lambrecht (1994), ou *broad focus* dans la terminologie de Ladd (1996), c'est-à-dire qu'aucun constituant n'est présupposé dans le discours. Elle peut tout aussi bien correspondre à une structure à "focus étroit" (*narrow focus*) ou à "focus contrastif" sur un des constituants de la phrase, répondant aux questions telles que *Qu'est-ce que Marie aime ?*, *Marie aime-t-elle le café ou le thé ?* ou *Qui aime le thé ?*. Tandis qu'une phrase comme *C'est le thé que Marie aime* peut généralement<sup>24</sup> correspondre à une seule structure informationnelle, celle à focus étroit (ou encore contrastif) sur le constituant *le thé*.

Cette définition<sup>25</sup> a une conséquence directe pour les études quantitatives et offre un outil (relativement) objectif pour déterminer l'ordre non marqué des constituants dans une langue donnée. En admettant cette définition, nous adoptons pour cette étude la prémisse selon laquelle l'ordre non marqué de la phrase entre deux alternatifs correspond à l'ordre qui, dans un échantillon arbitraire, apparaît avec la fréquence la plus élevée. C'est à l'aide de tests statistiques que nous déterminerons si la différence entre les fréquences observées est significative. Plus particulièrement, dans le problème qui nous intéresse ici, c'est-à-dire l'ordre relatif entre le COI et le COD, il s'agit de déterminer, dans un premier temps, quel est dans l'échantillon

24. La généralisation forte de Lambrecht (1994) relative aux phrases clivées du français a été depuis remise en question (cf. Rialland et al., 2002).

25. Il faut préciser que la définition pratique de l'ordre de base dans une langue donnée (*basic word order*, cf. §2.2.2) en termes de fréquence n'est pas une nouveauté due à Lambrecht (1994). Elle a été depuis longtemps pratiquée en typologie linguistique et en syntaxe générale universaliste. L'apport de cet auteur est de clarifier le rapport entre la fréquence et les fonctions pragmatiques, notamment quant à l'hypothèse implicite dans les travaux précédents qu'un ordre pourrait être "neutre" du point de vue pragmatique.

l'ordre le plus fréquent pour chacune des trois catégories du COD définies plus haut.

### 3.2.2.3 La variation de l'ordre dans le domaine préverbal

Une fois la définition de l'ordre canonique établie, concentrons-nous sur la variation de l'ordre entre les dépendants verbaux. Il s'agit plus précisément d'examiner les données du persan à la lumière des hypothèses supposées universelles quant aux préférences d'ordonnancement des mots (cf. §2.3.1). Bien qu'il existe un consensus sur le fait que le persan manifeste un ordre (relativement) libre des constituants – cela est en effet empiriquement incontestable –, cette question n'a jamais été abordée sous un angle quantitatif. Il s'agit d'une lacune à laquelle nous souhaitons remédier à travers l'étude des deux phénomènes suivants :

1. l'ordre relatif entre le COD et le COI dans une phrase ditransitive,
2. l'ordre relatif entre le sujet et le COD dans une phrase transitive.

Avant de poursuivre, il importe de préciser un point quant aux facteurs qui relèvent de la structure informationnelle de la phrase. Rappelons que dans la littérature, le bouleversement de l'ordre linéaire est principalement – et dans certains cas exclusivement – expliqué par des considérations discursives, telle que la topicalisation, etc. (cf. §1.3.3 au chapitre 1). Cette dimension, intéressante à plus d'un titre et relativement complexe, mérite une investigation qui dépasse le cadre de ce travail. Dans cette thèse, nous avons pris la décision, dans la mesure du possible, d'en faire abstraction, car, comme nous l'avons annoncé précédemment, nous souhaitons nous concentrer sur les facteurs extra-syntaxiques susceptibles d'intervenir dans le choix de l'ordonnancement toutes choses étant égales par ailleurs.

Dans la lignée des travaux mentionnés dans le chapitre précédent, nous effectuerons des études multifactorielles en prenant en compte les facteurs énumérés ci-dessous :

- le poids grammatical en termes de longueur (en nombre de mots),
- la définitude,
- le statut informationnel (dimension référentielle) sur une échelle binaire (donné vs. nouveau),
- le caractère animé sur une échelle binaire (humain vs. non humain),
- le lemme verbal,
- la dépendance sémantique.

Notons que tous ces facteurs ne sont pas pertinents pour les deux phénomènes étudiés. Toutefois, à ce stade, nous ne préciserons que partiellement le choix des facteurs pris en compte pour chaque phénomène. Nous aborderons l'effet du poids gramma-

tical dans la sous-section suivante (§3.2.3), où nous présenterons les prédictions des principes concurrents quant à l'effet attendu de ce facteur en persan.

Par la suite, nous esquisserons les questionnements soulevés par ces phénomènes en nous basant sur les observations exposées dans le chapitre 1. Nous nous gardons de faire une formulation explicite de nos questions de recherche, à dessein, car celles-ci devront se préciser au fur et à mesure en fonction des résultats des étapes précédentes.

**Ordre relatif entre le COD et le COI.** Force est de constater que l'ordre relatif entre ces compléments verbaux dépend du degré de la définitude du COD. En effet, indépendamment des résultats de la première étape de notre recherche, nous pouvons d'ores et déjà affirmer que le marquage différentiel intervient dans le choix de l'ordre relatif entre le COD et le COI.

Bien qu'en ce qui concerne les COD non marqués de  $=r\bar{a}$ , nous émettons des réserves sur l'hypothèse existante, nous sommes encline à affirmer que le marquage du COD par  $=r\bar{a}$  favorise sa séparation du verbe. Or, en persan, le marquage différentiel de l'objet est sensible à la hiérarchie de la définitude et/ou au caractère donné (en tant que statut informationnel). Rappelons que ces deux propriétés, bien que distinctes, sont étroitement liées.

Cet ordre ( $\text{COD}_{[+r\bar{a}]} < \text{COI}$ ) est conforme aux prédictions courantes dans la littérature quant à l'effet de la définitude en tant que facteur contribuant à l'accessibilité conceptuelle, cf. la hiérarchie du statut discursif en (16) (p. 88, §2.3.1). Par ailleurs, il y a lieu de penser que les COD nus ont une forte propension pour la position adjacente au verbe. L'ordre ( $\text{COD}_{[+r\bar{a}]} < \text{COI}$ ) est pareillement conforme à ladite hiérarchie, car les objets nus en persan se trouvent en bas de l'échelle de la définitude.

Rappelons que, d'après la hiérarchie d'accessibilité des rôles grammaticaux, cf. (13) (p. 87), l'ordre relatif entre ces deux compléments devrait correspondre à  $\text{COD} < \text{COI}$ . Ainsi, pour les COD hauts dans la hiérarchie de la définitude, l'ordre linéaire est conforme aux deux hiérarchies, tandis que pour les COD bas dans cette hiérarchie, il l'est seulement avec la hiérarchie du statut discursif.

**Ordre relatif entre le sujet et le COD.** Dans les rares travaux en syntaxe formelle sur le persan qui se penchent sur le bouleversement de l'ordre canonique entre le sujet et le COD, celui-ci est généralement évoqué en rapport avec le marquage différentiel de l'objet. Le seul cas de figure envisagé est la topicalisation du COD par  $=r\bar{a}$  qui provoquerait son positionnement en début de phrase, c'est-à-dire avant le sujet. Rappelons que ce procédé n'est pas réservé au COD et que d'autres

constituants, sous certaines conditions, peuvent être topicalisés par  $=r\bar{a}$  et placés en position initiale (cf. §1.1.3)

L'ordre canonique des constituants en persan, le SOV, respecte la hiérarchie d'accessibilité des rôles grammaticaux. Nous pouvons faire l'hypothèse, à l'instar de l'ordre relatif entre le COD et le COI, que les facteurs contribuant à l'accessibilité conceptuelle, tels que la définitude et/ou le caractère donné, pourrait favoriser le placement d'un COD en position initiale. En l'occurrence, parmi différentes catégories du COD, il semble que le COD<sub>[+rā]</sub> soit le plus disposé à précéder le sujet. Il en va de même pour un COD haut dans la hiérarchie de personne.

### 3.2.3 Questions relatives aux débats théoriques sur les préférences

Dans les recherches portant sur le fonctionnement du système langagier, que ce soit du point de vue de la production ou celui de la compréhension, la nécessité des données translinguistiques (*crosslinguistic*) (ex. Hemforth *et al.*, 2014; Jaeger & Norcliffe, 2009) est relevée à maintes reprises et la faible diversité de langues dans les études existantes est déplorée par de nombreux chercheurs.

Dans le chapitre précédent, nous avons eu l'occasion d'apprécier l'intérêt des données des langues diverses dans les recherches sur les préférences dans l'ordre des mots. À titre d'exemple, nous avons évoqué l'apport des langues à ordre flexible pour décorréler entre l'effet de l'accessibilité conceptuelle, notamment l'animéité, sur l'ordre linéaire et celui sur l'appariement des rôles grammaticaux (cf. §2.3.1). Dans cette perspective, le persan offre un terrain propice. En effet, la question de recherche consistant en l'examen de l'effet du caractère animé sur l'ordre relatif entre le sujet et le COD, que nous venons d'évoquer dans la section précédente, contribue directement à cette ligne de recherche.

Plus généralement, l'étude multifactorielle proposée par ce travail (cf. §3.2.2.3), rejoint les questions relatives aux débats théoriques sur les préférences d'ordonnement linéaire. Il s'agit, notamment, de l'effet des facteurs tels que la longueur, la définitude, le caractère défini ou animé, leur interactions et la façon dont ils interviennent pour déterminer le choix de l'ordre linéaire.

Dans cette partie, nous expliciterons la question de recherche qui consiste, plus particulièrement, en la comparaison entre des analyses concurrentes proposées pour rendre compte de l'effet du poids grammatical. Nous avons vu dans le chapitre 2 que les données d'une langue SOV, comme le japonais, ont remis en cause l'universalité du modèle de la production incrémentale, telle qu'il a été supposé rendre compte de HNPS en anglais, notamment la préférence "court-avant-long", tandis que les

analyses adoptant le point de vue du traitement (*processing*) se sont montrées compatibles avec les données des deux langues (cf. §2.3.2). Néanmoins, ultérieurement, à la suite des expériences confirmant la préférence “long-avant-court” en japonais, une analyse adoptant le point de vue de la production a été aussi proposée.

Dès lors, nous proposons de départager ces analyses concurrentes – mais complémentaires, au moins, en ce qui concerne le japonais –, en comparant entre les prédictions de chacune pour le persan, que nous expliciterons dans ce qui suit.

### 3.2.3.1 Prédications du principe *EIC/MiD* pour le persan

Le persan est une langue SOV, mais, contrairement au japonais, il manifeste une structure mixte. Comme nous l’avons vu dans le premier chapitre, en persan le SN, le SP et le CP sont à tête initiale. Nous illustrons les prédictions du modèle *EIC*, quant à l’ordre relatif entre le COD et le COI en persan, à l’aide de deux paires d’exemples. Précisons qu’en ce qui concerne l’ordre relatif entre le COD et le sujet dans une phrase transitive, les prédictions sont sensiblement identiques.

Dans le premier, nous prenons un COD marqué de =*rā*, ex. (6), et, dans le second, un COD non marqué, ex. (7).

- (6) a. in ketāb=e āmuzeš=e akkāsi=rā az ketābxāne  
ce livre=EZ apprentissage=EZ photographie=RĀ de bibliothèque  
gereft-am  
prendre.PST-1SG  
‘J’ai emprunté ce livre d’apprentissage de la photographie de la bibliothèque.’
- b. [VP [NP in ketāb=e āmuzeš=e akkāsi=rā] [PP az ketābxāne] gereft-am]  
1 2 3 4 5 6 7
- c. [VP [PP az ketābxāne] [NP in ketāb=e āmuzeš=e akkāsi=rā] gereft-am]  
1 2 3 4 5 6 7
- (7) a. yek ketāb=e āmuzeš=e akkāsi az ketābxāne  
un livre=EZ apprentissage=EZ photographie de bibliothèque  
gereft-am  
prendre.PST-1SG  
‘J’ai emprunté un livre d’apprentissage de la photographie de la bibliothèque.’
- b. [VP [NP yek ketāb=e āmuzeš=e akkāsi] [PP az ketābxāne] gereft-am]  
1 2 3 4 5 6 7
- c. [VP [PP az ketābxāne] [NP yek ketāb=e āmuzeš=e akkāsi] gereft-am]  
1 2 3 4 5 6 7

Comme l’illustrent ces paires d’exemples, il n’est pas possible de trancher en faveur

d'un ordre plutôt que l'autre, en fonction seulement du domaine de la reconnaissance des constituants (*CRD*, voir §2.3.2.1). En revanche, la mesure en temps réel (*MaOP*), ou la moyenne des rapports *IC-to-Word* pour chaque mot du *CRD*, (8) et (9), plaide en faveur des phrases (b), dont le score est légèrement plus élevé (54% vs. 46%). Ainsi, ce principe prédit un léger effet de la longueur correspondant à la préférence "court-avant-long" dans le domaine préverbal pour le persan.

- (8) a. [VP [NP in ketāb=e āmuzeš=e akkāsi=rā] [PP az ketābxāne] gereft-am]  
           1/1 1/2       1/3       1/4                   2/5 2/6  
           3/7
- b. [VP [PP az ketābxāne] [NP in ketāb=e āmuzeš=e akkāsi=rā] gereft-am]  
           1/1 1/2                   2/3 2/4       2/5       2/6  
           3/7
- (9) a. [VP [NP yek ketāb=e āmuzeš=e akkāsi] [PP az ketābxāne] gereft-am]  
           1/1 1/2       1/3       1/4                   2/5 2/6       3/7
- b. [VP [PP az ketābxāne] [NP yek ketāb=e āmuzeš=e akkāsi] gereft-am]  
           1/1 1/2                   2/3 2/4       2/5       2/6       3/7

N'omettons pas de mentionner qu'il aurait été possible de commencer la séquence *CRD* à partir du =*rā* et non du déterminant *in*. Cependant, comme en persan le marquage accusatif est différentiel, il n'y pas lieu de penser que l'analyseur traite ces séquences différemment avant qu'il ne parvienne à la fin du syntagme nominal. De ce fait, il semble judicieux de postuler un *CRD* identique dans les deux cas de figure. Notons que, dans le cas contraire, le modèle ne prédit aucun effet de la longueur. En effet, dans ce cas, il prédirait l'ordre  $COD_{[+rā]}$  - *COI*, quelle que soit la différence de longueur entre les deux constituants. Notons que cette prédiction irait dans le sens des hypothèses courantes pour le persan.

### 3.2.3.2 Prédications du modèle *DLT* pour le persan

Nous avons vu que le modèle *DLT* s'appuie sur la distance cumulée entre les dépendants verbaux et le verbe. Par conséquent, à la différence au modèle de Hawkins (2004), la direction mixte de la tête en persan n'altère pas les prédictions de ce modèle pour les langues *SOV*. En revanche, pour calculer la distance, le modèle, dans sa version étroite, ne tient compte que des expressions référentielles nouvelles. Toutefois, comme nous l'avons mentionné lors de la présentation de ce modèle, l'auteur n'exclut pas la possibilité de compter tous les mots. De plus, des travaux ultérieurs abordant ce modèle pour expliquer les préférences de l'ordre linéaire s'appuient sur

cette définition plus large (par ex. Temperley, 2007). Nous ferons de même. Notons toutefois que dans la définition étroite, un SN modifié par plusieurs syntagmes adjectivaux ne sera pas considéré comme un constituant long.

D’après ce modèle, en ce qui concerne l’ordre relatif entre le COD et le COI, comme le montre la paire d’exemples en (10) (cf. (7) ci-dessus), pour minimiser la distance entre les deux compléments verbaux et le verbe il faut déplacer le constituant long, c’est-à-dire le SN, vers la gauche<sup>26</sup>. Cela correspond donc à la préférence “long-avant-court”. Précisons que le modèle prédit la même préférence quant à l’ordre relatif entre le COD et le sujet dans une phrase transitive.

- (10) a. [<sub>NP</sub> yek ketāb=e āmuzeš=e akkāsi] [<sub>PP</sub> az ketābxāne] gereft-am]
- b. [<sub>PP</sub> az ketābxāne] [<sub>NP</sub> yek ketāb=e āmuzeš=e akkāsi] gereft-am]

### 3.2.3.3 Prédications de l’hypothèse d’accessibilité conceptuelle pour le persan

Le persan, comme le japonais, est une langue SOV, *pro drop* et à ordre flexible. Par conséquent, l’argumentation de Yamashita & Chang (2001) quant à la sensibilité du japonais aux facteurs conceptuels est également valable pour le persan. De ce fait, l’analyse en termes de la production incrémentale proposée par ces auteurs est susceptible de s’appliquer également au processus de la production de la parole en persan. Dès lors, l’effet du poids grammatical dans cette langue devrait correspondre à la préférence “long-avant-court”.

## 3.3 Conclusion

Dans ce chapitre, nous avons, en premier lieu, justifié notre approche méthodologique. Nous avons ensuite présenté les phénomènes empiriques que nous avons décidé d’examiner et les questions qui orientent notre projet de recherche.

Ces questions s’organisent en deux parties à l’image des objectifs que nous poursuivons dans cette thèse. Premièrement, nous souhaitons apporter un point de vue quantitatif sur l’étude de l’ordre des constituants en persan et par là contribuer aux débats théoriques sur la structure syntaxique la mieux adaptée pour représenter la phrase. Deuxièmement, notre objectif est d’enrichir les débats en cours sur les pré-

26. En (10-a) les distances entre le verbe et ces deux compléments sont de 1 et de 3 unités, tandis qu’en (10-b), elles sont de 4 et de 2 unités.

férences dans l'ordre des mots avec les données du persan, rarement représenté dans cette littérature. Or, le persan est une langue SOV à ordre flexible qui partage un bon nombre de propriétés avec le japonais et l'anglais, mais qui se distingue également de ces derniers sur certains aspects essentiels. Il offre de ce fait un terrain empirique favorable pour mettre à l'épreuve les hypothèses disponibles sur les préférences.

Avec ce chapitre, nous achevons la première partie de cette thèse. Nous y avons exposé les phénomènes empiriques que nous proposons d'étudier et ainsi que les prémisses théoriques sur lesquelles se fonde notre étude quantitative. La partie suivante est consacrée à la présentation détaillée de cette dernière.



Deuxième partie  
Études empiriques



## Chapitre 4

# Variation de l'ordre des mots dans le corpus Bijankhan

Ce chapitre et le suivant constituent la deuxième partie de cette thèse dédiée à la présentation de l'étude empirique que nous avons menée afin de répondre aux questions de recherche présentées en conclusion de notre première partie. Cette étude adopte une approche quantitative et repose sur deux sources de données complémentaires, à savoir le corpus et les expériences psycholinguistiques. Notre étude empirique débute par l'étude de corpus, qui nous permet d'explorer nos différentes questions de recherche à travers des échantillons de données produites naturellement, afin de tester nos hypothèses théoriques et d'identifier les facteurs pertinents. Nous affinerons ensuite ces hypothèses et étudierons ces facteurs de manière contrôlée, au chapitre 5, par des études expérimentales ciblées.

Ce chapitre est consacré à la présentation de nos études de corpus, en nous focalisant, tout d'abord, sur l'ordre relatif entre le COD et le COI dans les phrases ditransitives, dans la section 4.2 et, ensuite, sur l'ordre relatif entre le sujet et le COD dans les phrases transitives, dans la section 4.3. Les études présentées dans ces deux chapitres répondent aux différentes questions de recherche détaillées dans le chapitre 3. Ces questions relatives à la syntaxe du persan, cf. §3.2.2, ou relatives aux débats théoriques sur les préférences, cf. §3.2.3, seront abordées au fur et à mesure de l'avancement de la recherche.

Comme nous l'avons évoqué précédemment (cf. §3.1.2), en syntaxe, la population peut être définie comme l'ensemble des énoncés d'une langue donnée présentant le phénomène étudié. Dans notre étude, nous nous focaliserons sur deux populations de phrases en persan : 1) les phrases transitives dans lesquelles nous nous intéressons à l'ordre relatif entre le sujet et l'objet et 2) les phrases ditransitives dans lesquelles nous nous intéressons à l'ordre relatif entre l'objet direct et l'objet indirect. Pour la

constitution de nos échantillons, nous nous sommes basée sur un corpus existant, le corpus de Bijankhan que nous présentons dans la première section de ce chapitre.

## 4.1 Le corpus utilisé

Les données utilisées dans cette étude sont extraites d'un corpus de référence, en accès libre, le plus souvent utilisé dans les travaux dans le domaine de la linguistique de corpus et du traitement automatique des langues sur le persan : le corpus de Bijankhan (Bijankhan, 2004), noté BK par abréviation. Il est extrait du quotidien *Hamšahri* et contient environ 2,6 million de tokens étiquetés manuellement pour les parties-de-discours. Il s'agit – à notre connaissance – du seul corpus annoté de taille considérable pour le persan que l'on peut librement télécharger en format brut pour l'utiliser à sa guise<sup>1</sup>.

Le corpus de Bijankhan a été créé au sein du groupe de recherche sur la base de données (*DataBase Research Group*<sup>2</sup>) du Département de l'ingénierie électrique et informatique de l'Université de Téhéran. Sur son site, le corpus est présenté comme suit :

Bijankhan corpus is a tagged corpus that is suitable for natural language processing research on the Persian (Farsi) language. This collection is gathered from daily news and common texts. In this collection all documents are categorized into different subjects such as political, cultural and so on. Totally, there are 4300 different subjects. The Bijankhan collection contains about 2.6 million manually tagged words with a tag set that contains 40 Persian POS tags.

Deux versions de ce corpus étaient disponibles pour le téléchargement, au moment où nous nous y sommes intéressée. La première, présentée dans le paragraphe ci-dessus, est une version traitée du corpus original. Comme précisé dans ce paragraphe, il s'agit des modifications apportées en vue de rendre le corpus plus apte aux besoins de traitement automatique, notamment la simplification des étiquettes. Le jeu d'étiquettes de cette version contient 40 étiquettes, contre 550 pour la version originale (cf. Oroumchian et al., 2006; Bijankhan et al., 2011). Cette version se télécharge comme un fichier au format texte brut (.txt) Unicode, dont la taille

---

1. Au moment où nous avons débuté cette recherche, le corpus était téléchargeable depuis son site <http://ece.ut.ac.ir/dbrg/bijankhan/> Mais il semble que le site n'est plus accessible. Une présentation du corpus est disponible à l'adresse <http://dadegan.ir/catalog/bijankhan> [dernière consultation le 23/09/2016].

2. <http://ece.ut.ac.ir/dbrg/>

s'élève à 2 597 937 lignes, dans lequel chaque ligne correspond à un token en persan suivi de son étiquette en script latin.

La seconde, la version originale du corpus, en revanche, est une collection de 814 fichiers en format LBL. Chaque fichier comprend un ou plusieurs morceaux de texte séparés par l'étiquette **SUBJ** qui indique le sujet (politique, sport, etc.) du texte, téléchargeable en un fichier compressé et décrit de la manière suivante :

This file is a compressed version of the whole corpus in LBL text format.

This file contains the original Bijankhan corpus without any changes that was manually tagged and prepared at Research Center of Intelligent Signal Processing (RCISP). Its tag set contains 550 tags and totally it contains 4300 subject categories.

Étant donné que l'annotation de la version originale est plus riche et que les modifications apportées dans la version traitée sont toutes des transformations automatiques qui risquent d'augmenter la part d'erreur systématique dans le corpus, sans y ajouter d'informations pertinentes en ce qui concerne notre usage du corpus, c'est cette première version que nous avons utilisée pour notre étude. Ainsi, nous avons compilé les 814 fichiers dans un seul fichier texte dont la taille s'élève à 2 602 536 lignes.

Afin de pouvoir utiliser le corpus pour la constitution de nos échantillons, nous avons dû y apporter certaines transformations à l'aide du langage de programmation Python, notamment la normalisation des caractères et de l'écriture, la translittération et, plus importante pour notre étude, la lemmatisation des verbes.

Il faut noter que le corpus Bijankhan n'est pas délimité en phrases et ne porte pas d'annotations sur les relations syntaxiques. De ce fait, l'extraction des phrases pour nos échantillons ne pouvait se faire de manière automatique. Par conséquent, nous avons recouru à la méthode de l'échantillonnage arbitraire et à la sélection manuelle et nous avons procédé à la lemmatisation des verbes pour faciliter cette démarche. En effet, contrairement aux langues comme l'anglais ou le français, le nombre de lemmes verbaux (simples) est restreint en persan (moins de 250, cf. §1.1.4) et il est donc possible de lister les verbes (potentiellement) ditransitifs, plus précisément ceux susceptibles de prendre un COD et un COI, ainsi que les verbes (potentiellement) transitifs.

Dans le processus de lemmatisation, les conjugaisons informelles, ainsi que les orthographes alternatives, ont également été répertoriées et regroupées dans la liste afin de pouvoir attribuer le même lemme aux différentes variantes stylistiques d'un même verbe. Précisons qu'une information sur le registre (formel vs. informel) a été ajoutée à l'étiquette des tokens. De plus, celui-ci visait également les conjugaisons

périphrastiques dans la mesure du possible. Nous présentons brièvement l'algorithme que nous avons utilisé pour la lemmatisation dans les annexes (cf. Annexe A).

À la suite de cette lemmatisation, nous avons répertorié les verbes lemmatisés, plus précisément les tokens verbaux, dans une base de données, comportant 185 015 lignes de données au total.

Il faut noter que le processus de la lemmatisation n'a pas été effectif pour tous les verbes du corpus. En effet, certaines occurrences n'ont pas pu être identifiées par l'algorithme en raison notamment des erreurs d'étiquetage, des erreurs d'orthographe ou de leur forme irrégulière (non répertoriée). Ces occurrences n'ont pas été extraites et ne sont pas incluses dans la base de données. Aussi, dans le cas où la forme verbale est périphrastique, c'est-à-dire comportant un ou plusieurs auxiliaires, c'est l'ensemble de la périphrase qui constitue l'entrée verbale, l'étiquette étant mise à jour en conséquence.

Le nombre de lemmes verbaux simples identifiés dans le corpus est de 228 types. Ce nombre s'élève à 322 quand nous considérons la combinaison d'un verbe simple et d'une particule comme un lemme à part entière. Cela pourrait se justifier par le fait que ces particules n'ont pas d'autonomie syntaxique ni sémantique et ne présentent qu'une mobilité restreinte, dans la mesure où ils peuvent se séparer du verbe seulement par les enclitiques ou les préfixes flexionnels. En outre, le sens lexical et la structure argumentale sont attribuées à l'ensemble en tant qu'une unité lexicale, voir (1) et (2). Cela est pertinent dans notre étude, car, pour constituer nos tables de données, nous devons extraire les phrases en fonction du nombre de compléments verbaux présents qui dépendent de la structure argumentale du verbe.

- (1) a. āvardan 'emmener' → āvardan(x,y) : 'x emmène y'  
 b. āvardan 'emmener' → āvardan(x,y,z) : 'x emmène y à/de z'  
 c. dar-āvardan 'sortir (tr.)' → dar-āvardan(x,y,z) : 'x sort y de z'
- (2) a. dāštan 'avoir' → dāštan(x,y) : 'x a y'  
 b. bāz-dāštan 'empêcher' → bāz-dāštan(x,y,z) : 'x empêche y de (faire) z'

Les verbes lemmatisés ont été répertoriés dans une table où chaque verbe conjugué constitue une ligne. En plus du token et son lemme, chaque ligne contient un certain nombre d'informations automatiquement encodées, dont :

- **REF** : Index du token dans le corpus
- **TOKEN** : Verbe
- **INF** : Variable binaire précisant s'il s'agit d'un infinitif ou d'une forme fléchie
- **TAM** : Étiquette originale avec les mises à jour lors de la lemmatisation
- **NEG** : Variable binaire précisant la polarité du verbe

- **PER** : Personne du verbe
- **CL** : Enclitique personnel le cas échéant
- **LEMME** : Lemme verbal (avec particule, le cas échéant)
- **LEMME-SIMPLE** : Lemme verbal sans particule
- **PART** : Particule (s'il s'agit d'un verbe à particule)
- **LEM-FREQ** : Fréquence du lemme dans le corpus<sup>3</sup>
- **LEMSIMPLE-FREQ** : Fréquence du lemme simple le cas échéant
- **REG** : Registre lorsque cette information est disponible

La phrase dans laquelle se trouve le token verbal est également extraite avec une fenêtre de  $[-20,+5]$ , sachant qu'il est toujours possible de situer chaque token dans le corpus à l'aide de la variable **REF**. C'est à partir de cette base de données que nous avons extrait les tables de données pour nos différentes études de corpus.

## 4.2 L'ordre relatif entre le COD et le COI dans les phrases ditransitives

Nous commençons par l'étude de l'ordre relatif entre le COD et le COI dans les phrases ditransitives. Une des questions principales à laquelle cette thèse s'attelle est l'ordre canonique entre le COD et le COI dans une phrase ditransitive, qui constitue une question essentielle dans l'étude de la structure syntagmatique de la phrase en persan.

L'objectif premier de notre étude de corpus est d'examiner la validité empirique de l'hypothèse théorique que nous avons détaillée dans la première partie (cf. §3.2.2.1) que l'on désigne comme Hypothèse MDO ci-après. Pour rappel, selon cette hypothèse, dans l'ordre canonique dans une phrase ditransitive, le COD marqué de  $=r\bar{a}$  précède le COI tandis que le COD non marqué le suit.

- (3) Hypothèse MDO :
- a. COD<sub>[+RĀ]</sub> COI V
  - b. COI COD<sub>[-RĀ]</sub> V

---

3. Le calcul de fréquence n'est pas tout à fait exact pour certains lemmes verbaux pour lesquels la lemmatisation a été problématique. Il s'agit des lemmes qui connaissent des formes homographes. C'est le cas, à titre d'exemple, des lemmes tels que *kešidan* 'tirer' et *koštan* 'tuer'. En effet, les radicaux du présent des deux verbes s'écrivent de manière identique : *kš*. Lors de la lemmatisation automatique – uniquement basée sur la forme – il était impossible de distinguer entre les deux types pour les verbes formés sur la base du radical présent. Nous avons pris la décision dans ce cas de lemmatiser toutes ces formes comme *kešidan*, car c'est le verbe le plus fréquent entre les deux lemmes en ce qui concerne les formes basées sur le radical passé (422 contre 388) et au niveau de l'infinitif (113 contre 44).

Pour cela, nous étudierons la distribution de l'ordre relatif entre ces trois constituants dans un échantillon arbitraire extrait du corpus Bijankhan que nous avons décrit dans la section précédente. Selon la définition que nous adoptons pour l'ordre canonique ou non marqué, cf. §3.2.2.2, cette hypothèse sera validée si les fréquences observées pour l'ordre prédit dans chacune des deux cas de figure en (3) sont significativement plus élevées que l'ordre alternatif.

Nous testerons cette hypothèse en évaluant tout d'abord sa force de prédiction à l'aide de la méthode de la matrice de confusion, qui est une table de contingence permettant d'opposer les occurrences prédites (ou théoriques) aux occurrences observées et, ensuite, par des modèles de régression.

Mis à part le marquage par  $=r\bar{a}$ , comme nous avons pu également le constater dans la première partie de la thèse (cf. §3.2.2.1), un autre paramètre qui semble de prime abord pertinent dans la détermination de l'ordre relatif entre les deux objets est la “nudité” du COD. En effet, la dépendance sémantique entre le COD nu et le verbe, qui rappelle celle entre l'élément nominal et le verbe dans un prédicat complexe, devrait entraîner l'adjacence et se traduire dans une forte inclination de ces COD nus pour se placer à côté du verbe après le COI.

Rappelons qu'outre la question de l'ordre canonique particulièrement intéressante pour l'étude de la syntaxe du persan, nous nous intéressons aux facteurs susceptibles de provoquer de la variation dans l'ordre linéaire relatif entre les constituants, en premier lieu la longueur en nombre de mots.

Dans la suite de cette section, nous détaillons la façon dont nous avons constitué notre échantillon et préparé notre première table de données dont nous avons été amenée à augmenter la taille par la suite. Nous présenterons une analyse statistique multifactorielle de cette table de données et discuterons des résultats de celle-ci.

### 4.2.1 Constitution de l'échantillon

La population dont nous voulons étudier l'ordre canonique est celle des phrases ditransitives. Comme nous l'avons précisé plus haut, le corpus dont nous disposons n'est pas annoté pour les relations syntaxiques. Par conséquent, l'extraction des phrases ditransitives ne peut se faire sur la base de la structure syntaxique ou argumentale des phrases. Par contre, du fait du nombre restreint des verbes (simples) en persan, nous pouvons lister les verbes potentiellement ditransitifs et procéder à un échantillonnage sur la base de lemmes verbaux. Pour ce faire, nous utilisons la base de données dans laquelle nous avons répertorié les verbes lemmatisés du corpus.

#### 4.2.1.1 Verbes (potentiellement) ditransitifs

Une fois dressée la liste des lemmes verbaux de la base de données, nous avons identifié les verbes (verbes simples et verbes à particule confondus) qui pourraient potentiellement projeter une construction ditransitive. Le critère de sélection avait une portée large et nous avons inclus les verbes qui pouvaient, même très marginalement, correspondre à un schéma ditransitif.

Nous avons sélectionné 122 lemmes verbaux qui font un total de 45 134 tokens. Le Tableau 4.1 présente la liste des quinze verbes les plus fréquents qui constituent 90% du total des tokens identifiés<sup>4</sup>.

Lemme	Traduction <sup>5</sup>	Occurrences
dādan	‘donner’	10 494
goftan	‘dire’	9 418
gereftan	‘prendre’	6 849
afzudan	‘ajouter’	2 394
āvardan	‘emmener’	1 915
bordan	‘amener’ ou ‘gagner’	1 560
gozāštan	‘mettre’	1 496
neveštan	‘écrire’	1 457
zadan	‘frapper’	1 304
pardāxtan	‘payer’	1 045
kešidan <sup>6</sup>	‘tirer’	734
resāndan	‘faire arriver’	693
xordan	‘manger’ ou ‘heurter’	636
paziroftan	‘accepter’	535
andāxtan	‘jeter’	310
šenidan	‘entendre’	303
bar-dāštan	‘enlever’	294
baxšidan	‘pardonner’ ou ‘octroyer’	276
nahādan	‘mettre’ (soutenu)	254
rixtan	‘verser’	254
bastan	‘fermer’ ou ‘attacher’	247
porsidan	‘demander’	231
ferestādan	‘envoyer’	219

TABLE 4.1 – Corpus BK : Liste partielle des verbes potentiellement ditransitifs

4. Le verbe *kardan* ‘faire’ est le “verbalisateur” par excellence en persan (c’est-à-dire en tant qu’élément verbal des prédicats complexes) et, sans surprise, le verbe le plus fréquent (avec un total de 33 454 occurrences). Bien que, dans son emploi lexical, ce verbe ne projette pas une structure argumentale ditransitive, son emploi en tant que “verbalisateur” pourrait correspondre à un schéma ditransitif. Nous avons écarté ce lemme, car, étant donné sa fréquence très élevée dans le corpus, nous risquions d’en trouver une portion très élevée dans notre échantillon.

5. La plupart de ces lemmes sont polysémiques. La glose fournie ici correspond à la première

#### 4.2.1.2 Échantillonnage arbitraire

Nous avons extrait un échantillon arbitraire de 2 000 tokens, c'est-à-dire des verbes fléchis ditransitifs à partir de notre base de données<sup>7</sup>. Dans un tableur *Excel*, nous avons inventorié les 2 000 occurrences avec les phrases dans lesquelles ils se trouvent, c'est-à-dire une fenêtre de [-20,+5] tokens entourant le verbe, et nous avons procédé à l'identification manuelle des occurrences correspondant à notre objet d'étude, c'est-à-dire les propositions dans lesquelles le verbe est précédé d'(au moins) un complément nominal et d'(au moins) un complément prépositionnel.

#### 4.2.1.3 Sélection manuelle

Les occurrences de construction ditransitive de notre échantillon arbitraire présentent une variété de cas de figure, tels que propositions matrices ou principales, propositions avec sujet implicite ou explicite, présence de modificateurs ou d'autres compléments prépositionnel entre les constituants principaux. Nous avons décidé d'écarter un certain nombre de cas de figure de notre sélection et d'ignorer les autres sources de variation, car, étant donné la rareté des occurrences dans chaque configuration possible, il n'était pas envisageable d'inclure ces variables dans notre étude. Ainsi, nous avons restreint notre sélection aux seules occurrences correspondant exactement à l'un des patrons SP-SN-V ou SN-SP-V sans se préoccuper du reste de la phrase, par exemple la présence de modificateurs ou d'autres compléments prépositionnels, la réalisation du sujet et le type de proposition, cf. (4).

- (4) a. *bā satl* [<sub>SP</sub> *bar ru=ye ham*] [<sub>SN</sub> *āb*] [<sub>V</sub>  
avec seau à sur=EZ l'un l'autre eau  
*mi-pāš-and*]  
IPFV-asperger.PRS-3PL  
'... avec un seau ils s'aspergent d'eau'
- b. *kāregar=e exrāji* [<sub>SN</sub> *asid*] [<sub>SP</sub> *be rudxāne*] [<sub>V</sub> *rixt*]  
ouvrier=EZ licencié acide à rivière jeter.PST.3SG  
'... l'ouvrier licencié jeta de l'acide dans la rivière'
- c. *be šahr āmad tā* [<sub>SP</sub> *barāye kamān=aš*] [<sub>SN</sub> *čand*  
à ville venir.PST.3SG jusqu'à pour arc=3SG quelques  
*tir*] [<sub>V</sub> *be-xar-ad*]  
flèche SBJV-acheter.PRS-3SG

---

traduction donnée par le dictionnaire bilingue persan-français de Lazard (Lazard & Ghavam-Nejad, 1990). Dans le cas de l'homonymie seulement, les acceptions multiples sont fournies.

6. Le nombre 734 représente les 422 occurrences de la forme passée du verbe *kešidan* plus l'ensemble des occurrences de la forme présente des verbes *kešidan* et *koštan* (cf. la note de bas de page 3, p. 133 dans la section 4.1).

7. Pour ce faire, nous avons utilisé la fonction `random.sample()` du langage de programmation de Python pour générer 2 000 nombres arbitraires.

- ‘... il est venu en ville pour acheter quelques flèches pour son arc’
- d. *kāregar-hā be surat=e eštabāhi dar zamān=e hādese* [SN 16  
ouvrier-PL à façon=EZ faux dans temps=EZ incident 16  
*kilugeram orāniyom*] [SP *dar tānker*] [V *rixte=and*]  
kilogramme uranium dans réservoir jeter.PP=COP.3PL  
‘... au moment de l’incident, les ouvriers ont jeté par erreur seize kilos  
d’uranium dans le réservoir’
- e. *tā šānzdah=e māh=e Novāmr* [SN *vām-hā=ye daryāfti=ye*  
jusqu’à seize=EZ mois.EZ Novembre crédit-PL=EZ reçu=EZ  
*xod=rā*] [SP *be bānk*] [V *bāz-gardān-and*]  
soi=RĀ à banque retourner.PRS-3PL  
‘... (ils doivent) retourner les crédits reçus à la banque avant le seize  
novembre’
- f. *va* [SP *az in čāh=e margbār=i ke xod ijād karde*  
et de ce puits=EZ funeste=RSTR que soi création faire.PP  
*bud*] [SN *ānhā=rā*] [V *be-rahān-ad*]  
être.PST.3SG eux=RĀ SBJV-libérer.PRS-3SG  
‘... et les libérer de ce puits funeste qu’il avait créé lui-même’

Il importe de s’attarder sur la façon dont nous avons traité les combinaisons formant des prédicats complexes. Comme nous l’avons mentionné auparavant, nous adhérons à la vision qui, sur le plan syntaxique, traite l’élément nominal d’un prédicat complexe comme un complément nominal du verbe simple (cf. §1.1.4). C’est-à-dire concrètement que, pour identifier les occurrences correspondant aux patrons SP SN V ou SN SP V, nous n’avons pas pris en compte le lien sémantique ou idiomatique entre ces trois constituants. Ainsi, comme le montre la paire d’exemples en (5), nous ne tenons pas compte du fait que *ta’sir gozāštan* ‘impacter (Lit. impact mettre)’ est un prédicat complexe.

- (5) a. [SP *dar zendegi=ye ānhā*] [SN *ta’sir=e manfi*] [V  
dans vie=EZ eux impact=EZ négatif  
*mi-gozār-ad*]  
IPFV-mettre.PRS-3SG  
‘... il a un mauvais impact sur leurs vies’
- b. [SN *ta’sir=e manfi*] [SP *bar dānešgāh-hā*] [V *mi-gozār-ad*]  
impact=EZ négatif sur université-PL IPFV-mettre.PRS-3SG  
‘... il a un mauvais impact sur les universités’

Ce choix est discutable. Dans le cas de certains prédicats complexes fortement lexicalisés, il est n’est pas (ou plus) évident, même d’un point de vue purement syntaxique, de traiter l’élément nominal comme un complément d’objet direct du verbe simple. C’est le cas de l’exemple (6), dans lequel la séquence N-V forme un prédicat

complexe *qarār gereftan* 'se placer, s'installer (Lit. stabilité/calme prendre)'<sup>8</sup>.

- (6) tim=e Tehran dar dore=ye pišin=e in reqābat-hā dar  
 équipe=EZ Téhéran dans cercle=EZ dernier=EZ ce compétition-PL dans  
 rade=ye noxost **qarār gerefte bud**  
 rang=EZ premier stabilité prendre.PP être.PST.3SG  
 '... l'équipe de Téhéran s'était placée première dans le dernier tournoi.'

En effet, à cause du figement de la séquence, cet élément ne présente pas les propriétés typiques d'un complément d'objet direct. Non seulement il ne peut pas prendre l'enclitique =*rā*, ni être passivé, mais aussi il ne peut pas être modifié. Il a une mobilité très réduite s'approchant de celle des particules : il ne peut se séparer du verbe que par des suffixes flexionnels, les enclitiques personnels et l'auxiliaire de futur (cf. §1.1.4). Ce dernier pourrait effectivement rendre l'intégration de ce type d'occurrences problématique en introduisant un biais dans notre étude sur l'ordre linéaire, du fait que l'ordre relatif entre les deux constituants est restreint comme par une contrainte de grammaticalité<sup>9</sup>.

Or, le choix d'exclure ces cas n'est pas pour autant moins problématique. La question qui se pose alors est comment définir le critère d'exclusion, en particulier, sachant que les prédicats complexes fortement lexicalisés pour lesquels l'ordre linéaire présente une variation, comme en (5) ci-dessus, ne sont pas rares, cf. (7)<sup>10</sup>. En l'absence d'un ensemble solide de critères formels pour distinguer entre les deux types de combinaisons, autrement dit pour décider à quel niveau du continuum s'arrêter, le choix d'exclusion pourrait aussi falsifier les données en y injectant une part de subjectivité.

- (7) a. [be e'tesāb] [dast] zad-and vs. [dast] [be e'tesāb] zad-and  
 à grève main frapper.PST-3PL  
 'ils ont entrepris une grève.' : *dast zadan* 'entreprendre (Lit. main frapper)'
- b. [az hāl=e ānhā] [xabar] dād vs. [xabar] [az hāl=e  
 de état=EZ eux nouvelle donner.PST.3SG  
 ānhā] dād

8. Notons que le verbe *gereftan* 'prendre' a un emploi lexical avec une structure argumentale ditransitive : prendre(x,y,z) : x prend y de z. Aussi, *qarār gereftan* peut avoir une interprétation compositionnelle, 'prendre du calme ou se calmer', dans laquelle *qarār* maintient son sens littéral 'calme'.

9. Cela étant dit, nous nous plaçons dans une approche basée sur l'usage. Dans une telle approche, nous ne pouvons affirmer cela de manière intuitive et sans passer par les données d'usage à grande échelle. Une raison de plus pour laquelle nous n'avons pas exclu ces cas de notre étude.

10. Dans ces paires de phrase, la variante dans laquelle les deux composantes sont séparées est tirée du Corpus Bijankhan.

- 'il (les) a informé de leur état' : *xabar dādan* 'informer (Lit. nouvelle donner)'
- c. [az in] [nešān] dār-ad ke ... vs. [nešān] [az in] dārd ke ...  
de ce signe avoir.PRS-3SG que  
'il montre que...' : *nešān dādan* 'montrer (Lit. signe donner)'
- d. [be in kār] [edāme] dād vs. [edāme] [be in kār] dād  
à ce travail suite donner.PST.3SG  
'il a poursuivi ce travail/cette activité' : *edāme dādan* 'poursuivre, continuer (Lit. suite donner)'

Par conséquent, nous avons pris la décision de n'exclure aucun cas de figure sur la base du lien sémantique ou collocationnel avec le verbe. Toutefois, afin de contrôler le biais que cela pourrait introduire dans les données, nous tenons compte du degré de figement, plus précisément de collocationalité, à travers une variable numérique calculée automatiquement sur l'ensemble du corpus, que nous présentons dans le paragraphe suivant.

Ainsi, dans le cas des occurrences comme l'exemple (6), nous délimitons la proposition en prenant en considération les deux constituants qui précèdent immédiatement le verbe, cf. (8).

(8) ... [SP dar rade=ye noxost] [SN qarār] [V gerefte bud]

**Mesure du degré de collocationalité.** Pour rendre compte du lien collocationnel entre un verbe et son complément, nous avons défini une variable numérique, **PREDPROT-FREQ**, basée sur la fréquence de schéma prototypique des prédicats complexes. C'est-à-dire celui d'un verbe simple précédé immédiatement par un nom nu (N-V), ex. *harf zadan* 'parler (Lit. parole frapper)', ou par un syntagme prépositionnel minimal (P-N-V), ex. *be ohde gereftan* 'prendre en charge (Lit. à responsabilité prendre)', *be kār bordan* 'utiliser (Lit. à travail prendre)', *az dast dādan* 'perdre (Lit. de main donner)' ou *az beyn raftan* 'disparaître (Lit. de entre aller)'.

Pour ce faire, nous avons défini **PREDPROT** pour extraire des bigrammes N-V et des trigrammes P-N-V selon l'algorithme ci-dessous (le verbe comptant comme token<sub>0</sub>), pour lesquels nous avons calculé la fréquence totale sur l'ensemble du corpus, c'est-à-dire 185 015 verbes, **PREDPROT-FREQR**<sup>11</sup>.

11. Nous n'avons pas pondéré ces fréquences en fonction des fréquences des lemmes verbaux, car, étant donné le nombre restreint de lemmes verbaux et leur distribution particulièrement déséquilibrée, nous nous trouvons avec des ratios trompeurs. Pour être plus concret, un verbe comme *dādan* constitue, avec une fréquence de 10 494, environ un quart du total des 45 134 verbes ditransitifs, alors que nous avons seulement 10 hapax. La fréquence absolue nous est apparue plus pertinente

- si le  $token_{-1}$  est un Nom, alors regarder le  $token_{-2}$  :
  - si le  $token_{-2}$  est une Préposition, alors  $PREDPROT = token_{-2} token_{-1}$

*Lemme*

- si non,  $PREDPROT = token_{-1}$  *Lemme*
- si non,  $PREDPROT = NA$

Les éléments nominaux des prédicats complexes sont susceptibles d'avoir une valeur élevée pour cette variable, tandis que les COD nus ordinaires, c'est-à-dire qui n'entretiennent pas de rapport collocationnel avec le verbe, sont attendus d'avoir une valeur proche de zéro. Par exemple, la séquence *qarār gereftan* a la fréquence la plus élevée, 2 308 occurrences, parmi tous les bigrammes et trigrammes, mais des séquences comme *sorxāb mālidan* 'appliquer du rouge' ou *šen rixtan* 'verser du sable' n'apparaissent qu'une seule fois dans le corpus.

Cette mesure a l'avantage d'être complètement indépendante du jugement d'annotateur, mais elle présente également plusieurs inconvénients. Le problème le plus important est que cette mesure est approximative dans le sens où elle ne prend pas en compte la structure de la phrase, mais seulement les deux derniers mots avant le verbe<sup>12</sup>. De ce fait, elle n'est efficace que pour identifier : 1) les cas où le SN est un nom nu et entretient un rapport collocationnel avec le verbe, et 2) les cas où le SP a une structure minimale (c'est-à-dire Prép + Nom nu) et entretient un rapport collocationnel avec le verbe. Ces cas correspondent aux occurrences pour lesquelles la valeur de cette variable est relativement élevée. Dans les autres cas de figure, celle-ci n'est pas informative.

Un autre problème est que cette mesure est relative et dépendante du corpus. Ainsi, l'on peut obtenir des bigrammes ou trigrammes ayant des fréquences basses, mais qui toutefois forment des prédicats complexes, ex. *gāz zadan* 'mordre, croquer' (Lit. morsure frapper). La valeur *NA* et les valeurs basses de cette variable ne signifient pas forcément l'absence de lien collocationnel.

Pour cette raison, nous avons défini une variable catégorique pour représenter le degré de collocationalité. Après avoir passé en revue les fréquences de différentes séquences et avoir essayé différentes modalités pour cette variable, nous avons opté pour la conservation d'une seule catégorie et avons annoté les séquences dont la

---

à ce sujet.

12. À titre d'exemple, dans l'occurrence ci-dessous, cet algorithme repère le bigramme *e'temād dādan*, dont la fréquence totale dans le corpus est de 2, tandis que le nom *e'temād* fait partie ici du SN *hess=e e'temād*.

- (i) ... [be u] [hess=e e'temād] be-dah-id  
 à lui sentiment=EZ confiance IMP-donner.PRS-2PL (...)  
 '... donnez-lui un sentiment de confiance...'

fréquence est supérieure à 10 comme collocationnelles. Nous obtenons donc trois niveaux en prenant en compte le type de séquence : SN-Coll, SP-Coll et Nul<sup>13</sup>.

## 4.2.2 Table de données préliminaire

Notre première table de données, désignée ci-après par l'abréviation DS1, comporte 540 occurrences<sup>14</sup> qui correspondent chacune au patron SP-SN-V ou SN-SP-V. Ces occurrences sont distribuées sur 82 lemmes verbaux. La liste des 20 lemmes les plus fréquents se trouve dans le Tableau 4.2 ci-dessous<sup>15</sup> et la distribution des verbes en fonction de la fréquence du lemme dans l'ensemble du corpus en Figure 4.1.

Lemme	Traduction	Occurrences en DS1
gereftan	'prendre'	66
āvardan	'amener'	28
dādan	'donner'	27
gozāštan	'mettre'	25
bordan	'emmener' ou 'gagner'	23
andāxtan	'jeter'	20
zadan	'frapper'	20
resāndan	'amener'	18
bardāštan	'prendre'	15
ferestādan	'envoyer'	14
resānidan	'amener'	14
nahādan	'mettre' (soutenu)	13
kešāndan	'pousser'	12
gozārdan	'mettre' ou 'accomplir' (soutenu)	12
rixtan	'jeter'	11
gozidan	'sélectionner'	11
kešidan	'tirer'	10
rāndan	'faire partir'	10
dar-āvardan	'sortir (tr.)'	9
bar-gardāndan	'rendre' ou 'faire retourner'	9

TABLE 4.2 – DS1 : Lemmes verbaux et leurs fréquences

13. Il faut noter que nous avons corrigé cette variable pour les bigrammes N-V avec *čiz=i* 'quelque chose', annoté comme Nom dans le corpus, qui, malgré leurs fréquences élevées, pour des raisons compréhensibles, ne font pas partie des schémas prototypiques des prédicats complexes.

14. Nous avons écarté deux occurrences relevant de la poésie, ainsi que l'occurrence isolée ci-dessous dans laquelle nous avons un COD pronominal non marqué.

(i) ... ān-če az man porsid-and (...)   
 ce-que de moi demander.PST-3PL (...)   
 '... ce qu'on m'a demandé ...'

15. Voir Annexe B pour la liste complète.

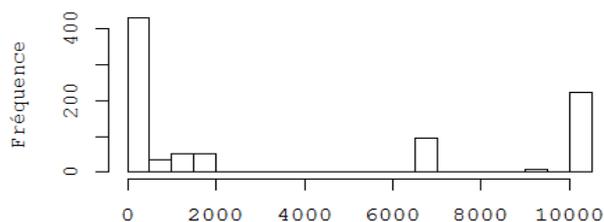


FIGURE 4.1 – DS1 : Histogramme des verbes par fréquence totale du lemme en BK

Pour chacune des occurrences, nous avons annoté manuellement les variables suivantes :

- **ORDRE** : variable binaire avec les valeurs **SN-SP** et **SP-SN**.
- $[\pm R\bar{A}]$  : variable binaire avec les valeurs  $[+R\bar{A}]$  et  $[-R\bar{A}]$ .
- $[\pm Nu]$  : variable binaire avec les valeurs  $[+Nu]$  et  $[-Nu]$ , sachant que par  $[+Nu]$  nous entendons COD nu monolexémique.

La distribution de l'ordre dans DS1, cf. Tableau 4.3, montre que l'ordre prédominant est l'ordre **OD-OI** à 59,5%. Quant à la distribution de l'ordre en fonction de nos deux variables, nous constatons que 96,1% des COD marqués de  $=r\bar{a}$  précèdent le complément prépositionnel, 87,1% des COD nus suivent le complément prépositionnel et 63,4% des COD “intermédiaires”, c'est-à-dire  $COD_{[-R\bar{A}]} COD_{[-Nu]}$ , précèdent le complément prépositionnel.

	COD $+R\bar{a}$	COD $+Nu$	COD intermédiaire	Total
SN-SP	249	27	45	321
SP-SN	10	183	26	219
Total	259	210	71	540

TABLE 4.3 – DS1 : Distribution de l'ordre par les variables  $[\pm R\bar{A}]$  et  $[\pm Nu]$ 

#### 4.2.2.1 Marquage différentiel de l'objet

Lorsque l'on considère le critère du marquage différentiel de l'objet pour déterminer l'ordre relatif entre les deux compléments, nous constatons que les COD non marqués ne constituent pas une classe homogène. En effet, bien que les  $COD_{[+R\bar{A}]}$  présentent un comportement fortement uniforme étant placés à 96,1% après le COI, comme prédit par l'hypothèse MDO (cf. §3.2.2.1), les  $COD_{[-R\bar{A}]}$  manifestent un comportement moins uniforme avec seulement 74,4% dans l'ordre prédit par cette hy-

pothèse.

	Prédit (SN-SP)		% correct
Observé (SN-SP)	0	1	
0	209	10	74,4%
1	<b>72</b>	249	96,1%
Total	281	259	84,8%
$\chi^2$ d'ajustement ( <i>goodness-of-fit</i> ) = 29,528			

N.B. : % correct avec une prédiction par défaut de SN-SP : 59,5%

TABLE 4.4 – DS1 : Matrice de confusion à partir de l'hypothèse MDO

Nous évaluons la force de prédiction de cette hypothèse à l'aide de la méthode de la matrice de confusion. L'exactitude (*accuracy*) d'un modèle de prédiction se mesure par le taux total des prédictions correctes, c'est-à-dire la proportion des occurrences pour lesquelles la prédiction du modèle correspond à l'observation. La qualité d'ajustement du modèle (*goodness-of-fit*) se mesure à l'aide du test d'ajustement de  $\chi^2$  : plus la valeur de  $\chi^2$  obtenue est petite, meilleure est la qualité de la prédiction.

Le tableau 4.4 ci-dessus montre qu'un modèle de prédiction de l'ordre basé sur le MDO permet d'obtenir une exactitude de 84,8%, mais la valeur du test  $\chi^2$  est relativement élevée.

Nous constatons que cette note élevée d'exactitude est due au fait que les COD<sub>[+R $\bar{A}$ ]</sub> présentent une variation triviale, qui est plus petite que le seuil d'erreur accepté dans beaucoup d'études scientifiques, *i.e.* 0,05. Or, les COD<sub>[-R $\bar{A}$ ]</sub> présentent une variation beaucoup plus élevée. Ainsi, bien que le marquage différentiel de l'objet permet de prédire l'ordre relatif dans 84,8% des cas dans nos données, nous constatons que ce critère occulte une partie de la réalité en ce qui concerne le comportement des COD non marqués.

La variable NU que nous avons annotée en parallèle avec la variable R $\bar{A}$  permet de voir que cette baisse d'uniformité provient des COD que nous avons qualifiés d'intermédiaires. En effet, alors que les COD<sub>[+NU]</sub> se comportent avec une homogénéité comparable au COD<sub>[+R $\bar{A}$ ]</sub>, les COD "intermédiaires" présentent une variation considérable.

Nous avons alors décidé d'affiner la catégorie COD<sub>[-NU]</sub> en séparant les COD ne portant aucune détermination mais portant des modifieurs, des COD portant une détermination (avec ou sans modifieurs). Ainsi, nous avons défini une nouvelle variable, COD-TYPE, que nous présentons ci-après.

#### 4.2.2.2 Différentes réalisations du COD

Pour rendre compte des différentes réalisations possibles du COD en affinant la catégorie des COD intermédiaires ci-dessus, nous définissons la variable **COD-TYPE** avec les quatre valeurs suivantes<sup>16</sup> :

- COD<sub>[Nu]</sub> : COD ne comportant aucune détermination ni modification.
- COD<sub>[Nu-MODIFIÉ]</sub> : COD ne comportant aucune détermination, mais comportant un ou plusieurs modifieurs.
- COD<sub>[INDÉFINI]</sub> : COD non marqué comportant une détermination, par ex. l'enclitique =*i* marqueur de l'indéfini, ex. (9-a), le déterminant *yek*, ex. (9-b), ou un quantificateur, ex. (9-c) et (9-d), sans faire de distinction entre ceux comportant des modifieurs et ceux n'en comportant aucun<sup>17</sup>.

- (9) a. [SN nām=i] [SP az Nina Anderiva] [V na-bar-ad]  
 nom=INDF de Nina Andreyeva NEG-porter.PRS-3SG  
 '... (qu'il) ne mentionne pas le nom de Nina Andreyeva'
- b. [SP be hezb=e mardom] [SN yek korsī] [V  
 à parti=EZ peuple un siège  
 mi-dah-ad]  
 IPFV-donner.PRS-3SG  
 '... il donne un siège au parti Peuple'
- c. be šahr āmad tā [SP barāye kamān=aš] [SN  
 à ville venir.PST.3SG pour-que pour arc=3SG  
 čand tir] [V be-xar-ad]  
 quelques flèche SBJV-acheter.PRS-3SG  
 '... il est venu en ville pour acheter quelques flèches pour son arc'

16. Dans un premier temps, nous avons établi un ensemble de traits binaires pour classifier les syntagmes nominaux, avec des valeurs 1, 0 ou NA. Ci-dessous la liste de ces traits :

- ±Nu
- ±Modifié
- ±Indéfini
- ±Quantifié
- ±Marqué (Quand la valeur de ce trait est 1, les autres traits auront la valeur NA).
- ±Enchâssé (contenant un enchâssé). Cette variable aurait pu ensuite servir pour caractériser la complexité, en termes de nombre de nœuds syntagmatiques, une des propriétés qui intervient dans la définition du poids.
- ±Pluriel
- ±Pronominal

Nous étions amenée à regrouper ces traits et à en supprimer certains à cause de nombre insuffisant d'occurrences qui satisfaisaient ces traits. Nous avons donc supprimé les traits ±Pluriel, ±Enchâssé et ±Pronominal, regroupé les deux traits ±Indéfini et ±Quantifié et abandonné le trait ±Modifié pour la classe des COD +Indéfini/Quantifié.

17. Comme nous allons le voir, nous aurions été obligée d'abandonner toute annotation plus fine et de nous tenir à la présente classification, à cause de la rareté des occurrences dans cette catégorie.

- d. kāregar-hā be surat=e eštabāhi dar zamān=e hādese [SN  
 ouvrier-PL à façon=EZ faux dans temps=EZ incident  
 16 kilugeram orāniyom] [SP dar tānker] [V  
 16 kilogramme uranium dans réservoir  
 rixte=and]  
 jeter.PP=COP.3PL  
 ‘... au moment de l’incident, les ouvriers ont jeté par erreur seize  
 kilos d’uranium dans le réservoir’

- COD<sub>[MARQUÉ]</sub> : COD portant l’enclitique =rā. Il faut souligner que nous n’avons pas fait de sous-catégorie à l’intérieur de cette catégorie, car les premières observations ont mis en avant un comportement homogène et une variation presque égale au seuil d’erreur généralement admis dans les études statistiques.

La distribution de l’ordre en fonction de ces quatre catégories de COD est donnée dans le Tableau 4.5. À l’aide du test  $\chi^2$ , nous constatons que les deux variables **ORDRE** et **COD** sont étroitement liées ( $\chi^2=348,7374$ ,  $df=3$ ,  $p<0,0001$ ).

	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-SP	27	10	<b>35</b>	<b>249</b>	321
SP-SN	<b>183</b>	11	15	10	219
Total	210	21	50	259	540

TABLE 4.5 – DS1 : Distribution de l’ordre par la variable **COD-TYPE**

Nous avons peu de données dans la classe COD<sub>[NU-MODIFIÉ]</sub> et la distribution équitable de cette classe entre les deux ordres possibles ne nous permet pas nous prononcer sur cette classe. Mais, en ce qui concerne la classe COD<sub>[INDÉFINI]</sub>, même si le nombre d’occurrences est limité, une tendance se dégage et nous constatons que, contrairement à l’hypothèse MDO, le comportement de cette classe est plus proche de celui de la classe COD<sub>[MARQUÉ]</sub> que celui de la classe COD<sub>[NU]</sub>, avec 70% dans l’ordre SN-SP.

	Prédit (SN-SP)		% correct
Observé (SN-SP)	0	1	
	0	194	25
	1	37	284
Total	231	309	88,5%
$\chi^2$ d’ajustement ( <i>goodness-of-fit</i> ) = 1,106			

N.B. : % correct avec une prédiction par défaut de SN-SP : 59,5%

TABLE 4.6 – DS1 : Matrice de confusion à partir de la variable **CODTYPE**

Comme le montre la matrice de confusion dans le Tableau 4.6, l'exactitude de la prédiction basée sur cette variable est de 3,7 points supérieure à l'exactitude de la prédiction basée sur le marquage différentiel de l'objet. De plus, la valeur du test  $\chi^2$  a fortement baissée, signifiant une meilleure qualité de prédiction pour cette variable<sup>18</sup>. Précisons que dans ce modèle de prédiction, l'ordre prédit pour les classes  $\text{COD}_{[\text{MARQUÉ}]}$  et  $\text{COD}_{[\text{INDÉFINI}]}$  est le SN-SP et pour les deux autres classes le SP-SN<sup>19</sup>.

Étant donné le nombre limité de données dans les classes  $\text{COD}_{[\text{NU-MODIFIÉ}]}$  et  $\text{COD}_{[\text{INDÉFINI}]}$  de notre table DS1, nous avons par la suite augmenté la taille de notre échantillon. Nous allons présenter la nouvelle table et la façon dont nous l'avons constituée dans la sous-section suivante, §4.2.3. Il faut préciser que toutes nos analyses ultérieures seront effectuées sur ce dernier jeu de données.

#### 4.2.2.3 Degré de collocationalité

Tout d'abord, il faut préciser que les deux variables **CODTYPE** et **COLL** sont étroitement liées dans nos données ( $\chi^2=353,4$ ,  $df=6$ ,  $p<0,0001$ ), étant donné la façon dont la dernière a été annotée. Le Tableau 4.7 donne la distribution de cette variable pour les quatre catégories du COD. Cette distribution n'est pas surprenante, du fait que les COD nus se trouvent à 87,1% dans l'ordre SP-SN-V et les COD marqués à 96,1% dans l'ordre SN-SP-V.

	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-Coll	128	0	1	0	129
SP-Coll	3	0	9	40	52
Nul	79	21	40	219	359
Total	210	21	50	259	540

TABLE 4.7 – DS1 : Distribution de la variable **COLL** par la variable **COD-TYPE**

Si l'on croise cette distribution avec la distribution de l'ordre, cf. Tableau 4.8, nous constatons que, dans 70% des cas, l'ordre SP-SN des  $\text{COD}_{[\text{Nu}]}$  est dû à la collocationalité, ce qui est attendu étant donné la façon dont nous avons annoté les données. Par ailleurs, pour 16,1% des occurrences de l'ordre SN-SP des  $\text{COD}_{[\text{MARQUÉ}]}$  et 25,7% des  $\text{COD}_{[\text{INDÉFINI}]}$ , le SP entretient un lien collocational avec le verbe. Notons

18. La comparaison des modèles de régression calculés avec ces deux variables par la fonction *anova* du logiciel R, basée sur le test du rapport de vraisemblance (*Likelihood ratio test*), confirme que la variable **CODTYPE** est un meilleur prédicteur que  $[\pm\text{RA}]$  ( $\chi^2=68,531$ ,  $p<0,0001$ ).

19. En effet, les données ne permettent pas de dégager une tendance pour la classe  $\text{COD}_{[\text{NU-MODIFIÉ}]}$  et nous avons donc gardé l'ordre par défaut fourni par le marquage différentiel de l'objet pour cette classe, sachant qu'étant donné la distribution des données dans cette classe, la valeur de l'exactitude ne change point avec le choix d'ordre prédit pour cette catégorie.

que les deux variables **COLL** et **ORDRE** sont également étroitement liées ( $\chi^2=281,2$ ,  $df=2$ ,  $p<0,0001$ ).

Ordre=SN-SP :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-Coll	0	0	0	0	0
SP-Coll	3	0	9	40	52
Nul	24	10	26	209	269
Total	27	10	35	249	321
Ordre=SP-SN :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-Coll	128	0	1	0	128
SP-Coll	0	0	0	0	0
Nul	55	11	14	10	90
Total	183	11	15	10	219

TABLE 4.8 – DS1 : Distribution de la variable **COLL** par la variable **COD-TYPE**

Or, en dépit du fait que ces constats soient conformes aux attentes en ce qui concerne le lien entre l'ordre linéaire et la dépendance sémantique, cette variable ne semble pas avoir une grande force explicative quant à la variation de l'ordre dans les données. En effet, sur les 27 occurrences de  $COD_{[NU]}$  dans l'ordre SN-SP, le SP entretient un rapport collocationnel avec le verbe dans seulement 3 (11,1%).

### 4.2.3 Table de données finale

Pour rappel, nous augmentons la taille de notre première table de données afin d'élever le nombre d'occurrences dans les classes de  $COD_{[NU-MODIFIÉ]}$  et  $COD_{[INDÉFINI]}$ , tout en veillant à ce que la répartition des données dans les catégories reste arbitraire pour garantir un échantillonnage non biaisé.

L'expérience précédente de la constitution du corpus sur la base d'échantillonnage arbitraire se révèle peu efficace. Nous avons dû passer en revue 2 000 occurrences pour n'en sélectionner que 540, dont seulement 71 dans les catégories qui présentent le plus de variation. Par conséquent, nous avons décidé de changer la méthode d'échantillonnage et, au lieu de faire des extractions arbitraires, nous avons décidé de faire des extractions ciblées, tout en restant non biaisées. Pour ce faire, nous avons extrait toutes les occurrences de deux lemmes typiquement ditransitifs peu fréquents dans le corpus, *rixtan* 'verser' et *ferestādan* 'envoyer', avec des occurrences totales de 254 et 219 respectivement, ainsi que deux échantillons arbitraires prélevés sur les deux lemmes les plus fréquents, *dādan* 'donner' et *gereftan* 'prendre', avec des fréquences de 10 494 et 6 849 respectivement. De plus, sur la base de la

variable que nous avons définie pour la mesure de collocationalité, nous avons écarté de cet échantillon les occurrences hautement collocationnelles ayant une valeur de **PREDPROT-FREQR** supérieure à 50 : *qarār gereftan*, 2 308, *qarār dādan*, 1 231, *edāme dādan*, 350, *az dast dādan*, 281, *tasmim gereftan*, 180, *xabar dādan*, 157, *gozāreš dādan*, 145, et *pāyān dādan*, 95.

Nous avons identifié les occurrences correspondant au patron SN-SP-V ou SP-SN-V avec les mêmes critères que ceux présentés plus haut. Cela nous fournit 364 occurrences supplémentaires et nous atteignons un total de 894 occurrences<sup>20</sup>. Les Tableaux 4.9 et 4.10 présentent la distribution de l'ordre dans cette table de données, que nous désignons ci-après par DS2, en fonction des deux variables **RĀ** et **NU**, ainsi qu'en fonction de la variable **CODTYPE** respectivement.

	COD +Rā	COD +Nu	COD intermédiaire	Total
SN-SP	<b>403</b>	43	<b>124</b>	570
SP-SN	21	<b>228</b>	75	324
Total	424	271	199	894

TABLE 4.9 – DS2 : Distribution de l'ordre par les variables  $[\pm R\bar{A}]$  et  $[\pm Nu]$

	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-SP	43	21	33	<b>403</b>	324
SP-SN	<b>228</b>	<b>42</b>	<b>103</b>	21	570
Total	271	63	136	424	894

TABLE 4.10 – DS2 : Distribution de l'ordre par la variable **COD-TYPE**

Nous constatons que les tendances dégagées dans la première table de données (DS1) se confirment. Les COD marqués apparaissent à 95% dans l'ordre SN-SP, les COD nus à 84,1% dans l'ordre inverse, et les COD intermédiaires présentent une variation relativement considérable justifiant leur séparation en deux classes, avec la variable **CODTYPE**. Au sein de cette catégorie, les COD indéfinis apparaissent à 75,7% dans l'ordre SN-SP, comme les COD marqués. En ce qui concerne la classe  $COD_{[Nu-MODIFIÉ]}$  nous pouvons à présent tracer une tendance, étant donné qu'ils apparaissent à 66,7% dans l'ordre SN-SP, comme les COD nus. Notons que, pour toutes les catégories, le test  $\chi^2$  montre que les taux sont différents de 50% ( $ps < 0,001$ ).

20. Notons que les données que nous présentons ici diffèrent légèrement d'une version antérieure de cette étude présentée dans Faghiri & Samvelian (2014), dans laquelle nous avons 905 occurrences. Précisons que cette mise à jour des données n'a aucun impact substantiel sur les résultats, ni sur les fréquences et leur significativité, *i.e.* (valeurs de p), ni sur les paramètres de notre modèle d'analyse multifactorielle.

Par ailleurs, sur ces 894 occurrences, l'exactitude d'un modèle de prédiction basé sur la variable **CODTYPE**, cf. Tableau 4.11, est de 86,8%, contre 79,0% pour un modèle basé sur la variable  $[\pm\bar{R}\bar{A}]$  et 84,4% pour un modèle basé sur la variable  $[\pm\bar{N}\bar{U}]$ , cf. Tableaux 4.12 et 4.13 respectivement. De plus, la comparaison entre les valeurs de  $\chi^2$  montre que la meilleure qualité s'obtient par la variable **CODTYPE** et qu'entre les deux variables binaires,  $[\pm\bar{N}\bar{U}]$  fournit une meilleure prédiction<sup>21</sup>.

	Prédit (SN-SP)		% correct
Observé (SN-SP)	0	1	
0	270	54	80,8%
1	64	506	90,4%
Total	334	560	86,8%
$\chi^2$ d'ajustement ( <i>goodness-of-fit</i> ) = 0,487			

N.B. : % correct avec une prédiction par défaut de SN-SP : 63,8%

TABLE 4.11 – DS2 : Matrice de confusion à partir de la variable **CODTYPE**

	Prédit (SN-SP)		% correct
Observé (SN-SP)	0	1	
0	303	21	64,5%
1	167	403	95,0%
Total	470	424	79,0%
$\chi^2$ d'ajustement ( <i>goodness-of-fit</i> ) = 95,627			

N.B. : % correct avec une prédiction par défaut de SN-SP : 63,8%

TABLE 4.12 – DS2 : Matrice de confusion à partir de la variable  $\bar{R}\bar{A}$

	Prédit (SN-SP)		% correct
Observé (SN-SP)	0	1	
0	228	96	84,1%
1	43	527	84,6%
Total	271	623	84,4%
$\chi^2$ d'ajustement ( <i>goodness-of-fit</i> ) = 14,874			

N.B. : % correct avec une prédiction par défaut de SN-SP : 63,8%

TABLE 4.13 – DS2 : Matrice de confusion à partir de la variable **NU**

Ces résultats préliminaires fragilisent l'hypothèse MDO et suggèrent que le marquage par  $=r\bar{a}$  ne peut à lui seul rendre compte de l'ordre canonique entre le COD et le COI. Ils montrent, particulièrement, la pertinence d'affiner la catégorie des

21. La comparaison des modèles de régression, calculés avec ces trois variables par la fonction *anova* du logiciel R, confirme que : 1) la variable  $[\pm\bar{N}\bar{U}]$  est un meilleur prédicteur que la variable  $[\pm\bar{R}\bar{A}]$  ( $\chi^2=6,234$ ,  $p<0,0001$ ), et 2) la variable **CODTYPE** est un meilleur prédicteur que la variable  $[\pm\bar{N}\bar{U}]$  ( $\chi^2=137,380$ ,  $p<0,0001$ ).

COD non marqués au sein de laquelle le COD nu se distingue par une distribution fortement biaisée et peu variable conformément à l'hypothèse MDO, tandis que le COD indéfini se distingue par sa proximité avec le COD marqué, plutôt que le COD nu, contrairement à cette hypothèse.

Précisons enfin que, du fait de nos critères d'échantillonnage, la distribution des données en ce qui concerne la variable **COLL** ne diffère guère dans cette table par rapport à la table DS1, étant donné que la presque totalité des occurrences ajoutées sont annotées [**NUL**] pour cette variable, cf. Tableau 4.14. En revanche, il est intéressant de noter que, si l'on ne prend en compte que les occurrences annotées [**NUL**], le taux moyen de l'ordre SP-SN baisse de plus de 10 points pour les COD nus, de 84,1% à 71,4%<sup>22</sup>, et que, dans ce cas de figure, la distribution de l'ordre pour cette catégorie ne diffère pas de celle des COD nu-modifiés ( $\chi^2=0,2695$ ,  $df=1$ ,  $p>0,5$ ).

Ordre=SN-SP :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-Coll	0	0	0	0	0
SP-Coll	3	0	9	41	53
Nul	40	21	94	362	517
Total	43	21	103	403	570
Ordre=SP-SN :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SN-Coll	128	0	1	0	129
SP-Coll	0	0	0	0	0
Nul	100	42	32	21	195
Total	228	42	33	21	324

TABLE 4.14 – DS2 : Distribution de l'ordre par les variables **COLL** et **COD-TYPE**

Par la suite, nous nous intéressons à l'étude de l'impact du poids grammatical sur l'ordre. La question, entre autres, est de savoir à quel point la variation présente dans les données peut être capturée par ce facteur.

#### 4.2.3.1 La longueur relative

Nous avons annoté chaque occurrence pour la longueur en nombre de mots des deux syntagmes (cf. §2.3.1.1)<sup>23</sup>, afin d'étudier l'effet de la longueur relative entre les deux syntagmes sur leur ordre linéaire.

22. Dans les autres catégories, la différences reste triviale (<1%).

23. Dans un premier temps, nous avons annoté nos constituants pour la complexité syntaxique en comptant le nombre de propositions enchâssées, mais nous avons abandonné cette variable en raison d'un nombre suffisant d'occurrences des constituants syntaxiquement complexes.

Les différentes catégories de COD ne présentent pas la même sensibilité à ce facteur. Tout d'abord, en ce qui concerne la catégorie COD<sub>[MARQUÉ]</sub>, l'ordre linéaire y présente une variation triviale, qui, comme le montre la Figure 4.2, ne peut être expliquée par la longueur relative. La longueur relative moyenne dans les ordres (0,7566 et 0,1905) diffère seulement numériquement, mais pas significativement ( $t=0,5052$ ,  $df=21,29$ ,  $p>0,5$ ).

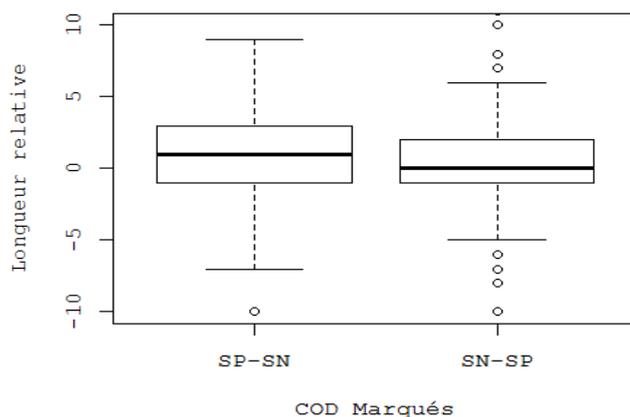


FIGURE 4.2 – DS2 : Boîte à moustache pour la longueur relative par l'ordre - COD Marqués

Par ailleurs, la longueur relative n'est pas pertinente pour la catégorie COD<sub>[NU]</sub>. Dans cette catégorie, le SN, par définition, ne contient qu'un seul mot, tandis que le SP en contient normalement au moins deux<sup>24</sup>. Nous n'avons donc qu'un seul cas de figure :  $SP \geq SN$  et, de ce fait, la longueur relative entre le SN et le SP a par définition une valeur négative pour ces COD. En d'autres termes, dans ce cas, la variable **CODTYPE** détermine l'essentiel, c'est-à-dire le signe de la longueur relative.

Toutefois, si l'on considère la longueur du SP isolément, comme le montre la Figure 4.3, elle ne serait pas sans effet sur l'ordre linéaire. En effet, la moyenne de la longueur des SP dans l'ordre SP-SN, à savoir 4,140, est significativement supérieure à celle des SP dans l'ordre inverse, à savoir 2,488 ( $t=6,158$ ,  $df=256,4$ ,  $ps<0,0001$ ). En d'autres termes, la longueur du SP favorise l'ordre SP-SN, ce qui converge avec la préférence "long-avant-court".

Enfin, dans la catégorie des COD intermédiaires, c'est-à-dire COD<sub>[NU-MODIFIÉ]</sub> et COD<sub>[INDÉFINI]</sub>, l'ordre linéaire semble être sensible à la longueur relative, cf. Figure

24. Il existe, en effet, un cas de figure dans lequel l'on peut avoir  $SP=SN$ , où le COI est constitué d'un seul mot graphique et phonologique, lors de la réalisation enclitique de l'argument sur la préposition, ex. *barāy=aš* 'pour lui'. Nous avons deux occurrences de ce type dans notre échantillon et les deux présentent l'ordre SP-SN.

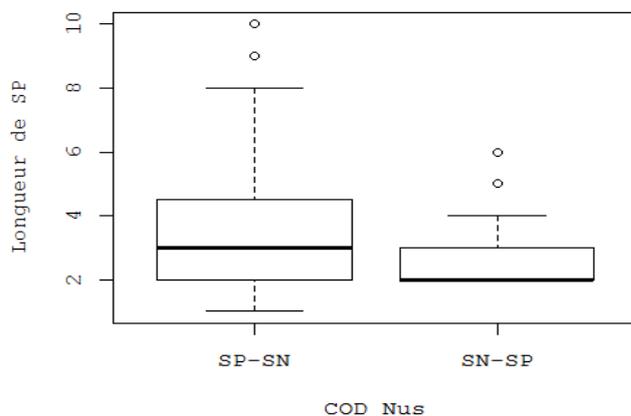


FIGURE 4.3 – DS2 : Boîte à moustache pour la longueur du SP par l'ordre - COD Nus

4.4. En effet, la moyenne de la longueur relative dans l'ordre SN-SP, à savoir -2,2148, est significativement inférieure à celle dans l'ordre inverse, à savoir -0,6826 ( $t=-3,74$ ,  $df=132,1$ ,  $p<0,0005$ ). Cela converge également avec la préférence “long-avant-court”.

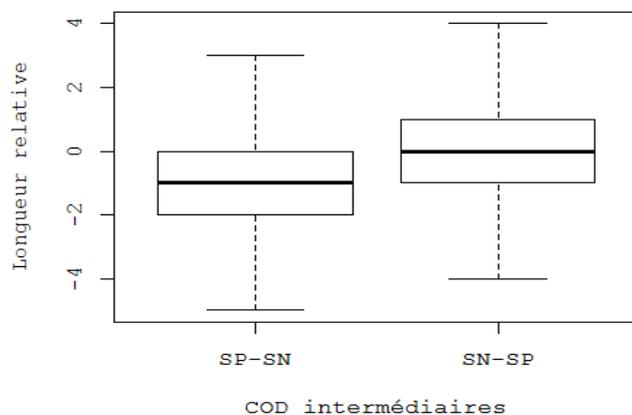


FIGURE 4.4 – DS2 : Boîte à moustache pour la longueur relative par l'ordre - COD Intermédiaires

#### 4.2.4 Analyse multifactorielle

Pour une analyse multifactorielle de ces données, nous avons construit des modèles de régression logistique pour prédire l'ordre linéaire entre les deux constituants,

avec les différentes variables prédictrices (ou variables explicatives) dont nous disposons.

La variable à prédire dans ces modèles est donc une variable binaire représentant l'ordre relatif, sachant que le succès est défini comme l'ordre SN-SP. Autrement dit, nous modélisons la probabilité d'obtenir l'ordre SN-SP en fonction des différentes valeurs de nos variables prédictrices.

Nous limitons nos données aux COD intermédiaires, qui présentent une variation intéressante à étudier, 199 occurrences au total, en laissant de côté les COD marqués et les COD nus monolexémiques<sup>25</sup>.

Ces 199 occurrences sont distribuées sur 31 différents lemmes verbaux. Rappelons que le lemme verbal, plus précisément le biais lexical du verbe, est susceptible d'avoir un effet sur l'ordre linéaire (cf. §2.3.1.5). L'effet de ce biais sur l'ordre est généralement traité en l'incluant dans le modèle de régression à effet mixte en tant que variable à effet aléatoire (Bresnan *et al.*, 2007). Cela permet d'obtenir une meilleure estimation pour les variables prédictrices tout en capturant la variation liée à ce biais.

Les variables prédictrices disponibles pour inclure dans le modèle en tant qu'effet fixe sont :

1. Type du COD (**CODTYPE**) : variable binaire avec deux valeurs **NU-MODIFIÉ** et **INDÉFINI**
2. Longueur relative (**LONGRLOG**) :  $\text{LOG}(\text{SN}_{\text{Nb-de-mots}}) - \text{LOG}(\text{SP}_{\text{Nb-de-mots}})$ <sup>26</sup>
3. Degré de collocationalité (**COLL**) : variable catégorique avec trois valeurs **SN-COL**, **SP-COL**, **NUL**

Le modèle maximal<sup>27</sup> que nous avons calculé comprend les trois variables, en incluant également les interactions entre elles. L'effet de la variable **COLL** ainsi que l'interaction entre les deux variables **CODTYPE** et **LONGRLOG** n'étant pas significatif ( $p > 0,11$  et  $p > 0,88$ ), nous avons calculé un modèle réduit (Mod-DS2) avec seulement les effets individuels de **CODTYPE** et **LONGRLOG**, dont nous présentons les résultats dans le Tableau 4.15.

Il importe de rappeler, quant à l'effet de la variable **COLL**, que celle-ci prise isolément est corrélée avec l'ordre. Toutefois, étant fortement dépendante du type

25. Bien que, dans le cas des COD nus monolexémiques, l'ordre linéaire soit sensible à la longueur du SP, la valeur de la longueur relative étant fortement biaisée pour cette catégorie, leur inclusion dans le modèle perturbe l'analyse sans y apporter d'élément supplémentaire.

26. Puisque cela est habituel, nous avons utilisé une transformation logarithmique pour minimiser l'effet des valeurs extrêmes.

27. Précisons que nous avons réduit notre modèle selon la méthode de comparaison des modèles à l'aide du fonction *anova* du R basée sur le test du rapport de vraisemblance (*Likelihood ratio test*).

de COD, **CODTYPE**, elle ne permet pas de capturer davantage de variation dans les données que ce que l'on peut capturer par cette dernière variable, comme le montre la comparaison des modèles ( $\chi^2=4,31$ ,  $df=2$ ,  $p>0,1$ ). Rappelons que sur les 64 occurrences (sur 334) de  $COD_{[NU]}$  et  $COD_{[NU-MODIFIÉ]}$ , qui présentent l'ordre SN-SP-V, seules 3 sont annotées comme **SP-COLL**, et sur les 54 occurrences (sur 560) de  $COD_{[MARQUÉ]}$  et  $COD_{[INDÉFINI]}$ , qui présentent l'ordre SP-SN-V, 1 seule est annotée comme **SN-COLL**, cf. Tableau 4.14, p. 150.

Effets aléatoires :				
	Groupes	Nom	Variance	Erreur-type
	VERBE	(Intercepte)	0,217	0,466
Nombre d'obs. : 199, groupes : VERBE, 31				
Effets fixes :				
	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte	1,556	0,300	5,19	<0,0001
<b>CODTYPE=NU-MODIFIÉ</b>	-2,005	0,357	-5,61	<0,0001
<b>LONGREL</b>	0.867	0,264	3,28	0,001

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

TABLE 4.15 – Mod-DS2 : Résumé des résultats

Les résultats montrent que les  $COD_{[NU-MODIFIÉ]}$  et les  $COD_{[INDÉFINI]}$  s'opposent. Le premier se joint au COD nu en favorisant l'ordre SP-SN, tandis que le second se joint au COD marqué en votant pour l'ordre SN-SP. Ils montrent également que la longueur relative a un effet significatif sur l'ordre linéaire. De plus, elle permet de capturer la variation dans les données au-delà de la réalisation du COD, comme le montre la comparaison des modèles ( $\chi^2=11,68$ ,  $df=1$ ,  $p<0,001$ ).

Le coefficient positif de **LONGREL** signifie que plus le SN>SP, plus le SN tend à précéder le SP. En d'autres termes, cet effet correspond effectivement à la préférence "long-avant-court".

### 4.3 L'ordre relatif entre le sujet et le COD dans les phrases transitives

L'ordre SOV est unanimement désigné dans la littérature comme l'ordre canonique des constituants dans une phrase transitive en persan. À l'aide de cette étude de corpus, nous nous proposons d'observer la validité empirique de cette affirmation.

Notre corpus relevant seulement de la modalité écrite, nous nous attendons à y trouver une forte proportion de l'ordre SOV par rapport aux ordres alternatifs. Par ailleurs, au chapitre 3 (cf. §3.2.2.3), en nous basant sur l'hypothèse de l'accessibilité

conceptuelle, et du fait que l'ordre SOV respecte la hiérarchie d'accessibilité des rôles grammaticaux, nous avons avancé que les COD marqués constituent la catégorie du COD pour lequel l'on s'attend à trouver le plus de bouleversements de cet ordre.

Nos explorations préliminaires sur le corpus allant dans le sens de ces attentes, nous n'avons pas procédé avec la même méthode d'échantillonnage que pour les phrases ditransitives. Après une étape préliminaire, nous avons constitué une table de données par un échantillonnage arbitraire du verbe *didan* 'voir', un verbe simple typiquement transitif, pour étudier la distribution de l'ordre en fonction de différentes catégories du COD. Nous avons ensuite constitué une table de données par échantillonnage arbitraire sur des constituants marqués de  $=r\bar{a}$ , étant donné que cet enclitique est annoté en tant que tel dans le corpus, pour constater quelle est la marge de variation la plus élevée.

Dans l'étape préliminaire, nous avons extrait un échantillon arbitraire limité au verbe fléchi, ayant comme lemme un des huit verbes potentiellement ditransitifs suivants, dans l'ordre décroissant de fréquence : *goftan* 'dire', *didan* 'voir', *koštan* 'tuer', *fahmidan* 'comprendre', *neveštan* 'écrire', *āfaridan* 'créer', *āzordan* 'contrarier', *āzmudan* 'évaluer'. Sur les 529 occurrences de propositions que nous avons annotées dans cet échantillon, 427 contiennent, entre autres choses, un sujet explicite et un complément verbal non phrastique, 78 contiennent des compléments phrastiques et suivent tous un ordre VO, et les 24 restantes suivent un ordre verbe final.

Sur les 427 occurrences qui nous intéressent, 406 correspondent à l'ordre SOV et 20 à l'ordre OSV, 2 relèvent du registre poétique et 1 de l'ordre SVO, c'est-à-dire une proportion de 95,1% pour l'ordre SOV, de 0,047% pour le OSV et de 0,002% pour le SVO.

Pour une annotation plus fine des données, nous nous sommes limitée au verbe *didan* 'voir'. Nous avons extrait un échantillon arbitraire de 1 500 occurrences de ce verbe et sélectionné les occurrences, dans lesquelles le sujet et le COD ont une réalisation lexicale, que nous avons ensuite annotées pour l'ordre ainsi que pour le type du COD, **COD-TYPE**, avec les mêmes valeurs que précédemment. Un total de 245 occurrences ont été identifiées et annotées, cf. Tableau 4.16, dont 126 sont des occurrences correspondant exactement à l'enchaînement des trois constituants sans qu'aucun autre élément ne s'y intercale<sup>28</sup>.

L'ordre SOV apparaît dans les deux tables avec une fréquence écrasante de 96,4% et 97,6% respectivement. La seule catégorie de COD qui permet une variation de

---

28. Nous avons écarté une occurrence relevant du registre poétique dans laquelle un COD nu est postposé au verbe, ainsi que deux autres occurrences dans l'ordre SOV dans lesquelles le COD étant un mot Qu- non marqué ne correspond à aucune de nos quatre catégories.

Sélection large :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SOV	46	7	41	142	236
OSV	0	0	0	9	9
Total	46	7	41	151	245
Sélection restreinte :					
	Nu	Nu-Modifié	Indéfini	Marqué	Total
SOV	28	5	24	66	123
OSV	0	0	0	3	9
Total	28	5	24	69	126

TABLE 4.16 – Verbe *didan* ‘voir’ : Distribution de l’ordre par la variable COD-TYPE

l’ordre est le COD marqué avec 6% et 4,3% de l’ordre OSV. Dans les autres catégories, l’ordre OSV ne se produit jamais.

Notons que cela appuie fortement notre hypothèse quant au fait que le niveau de la détermination du COD joue un rôle décisif sur la possibilité de celui-ci de précéder le sujet (cf. §3.2.2.3). Suite à cette observation, nous nous sommes concentrée sur les COD marqués afin de déterminer la marge de variation la plus élevée qui peut être atteinte.

À cette fin, nous avons extrait un échantillon arbitraire de 900 occurrences de constituants marqués de  $=r\bar{a}$ , en nous basant sur l’annotation de ce dernier dans le corpus. Selon les mêmes critères que pour la table de données du verbe *didan*, nous avons identifié 190 occurrences dans cet échantillon que nous avons annotées selon l’ordre : 177 (93,2%) suivent l’ordre SOV, 12 (6,3%) l’ordre OSV et seulement 1 (0,05%) l’ordre SVO. Sur ces 190 occurrences, 100 correspondent à une sélection restreinte, dont 9 suivent l’ordre OSV et 1 l’ordre SVO.

Étant donné les proportions très élevés de l’ordre SOV, dont même la plus basse est supérieure à 90%, et l’absence d’une variation substantielle, nous n’avons pas annoté les données pour d’autres facteurs.

## 4.4 Conclusions

Tout d’abord, il faut noter que la distribution de l’ordre dans nos échantillons ne laisse aucun doute quant au fait que l’ordre SOV est l’ordre canonique des constituants en persan. Dans nos données, cet ordre est en effet quasi systématiquement préféré aux ordres alternatifs.

Quant à l’ordre relatif “canonique” entre les deux objets, une comparaison entre la distribution de l’ordre dans notre corpus pour les deux types de phrases révèle un

écart non négligeable entre la variation de l'ordre entre le sujet et le COD au sein des phrases transitives par rapport aux phrases ditransitives. Indépendamment de l'hypothèse soutenue quant à l'ordre relatif "canonique" entre le COD et le COI, le taux des occurrences qui ne suivent pas l'ordre prédit (22%, 15,6% ou 14,6%) demeure plus élevé que ceux entre le sujet et le COD (4,6% et 6,8% pour les COD marqués). Nous tâcherons d'approfondir cet écart à travers nos données expérimentales au chapitre 5.

En ce qui concerne le critère MDO pour déterminer l'ordre relatif entre les deux objets, les COD marqués se comportent, et cela de manière presque systématique, selon la prédiction de l'hypothèse MDO. Les COD non marqués, en revanche, constituent une classe non homogène au sein de laquelle seuls les objets nus suivent la prédiction de l'hypothèse MDO. Le comportement des COD non marqués portant une détermination contrevient à cette hypothèse.

De plus, cette distribution de l'ordre converge avec l'hypothèse que nous avons formulée au chapitre 3 (cf. §3.2.2.3) sur la base de la hiérarchie de définitude, suggérant que la définitude du COD favorise son placement plus tôt dans la phrase.

Quant aux effets des différents facteurs fonctionnels sur l'ordre linéaire, nous trouvons un effet significatif du poids grammatical. Mais cet effet n'est pas systématique et dépend de la réalisation du COD, dans la mesure où la longueur relative entre le COD et le COI ne joue aucun rôle dans la détermination de l'ordre linéaire pour les COD marqués.

Le rôle que ce facteur joue dans le cas des autres COD correspond à la préférence "long-avant-court", prédite par l'hypothèse de l'accessibilité conceptuelle de Yamashita & Chang (2001), ainsi que par le modèle *DLT* de Gibson (2000) pour le japonais comme pour le persan (cf. §3.2.3.3 et §3.2.3.2 respectivement), mais qui va à l'encontre des prédictions de la théorie de Hawkins (1994, 2004) (cf. §3.2.3.1).

En ce qui concerne l'effet de la dépendance sémantique ou collocationnelle, nos données illustrent le fait que, conformément à ce qui était attendu, le lien collocationnel favorise l'adjacence. Toutefois, nous n'avons pas pu explorer ce facteur de manière indépendante, car, tel que nous l'avons opérationnalisé, il est en forte corrélation avec le type de COD dans nos données et se voit éclipsé par l'effet de ce dernier. Nous ne prolongerons pas l'étude de ce facteur dans la suite de ce travail, où nous nous concentrerons, avant tout, sur les constats inattendus ou discutables de nos études de corpus, et la laisserons pour les recherches futures.

Dans le chapitre suivant, nous détaillerons les expériences que nous avons conduites pour l'approfondissement de ces constats. Notre premier objectif sera de tester nos résultats quant à l'hypothèse MDO et le second de tester l'effet de la longueur sur

l'ordre linéaire. En outre, nous tâcherons d'étudier l'effet du statut informationnel ainsi que le caractère animé du référent sur l'ordre que nous n'avons pas abordé dans nos études de corpus. Nous nous efforcerons également de comparer l'ampleur de la variation de l'ordre linéaire entre les phrases transitives et ditransitives afin de décider s'il est judicieux de désigner un ordre canonique entre le COD et le COI.

# Chapitre 5

## Études expérimentales de la variation de l'ordre des mots

Ce chapitre est consacré à la présentation des données de nos études expérimentales. Nous avons conduit cinq expériences, dont quatre concernent les phrases ditransitives et la dernière les phrases transitives. Nous suivons cette organisation et regroupons les expériences selon qu'elles portent sur les phrases ditransitives ou sur les phrases transitives, abordées respectivement dans les sections 5.3 et 5.4 de ce chapitre.

Nos expériences relèvent toutes du même paradigme expérimental, que nous présenterons dans la première section de ce chapitre, et suivent la même méthode que nous décrirons en détail dans la deuxième section. La présentation de chaque expérience se limite, par conséquent, à la présentation des hypothèses expérimentales et du plan de l'expérience, des stimuli<sup>1</sup>, des participants, et, enfin, des résultats.

En ce qui concerne la présentation des résultats de nos analyses statistiques pour faciliter la lecture et éviter d'alourdir le texte outre mesure, nous ne rapporterons dans ce chapitre que les résultats du modèle de régression mixte "optimal" obtenue par la réduction du modèle maximal<sup>2</sup> selon la méthode de comparaison progressive des modèles (Bates *et al.*, 2015, *cf.*). Par ailleurs, dans ce chapitre, nous nous limiterons à la présentation des résultats pour les variables à effet fixe et présenterons les résultats complets de chaque modèle dans les annexes (*cf.* Annexe D). Le cas échéant, les résultats des modèles calculés avec des jeux de données alternatifs y sont également présentés<sup>3</sup>.

---

1. La liste complète de l'ensemble de nos stimuli se trouve dans les annexes (*cf.* Annexe C).

2. Précisons que tout au long de ce travail nous calculons tout d'abord un modèle maximal, en prenant en compte les pentes et les interceptes individuels pour toutes nos variables aléatoires.

3. L'ensemble de nos données brutes est disponible à l'adresse suivante :  
<http://perspred.cnrs.fr/fr/Faghiri-PhD>.

## 5.1 Le choix de notre paradigme expérimental

Dans les paragraphes suivants, nous passons en revue les différents paradigmes expérimentaux mis en place dans les travaux existants similaires avant d'aborder les contraintes qui ont guidé le choix du paradigme que nous avons utilisé dans nos expériences.

### 5.1.1 Paradigmes expérimentaux existants

Différents paradigmes expérimentaux *off-line* et *on-line* sont utilisés dans les travaux en psycholinguistique et en linguistique expérimentale. Les études expérimentales traitent de la variation de l'ordre des mots ou plus généralement l'alternance de structure recourant majoritairement aux paradigmes *off-line*. En effet, la variable dépendante dans ces travaux est généralement la réponse des sujets et rarement leur temps de réaction, ou de fixation oculaire ou d'autres mesures relevant du traitement en temps réel. Cette réponse consiste soit en l'ordre relatif des constituants dans la phrase formulée ou choisie par le répondant, opérationnalisé par une variable catégorielle, généralement mais pas obligatoirement à deux modalités.

Le paradigme expérimental le plus couramment utilisé par les chercheurs dans le domaine de la syntaxe est sans aucun doute celui de l'élicitation de jugement de grammaticalité des locuteurs, plus judicieusement appelé jugement d'acceptabilité (voir Schutze, 1996; Cowart, 1997). Il s'agit effectivement d'un paradigme à portée de main dont la mise en place ne nécessite aucun équipement particulier et qui peut se faire à l'aide d'un questionnaire<sup>4</sup>. Ce paradigme repose sur une tâche métalinguistique qui invite le locuteur à se prononcer sur l'acceptabilité d'une phrase de sa langue native, moyennant une note sur une échelle de type Likert – à nombre de points pair ou impair – ou en opérant un choix binaire (ex. acceptable vs. inacceptable), c'est-à-dire un choix forcé<sup>5</sup>. Pour l'élicitation des préférences en ordre linéaire et l'étude des facteurs fonctionnels, certains travaux empiriques en syntaxe utilisent des paradigmes similaires (Rosenbach, 2002; Gries, 2003; Rosenbach, 2005; Bresnan, 2007; Bresnan & Ford, 2010).

Les tâches de jugement s'approchent plutôt des paradigmes de compréhension (et non de production). Des travaux se situant dans le domaine de la psycholinguistique font appel aux différents paradigmes relevant de la production dans lesquels les participants sont amenés à formuler (ou à reformuler) des phrases avec plus ou moins

---

4. Nous avons employé ce paradigme dans deux expériences complémentaires, sans rapport avec l'ordre relatif des constituants, que nous présenterons dans le chapitre 6.

5. Précisons qu'une échelle de Likert à nombre de points impair, à la différence d'une échelle paire ou un paradigme à choix forcé, inclut une position neutre au milieu.

de contraintes (Bock & Warren, 1985; Bock, 1986; Bock et al., 1992; McDonald et al., 1993; Hartsuiker et al., 1999; Arnold et al., 2000; Yamashita & Chang, 2001; Tanaka et al., 2005, 2011a; Stallings & MacDonald, 2011; Ros et al., 2015, entre autres).

Mis à part la pertinence de ces paradigmes dans de tels travaux dont le but est précisément l'étude du système de la production langagière, ces derniers partagent l'avantage d'éviter une tâche métalinguistique directe, au demeurant difficile à instruire<sup>6</sup>. En revanche, ces paradigmes nécessitent la participation présenteielle des locuteurs et ne peuvent être réalisés en bonne et due forme par des questionnaires.

Par la suite, nous présentons brièvement l'ensemble des paradigmes mis en place dans les travaux étudiant l'impact des facteurs fonctionnels sur le choix entre des structures alternatives, notamment en ce qui concerne l'ordre relatif entre constituants. Puis, nous nous attardons sur les contraintes qui ont guidé le choix du paradigme que nous avons adopté pour nos études expérimentales.

Dans son étude de l'alternance génitive en anglais, Rosenbach (2002, 2005) utilise un questionnaire à choix forcé où les participants lisent un petit paragraphe de texte à la fin duquel ils doivent choisir entre une structure génitive et une structure prépositionnelle (*the England's king* vs. *the king of England*), celle qui leur paraît mieux "*sounded better to them and not to take into account what grammar books would tell them*" selon les termes de l'auteur (Rosenbach, 2002, p. 130).

Gries (2003) utilise une tâche d'élicitation de jugement d'acceptabilité sur une échelle de type Likret à nombre de points impair (-3 à +3) pour tester, entre autres choses, les résultats de sa modélisation basée sur les données de corpus pour la reconnaissance des constructions datives prototypiques en anglais. Dans cette étude, chaque construction est notée individuellement et non en comparaison avec sa construction alternative. Bresnan (2007) et Bresnan & Ford (2010), quant à eux, testent la précision de leur modèle de prédiction de l'alternance dative obtenu – par régression logistique – à partir de données de corpus, en demandant à des locuteurs natifs de différentes variétés d'anglais de noter simultanément les deux constructions alternatives de manière respective en distribuant 100 points sur chacune des deux (ex. 0-100,30-70,50-50,70-30,100-0). Thuilier (2012) utilise des questionnaires similaires pour l'étude de l'ordre des compléments verbaux en français.

Pour étudier l'effet de l'accessibilité discursive et la longueur sur l'ordre relatif des compléments en anglais, Arnold et al. (2000) emploient un paradigme de production relativement libre dans la mesure où les stimuli ne sont pas constitués par avance. Dans leur expérience, sous forme d'un jeu en binôme, les auteurs provoquent une situation où les participants, à tour de rôle, sont amenés à produire, à l'aide de cartes

---

6. Pour une discussion, voir par ex. Schutze (1996), Cowart (1997)

portant des images de différents objets, des phrases à construction ditransitive dans lesquelles l'accessibilité discursive ainsi que la longueur des arguments bénéficiaire et thème est susceptible de varier. En effet, les auteurs se donnent comme objectif de comparer le point de vue du locuteur et celui de l'interlocuteur pour évaluer la pertinence des modèles guidés par la compréhension, notamment celui de Hawkins (1990, 1994).

Quant aux travaux en psycholinguistique, Bock & Warren (1985) et McDonald et al. (1993), par exemple, utilisent une tâche de rappel de phrase pour examiner l'impact direct de l'accessibilité (conceptuelle) du référent, en termes, respectivement, du niveau de concrétisation (*concreteness*) et de l'animéité, sur l'ordre des constituants – indépendamment de son impact sur l'assignation des rôles grammaticaux.

Cette tâche consiste grossièrement à faire écouter un ensemble de phrases aux participants et à leur demander, après un laps de temps, de les écrire de mémoire sur un papier blanc en entendant un indice précis (par exemple, le préambule ou le verbe principal de la phrase pour chaque item).

Une variante de cette procédure, améliorée par une tâche de description d'image effectuée à l'oral et transcrite ensuite, est également utilisée dans des travaux similaires pour l'étude de l'impact de l'amorçage (*priming*) sur la structure de la phrase (ex. passive vs. active, dative prépositionnelle vs. dative à double objet) (par ex. Bock, 1986; Bock et al., 1992) ou sur l'ordre des constituants (e.g. Hartsuiker et al., 1999).

Stallings et al. (1998), quant à eux, emploient également une tâche de rappel de phrase, présentée comme une expérience étudiant la mémoire, pour étudier l'effet de la longueur du SN sur l'ordre relatif entre des compléments verbaux. Dans leur expérience, les constituants d'une phrase sont affichés sur un écran pendant un certain temps sous forme de blocs de texte, et les participants ont pour tâche de construire la phrase à partir de ces constituants, qu'ils auront à formuler (oralement) de mémoire après un laps de temps. Ceci est fait dans le but d'encourager une production de phrases sur la base du contenu (sémantique) ou du sens. Cette tâche est reprise avec de légères modifications dans les recherches ultérieures sur d'autres langues, notamment par Yamashita & Chang (2001) sur le japonais et Ros et al. (2015) sur le basque. La Figure 5.1 illustre schématiquement la façon dont cette tâche est administrée pour les phrases transitives, chez Yamashita & Chang (2001, P. B49).

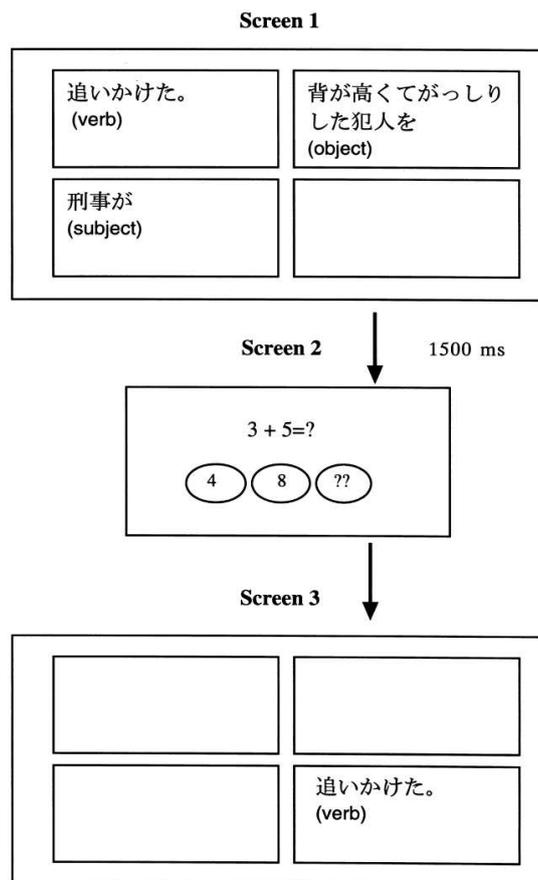


FIGURE 5.1 – Exemple d’écran pour la tâche de rappel de phrase (Yamashita et Chang, 2001)

### 5.1.2 Paradigme adopté pour cette étude

Pour des raisons d’ordre pratique, notamment par le fait que cette recherche s’effectue en France, nous ne pouvions pas compter sur le recrutement d’un nombre suffisant de locuteurs natifs monolingues ou de locuteurs persanophones dont le persan reste la langue dominante. Par conséquent, nous avons écarté les paradigmes présentiels et avons décidé de nous servir d’une plateforme d’accueil d’expériences psycholinguistiques en ligne, notamment *Ibex Farm* (Drummond, 2013)<sup>7</sup>. Elle nous permet de solliciter la participation des locuteurs indépendamment de leur localisation géographique en leur envoyant simplement un hyperlien.

La plateforme *Ibex Farm* accueille gratuitement les expériences psycholinguistiques et assure leur bon fonctionnement méthodologique<sup>8</sup>. Les paradigmes expé-

7. [www.spellout.net/ibexfarm/](http://www.spellout.net/ibexfarm/)

8. Cette plateforme utilise les langages de programmation *JavaScript* et *HTML* et offre une bonne flexibilité en ce qui concerne la conception des paradigmes, à condition de disposer d’une certaine maîtrise de ces langages et des connaissances informatiques. Elle propose des modèles prêts à l’emploi pour certains types de questionnaires (administrés en anglais) dont l’on peut se servir,

mentaux pouvant se conduire d'une façon auto-administrée via l'Internet, tels que le jugement d'acceptabilité, la lecture à vitesse autogérée (*self-paced reading*) et la phrases à trous (*cloze tasks*) sont réalisables sur cette plateforme.

Toutefois, cette méthode nous impose un certain nombre de contraintes dans la conception de notre paradigme, car le fait que les expériences s'effectuent de manière auto-administrée nous laisse un contrôle limité sur les conditions dans lesquelles les sujets accomplissent la tâche demandée.

De plus, du fait d'un recrutement effectué sur la base du volontariat, nous devons concevoir des expériences qui ne soient pas fastidieuses. En effet, bien qu'il y ait lieu de penser que beaucoup d'internautes se portent volontaires pour participer à de tels questionnaires, ne serait-ce que par curiosité, il est hautement probable qu'une bonne partie abandonne à mi-chemin si cela ne les intéresse pas suffisamment ou les ennue. En plus de son impact sur le nombre de participants de chaque expérience, cela peut compromettre la bonne gestion des listes par la plateforme.

Afin de se rapprocher des paradigmes de production décrits plus haut, en nous inspirant notamment de la tâche utilisée par Stallings *et al.* (1998), reprise et détaillée par Yamashita & Chang (2001), nous avons élaboré un paradigme similaire de production de phrases avec constituants affichés sur l'écran. À la différence que celui-ci ne se présente pas comme une tâche de rappel de phrase, mais comme une tâche de complétion (semi-guidée) de phrases. Cette tâche est similaire à celle de phrases à trous, sauf que les participants doivent remplir les trous avec des éléments donnés.

Dans notre paradigme, les participants doivent compléter une phrase par des constituants affichés sur l'écran, en les écrivant dans des cases vides (voir Figure 5.2). Pour amener les participants à formuler la phrase à partir de son contenu sémantique et lexical, nous ajoutons un constituant supplémentaire formellement identique à un des autres constituants mais, lexicalement différent de celui-ci. Ainsi, compléter la phrase n'est pas équivalent d'un simple remplissage des trous et ne se résume pas au choix de l'ordre relatif entre des constituants affichés sur un écran, puisque les participants doivent se pencher sur son contenu.

L'instruction donnée aux participants est de formuler avec les éléments donnés la phrase qui leur semble la mieux convenir pour compléter le préambule, sachant que le temps qu'ils y consacrent est enregistré. Cette instruction nous permet non

---

en suivant les instructions, même en étant peu familier avec le codage. Il est également possible de manipuler les modules disponibles afin d'élaborer son propre questionnaire. Nous avons apporté des modifications nécessaires pour : 1) adapter le format à la langue persane qui s'écrit de droite à gauche et 2) transformer une des formes disponibles pour notre paradigme expérimental qui n'est pas pré-conçu dans cette plateforme.

نمایشگاه بسیار شلوغ بود. بچه‌ها با پدرمادر هایشان سر و کله می‌زدند. کارمند نمایشگاه که از سر و صدا کلافه شده بود

داد.

چندتا آدامس

چندتا آبنبات

به بچه‌ها که بی‌تابی می‌کردند

ادامه →

FIGURE 5.2 – Exemple d’écran pour la tâche de complétion de phrases dans notre paradigme

seulement de dissimuler la question de recherche, mais a aussi la vertu d’inciter les participants à se focaliser sur le contenu conceptuel de la phrase et, de ce fait, de provoquer potentiellement une formulation relativement spontanée, du point de vue de la forme, voire de nous approcher d’une situation de production telle qu’obtenue dans les études précédentes.

Dans toutes nos expériences, les réponses sont annotées pour l’ordre relatif qui constitue toujours notre variable dépendante. Nos expériences croisent, au plus, deux variables indépendantes. Les variables indépendantes étudiées se rapportent à la longueur, l’animéité, le caractère donné et le niveau de la détermination.

## 5.2 Mise en place des expériences et récolte des données

Dans les paragraphes suivants, nous détaillerons le déroulement de nos expériences, menées par des questionnaires (auto-)administrés sur la plateforme *Ibex Farm*. Ces questionnaires se déroulent de manière identique. Les participants sont sollicités, à chaque fois, selon la même démarche, et les résultats sont recueillis et annotés d’après les mêmes principes. Nous avons recouru à trois questionnaires au total, étant donné que trois de nos cinq expériences ont été conduites avec le même questionnaire<sup>9</sup>.

La plateforme *Ibex Farm* assure la génération des listes ordonnées d’items expérimentaux, combinés aux items distracteurs (*fillers*), à présenter à chaque participant, suivant la procédure du carré latin. Chaque participant voit ainsi tous les items et

9. Ces questionnaires sont consultables aux adresses suivantes :

Q1 (Exp. D1) : <http://spellout.net/ibexexps/pegah/autobus/experiment.html>

Q2 (Exp. D2) : <http://spellout.net/ibexexps/pegah/autobus2/experiment.html>

Q3 (Exp. D3, D4 et T1) : <http://spellout.net/ibexexps/pegah/panaahju/experiment.html>

toutes les conditions expérimentales, mais jamais le même item expérimental dans plusieurs conditions. De plus, la plateforme prend en charge l'ordonnancement des items dans chaque liste. L'ordre de présentation des items est randomisé individuellement pour chaque participant, avec cette contrainte que, sur chaque liste, les items expérimentaux, appartenant à une même liste, soient espacés d'au moins deux distracteurs. Précisons que, dans le questionnaire combinant trois expériences, pour les items expérimentaux d'une liste donnée, les items appartenant aux deux autres listes font office de distracteur.

### 5.2.1 Recrutement de participants

Nous avons proposé les expériences sous formes de lien Ibex, accompagné d'une description succincte sur le sujet et la finalité de l'expérience, incluant un bref *modus operandi* et l'estimation de la durée, transmis à travers des réseaux sociaux et par envoi de messages.

Aucun critère de sélection n'a pas été précisé en aval. C'est seulement en fonction des réponses aux questions générales que nous avons écarté certains répondants dont le profil linguistique n'était pas conforme à nos critères<sup>10</sup>.

### 5.2.2 Déroulement des questionnaires

Les questionnaires débutent par une présentation générale de l'expérience, précisant notamment le fait qu'il fait partie d'un projet de recherche académique dont le sujet est l'étude des comportements linguistiques des locuteurs du persan. Cette présentation inclut le déroulement du questionnaire et une estimation du temps nécessaire pour l'accomplir. Elle informe les volontaires que leurs réponses ne seront enregistrées qu'une fois le questionnaire terminé et que l'utilisation optimale des résultats, étant donné la prise en compte du temps, exigeait la concentration sur la tâche, ce qui impliquerait de réserver le temps nécessaire pour le questionnaire et de ne rien faire de concomitant<sup>11</sup>.

Les questionnaires sont anonymes, mais contiennent des questions générales, qui, mise à part l'âge et le sexe, permettent de situer le profil linguistique des répondants. En effet, étant donné la forme immatérielle des questionnaires et le fait qu'habiter en Iran et/ou être persanophone monolingue ne constituaient pas des conditions préalables, il était important de s'assurer que le persan soit la langue dominante

---

10. Précisons que les participants ont répondu aux questionnaires de leur plein gré sur la base de volontariat et sans contrepartie de rémunération ou d'autres formes de récompense.

11. Par ailleurs, dès les premières lignes de cette introduction, les volontaires sont priés de ne pas refaire le questionnaire ou le recommencer s'ils l'ont abandonné au milieu.

des répondants. Un courriel de contact est également fourni aux participants pour leur permettre de signaler une anomalie ou simplement partager leurs éventuels commentaires. Notons pour finir que les questionnaires sont présentés entièrement en persan.

Afin de démarrer l'expérience, le répondant doit accepter de déclarer son consentement express pour sa participation dans l'expérience. Les instructions détaillées sur la tâche demandée sont ensuite présentées aux répondants accompagnées, en bas de la même page, d'un exemple en guise d'entraînement.

Une nouvelle page annonce au répondant qu'il est prêt pour commencer et qu'il peut le faire en cliquant sur "*Continuez*".

### 5.2.3 Exécution de la tâche

Après le démarrage, chaque item apparaît sur l'écran sur une page à part (voir par ex. Figure 5.2, p. 165). Cette page contient en haut, aligné à droite, un préambule écrit en noir ; au milieu, aligné au centre, une liste (non numérotée et sans puces) de syntagmes, écrits avec la même taille de caractère mais en bleu ; en bas, aligné à droite, une phrase incomplète dans laquelle les constituants manquants sont représentés par des cases vides de taille égale. La liste comporte un élément de plus que le nombre des constituants manquants.

La tâche consiste à compléter la phrase, c'est-à-dire à remplir les cases vides, avec le choix des syntagmes proposés, en les utilisant tels quels et sans y apporter de modification. Pour faciliter l'exécution de cette tâche, mais aussi pour éviter les erreurs de frappe qui pourraient gêner l'annotation automatique des résultats, les participants sont invités à remplir les cases par un copier-coller (ou par *drag-and-drop*), sachant que la sélection du contenu entier de chaque syntagme peut se faire en double-cliquant dessus.

Il faut noter que, pour contrôler l'effet d'amorçage de l'ordre d'apparition des syntagmes proposés sur l'écran, nous avons neutralisé ce paramètre. Dans le premier questionnaire, nous avons randomisé l'ordre d'apparition des éléments pour chaque item de manière individuelle, c'est-à-dire que l'ordre dans lequel les options devaient apparaître sur l'écran n'était pas fixé par avance, mais un ordre arbitraire y était assigné à l'aide un script Java. Nous avons abandonné cette méthode par la suite et nous avons opté pour la neutralisation par la contrebalance.

Une fois les cases remplies, le participant doit cliquer sur "*Continuez*" pour passer à l'item suivant. Pour cela, toutes les cases doivent être remplies, sinon le participant reçoit un message d'erreur. Une fois passé à l'item suivant, aucun retour en arrière n'est possible. Les items ne sont pas numérotés et leur nombre total

n'est pas indiqué aux participants. Néanmoins, une barre de progression purement graphique est affichée en haut des pages.

#### 5.2.4 Annotation des résultats

Les réponses de chaque participant sont envoyées au serveur de manière regroupées et seulement une fois que le questionnaire est terminé. Les réponses de tous les participants sont inventoriées dans un même fichier texte sur la plateforme.

À l'aide d'un filtre en langage de programmation Python, nous avons transformé en table de données le fichier de résultat pour chaque expérience. Ainsi, pour chaque expérience, nous avons constitué une table de données, dans laquelle chaque ligne correspond à une réponse recueillie et les colonnes aux différentes variables, et une table contenant les profils de répondants.

Les réponses sont annotées sous forme d'une variable catégorique, prédéfinie en fonction de l'expérience, représentant l'ordre dans lequel les éléments apparaissent dans la phrase complétée par le répondant. Cette annotation est faite dans un premier temps à l'aide d'un script Python et est basée sur la correspondance exacte entre les options proposées et les textes inscrits dans les cases, complétée ensuite manuellement dans les cas où l'annotation automatique avait échoué<sup>12</sup>. Les réponses erronées, somme toute rares, ont été annotées comme "NA" et écartées dans les analyses statistiques.

### 5.3 Les phrases ditransitives

Nous avons conduit quatre expériences pour étudier l'ordre relatif entre le COD et le COI dans les phrases ditransitives en persan.

Nos deux premières expériences, Exp. D1 et Exp. D2, sont conçues principalement dans le but de vérifier les résultats de notre étude de corpus, notamment ceux qui fragilisent les hypothèses existantes, c'est-à-dire le positionnement canonique des COD non marqués et l'impact de la longueur relative sur l'ordre linéaire. À l'aide de ces expériences, nous étudions également l'impact du statut discursif que nous n'avons pas pu traiter dans l'étude de corpus.

La troisième expérience, Exp. D3, est conçue pour comparer l'ordre par défaut de trois types de COD non nu : 1) le COD marqué de  $=r\bar{a}$ , 2) le COD non marqué portant l'enclitique indéfini  $=i$ , et 3) le COD non marqué portant le déterminant

---

12. Cela peut survenir lorsque les textes sont altérés, par exemple à cause d'une erreur de frappe, ou lorsqu'ils sont incomplets ou hors propos – dans les cas où les répondants ont inscrit directement les réponses au lieu de faire un copier-coller.

$ye(k)$ . La quatrième expérience, Exp. D4, est destinée à comparer l'ordre par défaut des COD nus avec ou sans ajouts. Ces deux expériences sont conduites via un même questionnaire<sup>13</sup>, qui permet, ensuite, de comparer tous ces COD ensemble.

### 5.3.1 Expérience D1 : Les COD non marqués indéfinis

Le but de cette expérience est d'étudier : 1) le positionnement préféré du COD non marqué indéfini par rapport au COI et 2) l'effet de la longueur et l'accessibilité discursive des constituants sur l'ordre relatif. En ce qui concerne la longueur, nous nous sommes basée sur la longueur relative, la différence de longueur en termes de nombre de mots entre les deux constituants, et avons défini deux niveaux, COD plus long que le COI et *vice versa*.

Par accessibilité discursive, nous entendons le fait que le référent d'un constituant est évoqué auparavant dans le discours ou non (*givenness* en anglais). Dans ce cas, il aura le statut discursif "donné". Ainsi, nous avons défini deux niveaux pour le statut discursif d'un constituant : donné *vs.* nouveau. Le référent d'un constituant donné se trouve dans le contexte qui le précède, tandis que le référent d'un constituant nouveau en est totalement absent, ainsi que non inférable.

Notons que, dans cette expérience, le COD ne peut avoir le statut discursif donné. En effet, si le référent du COD est déjà introduit par le contexte, le COD ne peut avoir une réalisation indéfinie. Nous ne pouvons donc que faire varier le statut discursif du COI.

Ainsi, dans cette expérience, nous nous sommes limitée aux COD non marqués indéfini et avons manipulé deux facteurs : 1) la longueur relative et 2) le statut discursif du COI, selon une conception 2 x 2 (2 facteurs x 2 niveaux).

Nos hypothèses expérimentales, basées sur les résultats de notre étude de corpus, sont les suivantes :

**Hypothèse 1** Les COD indéfinis se positionnent de préférence avant le COI (la fréquence totale de l'ordre OI-OD est significativement plus élevée que 50%).

**Hypothèse 2** La longueur favorise le positionnement plus en amont de la phrase (la fréquence de l'ordre OI-OD pour COI>COD est significativement supérieure à celle pour COI<COD).

**Hypothèse 3** L'accessibilité discursive (c'est-à-dire *givenness*) favorise le positionnement plus en amont de la phrase (la fréquence de l'ordre OI-OD pour COI donné est significativement supérieure à celle pour COI nouveau).

---

13. Ce questionnaire inclut également notre cinquième expérience, Exp. T1, sur les phrases transitives.

### 5.3.1.1 Conception

Nous résumons ci-dessous la conception de l'expérience.

#### Variables indépendantes :

##### 1. Longueur relative entre les deux constituants :

Pour manipuler la longueur, nous avons essentiellement ajouté des modifieurs. Pour le COD, les modifieurs sont de type adjectival – nous avons également utilisé la coordination, cf. ex. (2) –, et, pour le COI, des subordonnées relatives. La différence de longueur entre les deux constituants est d'au moins 5 syllabes (en moyenne 7) et toujours clairement visible sur le plan graphique.

##### 2. Statut discursif du COI :

Le référent d'un constituant donné se trouve dans le contexte qui précède la phrase cible tandis que le référent d'un constituant nouveau est absent du contexte précédant celle-ci. Pour manipuler l'accessibilité, chaque phrase cible est précédée d'une phrase de contexte en guise de préambule. Dans le cas des COI donnés, le référent est mentionné dans la phrase de contexte.

La reprise du COI dans la phrase à compléter se fait par un syntagme non pronominal et, pour que cela ne paraisse pas étrange aux répondants, nous avons essayé de constituer la phrase contexte de manière à ce que le référent y apparaisse le plus tôt possible.

#### Conditions :

1. COI donné / COI < COD : IDC (IO-donné-court)
2. COI donné / COI > COD : IDL (IO-donné-long)
3. COI nouveau / COI < COD : INC (IO-nouveau-court)
4. COI nouveau / COI > COD : INL (IO-nouveau-long)

#### Variables dépendantes :

- Ordre relatif entre le COD et le COI dans la phrase complétée par le participant, opérationnalisé par une variable binaire.

#### Variable de contrôle :

- Caractère animé des constituants :

Dans tous les items, le COD est un nom concret inanimé et le COI un nom animé humain.

- Rôle sémantique des arguments :

Dans tous les items, le COD correspond à l'argument thème. Le rôle du COI, en revanche, varie en fonction du verbe. Il peut correspondre à l'argument bénéficiaire, source ou patient.

### 5.3.1.2 Stimuli

Nous avons préparé quatre séries de 20 phrases ditransitives, correspondant à nos quatre conditions expérimentales. Ces phrases sont réparties sur 7 lemmes verbaux (simples), dont nous détaillerons la liste plus bas. Toutes les phrases sont à la troisième personne et aucun déictique n’y est utilisé.

Chaque phrase à compléter comporte deux choix de COD, un COI et un verbe. Les COD sont toujours construits comme des syntagmes nominaux indéfinis, portant soit le déterminant indéfini *yek* ‘un’, soit le quantifieur *čand(=tā)* (ou sa variante *čandin*) ‘quelques’. Au niveau sémantique, le COD est toujours un nom concret correspondant à l’argument thème du verbe. Dans chaque phrase, les deux choix du COD partagent les mêmes traits formels, notamment le déterminant et, approximativement, la longueur, ex. (1) et (2). Ils sont tous les deux possibles dans la phrase, c’est-à-dire qu’aucun ne rend la phrase agrammaticale ou inacceptable, mais ils n’ont pas forcément la même pertinence dans le contexte.

- (1) a. *čand=tā ābnabāt*  
quelques=CLF bonbon  
'quelques bonbons'
- b. *čand=tā ādāms*  
quelques=CLF chewing-gum  
'quelques chewing-gums'
- (2) a. *yek varaq kāqaz va čand=tā medād-rangi*  
un feuille papier et quelques=CLF crayon-de-couleur  
'Une feuille de papier et quelques crayons de couleur'
- b. *yek tofang=e bādi va čand=tā fešang*  
un fusil=EZ à-vent et quelques=CLF cartouche  
'Un fusil à vent (pour enfant) et quelques cartouches'

Le COI, quant à lui, dénote un être humain, dont le rôle sémantique varie en fonction du lemme verbal, tout comme la préposition. Le syntagme nominal du COI a toujours une forme lexicale, c’est-à-dire un nom propre ou un nom commun, *bačče* ‘l’enfant’, *mariz* ‘le malade’, etc. Dans la condition où le COI est donné et son référent est présent dans la phrase de contexte, cela implique une reprise non pronominale. Pour que cette reprise ne paraisse pas étrange aux participants, nous avons essayé de constituer le préambule de manière à augmenter, dans la mesure de possible, la distance entre les deux. Nous reviendrons sur le préambule plus bas.

Le choix des verbes pour construire l’ensemble des stimuli de cette expérience n’a pas été aisé. Nous avons rencontré deux types de contraintes :

**Contraintes lexicales** L'ordre SOV est l'ordre canonique et caractéristique du persan standard, tandis que le registre moins formel connaît une plus grande variation au niveau de l'ordre. De plus, les stimuli sont présentés aux sujets dans le mode écrit et il était logique et attendu qu'ils soient conformes aux normes du persan écrit, formel et standard.

En persan, le nombre de verbes simples ne dépasse pas les 250 dont environ 150 sont d'un usage courant. En effet, comme nous l'avons déjà vu auparavant, une majorité des prédicats verbaux à trois arguments sont exprimés par des prédicats complexes (cf. §1.1.4). Dans notre corpus, nous avons identifié les verbes simples potentiellement ditransitifs, mais ces verbes ne sont pas tous neutres de point de vue du registre. Certains véhiculent une connotation soutenue, littéraire (voire poétique) ou même archaïque, d'autres vulgaire. Il était important de contrôler le registre et d'éviter les verbes marqués. Nous avons choisi des verbes qui s'utilisent dans le langage courant, mais qui n'ont pas pour autant de connotation informelle.

**Contraintes sémantiques** Étant donné que nous contrôlons le caractère animé, notre choix de verbes est contraint. Pour neutraliser ce facteur, il fallait soit le contrebalancer, soit lui donner une valeur fixe. Nous avons adopté la seconde option, étant donné la rareté des verbes ditransitifs (simples) prenant un COD humain. Nous avons donc formé tous les stimuli selon le schéma :  $OI_{+hum}/OD_{-anim}$ .

Malgré ces contraintes, nous avons essayé de diversifier le plus possible les lemmes du point de vue de leur sens. Nous avons inclus des verbes causatifs ainsi que des verbes de transfert. L'événement décrit implique un agent humain, un thème concret inanimé et, en fonction du verbe, un bénéficiaire (*dādan*, *baxšīdan*, *ferestādan*, *foruxtān*), une source (*gereftān*, *xarīdan*) ou un patient humain (*xorāndan*).

Les prédicats utilisés sont les suivants :

- *dādan* (*be*) 'donner (à)' (x donne y à z) : 3 items
- *foruxtān* (*be*) 'vendre (à)' (x vend y à z) : 3 items
- *baxšīdan* (*be*) 'octroyer (à)' (x octroie y à z) : 3 items
- *ferestādan* (*barāye*) 'envoyer (pour)' (x envoie y à z) : 2 items
- *xarīdan* (*az*) 'acheter (de)' (x achète y de z) : 3 items
- *gereftān* (*az*) 'prendre (de)' (x prend y de z) : 3 items
- *xorāndan* (*be*) 'faire manger (à)' (x fait manger y à z) : 3 items

Dans tous les items, le préambule comporte une phrase de contexte indépendante, c'est-à-dire séparée du reste du préambule par un point. Ensuite, la phrase à compléter est introduite, soit par un cadrant simple incluant le sujet, ex. (3), soit en tant que complément phrastique, ex. (4).

- (3) ānhā har sāl nazdik=e eyd ...  
 eux chaque an près.EZ nouvel-an  
 ‘Chaque année à l’approche de nouvel an, ils ...’
- (4) ... āqā=ye Parsa movaffaq šod ...  
 Monsieur Parsa réussi devenir.PST.3SG  
 ‘... Monsieur Parsa réussit ...’

Un exemple d’item dans ces quatre conditions est présenté dans le Tableau 5.1. Nos 20 items expérimentaux sont combinés avec 40 distracteurs, dont aucun ne comporte de phrases ditransitives à compléter. À titre d’exemple, les différents constituants en jeu dans les distracteurs sont :

- l’agent et le patient d’un verbe passif,
- l’agent et le patient d’un verbe transitif,
- des ajouts, par exemple différents types de modifieur verbal (temps, manière, lieu, etc.).

### 5.3.1.3 Procédure

Chaque item de cette expérience est présenté par un préambule, une phrase incomplète, où il manque le COD et le COI, représentés par deux cases vides suivies par le verbe, et, sous ces deux cases, trois propositions de syntagme pour compléter la phrase, incluant deux choix de COD et un choix de COI, présentées selon un ordre arbitraire. Tous ces éléments sont présentés simultanément sur l’écran :

نمایشگاه بسیار شلوغ بود. بچه‌ها با پدر مادر هایشان سروکله می‌زدند. کارمند نمایشگاه که از سروصدا کلافه شده بود

نام:

چندتا آدامس

چندتا آبنبات

به بچه‌ها که بی‌تابی می‌کردند

نام →

FIGURE 5.3 – Exp. D1 : Copie d’écran pour un item donné

Les participants complètent la phrase en remplissant les cases par les options de leur choix et dans l’ordre qu’ils souhaitent. La variable dépendante est l’ordre relatif entre le COD et le COI dans la phrase complétée, une variable binaire avec deux valeurs possibles (DO-IO et IO-DO). Le choix entre les deux COD est ignoré puisque les deux présentent les mêmes propriétés formelles. Les réponses dans lesquelles les cases ne sont pas complétées selon ces consignes seront écartées.

<b>COI donné</b>			
<b>Phrase de contexte</b>			
<i>Xānom=e Soltāni zan=e dastodelbāz o mehrabāni bud. Vaqti ba'd az marg=e šohar servat ziādi be ers bord,</i> Madame S. était une femme gentille et généreuse. Lorsqu'elle hérita des propriétés de son mari,			
<b>Phrase à compléter</b>			
	<b>COI &gt; COD</b>	<b>COI &lt; COD</b>	<b>Verbe</b>
<b>COI</b>	<i>be yeki az xāhar-hā=yaš ke tāze navedār šode bud</i> à une de ses sœurs qui a eu un petit-enfant récemment	<i>be yeki az xāhar-hā=yaš</i> à une de ses sœurs	<i>baxšid</i> octroya
<b>COD1</b>	<i>yek āpārtemān</i> un appartement	<i>yek āpārtemān=e sad-o-panjāh-metri=ye do-xābe</i> un appartement de deux pièces de 150 m2	
<b>COD2</b>	<i>yek gardanband</i> un collier	<i>yek gardanband=e morvārid=e dānedorošt=e asl</i> un collier de perles de gros calibre authentique	
<b>COI nouveau</b>			
<b>Phrase de contexte</b>			
<i>Xānom=e Soltāni zamān=e ezdevāj=e xāharzāde=aš, servati nadāšt. Vaqti b'ad az marg=e šohar servatmand šod,</i> Madame S. n'était pas riche quand son neveu (sa nièce) se maria. Lorsqu'elle hérita des propriétés de son mari,			
<b>Phrase à compléter</b>			
	<b>COI &gt; COD</b>	<b>COI &lt; COD</b>	<b>Verbe</b>
<b>COI</b>	<i>be xāharzāde=aš ke tāze baččedār šode bud</i> à son neveu (sa nièce) qui a eu un enfant récemment	<i>be xāharzāde=aš</i> à son neveu (sa nièce)	<i>baxšid</i> octroya
<b>COD1</b>	<i>yek āpārtemān</i> un appartement	<i>yek āpārtemān=e sad-o-panjāh-metri=ye newsāz</i> un appartement de deux pièces neuf	
<b>COD2</b>	<i>yek gardanband</i> un collier	<i>yek gardanband=e morvārid=e dānedorošt=e asl</i> un collier de perles de gros calibre authentique	

TABLE 5.1 – Exp. D1 : Exemple d'un item dans ses quatre conditions

### 5.3.1.4 Participants

33 locuteurs natifs du persan, âgé de 19 à 59 ans (avec des âges moyen de 30,96 et médian de 29 ans), dont 14 femmes et 19 hommes, ont rempli le questionnaire. Aucune réponse n'a été écartée de la table de données finale.

Il faut noter que 54 locuteurs natifs avaient rempli un premier questionnaire que nous avons dû écarter à cause d'une erreur de codage. En raison de cela, chaque participant a rempli une sélection de 16 items sur les 20 prévus. Après avoir résolu cette erreur, nous avons recruté de nouveaux participants sans qu'il y ait de recoupement entre les deux groupes. Les résultats obtenus à partir des données de ce premier groupe s'accordent avec les résultats des données des 33 participants que nous rapportons ici. Les résultats du modèle calculé à partir de ce jeu de données se trouvent dans les annexes (cf. Annexe D.1.1).

### 5.3.1.5 Résultats

La distribution de l'ordre dans l'ensemble des données, cf. Tableau 5.2, démontre une préférence claire pour l'ordre OD-OI-V (68%), comme prédite par notre première hypothèse. Notons que la proportion de l'ordre OD-OI reste supérieure à la moyenne dans toutes nos conditions expérimentales.

Conditions	OD-OI-V	OI-OD-V	Total
IDC	142	25	165
IDL	95	70	165
INC	123	42	165
INL	89	76	165
Total	449	211	660

TABLE 5.2 – Exp. D1 : Distribution de l'ordre pour chaque condition

La distribution de l'ordre en fonction du lemme verbal, cf. Figure 5.4, montre qu'il y a effectivement une variation liée au verbe. Le verbe *xorāndan* 'faire manger' présente un biais significatif pour l'ordre OD-OI-V (81,8% vs. 68%;  $\chi^2(1)=8,1467$ ,  $p<0,005$ ). Quant aux autres, il n'y pas de différence significative ( $ps>0,1$ )<sup>14</sup>.

La distribution de l'ordre en fonction de nos deux facteurs expérimentaux, cf. Figure 5.5, montre que lorsque le COD est plus long que le COI, la préférence pour l'ordre OD-OI est beaucoup plus importante que lorsqu'il en est plus court (80,3% vs. 55,7%). En ce qui concerne le statut discursif du COI, la préférence pour l'ordre

14. Nous avons également analysé ce biais en calculant un modèle mixte incluant ce facteur comme effet fixe (avec le sujet et l'item comme effets aléatoires). Les résultats de ce modèle s'accordent avec ceux obtenus à l'aide du test  $\chi^2$ .

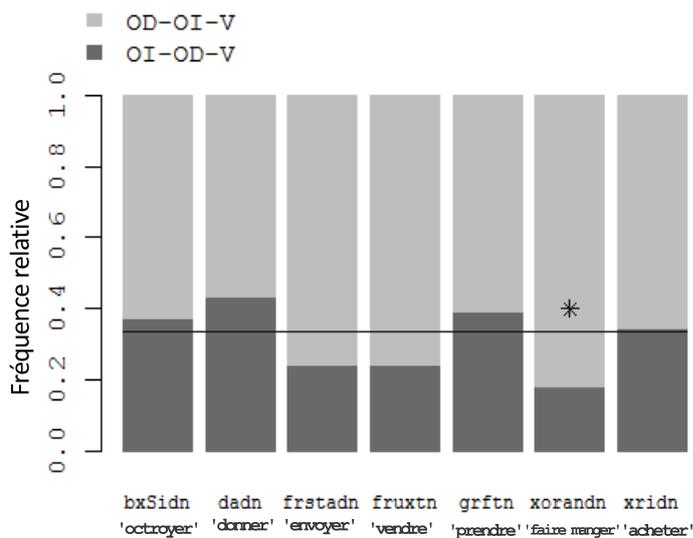


FIGURE 5.4 – Exp. D1 : Distribution de l'ordre pour le lemme verbal

DO-IO est plus importante pour les COI donnés, ce qui est à l'opposé de la préférence attendue, c'est-à-dire la préférence “donné-avant-nouveau”. Toutefois, la différence entre les deux taux est relativement petite (71,8% vs. 64,2%).

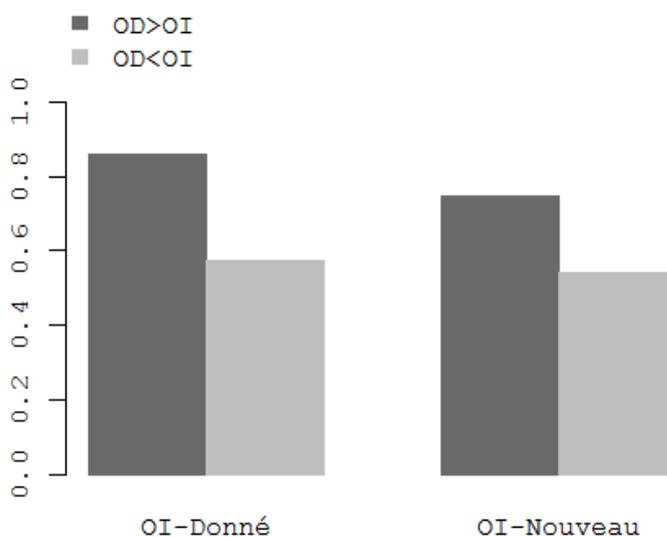


FIGURE 5.5 – Exp. D1 : % de l'ordre OD-OI-V

Les résultats du modèle optimal de régression logistique mixte que nous avons calculé, avec les deux variables dépendantes en tant que facteurs à effet fixe<sup>15</sup> et les variables sujet, item et lemme verbal<sup>16</sup> en tant que facteurs à effet aléatoire, cf. Ta-

15. Le modèle est codé selon le contraste de type “somme” pour les deux facteurs. Ainsi, l'intercepte représente la moyenne générale et les coefficients les écarts de chaque groupe par rapport à cette valeur.

16. Notons que la variable item est nichée dans la variable verbe, mais cela n'a aucun impact sur

	Estimation	Erreur-type	Valeur de z	Valeur de p
(Intercepte)	2,405	0,397	6,080	<0,0001
OI>OD	-1,972	0,430	-4,587	<0,0001
OI-NOUVEAU	-0,992	0,324	-3,059	0,0022
OI>OD :OI-NOUVEAU	0,825	0,407	2,027	0,0426

N.B. : Le succès correspond à ORDRE=OD-OI

TABLE 5.3 – Exp. D1 : Résumé des résultats du modèle optimal pour les effets fixes

bleau 5.3, montrent une préférence générale claire pour l'ordre OD-OI-V prédite par notre première hypothèse.

Ensuite, ils montrent que la longueur a un effet significatif. Celle-ci, avec un coefficient négatif pour l'ordre OD-OI dans la condition OI-nouveau, correspond à la préférence “long-avant-court” prédite par notre deuxième hypothèse. Le statut discursif du COI montre également un effet significatif ( $p < 0,005$ ). Or, cet effet, avec un coefficient négatif pour l'ordre OD-OI dans la condition OI>OD, correspond à la préférence “nouveau-avant-donné” et va ainsi dans le sens inverse de la préférence prédite par notre troisième hypothèse. Il existe également une interaction entre les deux variables ( $p < 0,05$ ) avec un coefficient positif, ayant presque la même valeur que le coefficient du statut discursif du COI. Cette interaction signifie que la préférence “nouveau-avant-donné” (des COD nouveaux) s'annule dans la condition OI>OD. Autrement dit, quand le COI est long, son statut discursif ne joue pas d'effet, tandis que lorsqu'il est court, le fait d'être nouveau favorise l'ordre OI-OD ou le fait d'être donné favorise l'ordre OD-OI.

### 5.3.1.6 Discussion

Ces résultats nous permettent de confirmer nos deux premières hypothèses : 1) les COD indéfinis préfèrent se placer avant le COI et 2) la longueur relative a un effet significatif sur l'ordre correspondant à la préférence “long-avant-court”. En ce qui concerne notre dernière hypothèse, les résultats ne vont dans le sens de ce qui est attendu, mais ne permet pas non plus d'établir une préférence “nouveau-avant-donné” générale. Nous attendons les résultats de l'expérience suivante, qui a la même configuration quant au statut discursif du COI, pour nous prononcer sur cette hypothèse. En ce qui concerne notre dernière hypothèse, bien que les résultats ne vont pas dans le sens attendu, ils ne permettent pas non plus de trancher en faveur de la préférence inverse.

Par ailleurs, une considération d'ordre stylistique pourrait expliquer la tendance la modélisation. Le modèle se calcule dans R de la même manière que pour les variables aléatoires croisées.

à mettre les COI donnés courts après le COD. Dans la condition où le référent du COI est mentionné dans le contexte, la reprise n'est pas pronominale, ce qui, n'étant pas justifié par la distance entre les deux, alourdit la phrase et n'est pas préconisé du point de vue stylistique. Cela est notamment le cas dans la condition COI<COD où les deux expressions sont (presque) identiques, ce qui pourrait amener les locuteurs à placer le COI le plus loin possible pour maximiser leur distance. Dans la condition COI>COD, en revanche, l'ajout d'une subordonnée relative pourrait justifier la répétition.

### 5.3.2 Expérience D2 : Les COD non marqués nus modifiés

Nous avons adapté l'expérience D1 pour l'appliquer aux COD nus modifiés. Cette deuxième expérience reprend ainsi la conception et les items de l'expérience D1, avec une légère modification concernant les conditions **INC** (OI-nouveau-court) et **IDC** (OI-donné-court), car, étant donné la difficulté de manipuler la longueur de ce type de COD, nous avons gardé ce dernier identique dans les quatre versions. Ainsi, dans chaque item, le COD est constitué d'un nom nu et d'un ajout, ex. COD1 : *gardanband=e morvārid* 'collier de perle', COD2 : *āpārtēmān=e nowsāz* 'appartenance neuf (nouvellement construit)'. Cette conception permet d'étudier la position préférée de ces COD par rapport au COI, mais aussi de tester si nos hypothèses 2 et 3 formulées plus haut sont valides indépendamment du type de COD.

Les hypothèses expérimentales sous-jacentes sont les suivantes. Mis à part la première, elles sont identiques à celles de l'expérience précédente.

**Hypothèse 1** Les COD nus modifiés se positionnent de préférence après le COI.

**Hypothèse 2** La longueur favorise le positionnement plus en amont de la phrase.

**Hypothèse 3** L'accessibilité discursive favorise le positionnement plus en amont de la phrase.

#### 5.3.2.1 Conception

Ci-dessous la conception de l'expérience en résumé :

**Variables indépendantes** : voir Exp. D1.

**Conditions** :

1. COI donné / COI = COD : **IDC** (IO-donné-court)
2. COI donné / COI > COD : **IDL** (IO-donné-long)
3. COI nouveau / COI = COD : **INC** (IO-nouveau-court)

## 4. COI nouveau / COI &gt; COD : INL (IO-nouveau-long)

**Variables dépendantes** : voir Exp. D1.

**Variable de contrôle** : voir Exp. D1.

### 5.3.2.2 Stimuli

Les stimuli sont identiques à ceux de l'expérience précédente, excepté quelques légères modifications pour adapter les items aux nouvelles conditions. Il en va de même pour les items distracteurs qui sont tous identiques à ceux utilisés dans le questionnaire de D1.

### 5.3.2.3 Procédure

La procédure de cette expérience est entièrement identique à celle de l'expérience précédente (cf. §5.3.1.3), sauf que la liste des options est présentée dans un ordre contrebalancé. Dans la première moitié des stimuli, la liste suit, de haut en bas, l'ordre COD1, COD2, COI, et, dans la seconde moitié COI, COD1, COD2.

### 5.3.2.4 Participants

26 locuteurs natifs du persan, âgés de 24 à 42 ans (avec des âges moyen de 31,42 et médian de 31 ans), dont 16 femmes et 10 hommes, ont rempli le questionnaire. Aucune réponse n'a été écartée de la table de données finale.

### 5.3.2.5 Résultats

La distribution de l'ordre dans l'ensemble des données, cf. Tableau 5.4, démontre que l'ordre OI-OD-V a été choisi dans 89,9% des cas, ce qui reflète une préférence quasi-systématique pour cet ordre, prédite par notre première hypothèse.

Conditions	OD-OI-V	OI-OD-V	Total
IDC	19	111	130
IDL	8	122	130
INC	18	107	125
INL	7	123	130
Total	52	463	515

TABLE 5.4 – Exp. D2 : Distribution de l'ordre pour chaque condition

La distribution de l'ordre en fonction du lemme verbal, cf. Figure 5.6, montre qu'il y a effectivement une variation liée au verbe. Le biais du verbe *xorāndan*

'faire manger' pour l'ordre OD-OI-V se confirme (74,4% vs. 89,9%;  $\chi^2(1)=13,9204$ ,  $p<0,005$ ). En outre, nous constatons que, dans une moindre mesure, les verbes *gereftan* 'prendre' (100% vs. 89,9%;  $\chi^2(1)=7,4167$ ,  $p<0,01$ ) et *dādan* 'donner' (98,7% vs. 89,9%;  $\chi^2(1)=5,4297$ ,  $p<0,05$ ) ont un biais pour l'ordre inverse, c'est-à-dire IO-DO-V<sup>17</sup>.

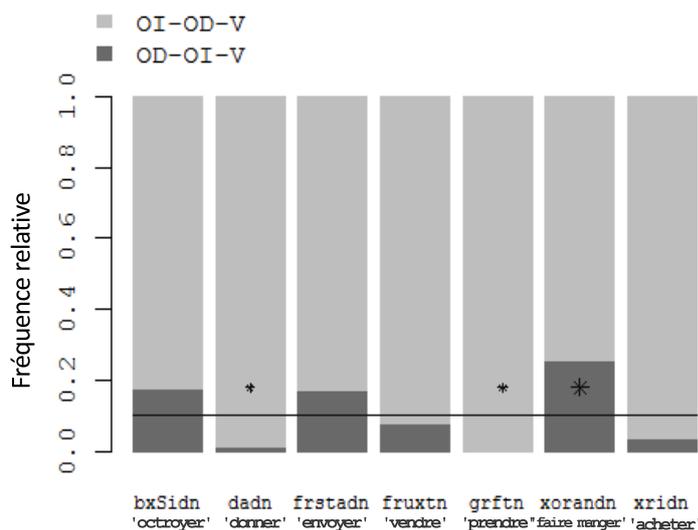


FIGURE 5.6 – Exp. D2 : Distribution de l'ordre pour le lemme verbal

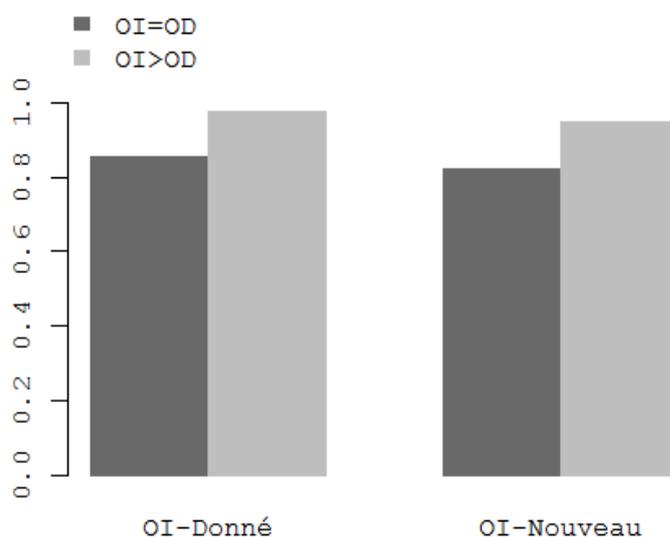


FIGURE 5.7 – Exp. D2 : % de l'ordre OI-OD-V

La distribution de l'ordre en fonction de nos deux facteurs expérimentaux, cf. Figure 5.7, montre que, lorsque le COD est plus long que le COI, la préférence pour

17. Les résultats du modèle mixte incluant le verbe comme effet fixe appuient l'effet significatif du verbe *xorāndan* 'faire manger'.

l'ordre OD-OI est plus importante dans le cas inverse (94,23% vs. 85,49%). En ce qui concerne le statut discursif du COI, la proportion de l'ordre OI-OD est presque identique dans les deux conditions (90,20% vs. 89,62%).

	Estimation	Erreur-type	Valeur de z	Valeur de p
(Intercepte)	4,299	0,830	5,175	<0,0001
OI>OD	1,479	0,308	4,802	<0,0001
OI-DONNÉ	-0,040	0,203	-0,197	0,844
OI>OD :OI-DONNÉE	-0,059	0,202	-0,294	0,769

N.B. : Le succès correspond à ORDRE=OI-OD

TABLE 5.5 – Exp. D2 : Résumé des résultats du modèle optimal pour les effets fixes

Les résultats du modèle optimal de régression logistique mixte que nous avons calculé, avec les deux variables dépendantes en tant que facteurs à effet fixe<sup>18</sup> et les variables sujet, item et lemme verbal en tant que facteurs à effet aléatoire, cf. Tableau 5.5, montrent une forte préférence générale pour l'ordre OI-OD-V prédite par notre première hypothèse. De plus, ils montrent que la longueur a un effet significatif, qui, avec un coefficient positif pour l'ordre OI-OD dans la condition OI>OD, correspond à la préférence “long-avant-court”, prédite par notre deuxième hypothèse. En revanche, ni le statut discursif du COI, ni l'interaction entre les deux variables ont un effet significatif ( $ps > 0,75$ ).

### 5.3.2.6 Discussion

Ces résultats confirment nos deux premières hypothèses : 1) les COD nus modifiés préfèrent se placer après le COI et 2) la longueur relative a un effet significatif sur l'ordre correspondant à la préférence “long-avant-court”. En ce qui concerne notre troisième hypothèse, les résultats de cette expérience ne montrent aucun effet du statut discursif du COI sur l'ordre et donc n'appuient pas la préférence “nouveau-avant-donné” (plus précisément, OI-nouveau-court avant OI-donné-court), que l'on a pu observer avec les données de l'expérience D1.

Une absence d'effet du statut discursif du COI – plus précisément de son caractère donné vs. nouveau – pourrait s'expliquer par le fait que, sur la hiérarchie de l'accessibilité des rôles grammaticaux, le COI occupe une place inférieure au COD et donc son statut discursif joue un rôle moins important que celui du COD. Nos données à ce stade ne permettent pas de se prononcer sur cette hypothèse. Premièrement, parce que nous n'avons seulement manipulé que le statut du COI et ne disposons pas de données comparables pour le statut discursif du COD. Deuxièmement, car

18. Du même que dans le modèle précédent, l'intercepte représente la moyenne générale et les coefficients les écarts de chaque groupe par rapport à cette valeur.

il s'agit d'un non effet et il est possible qu'avec d'autres paradigmes expérimentaux l'on puisse détecter un effet du statut discursif potentiellement en interaction avec d'autres facteurs comme la prosodie. Monter des expériences avec de tels paradigmes dépasse le cadre de cette thèse. Ainsi, nous laissons l'étude approfondie de ce facteur pour de recherches futures.

### 5.3.3 Expérience D3 : Les COD non nus

Dans cette expérience, nous nous concentrons sur les COD non nus, avec un double objectif. Notre premier objectif est de savoir si les COD marqués ont effectivement une préférence forte pour l'ordre OD-OI-V, et si les COD (non marqués) indéfinis ont une préférence moins forte pour cet ordre.

Ensuite, nous voulons comparer le comportement des deux différents types de COD (non marqué) indéfini, à savoir l'indéfini portant l'enclitique  $=i$ , ex. *ketāb=i*, ou portant le déterminant *ye(k)*, ex. *yek ketāb*. Ces deux COD ont un même niveau de détermination, bien qu'ils ne soient pas toujours interchangeables (cf. §1.1.2.2). De ce fait, nous pensons qu'au niveau de leur position linéaire vis-à-vis du COI, le seul facteur qui pourrait entrer en jeu est le poids grammatical. D'un côté, l'on peut dire que, du point de vue purement syntaxique, les deux COD comportent le même nombre d'unités syntaxiques et ont, par conséquent, un poids grammatical identique. Dans ce cas, l'on ne prédit aucune différence dans leur ordonnancement linéaire par rapport au COI. De l'autre côté, sur le plan graphique et phonologique, l'un forme un seul mot, tandis que l'autre en forme deux (sur les deux plans). Dans ce cas, l'on prédit une légère avance des COD portant le déterminant *yek* sur ceux portant l'enclitique  $=i$ .

Les hypothèses expérimentales sous-jacentes sont les suivantes :

**Hypothèse 1** Les COD marqués de  $=rā$  ont une forte préférence pour se positionner avant le COI.

**Hypothèse 2** Les COD indéfinis (non marqués de  $=rā$ ) se positionnent de préférence avant le COI, mais avec une préférence moins forte que les COD marqués de  $=rā$ .

**Hypothèse 3** Il y a potentiellement une légère avance des COD indéfinis avec *yek* pour préférer l'ordre OD-OI-V sur les COD indéfinis marqués de  $=i$ .

#### 5.3.3.1 Conception

Nous résumons ci-dessous la conception de l'expérience.

**Variables indépendantes :**

— Type de COD.

**Conditions :**

1. COD marqué :  $\mathbf{R\bar{A}}$ , ex. (5-a)
2. COD indéfini déterminé par *yek* :  $\mathbf{YEK}$ , ex. (5-b)
3. COD indéfini marqué de  $=i$  :  $\mathbf{I}$ , ex. (5-c)

**Variables dépendantes :**

— Ordre relatif entre l'OD et l'OI dans la phrase complétée par le participant, opérationnalisé par une variable catégorique à deux modalités.

**Variable de contrôle :**

1. Longueur relative :

Le COD et le COI comportent le même nombre d'éléments lexicaux (hormis la préposition).

2. Caractère animé des constituants :

Tous les items respectent le schéma  $\text{OI}_{+anim}/\text{OD}_{-anim}$ .

- (5) a. ... [<sub>COD1</sub> ketāb=rā] [<sub>COI</sub> beyn=e bačče-hā] [<sub>V</sub> paxš  
livre=RĀ entre=EZ enfant-PL distribution  
kard]  
faire.PST.3SG  
'... il distribua le livre parmi les enfants.'
- b. ... [<sub>COD1</sub> yek ketāb] [<sub>COI</sub> beyn=e bačče-hā] [<sub>V</sub> paxš  
un livre entre=EZ enfant-PL distribution  
kard]  
faire.PST.3SG  
'... il distribua un livre parmi les enfants.'
- c. ... [<sub>COD1</sub> ketāb=i] [<sub>COI</sub> beyn=e bačče-hā] [<sub>V</sub> paxš  
livre=INDF entre=EZ enfant-PL distribution  
kard]  
faire.PST.3SG  
'... il distribua un livre parmi les enfants.'

### 5.3.3.2 Stimuli

Nous avons préparé trois séries de 15 phrases ditransitives, comme la série illustrée en ex. (5), correspondant à nos trois conditions expérimentales. Chaque phrase à compléter comporte deux choix formellement identiques de COD, un COI et un verbe, sachant que le sujet est présent dans le préambule. Toutes les phrases sont à la troisième personne et aucun déictique n'y est utilisé.

Dans cette expérience, nous ne nous sommes plus limitée aux verbes simples, parce que cette contrainte dans la préparation de nos stimuli (dans D1 et D2) ne constitue pas une nécessité méthodologique ou empirique incontournable. Il faut noter que, dans les recherches empiriques similaires sur le persan, les chercheurs utilisent les verbes simples et les prédicats complexes de manière indifférente. Nous avons donc utilisé des prédicats complexes, hautement lexicalisés, se comportant (du point de vue sémantique) comme des verbes (simples) ditransitifs (cf. §1.1.4).

Du point de vue de l'appariement des rôles grammaticaux, nous nous sommes limitée aux prédicats dénotant *grosso modo* un changement de possession, c'est-à-dire aux verbes de transfert (du type "donner"), notamment du fait du biais avéré pour le verbe causatif *xorāndan* dans les deux expériences précédentes. Au niveau des arguments, dans toutes les phrases, le sujet est un nom animé (humain) correspondant à l'argument agent, les deux choix de COD sont des noms inanimés correspondant à l'argument thème et le COI est un nom animé (humain) correspondant à l'argument bénéficiaire.

Les prédicats utilisés sont les suivants :

- *āvardan* (*barāye*) 'apporter (pour)' (x apporte y pour z) : 2 items
- *dādan* (*be*) 'donner (à)' (x donne y à z) : 5 items
- *ferestādan* (*barāye*) 'envoyer (pour)' (x envoie y pour z) : 2 items
- *hedīye dādan* (*be*) 'offrir (à) (Lit. cadeau donner)' (x offre y à z) : 2 items
- *paš kardan* (*beyn=e*) 'distribuer entre=EZ (Lit. distribution faire)' (x distribue y entre z) : 2 items
- *pišnahād kardan* (*be*) 'proposer (à) (Lit. proposition faire)' (x propose y à z) : 2 items

Il faut noter que les deux choix du COD sont possibles, du point de vue de la grammaticalité, c'est-à-dire qu'aucun ne rend la phrase agrammaticale ou inacceptable, mais ils peuvent ne pas avoir la même pertinence en contexte.

Nous avons également contrôlé la longueur et la structure des deux arguments, ex. (6), à comparer avec (5) ci-dessus, et (7).

- (6) a. [COD<sub>1</sub> (yek) ketāb]  
(un) livre
- b. [COD<sub>2</sub> (yek) jozve]  
(un) fascicule
- (7) a. [COD<sub>1</sub> (yek) jāyeze=ye viĵe]  
(un) prix=EZ spécial
- b. [COD<sub>2</sub> (yek) medāl=e eftexār]  
(un) médaille=EZ honneur
- c. [COI (be) kārgardān=e javān]  
(à) réalisateur=EZ jeune

Dans cette expérience, étant donné que notre plan expérimental n'exige pas de manipulation sur le plan discursif, nos phrases à compléter ne sont pas présentées par une phrase de contexte. Ainsi, en guise de préambule, nous présentons nos phrases cibles par un simple cadran suivi du sujet de la phrase ditransitive à compléter, ex. (8), ou en tant que complément phrastique, ex. (9).

- (8) dar daqāyeyq=e axar=e jalase=ye sevvom=e kelās ostād=e  
 dans minutes=EZ fin=EZ séance=EZ troisième=EZ cours professeur=EZ  
 riyāzi ...  
 mathématiques  
 'Quelques minutes avant la fin de la troisième séance de cours, le professeur  
 de mathématiques ...'
- (9) ba'd-az čand sāat mašrevat dāvarān tasmim gereft-and ...  
 après quelques heure concertation arbitres décision prendre.PST-3PL  
 'Après quelques heures de concertation, les arbitres ont décidé (qu'ils)...

Les 15 items de cette expérience ont été combinés avec 40 items distracteurs, dont 10 items expérimentaux de l'expérience D4 (cf. §5.3.4) et 20 items de l'expérience T1 (cf. §5.4.1). Les 10 distracteurs supplémentaires portent tous sur l'emploi de =rā avec des circonstants temporels, cf. ex. (10), pour une expérience pilote en dehors de cadre de cette thèse.

- (10) Mahsa sinemā ne-mi-rav-ad čon mi-xāh-ad [Op1  
 Mahsa cinema NEG-IPFV-aller.PRS-3SG car IPFV-vouloir.PRS-3SG  
 in aṅar=e hafte] [Op2 dar xāne] [Op3 in aṅar=e hafte=rā] film  
 ce fin=EZ semaine dans maison ce fin=EZ semaine=RA film  
 be-bin-ad  
 SBJV-regarder.PRS-3SG  
 'Mahsa ne va pas au cinéma, car, ce week-end, elle veut regarder des films  
 chez elle.'

### 5.3.3.3 Procédure

Chaque item de cette expérience est présenté par un préambule, à savoir une phrase incomplète où il manque le COD, le COI et le verbe représentés par trois cases vides, et quatre propositions de syntagme pour compléter la phrase, incluant deux choix de COD, un choix de COI et un choix de verbe, présentées dans un ordre contrebalancé. Dans la première moitié des stimuli, la liste suit, de haut en bas, l'ordre verbe, COD1, COD2, COI et, dans la seconde moitié, l'ordre verbe, COI, COD1, COD2. Tous ces éléments sont présentés simultanément sur l'écran, cf. Figure 5.8.



FIGURE 5.8 – Exp. D3 : Copie d'écran pour un item donné

Les participants complètent la phrase en remplissant les cases par les options de leur choix et dans l'ordre qu'ils souhaitent. La variable dépendante de l'expérience est l'ordre entre les trois constituants dans la phrase complétée, une variable catégorique qui peut théoriquement avoir six valeurs (DIV, IDV, DVI, IVD, VID et VDI). Seuls les deux ordres sur lesquels nous nous concentrons, c'est-à-dire DIV et IDV, se sont produits. Par souci de cohérence entre les expériences, nous les présentons par DO-IO et IO-DO.

Comme dans les deux expériences précédentes, le choix entre les deux COD est ignoré car les deux présentent les mêmes propriétés formelles. Les réponses dans lesquelles les cases ne sont pas complétées selon ces consignes seront écartées.

#### 5.3.3.4 Participants

60 locuteurs natifs du persan, âgés de 23 à 59 ans (avec un âge moyen et médian de 31,63 et 29,5 respectivement), dont 35 femmes et 25 hommes, ont rempli le questionnaire.

À cause d'une erreur, nous avons écarté les réponses de nos 21 premiers participants, pour lesquels la répartition des items dans les conditions ne s'est pas faite de manière unifiée. Nous avons eu aussi une réponse erronée dans la condition "Rā". Cela nous laisse un ensemble de 584 observations dans notre table de données finale. Il faut noter que les résultats en incluant les données de ces 21 participants, c'est-à-dire 898 observations au lieu des 584, ne diffèrent pas des résultats rapportés ici, dans leurs indications principales<sup>19</sup>.

#### 5.3.3.5 Résultats

Les données brutes de la distribution de l'ordre pour les trois conditions sont présentées dans le Tableau 5.6. Les proportions moyennes dans les trois conditions,

19. Les résultats incluant tous les participants se trouvent dans les annexes (cf. Annexe D.1.3).

dans l'ordre 83,6%, 60% et 65,5%, montrent que l'ordre OD-OI-V se produit plus fréquemment que l'ordre inverse pour les trois types du COD (cf. Figure 5.9). Nous observons que les COD non marqués portant le déterminant *yek* ont une préférence numériquement moins élevée que ceux portant l'enclitique *=i*.

Conditions	OD-OI-V	OI-OD-V	Autres ordres	Total
Rā	163	31	0	194
Yek	117	78	0	195
I	127	68	0	195
Total	407	177	0	584

TABLE 5.6 – Exp. D3 : Distribution de l'ordre pour chaque condition

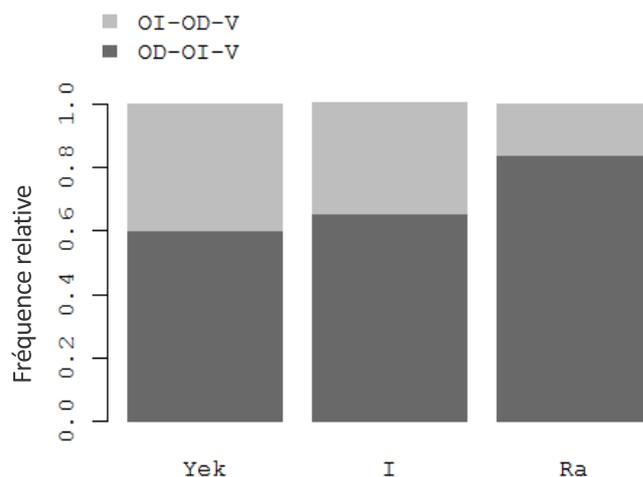


FIGURE 5.9 – Exp. D3 : Distribution de l'ordre

Par ailleurs, en ce qui concerne la distribution de l'ordre en fonction du lemme verbal, cf. Figure 5.10, nous constatons un faible écart par rapport à la moyenne, mis à part pour les deux verbes *ferestādan* 'envoyer' et *hediye dādan* 'offrir', qui ont un biais significatif ( $p < 0,005$ ) respectivement pour l'ordre OD-OI-V et l'ordre OI-OD-V (85,7% et 55,8% vs. 69,2%)<sup>20</sup>.

Les résultats du modèle optimal de régression logistique mixte que nous avons calculé avec notre variable indépendante, en tant que facteurs à effet fixe, et les variables sujet et item, en tant que facteurs à effet aléatoire, cf. Tableau 5.7, notamment la valeur positive de l'intercepte ( $p < 0,05$ ), confirment l'ordre OD-OI pour l'ensemble de ces COD. Comme attendu, la préférence des COD marqués pour l'ordre OD-OI

20. Les résultats du modèle mixte incluant le verbe comme effet fixe appuient seulement l'effet significatif du verbe *ferestādan* 'envoyer'.

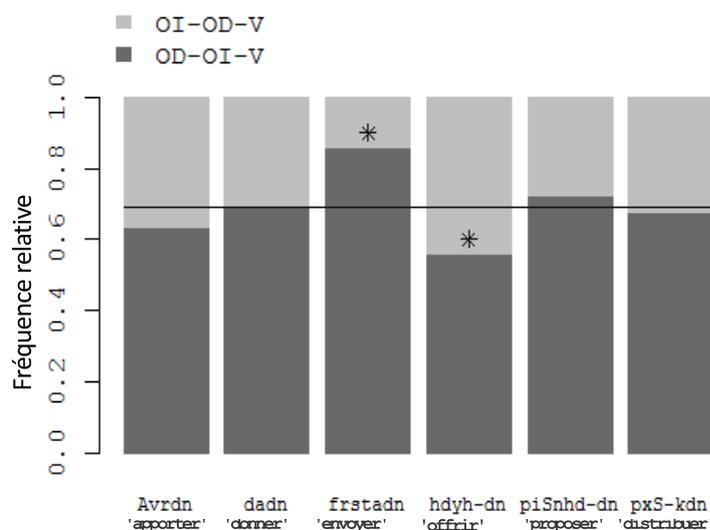


FIGURE 5.10 – Exp. D3 : Distribution de l'ordre pour le lemme verbal

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte (OD=YEK)	0,447	0,220	2,025	0,043
OD=I	0,473	0,334	1,417	0,156
OD=RA	1,661	0,394	4,210	<0,0001

N.B. : Le succès correspond à ORDRE=OD-OI

TABLE 5.7 – Exp. D3 : Résumé des résultats du modèle optimal pour les effets fixes

est plus forte que pour les deux autres types de COD. En revanche, l'on n'observe pas de différence significative entre les deux différents types de COD non marqué indéfini ( $p > 0,15$ ).

### 5.3.3.6 Discussion

Ces résultats valident nos deux premières hypothèses. Il faut néanmoins noter qu'en ce qui concerne la première hypothèse, la fréquence moyenne observée dans cette expérience est plus basse que ce à quoi nous nous attendions étant donné les résultats de notre étude de corpus dans lequel l'ordre OD-OI-V se produit avec un taux de 96% (contre 83.6% ici). Quant à notre troisième hypothèse, ces données privilégient le critère purement syntaxique ne prédisant aucune différence entre ces deux COD indéfinis.

### 5.3.4 Expérience D4 : Les COD nus

Dans cette expérience, nous nous concentrons sur les COD nus, dans un but également double. En premier lieu, nous voulons savoir si les COD nus ont effective-

ment une préférence forte pour l'ordre OI-OD-V et, en second lieu, si cette préférence décroît avec la longueur, c'est-à-dire avec la présence de modifieurs.

Les hypothèses expérimentales sous-jacentes sont les suivantes :

**Hypothèse 1** Les COD nus ont une forte préférence pour l'ordre OI-OD-V.

**Hypothèse 2** Les COD nus modifiés se trouvent moins fréquemment dans l'ordre OI-OD-V que les COD nus.

### 5.3.4.1 Conception

Nous résumons ci-dessous la conception de l'expérience.

**Variables indépendantes :**

— Type de COD ou présence de modifieurs.

**Conditions :**

1. COD nu : **NU**, ex. (11-a)
2. COD nu modifié : **NU-MOD**, ex. (11-b)

**Variables dépendantes :**

— Ordre relatif entre l'OD et l'OI dans la phrase complétée par le participant, opérationnalisé par une variable catégorique à deux modalités.

**Variable de contrôle :**

1. Longueur (et réalisation) du COI :

Tous les COI sont composés d'une préposition et d'un nom.

2. Caractère animé des constituants :

Dans cette expérience, contrairement aux précédentes, le COI est un argument inanimé locatif. Ainsi, tous les items respectent le schéma OI-*anim*/OD-*anim*. Ce choix nous permettrait éventuellement d'observer un effet de l'animéité par la comparaison entre cette expérience et l'expérience D2 (cf. §5.3.2).

- (11) a. ... [COD<sub>1</sub> gol] [COI sar=e miz] [V be-gozār-and]  
           fleur sur=EZ table SBJV-mettre.PRS-3PL  
           '... qu'ils mettent une/des fleurs sur la table.'
- b. ... [COD<sub>1</sub> gol=e orkide=ye sefid] [COI sar=e miz] [V  
           fleur=EZ orchidée=EZ blanc sur=EZ table  
           be-gozār-and]  
           SBJV-mettre.PRS-3PL  
           '... qu'ils mettent une/des (fleur(s) de) orchidée(s) blanche(s) sur la  
           table.'

### 5.3.4.2 Stimuli

Nous avons préparé deux séries de 10 phrases ditransitives, similaires aux deux phrases de l'exemple (11), correspondant à nos deux conditions expérimentales. Similaire à l'expérience précédente, chaque phrase à compléter comporte deux choix formellement identiques de COD, un COI et un verbe simple ou un prédicat complexe. Toutes les phrases sont à la troisième personne et aucun déictique n'y est utilisé.

Du point de vue de l'appariement des rôles grammaticaux, nous nous sommes limitée aux prédicats dénotant *grosso modo* un changement de position, *i.e.* verbes de déplacement (de type “mettre”), qui conviennent mieux avec un thème inanimé. Au niveau des arguments, dans toutes les phrases, le sujet est un nom animé (humain) correspondant à l'argument agent, les deux choix de COD sont des noms inanimés correspondant à l'argument thème et le COI un nom inanimé correspondant à l'argument but ou l'argument source.

Les prédicats utilisés sont les suivants :

- *āvardan* (*az*) ‘apporter (de)’ (x apporte y de z) : 2 items
- *ferestādan* (*be*) ‘envoyer (à)’ (x envoie y à z) : 2 items
- *gozāštan* (*sar/tu*) ‘mettre (sur/dans)’ (x met y sur/dans z) : 2 items
- *tozi' kardan* (*dar*) ‘distribuer (dans) (Lit. distribution faire)’ (x distribue y dans z) : 3 items
- *vāred kardan* (*az*) ‘importer (de) (Lit. entrée faire)’ (x importe y de z) : 1 item

Dans chaque item, nous avons ajouté deux modificateurs adjectivaux au COD nu pour obtenir un COD nu modifié. Dans chaque condition, nous avons contrôlé la longueur des deux choix du COD, ex. (12) (à comparer avec (11) ci-dessus). Notons enfin que, comme dans les trois expériences précédentes, du point de vue de la grammaticalité, les deux choix du COD sont possibles dans la phrase, mais ils peuvent ne pas avoir la même pertinence en contexte.

- (12) a. [COD<sub>1</sub> gol] vs. [COD<sub>1</sub> gol=e orkide=ye sefid]  
           fleur                  fleur=EZ orchidée=EZ blanc  
           ‘fleur(s)’ vs. ‘(fleur(s) de) orchidée(s) blanche(s)’
- b. [COD<sub>2</sub> āš] vs. [COD<sub>2</sub> āš=e rešte=ye Tabrizi]  
           soupe                  soupe=EZ vermicelle=EZ de-Tabriz  
           ‘(de la) soupe’ vs. ‘(de la) soupe de vermicelle Tabrizi’

Dans cette expérience, comme dans la précédente, nos phrases cibles sont présentées en tant que compléments phrastiques, ex. (13) et (14).

- (13) modir=e hotel sefāreš kard hatman...  
 directeur=EZ hôtel commande faire.PST.3SG impérativement  
 ‘Le directeur de l’hôtel recommanda (les employés) pour qu’ils ... impérati-  
 vement.’
- (14) Tarāne diruz goft ke Vahid qowl dāde ast...  
 Taraneh hier dire.PST.3SG que Vahid promesse donner.PP être.PRS.3SG  
 ‘Taraneh disait hier que Vahid a promis de ...’

Les 10 items de cette expérience ont été combinés avec 45 items distracteurs, dont 15 items expérimentaux de l’expérience D3 (cf. §5.3.3), 20 items de l’expérience T1 (cf. §5.4.1) et 10 distracteurs supplémentaires (détaillés en §5.3.3.2).

### 5.3.4.3 Procédure

La procédure de cette expérience est entièrement identique à celle de l’expérience précédente, cf. §5.3.3.3.

### 5.3.4.4 Participants

Cette expérience a été conduite via le même questionnaire que l’expérience précédente. Les participants des deux expériences sont identiques (voir §5.3.3.4 pour les détails), mais, contrairement à l’expérience D3, nous avons pu retenir les réponses de tous les participants, mises à part trois réponses erronées écartées de la table de données finale. Cela nous laisse 595 réponses au total.

### 5.3.4.5 Résultats

Les données brutes de la distribution de l’ordre pour les trois conditions sont fournies dans le Tableau 5.8. Nous avons écarté de nos analyses la réponse avec un ordre non verbe final (“Autres ordres” dans le tableau). Les proportions moyennes dans les deux conditions, cf. Figure 5.11, montrent une forte préférence pour les COD nus (71,8%) pour l’ordre OI-OD, mais ne montrent qu’une préférence numérique en ce qui concerne les COD nus modifiés (50,7%).

Conditions	OD-OI-V	OI-OD-V	Autres ordres	Total
Nu	84	214	1	299
Nu-mod	147	151	0	298
Total	231	365	1	597

TABLE 5.8 – Exp. D4 : Distribution de l’ordre pour chaque condition

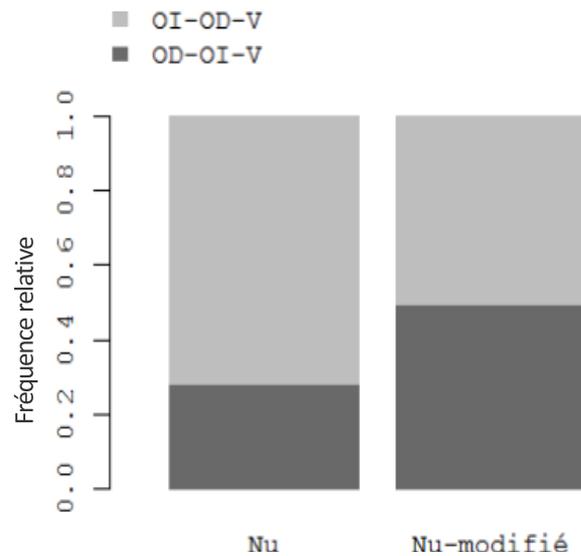


FIGURE 5.11 – Exp. D4 : Distribution de l'ordre

Quant à la distribution de l'ordre en fonction du lemme verbal, cf. Figure 5.12, l'on retrouve le biais du verbe *ferestādan* 'envoyer' pour l'ordre OD-OI-V (52,1% vs. 38,8% ;  $\chi^2(1)=6,7599$ ,  $p<0,01$ ). Nous trouvons également un biais important pour le verbe *āvardan* 'apporter' envers l'ordre OI-OD-V (17,6% vs. 38,8% ;  $\chi^2(1)=18,4569$ ,  $p<0,0001$ )<sup>21</sup>.

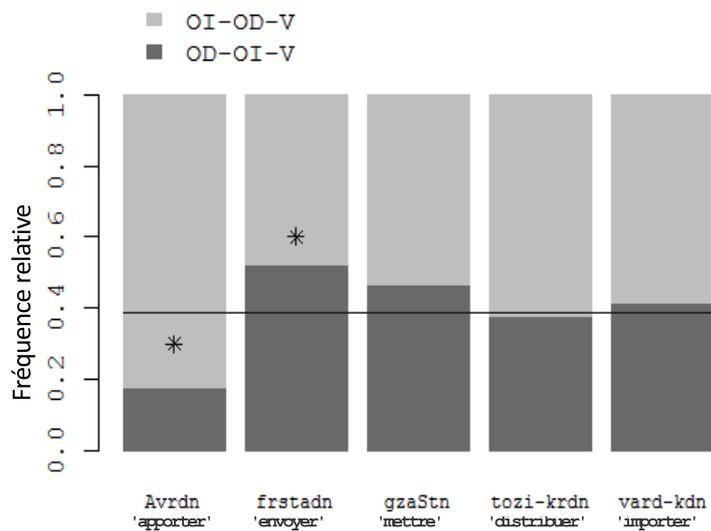


FIGURE 5.12 – Exp. D4 : Distribution de l'ordre pour le lemme verbal

Les résultats du modèle optimal de régression logistique mixte, calculé avec notre variable dépendante en tant que facteur à effet fixe et les variables sujet et item en

21. Les résultats du modèle mixte incluant le verbe comme effet fixe n'appuient par l'effet significatif de ces verbes.

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte (OD=NU)	1,455	0,483	3,009	0,0026
OD=NU-MOD	-1,476	0,266	-5,548	<0,0001

N.B. : Le succès correspond à ORDRE=OI-OD

TABLE 5.9 – Exp. D4 : Résumé des résultats du modèle optimal pour les effets fixes

tant que facteurs à effet aléatoire, cf. Tableau 5.9, confirment la forte préférence pour des COD nus de se placer après le COD et montrent que cette préférence s’annule pour les COD nus modifiés<sup>22</sup>.

### 5.3.4.6 Discussion

Ces résultats confirment globalement nos deux hypothèses. Cependant, les fréquences observées pour l’ordre concurrent, c’est-à-dire OD-OI-V, sont plus élevées que ce à quoi nous nous attendions. Elles sont notamment plus élevées que nos données de corpus (15,8% vs. 28,2% pour les OD nus, et 33,3% vs. 47,1% pour les OD nus modifiés). Plus particulièrement, en ce qui concerne les COD nus modifiés, elles sont beaucoup plus élevées que dans l’expérience D2, où le taux de l’ordre OD-OI-V – dans les conditions COI = COD – était de 14,5%. Or, ces deux expériences diffèrent sur un point crucial à savoir l’animéité du COI. La comparaison des deux résultats (D2 vs. D4) semble indiquer que le caractère animé du COI favorise son positionnement en amont de la phrase.

### 5.3.5 Combinaison des Expériences D3 et D4

Les deux expériences D3 et D4 ont été conduites via le même questionnaire et avec les mêmes participants. De ce fait, pour comparer les différents types de COD que nous avons définis, nous avons combiné les données des deux expériences.

Nous avons inclus les réponses des 60 participants, les réponses erronées et les réponses avec un ordre non verbe final mises à part. Cela nous laisse un ensemble de 1 494 observations réparties sur 25 items.

À cette fin, nous avons réuni les deux variables en une seule variable indépendante, le type de COD, et avons réduit le nombre de nos conditions à quatre en regroupant les deux types de COD indéfini en une seule catégorie. Nos quatre conditions sont donc les suivantes : 1) Nu, 2) Nu modifié, 3) Indéfini et 4) Marqué. La distribution de l’ordre pour chacune est donnée dans le Tableau 5.10.

22. Le coefficient de ces COD a une valeur très proche de l’intercepte avec le signe négatif. Par conséquent, le modèle retourne une valeur proche de zéro (-0,021) signifiant qu’il n’est pas possible de trancher en faveur d’un ordre ou un autre pour ces COD.

Conditions	OD-OI-V	OI-OD-V	Total
Nu	84	214	298
Nu-modifié	147	151	298
Indéfini	363	235	598
Marqué	258	42	300
Total	852	642	1 494

TABLE 5.10 – Exp. D3-D4 : Distribution de l'ordre pour chaque condition

Nous avons analysé ces données avec des modèles de régression logistique mixte, avec notre variable indépendante en tant que facteur à effet fixe et les variables sujet et item en tant que facteurs à effet aléatoire.

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte (OD=NU)	-1,213	0,367	-3,307	0,0009
OD=NU-MODIFIÉ	1,223	0,218	5,615	<0,0001
OD=INDEFINI	1,711	0,393	4,349	<0,0001
OD=MARQUÉ	3,192	0,392	8,136	<0,0001

N.B. : Le succès correspond à ORDRE=OD-OI

TABLE 5.11 – Exp. D3-D4-M1 : Résumé des résultats du modèle optimal pour les effets fixes

Étant donné que les valeurs de p sont toutes suffisamment petites ( $p < 0,005$ ), les coefficients sont directement comparables. Rappelons que, pour une valeur donnée de type de COD, le résultat, c'est-à-dire l'addition du coefficient correspondant avec l'intercepte, donne la probabilité du succès, c'est-à-dire l'ordre OD-OI-V, une fois transformée par la fonction inverse de *logit*. Pour simplifier, les montants positifs prédisent le succès et les montants négatifs l'échec, et le zéro, correspondant à une probabilité de 0,5, signifie l'impossibilité de trancher en faveur de l'un ou l'autre.

Tout d'abord, nous observons un changement de signe de la valeur des coefficients entre le COD nu et les autres types de COD, reflétant les biais des premiers pour l'ordre OI-OD-V et des derniers pour l'ordre inverse<sup>23</sup>. Toutefois, en ce qui concerne le COD nu modifié, bien que l'addition du coefficient et de l'intercepte résulte en une valeur positive (0,01), celle-ci reste proche du zéro et ne permet pas de trancher en faveur de l'ordre inverse<sup>24</sup>. Pour les COD indéfinis, en revanche, cette valeur positive (0,498) est plus importante et permet de trancher en faveur de l'ordre OD-OI-V<sup>25</sup>. Quant aux COD marqués, la valeur (1,979) est encore plus élevée, et impose l'ordre

23.  $\text{logit}^{-1}(-1, 213) = 0, 229$

24.  $\text{logit}^{-1}(0, 01) = 0, 502$

25.  $\text{logit}^{-1}(0, 498) = 0, 622$

OD-OI-V comme l'ordre par défaut<sup>26</sup>. Ainsi, ces résultats mettent en évidence le fait que la préférence pour l'ordre OD-OI-V, autrement dit la probabilité d'avoir l'ordre OD-OI-V, s'accroît à mesure que l'on s'éloigne des COD nus (monolexémiques) et que l'on s'approche des COD marqués de  $=r\bar{a}$ .

Dans un second temps, nous avons écarté la catégorie des COD nus modifiés pour ne garder que les principaux types de COD, qui s'obtiennent en croisant les deux traits binaires essentiels dans la réalisation d'un syntagme nominal occupant la position du COD en persan, c'est-à-dire le marquage par  $=r\bar{a}$ ,  $[\pm R\bar{a}]$  et la nudité,  $[\pm Nu]$  :

1.  $+R\bar{a}/+Nu$  :  $\emptyset$
2.  $+R\bar{a}/-Nu$  : COD Marqué
3.  $-R\bar{a}/+Nu$  : COD Nu
4.  $-R\bar{a}/-Nu$  : COD Indéfini

Cela nous permettra d'évaluer précisément l'ampleur d'effet de chaque trait, respectivement, en comparant la différence, d'un côté, entre les COD indéfini et marqué et, de l'autre côté, entre les COD nu et indéfini. Les résultats du modèle, plus précisément les coefficients et les valeurs de p, sont similaires (cf. Tableau 5.12).

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte (COD NU)	-1,294	0,413	-3,146	0,0016
COD INDÉFINI	1,795	0,436	4,118	<0,0001
COD MARQUÉ	3,281	0,434	7,558	<0,0001

N.B. : Le succès correspond à ORDRE=OD-OI

TABLE 5.12 – Exp. D3-D4-M1 : Résumé des résultats du modèle optimal pour les effets fixes

Ainsi, nous pouvons dire que la probabilité de l'ordre OD-OI-V s'accroît avec le degré de détermination. La Figure 5.13 représente les probabilités obtenues à partir des résultats du modèle (à l'aide fonction  $\text{logit}^{-1}$ ) pour les trois types de COD.

26.  $\text{logit}^{-1}(1,979) = 0,879$

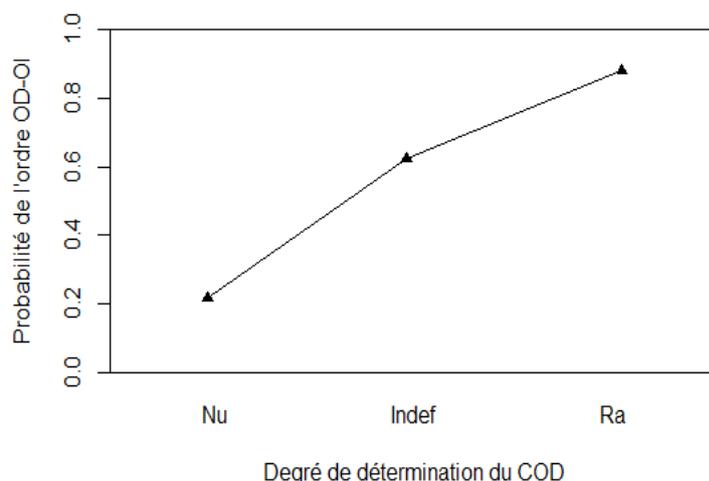


FIGURE 5.13 – La probabilité de l'ordre OD-OI en fonction du degré de détermination du COD

### 5.3.6 Bilan

Dans le Tableau 5.13<sup>27</sup>, nous résumons les préférences observées, en termes de pourcentage, pour les principaux types de COD dans les différentes configurations. Précisons que le “COD Nu” inclut ici les COD nus monolexémiques et les COD nus modifiés. Pour le COD indéfini, nous tenons compte des données rapportées dans le Tableau 5.10.

COD	COI Humain				COI Inanimé				M
	I≈D	I<D	I>D	M	I≈D	I<D	I>D	M	
Nu	14,5	-	5,7	10,1	28,1	49,3	-	38,8	24,5
Indef	60,7	80,3	55,7	64,4	-	-	-	-	64,4
Rā	83,6	-	-	83,6				-	83,6

TABLE 5.13 – Exp. D1-D4 : % de OD-OI-V

Tout d'abord, nos données permettent de confirmer le fait que les COD (non marqués) indéfinis se regroupent avec les COD marqués, contre les COD (non marqué) nus, pour préférer se placer avant le COI et que la préférence des COD pour se placer avant le COI s'accroît avec leur degré de détermination : COD nu < COD indéfini < COD marqué de = $r\bar{a}$ .

En ce qui concerne l'effet des différents facteurs fonctionnels sur l'ordre relatif

27. Dans la plupart des configurations où la longueur des deux objets est identique, les deux constituants ont exactement la même longueur en nombre de mots en tenant compte de la préposition. Mais, dans certains cas de figure, par exemple dans le cas des COD nus monolexémiques, l'on ne peut obtenir de longueurs identiques en tenant compte de la préposition. C'est pour cette raison que nous représentons le cas de figure dans laquelle la longueur relative entre les deux constituants est contrôlée par le signe de l'identité approximative.

entre les constituants, les résultats de nos expériences confirment la préférence “long-avant-court”. Ils nous permettent en outre de formuler l’hypothèse selon laquelle les propriétés sémantiques du COI ont un effet significatif sur l’ordre relatif par rapport au COD. Il semble que plus le COI est haut dans la hiérarchie de la personne, plus il tend à se placer avant le COD. Cette tendance se traduit dans la préférence “animé-avant-inanimé” bien connue dans la littérature (voir §2.3.1.3 dans le chapitre 2), établie pour l’ordre relatif entre les compléments verbaux dans des langues variées (cf. par ex. Bresnan *et al.*, 2007; Tanaka *et al.*, 2011a).

De plus, nos données appuient l’existence du biais lexical, dans la mesure où les tendances que nous avons relevées pour chaque lemme verbal sont constantes tout au long de nos expériences, bien qu’elles ne soient pas toujours (statistiquement) significatives. Par exemple, le verbe *ferestādan* ‘envoyer’ présente un taux de l’ordre OD-OI-V plus élevé que la moyenne, dans les quatre expériences où il apparaît, sachant que l’écart est significatif dans deux d’entre elles. De même, le verbe *xorāndan* ‘faire manger’ a un taux significativement plus élevé de l’ordre OD-OI-V dans les deux expériences où il apparaît. Les verbes *āvardan* ‘apporter’, *gereftan* ‘prendre’ et *dādan* ‘donner’, à l’inverse, semblent avoir un penchant pour l’ordre OI-OD-V.

Enfin, nos résultats ne nous permettent pas de nous prononcer sur l’effet du statut discursif du COI sur l’ordre relatif entre les deux objets.

## 5.4 Les phrases transitives

Les résultats de l’étude de corpus ne laissent aucun doute quant au statut particulier de l’ordre SOV en persan. Ainsi, en ce qui concerne l’ordre relatif entre le sujet et le COD, notre étude se concentre sur l’impact des facteurs tels que l’animéité et le poids grammatical. Par la même occasion, nous pouvons vérifier le niveau de la variation par rapport à l’ordre canonique dans les phrases transitives, en comparaison avec les phrases ditransitives.

À cette fin, nous avons conduit notre expérience sur les phrases transitives par le biais du même questionnaire que pour nos deux dernières expériences, D3 et D4, sur les phrases ditransitives qui englobent les quatre catégories du COD.

### 5.4.1 Expérience T1

Le but de cette expérience est double. Nous voulons étudier l’effet de la longueur ainsi que l’animéité sur l’ordre relatif entre le sujet et le COD, ainsi que leur interaction. Plus précisément, nous cherchons à savoir dans quelle mesure un COD plus accessible peut devancer le sujet.

Nous adaptons une conception 2 x 2 dans laquelle nous manipulons ces deux facteurs.

Étant donné les résultats de notre étude de corpus, nous nous sommes limitée au COD marqué de  $=r\bar{a}$  pour cette expérience. Pour augmenter la probabilité d'avoir de la variation dans les données, c'est-à-dire d'augmenter le taux de l'ordre OSV, le COD doit rester hautement accessible. Pour ce faire, nous nous limitons au COD humain.

Ainsi, nous manipulons le caractère animé seulement pour le sujet, avec deux valeurs : 1) [+humain] et 2) [-animé], et n'incluons que les COD [+humain]. De même, nous manipulons la longueur relative seulement en modifiant la longueur du COD, avec deux valeurs : 1) court et 2) long, en gardant le sujet court.

Cette conception nous permet d'évaluer dans quelle mesure un COD plus accessible peut devancer le sujet. Nous pensons que l'ordre OSV a plus de chance de se produire dans les configurations où le sujet est (conceptuellement) moins accessible que le COD. Cela peut être le cas lorsque le sujet est inanimé et, dans une moindre mesure, lorsque le COD est plus long, selon l'hypothèse de l'accessibilité conceptuelle des constituants longs de Yamashita & Chang (2001). Effectivement, nous faisons l'hypothèse que l'accessibilité (conceptuelle) relative entre le sujet et le COD est plus sensible au facteur de l'animéité qu'à celui de la longueur, car si l'on admet que cette dernière constitue une des dimensions de l'accessibilité conceptuelle, il est judicieux de considérer qu'elle y joue un rôle moins important que celui de l'animéité ou du rôle grammatical.

Les hypothèses expérimentales sous-jacentes sont les suivantes :

**Hypothèse 1** L'ordre SOV est l'ordre prédominant.

**Hypothèse 2** Plus le COD est (conceptuellement) accessible, plus il tend à devancer le sujet.

**Hypothèse 3** L'animéité joue un rôle plus important que la longueur sur l'accessibilité relative entre le sujet et le COD.

#### 5.4.1.1 Conception

Nous résumons ci-dessous la conception de l'expérience.

**Variables indépendantes :**

1. Animéité du sujet :

Cette variable prend deux valeurs dans notre expérience : animé et inanimé. Nous avons choisi les sujets inanimés de manière à ce qu'ils soient dénués de propriété agentive, dans le sens où l'on ne peut leur prêter d'intention. Nos

sujets animés sont des agents humains. Le COD, en revanche, est toujours un patient humain.

## 2. Longueur du COD :

Pour manipuler la longueur relative, nous avons modifié la longueur du COD en y ajoutant une subordonnée relative d'une longueur moyenne de 4 mots. Comme nous l'avons mentionné auparavant, en persan lorsque le COD est modifié par une relative, l'enclitique =*rā* peut être placé sur le COD ou sur la périphérie droite de la subordonnée relative. Cette dernière possibilité est toutefois proscrite par la grammaire. Nous avons inclus cette possibilité comme une condition supplémentaire, mais, comme nous allons le voir, nous l'écarterons par la suite dans les analyses de nos deux variables expérimentales.

### Conditions :

1. Sujet animé      Suj=OD            : **A**, ex. (15-a)
2. Sujet animé      Suj<OD    *rā*-Rel : **B**, ex. (15-b)
3. Sujet animé      Suj<OD    Rel-*rā* : **B'**, ex. (15-c)
4. Sujet inanimé    Suj=OD            : **C**, ex. (15-d)
5. Sujet inanimé    Suj<OD    *rā*-Rel : **D**, ex. (15-e)

### Variables dépendantes :

— Ordre des constituants dans la phrase complétée par le participant, opérationnalisé par une variable catégorique, qui peut théoriquement avoir six valeurs : SOV, OSV, SVO, OVS, VSO et VOS. Nous nous concentrons ensuite sur les deux ordres verbes finaux, sachant que les occurrences dans les autres ordres sont négligeables (voir Tableau 5.14).

### Variable de contrôle :

— Dans tous les items, le COD est un argument patient humain marqué de =*rā*.

- (15) a. ... [<sub>Suj</sub> Alirezā] [<sub>OD</sub> bačče=*rā*] [<sub>V1</sub> bidār karde ast]  
           Alireza        enfant=*RĀ*      réveillé faire.PP être.PRS.3SG  
           '... Alireza a réveillé l'enfant.'
- b. ... [<sub>Suj</sub> Alirezā] [<sub>OD</sub> bačče=*rā* ke yek-o-nim darje tab  
           Alireza        enfant=*RĀ* que une-et-demi degré fièvre  
           dār-ad]            [<sub>V</sub> bidār karde ast]  
           avoir.PRS-3SG    réveillé faire.PP être.PRS.3SG  
           '... Alireza a réveillé l'enfant qui fait une fièvre de 38,5°'

- c. ... [<sub>Suj</sub> Alirezā [<sub>OD</sub> bačče ke yek-o-nim darje tab  
Alireza enfant que une-et-demi degré fièvre  
dār-ad=rā] [<sub>V</sub> bidār karde ast]  
avoir.PRS-3SG=RĀ réveillé faire.PP être.PRS.3SG  
'... Alireza a réveillé l'enfant qui fait une fièvre de 38,5°']
- d. ... [<sub>Suj</sub> sarosedā [<sub>OD</sub> bačče=rā] [<sub>V</sub> bidār karde ast]  
bruit enfant=RĀ réveillé faire.PP être.PRS.3SG  
'... le bruit a réveillé l'enfant.']
- e. ... [<sub>Suj</sub> sarosedā [<sub>OD</sub> bačče ke yek-o-nim darje tab  
bruit enfant que une-et-demi degré fièvre  
dār-ad=rā] [<sub>V</sub> bidār karde ast]  
avoir.PRS-3SG=RĀ réveillé faire.PP être.PRS.3SG  
'... le bruit a réveillé l'enfant qui fait une fièvre de 38,5°']

#### 5.4.1.2 Stimuli

Nous avons préparé 20 phrases transitives, déclinées en cinq versions en fonction de nos cinq conditions expérimentales, similaires à celles illustrées en (15), comportant chacune un sujet, un COD et deux choix de verbes (simples ou de prédicats complexes) transitifs.

Nos verbes simples ou prédicats complexes sont des causatifs dénotant un changement d'état, par exemple *bidār kardan* 'réveiller', *majruh kardan* 'blesser', *xandāndan* 'faire rire', *tarsāndan* 'faire peur', *ārām kardan* 'calmer', *nejāt dādan* 'sauver'. Étant donné que le choix du verbe est laissé au participant, nous avons préparé nos phrases avec 40 verbes différents, afin qu'il n'y ait aucun verbe qui puisse apparaître à deux reprises.

Du point de vue sémantique, dans chaque phrase, les deux choix de verbe sont soit clairement différents, ex. 'fatiguer' vs. 'réveiller', voire opposés, ex. 'faire pleurer' vs. 'faire rire', soit relativement proches, ex. *komak kardan* 'aider' vs. *moāleje kardan* 'guérir', voire synonymes, ex. *majruh kardan* vs. *zaxmi kardan* 'blesser'.

Dans chaque phrase, le sujet correspond à l'argument agent et le COD au patient. Les sujets et les COD animés sont des noms propres, ex. *Alireza*, des liens de parenté, ex. 'Grand-mère', 'le frère de', des noms de profession, ex. 'le professeur', 'l'agent d'accueil', 'l'avocat', ou d'autres noms avec dénotation humaine, ex. 'l'enfant', 'les invités'. Dans un item donné, nous contrôlons pour la structure et, dans la mesure du possible, pour la longueur (en nombre de syllabes), d'un côté, entre sujets [+anim] vs. sujets [-anim], de l'autre côté, entre le sujet et le COD, ex. (16) à (18). Notons que nous avons choisi nos sujets inanimés de manière à ce que l'on ne puisse guère leur prêter une intention (pour causer l'événement), et cela surtout dans la configuration où ils prennent la structure 'N de N<sub>propre</sub>', ex. (17) et (18).

- (16) a. [Suj<sub>[+anim]</sub> moalleme=e varzeš] vs. [OD modir=e madrese]  
           professeur=EZ sport                    directeur=EZ école  
           ‘le professeur de sport’ vs. ‘le directeur d’école’
- b. [Suj<sub>[-anim]</sub> govāhi=ye pezešk] vs. [OD modir=e madrese]  
           certificat=EZ médecin                    directeur=EZ école  
           ‘le certificat médical’ vs. ‘le directeur d’école’
- (17) a. [Suj<sub>[+anim]</sub> xāhar=e Mahšid] vs. [OD barādar=e Māziyār]  
           soeur=EZ Mahshid                    frère=EZ Mazyiar  
           ‘la soeur de Mahshid’ vs. ‘le frère de Mazyiar’
- b. [Suj<sub>[-anim]</sub> ma’jun=e Mahšid] vs. [OD barādar=e Māziyār]  
           cocktail=EZ Mahshid                    frère=EZ Mazyiar  
           ‘le cocktail de Mahshid’ vs. ‘le frère de Mazyiar’
- (18) a. [Suj<sub>[+anim]</sub> xāhar=e Parastu] vs. [OD šohar=e Maryam]  
           sœur=EZ Parastou                    mari=EZ Mahshid  
           ‘la sœur de Parastou’ vs. ‘le mari de Maryam’
- b. [Suj<sub>[-anim]</sub> mehmāni=ye Parastu] vs. [OD šohar=e Maryam]  
           fête=EZ Parastou                    mari=EZ Maryam  
           ‘la fête de Parastou’ vs. ‘le mari de Mahshid’

Comme dans nos deux expériences précédentes, nos items expérimentaux n’ont pas de phrase de contexte en guise de préambule. Nos phrases transitives à compléter sont présentées comme des compléments phrastiques, introduits par des verbes ou des adjectifs appelant un complément phrastique, ex. (19) et ex. (20), ainsi que comme subordonnées adverbiales, ex. (21).

- (19) Mādarbozorg bā negarāni goft                    šenide                    ast                    ...  
       Grand-mère avec angoisse dire.PST.3SG entendre.PP être.PRS.3SG  
       ‘Grand-mère a dit avec angoisse qu’elle a entendu que ...’
- (20) Maryam mi-guy-ad                    qāzi omidvār ast                    ...  
       Maryam IPFV-dire.PST-3SG juge espérant être.PRS.3SG  
       ‘Maryam dit que le juge espère que ...’
- (21) Minā mi-guy-ad                    Azāde asabāni ast                    čon ...  
       Mina IPFV-dire.PST-3SG Azadeh en-colère être.PRS.3SG car  
       ‘Mina dit que Azadeh est en colère parce que ...’

Les 20 items de cette expérience sont combinés avec 35 distracteurs, dont 25 items expérimentaux de nos deux expériences précédentes (D3, cf. §5.3.3 et D4, cf. §5.3.4) et 10 distracteurs supplémentaires détaillés en §5.3.3.2.

### 5.4.1.3 Procédure

Chaque item de cette expérience est présenté par une phrase incomplète où manquent le sujet, le COD et le verbe représentés par trois cases vides, avec quatre options pour remplir les cases vides, incluant les deux choix du verbe et le sujet et le COD. Dans la première moitié des stimuli, cette liste suit, de haut en bas, l'ordre verbe1, COD, Sujet, verbe2 et, dans la seconde moitié, l'ordre verbe1, Sujet, COD, verbe2. Tous ces éléments sont présentés simultanément sur l'écran, cf. Figure 5.14.



FIGURE 5.14 – Exp. T1 : Copie d'écran pour un item donné

### 5.4.1.4 Participants

Cette expérience a été conduite via le même questionnaire que nos deux expériences précédentes portant sur les phrases ditransitives. Les participants à ces trois expériences en sont les mêmes (voir §5.3.3.4 pour les détails). Douze réponses erronées ont été écartées de la table de données finale de cette expérience, ce qui laisse un ensemble de 1188 observations.

### 5.4.1.5 Résultats

Les données brutes de cette expérience sont présentées dans le Tableau 5.14. Comme attendu, l'ordre SOV se produit avec une fréquence très élevée de 92,41%.

Tout d'abord, nous observons qu'il n'y pas de différence significative liée à la position de  $=r\bar{a}$  dans la relative, et cela ni au niveau de l'ordre (95,42% pour la cond. **B** ( $r\bar{a}$ -Rel) vs. 91,56% pour la cond. **B'** (Rel- $r\bar{a}$ ))<sup>28</sup>, ni au niveau du temps d'accomplissement de la tâche ( $t=1,0575$ ,  $df=242,199$ ,  $p>0,2$ ). Nous avons donc écarté cette

28. Une différence entre ces deux conditions, ex. (i-b) vs. ex. (i-c), serait attendue, selon le principe *EIC*, à condition de commencer le *CRD* à partir du  $=r\bar{a}$ , au lieu de *bačče*, (cf. §3.2.3.1) :

- (i) a. [[Alirezā] [bačče(= $r\bar{a}$ ) ke yek-o-nim darje tab dār-ad(= $r\bar{a}$ )] bidār-karde-ast]  
 1/1      2/2      2/3 2/4      2/5 2/6 2/7      3/8 → score=57%



que facteurs à effet aléatoire<sup>29</sup>, cf. Tableau 5.15, confirme la forte préférence pour l'ordre SOV. Les résultats montrent un effet significatif de l'animéité, qui avec un coefficient positif dans la condition Suj-animé correspond à la préférence “animé-avant-inanimé”. Ils ne montrent, en revanche, aucun effet pour la longueur. Il en va de même pour l'interaction entre les deux variables ( $ps > 0,5$ ).

	Estimation	Erreur-type	Valeur de z	Valeur de p
Intercepte	3,745	0,325	11,536	<0,0001
ANIM :Suj-ANIM	0,400	0,172	2,329	<0,0001
LONG :OD>Suj	-0,101	0,155	-0,654	0,513
OD>Suj :Suj-ANIM	-0,012	0,155	-0,078	0,938

N.B. : Le succès correspond à ORDRE=SOV

TABLE 5.15 – Exp. T1 : Résumé des résultats du modèle optimal pour les effets fixes

## Discussion

Ces résultats confirment le fait que l'ordre prédominant dans les phrases transitives en persan est SOV. Ils appuient également nos hypothèses concernant le lien entre l'ordre relatif et l'accessibilité conceptuelle, étant donné que l'effet de l'animéité l'emporte largement sur celui de la longueur.

Néanmoins, le fait que la longueur ne manifeste aucun effet est surprenant, d'autant plus que dans cette conception le COD comporte une enchâssée. Comme le montre ex. (22), déplacer le COD à gauche (ou encore à droite après le verbe puisque cela n'est pas impossible en persan) devrait faciliter le traitement et la compréhension de la phrase, selon le modèle *DLT* – et même dans sa définition étroite (cf. §3.2.3.2) –, ainsi que selon l'hypothèse de Wasow (2013), car cet ordre permet d'éviter une structure complexe au milieu, dont le traitement serait coûteux.

- (22) a. Alirezā [bačče=rā [<sub>CP</sub> ke yek-o-nim darje tab dār-ad]] bidār-karde-ast  
 b. [bačče=rā [<sub>CP</sub> ke yek-o-nim darje tab dār-ad]] Alirezā bidār-karde-ast  
 c. Alirezā bidār-karde-ast [bačče=rā [<sub>CP</sub> ke yek-o-nim darje tab dār-ad]]

Mentionnons enfin que le principe *EIC* prédit également un certain effet de la longueur, mais dans le sens inverse, comme le montre ex. (23), préconisant le placement

29. Le modèle est codé selon le contraste de type “somme” pour les deux variables. Ainsi, l'intercepte représente la moyenne générale et les coefficients les écarts de chaque groupe par rapport à cette valeur.

du COD long après le sujet. Toutefois, la fréquence déjà très haute de l'ordre SOV ne laisse pas vraiment de marge pour tester cette hypothèse.

- (23) a. [[Alirezā] [bačče=rā ke yek-o-nim darje tab dār-ad] bidār-karde-ast]  
 1/1      2/2      2/3 2/4      2/5 2/6 2/7      3/8 → score=57%
- b. [[bačče=rā ke yek-o-nim darje tab dār-ad] [Alirezā] bidār-karde-ast]  
 1/1      1/2 1/3      1/4 1/5 1/6      2/7      3/8 → score=39%

## 5.5 Conclusions

Les résultats de nos différentes études expérimentales sur l'ordre relatif entre le COD et le COI dans les phrases ditransitives confortent les hypothèses que nous avons formulées en nous basant sur les résultats de notre étude de corpus.

Ils confirment l'ordre canonique SOV pour les phrases transitives en persan, mais contredisent l'hypothèse communément admise sur l'ordre relatif canonique entre les deux objets dans les phrases ditransitives. Rappelons que cette hypothèse postule un ordre canonique dichotomique pour les COD en fonction de leur marquage par =rā. Nous reviendrons sur cette hypothèse et ses implications théoriques dans le premier chapitre de la dernière partie de notre thèse (ch. 6). L'hypothèse d'un tel ordre canonique dichotomique dans les phrases ditransitives est la pierre angulaire des analyses les plus répandues sur la structure du syntagme verbal en persan.

Premièrement, nos résultats montrent clairement que les COD non marqués indéfinis se joignent aux COD marqués pour préférer l'ordre OD-OI-V, tandis que les COD non marqués nus préfèrent l'ordre inverse. Cela signifie que le marquage par =rā ne constitue pas un critère valide pour déterminer l'ordre relatif entre le COD et le COI. *A contrario*, nos résultats montrent que cet ordre relatif dépend du degré de détermination du COD.

D'un côté, les COD nus manifestent une forte préférence à apparaître adjacents au verbe, ce qui est attendu étant donné le manque d'autonomie (sémantico-référentielle) de ces COD, leur lien sémantique étroit avec le verbe et leur légèreté. Les COD nus modifiés, en revanche, perdent cette légèreté et gagnent une certaine saillance par rapport à leurs contreparties non modifiées, et montrent, par conséquent, une moins forte préférence pour la position adjacente au verbe.

De l'autre côté, les COD marqués sont des entités hautement saillantes et disposant d'une autonomie (sémantico-référentielle) pleine par rapport au verbe, ce qui explique leur capacité à se séparer de celui-ci et leur forte tendance à apparaître plus tôt dans la phrase. Les COD indéfinis sont également des entités autonomes, du point de vue sémantico-référentiel, et, de ce fait, peuvent se séparer du verbe. Ils ne

disposent pas toutefois du même niveau de saillance et d'indépendance (sémantique) par rapport au verbe, et, par conséquent, présentent une moins forte préférence pour cette position. Notons que le lien entre l'ordre et le degré de détermination des COD n'est pas linéaire, dans la mesure où l'écart entre le COD nu (modifié et non modifié confondus) et le COD indéfini est plus important que l'écart entre ce dernier et le COD marqué.

Par ailleurs, des facteurs fonctionnels, tels que la longueur et l'animéité (la hiérarchie de la personne pour être précise), influent sur l'ordre linéaire entre les deux COD. Nos données nous ont permis d'établir les préférences "long-avant-court" et "animé-avant-inanimé", que l'on peut traiter comme deux contraintes faibles déterminant l'ordre relatif entre le COD et le COI. Le continuum présenté ci-dessus sur la base du degré de détermination du COD, qui est étroitement lié à une autre dimension de l'accessibilité, à savoir la saillance discursive, peut également être considéré comme une contrainte préférentielle : plus un COD est déterminé, plus il tend de se séparer du verbe et à se placer tôt dans la phrase ; ou encore, plus un COD est accessible dans le discours, plus il tend de se séparer du verbe et à se placer tôt. Cette préférence constitue alors la contrainte faible la plus prépondérante pour la détermination de l'ordre linéaire dans les phrases ditransitives, mais ne suffit pas à elle seule pour déterminer l'ordre linéaire.

Toutefois, il nous reste à déterminer, dans des recherches ultérieures, entre la longueur et l'animéité laquelle joue un rôle plus influent. Il y aura lieu également d'approfondir l'interaction entre les trois préférences afin de pouvoir proposer un modèle de prédiction global en fonction de la pondération de ces trois contraintes. Notons que ces trois contraintes convergent dans la mesure où elles correspondent à une préférence plus générale de placer les constituants les plus "accessibles" (conceptuellement ou discursivement) plus tôt dans la phrase.

Nous reviendrons sur ces préférences et la manière dont l'on peut en rendre compte dans le second chapitre de la troisième partie (ch. 7). Par la même occasion, nous discuterons de l'importance de l'ensemble de nos observations pour le débat sur les préférences de l'ordre linéaire du point de vue de la typologie. Comme nous le verrons, celles-ci, notamment en ce qui concerne l'effet de la longueur – plus particulièrement son absence dans certains cas de figure –, présentent un intérêt particulier en permettant de comparer différentes analyses possibles pour la préférence "long-avant-court" dans les langues (S)OV.

Deuxièmement, nous constatons que le niveau de la variation par rapport à l'ordre prédominant, plus précisément le taux d'ordre non canonique, diffère considérablement en fonction des constituants concernés. Le taux des ordres non SOV

est faible dans les phrases transitives et, même dans les cas de figure disposés à manifester de la variation, reste sous le seuil de 10%. Quant à l'ordre relatif entre le COD et le COI, en revanche, le taux des ordres qui ne suivent pas l'ordre prédit, dépasse les 15%, quel que soit le critère choisi pour le déterminer, et peut même s'approcher des 50% dans les cas de figure les plus enclins à la variation, notamment dans le cas des COD nus modifiés. Pour résumer, nous constatons un ordre presque figé avec le sujet, mais un ordre plus flexible pour les objets.

En outre, l'ordre relatif entre le COD et le COI se montre beaucoup plus sensible aux facteurs fonctionnels que l'ordre relatif entre le sujet et le COD, pour lequel l'ordre SOV semble s'imposer comme une contrainte forte peu influencée par des facteurs fonctionnels. Ainsi, nos résultats suggèrent qu'il n'est pas empiriquement justifié de poser un ordre relatif canonique entre le COD et le COI.

La dernière partie de notre thèse est consacrée aux retombées théoriques des résultats de nos études empiriques. Dans le chapitre 6, nous discuterons de l'impact de ces découvertes pour l'étude de la structure syntaxique du persan. Dans le chapitre 7, nous présenterons une discussion de l'ensemble de nos conclusions en ce qui concerne l'effet des différents facteurs sur l'ordre linéaire, en mettant en perspective les préférences attendues et les préférences observées, afin de contribuer aux débats théoriques en cours sur la façon dont il faut rendre compte de ces préférences.



## Troisième partie

# Discussion générale et apports théoriques



# Chapitre 6

## Renouveler l'analyse de la structure syntagmatique du persan

Dans la première partie de cette thèse, nous avons exposé les différents phénomènes de la variation de l'ordre en persan, qui concernent aussi bien le domaine préverbal que le domaine postverbal (§1.3.3). Nous en avons évoqué les différents traitements syntaxiques disponibles dans la littérature, majoritairement dans le cadre de la théorie générative chomskyenne, qui adoptent une vision configurationnelle du persan.

Par la même occasion, nous avons vu que la littérature en syntaxe formelle sur l'ordre linéaire des constituants en persan se fonde sur la position du complément d'objet direct. En effet, bien que les auteurs s'accordent pour postuler SOV comme l'ordre sous-jacent en persan, ils ne se rejoignent pas toujours en ce qui concerne la position syntaxique du COD marqué (de = $r\bar{a}$ ) vis-à-vis de sa contrepartie non marquée. En l'absence d'études empiriques montrant le contraire, l'ordre relatif canonique entre le COD et le COI a été considéré, de manière consensuelle, comme dépendant du marquage différentiel de l'objet (MDO) : le COD marqué (de = $r\bar{a}$ )<sup>1</sup> précède le COI, tandis que le COD non marqué suit le COI (Karimi, 1994; Browning & Karimi, 1994; Mahootian, 1997; Rasekhmahand, 2004; Ganjavi, 2007; Windfuhr & Perry, 2009; Roberts et al., 2009, entre autres) :

- (1) a. (S) OD= $r\bar{a}$  OI V  
b. (S) OI OD V

Le contraste présenté en (1) a conduit certains linguistes à suggérer que les

---

1. Rappelons que cette généralisation ne se formule pas toujours en termes de marquage différentiel de l'objet, mais en faisant appel aux propriétés que l'on suppose à l'origine de la présence de = $r\bar{a}$ , notamment la définitude (Mahootian, 1997; Roberts et al., 2009; Windfuhr & Perry, 2009) ou la spécificité (Karimi, 1994; Browning & Karimi, 1994).

deux types du COD, marqué et non marqué, ne partagent pas les mêmes propriétés syntaxiques et structurelles (Samvelian, 2001, est une exception notable).

Dans la première partie (cf. section 1.4), nous avons brièvement évoqué cette hypothèse désignée comme celle de la “position syntaxique double de l’objet” (PSDO). Dans ce chapitre, nous en effectuerons une présentation détaillée des analyses les plus répandues en distinguant les approches admettant une position de base identique (§6.1.2) de celles admettant deux positions de base distinctes (§6.1.1). Bien que l’ordre relatif entre le COD et le COI soit sans aucun doute l’argument fondamental, et le socle commun sur lequel reposent toutes ces contributions, il n’est pour autant pas le seul à être avancé pour soutenir cette hypothèse. Dans la section 6.2, nous effectuerons une revue critique des autres arguments évoqués dans ce sens. Enfin, à l’issue de cet examen, nous reviendrons sur la structure du syntagme verbal en persan et en proposerons une analyse davantage en phase avec les faits empiriques (section 6.3).

## 6.1 Hypothèse de la position syntaxique double de l’objet (PSDO)

Dans cette section, nous présentons l’hypothèse de la position syntaxique double du COD (PSDO) à travers notamment des travaux de Browning et E. Karimi (1994)<sup>2</sup>, Ghomeshi (1997) et Karimi (1990, 2003a, 2005)<sup>3</sup>.

Bien que cette hypothèse ne soit formulée comme telle que par Karimi (2003a, p. 105), qu’elle a nommée “*Two Object Positions Hypothesis (TOPH)*”, tous ces auteurs s’accordent pour postuler que le COD marqué de  $=r\bar{a}$  occupe une position structurelle plus élevée et se caractérise par une indépendance plus importante vis-à-vis du verbe que sa contrepartie non marquée. Les points de divergences portent sur le traitement de l’enclitique  $=r\bar{a}$ , ainsi que sur la structure syntagmatique sous-jacente et les mouvements correspondants postulés.

Il est à noter que ces travaux ne poursuivent pas les mêmes objectifs. Karimi (1990) étudie le fonctionnement de  $=r\bar{a}$  pour en écarter l’analyse en tant que marqueur d’accusatif, de COD défini et/ou spécifique ou de topicalité. Elle propose, en revanche, d’analyser  $=r\bar{a}$  comme marqueur du Cas spécifique oblique, ce qui permet, selon l’auteur, de rendre compte de son emploi canonique aussi bien que de

2. Précisons qu’il y a deux linguistes ayant le patronyme Karimi que nous citons ici, Ezat Karimi et Simin Karimi. Par souci de clarté, nous nous référons au premier par E. Karimi.

3. Rappelons que Karimi (1989, 1990, 1994, 2003a, 2005, entre autres) a consacré de nombreux travaux au marquage et au positionnement de l’objet direct en persan. Nous nous limiterons à trois d’entre eux.

son emploi discursif. Ce Cas spécifique oblique, réservé au  $SN_{[+spécifique]}$ , est assigné hors du domaine du gouvernement de la tête de ce dernier, en l'occurrence, selon les différents cas de figure, dans la position de l'adjoint du  $V''$  ou du  $I''$  ou dans la position du spécifieur du CP.

Karimi (2003a) se focalise sur les divergences syntaxiques et sémantiques des COD spécifique vs. non spécifique pour en proposer deux positions sous-jacentes distinctes. Elle s'intéresse également aux variations de l'ordre concernant le COD et en suggère une analyse en termes de *scrambling* (voir §2.2.1.3 dans le chapitre 2). C'est une analyse qu'elle approfondit par la suite (Karimi, 2005), en la développant notamment sous un angle typologique et en tenant compte des données d'autres langues, en vue de proposer une nouvelle analyse de *scrambling* dans le cadre de la théorie minimaliste (cf. Chomsky, 1995, entre autres).

Browning & Karimi (1994) s'intéressent également à la question du *scrambling* en persan. En effet, ces auteurs étudient  $=r\bar{a}$  dans la perspective de comparer trois différents mouvements dans cette langue : le déplacement du  $COD_{[+spécifique]}$  pour vérifier le filtre du Cas (dénommé *object shift*), le *scrambling* vers la position initiale (*clause initial scrambling*), et l'extraction ou le *scrambling* à distance (*long distance scrambling*). Cette comparaison s'effectue sur la base de la grille des critères communément utilisée pour départager entre, d'un côté, "mouvements A", c'est-à-dire le déplacement vers une position argumentale (*A-position*) motivé par le filtre du Cas, et, de l'autre, "mouvements A-barre", c'est-à-dire le déplacement vers une position non argumentale (*A'-position*) non motivé par le filtre du Cas. Elle permet, selon les auteurs, de mettre en lumière des comportements mixtes, observés également dans d'autres langues, et ainsi de contribuer à la mise en question de la dichotomie A-position vs. A'-position, évoquée antérieurement (Webelhuth 1989 pour l'allemand, Mahajan 1990 pour le hindi, et Déprez 1989 qui, sur la base d'une étude à travers des langues diverses, en propose une redéfinition).

Ghomeshi (1997), quant à elle, revient sur le fonctionnement de  $=r\bar{a}$  aussi bien en tant que marqueur du COD qu'en tant que marqueur de topicalité. À propos de ses propriétés sémantiques, elle rejoint Lazard (1982) et Dabir-Moghaddam (1992) pour mettre en question le traitement de  $=r\bar{a}$  par un seul trait binaire, notamment le trait  $[\pm\text{spécifique}]$ . Du point de vue syntaxique, elle propose d'analyser  $=r\bar{a}$  comme un marqueur de Cas assigné dans la position de l'adjoint du  $V''$  et autorisé selon le critère thématique, dont la définition établie par cet auteur permettrait de rendre compte des différents emplois de  $=r\bar{a}$ .

Ces travaux diffèrent également quant aux arguments avancés pour soutenir l'hypothèse de la PSDO (cf. section 6.2). Browning & Karimi (1994) évoquent seulement

l'ordre canonique linéaire. Karimi (1990, 1994) mentionne, en outre, la liberté de positionnement des COD marqués, notamment la possibilité pour eux d'apparaître dans une position postverbale, contrairement aux COD non marqués, qui seraient confinés à la position adjacente au verbe. Ghomeshi (1997) abonde également dans ce sens en se référant à ces travaux. Elle considère notamment cette dernière observation comme révélatrice de l'indépendance des COD marqués vis-à-vis du verbe, dont ne bénéficient pas leurs contreparties non marquées. Elle souligne, en outre, que l'hypothèse d'une différence structurelle entre les COD définis/spécifiques/présupposés, d'un côté, et les COD indéfinis, de l'autre, est déjà suggérée dans la littérature sur la base des données d'autres langues, telles que le hindi, le turc, le néerlandais, etc. (Mahajan, 1990; Diesing, 1992; Koopman & Sportiche, 1991, entre autres), et, de ce fait, le persan ne serait pas un cas isolé. Par la suite, Karimi (2003a, 2005) apportera des arguments supplémentaires pour étayer cette hypothèse, notamment en ce qui concerne le lien sémantique avec le verbe (§6.2.1), la portée des quantifieurs (§6.2.2), le liage (§6.2.3), les vides parasites (§6.2.4) et la coordination (§6.2.5).

Nous présentons dans ce qui suit le traitement proposé dans chacun de ces travaux, tout d'abord ceux qui postulent deux positions sous-jacentes distinctes à la base (§6.1.1) et ensuite ceux qui postulent une même position de base (§6.1.2). Nous examinerons les arguments avancés pour soutenir l'hypothèse de la PSDO dans la section 6.2.

### 6.1.1 Objets générés dans deux positions distinctes

Dans son article sur le fonctionnement de  $=r\bar{a}$ , Karimi (1990, p. 163) analyse ce dernier, dans le cadre de la théorie du Cas (Chomsky, 1981, 1986), comme le marqueur du cas spécifique-oblique qui se définit selon la configuration schématisée en (2).

- (2) [ ... CP ... NP ... [ $\beta$  ... ] ... ] où  $\alpha$  est la tête du  $\beta$ <sup>4</sup>  
           [+Spécifique]  
           [-Nominal]

Karimi (1990, p. 163)

L'auteur postule qu'un SN spécifique-oblique est marqué de  $=r\bar{a}$  lorsqu'il se trouve en dehors du domaine minimal de gouvernement de sa tête lexicale, en l'occurrence la position adjointe au  $V''$  ou au  $I''$ , comme illustré en (4). Karimi souligne

---

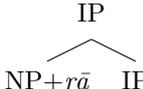
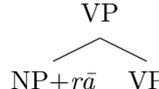
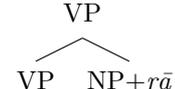
4. Rappelons que, dans la définition donnée au préalable par Karimi (1990, p. 160), le nœud  $\beta$  est gouverné par le nœud  $\alpha$ .

que cette analyse peut s'étendre aux compléments nominaux circonstanciels (*nominalized adverbs*) marqués de =*rā*, ex. (3) :

- (3) šab=e piš=o aslan na-xābid-am  
 nuit=EZ dernier=RĀ absolument NEG-dormir.PST-1SG  
 'Quant à hier soir, je n'ai pas du tout dormi.'

Karimi (1990, pp. 167-168)

Postulant qu'en persan la position sous-jacente du COD marqué est adjacente au verbe, Karimi (1990) en conclut que celui-ci n'apparaît pas dans sa position de base. Cette hypothèse serait confortée par le fait que le COD marqué de =*rā* dispose d'une liberté de positionnement dont est privée sa contrepartie non marquée. Les phrases en (5), correspondant respectivement aux trois structures schématisées en (4), attestent de différents ordonnancements possibles pour le COD marqué.

- (4) a.  b.  c. 

Karimi (1990, p. 163)

- (5) a. [IP Sepida=ro [IP man tu kelās did-am]]  
 Sepide=RĀ moi dans classe voir.PST-1SG  
 'Quant à Sepideh, je l'ai vue en classe.'  
 b. man [VP Sepida=ro [VP tu kelās did-am]]  
 c. man [VP [VP tu kelās did-am ] Sepida=ro]

Karimi (1990, p. 164)

Karimi (1990) rend compte de manière similaire de la fonction discursive de =*rā*, notamment dans le cas de l'extraction depuis la proposition matrice (*long distance extraction*), ex. (6), ou depuis un syntagme nominal, ex. (7-a), ou prépositionnel, ex. (7-b), appelé "*clitic binder*" par l'auteur du fait de la présence d'un pronom clitique résomptif. Pour ce faire, elle postule que le cas spécifique-oblique est également assigné dans une position non argumentale (A-barre), en l'occurrence celle de specifieur du CP, comme schématisé en (8-a) et (8-b), pour l'extraction à distance et l'extraction locale (*clitic binder*) respectivement <sup>5</sup>.

- (6) Ali=ro Maryam mi-xāst (ke) be-bin-e  
 Ali=RĀ Maryam IPFV-vouloir.PST.3SG (que) SBJV-voir.PRS-3SG

5. Karimi (1990, p. 185) souligne que son analyse va dans le sens de l'analyse proposée par Lazard (1982) en termes de polarisation (cf. §1.1.3).

‘C’est Ali que Maryam voulait voir.’

cf. ex. (100), section 1.3 (p. 41)

- (7) a. māšin=o dar=eš=o bast-am  
voiture=RĀ porte=3SG=RĀ fermer.PST-1SG  
‘Quant à la voiture, j’ai fermé sa porte.’  
b. Sāsān=o bāh=aš raqsid-am  
Sasan=RĀ avec=3SG danser.PST-1SG  
‘Sasan, j’ai dansé avec lui.’
- (8) a. [CP NP+rā<sub>i</sub> ... [IP ... t<sub>i</sub> ... V]]  
b. [CP NP+rā<sub>i</sub> ... [IP ... V [CP t<sub>i</sub> [IP ... t<sub>i</sub> ... V]]]]

Karimi (1990, p. 163)

Dans son article intitulé “*On Object Positions, Specificity, and Scrambling in Persian*” (2003a), Karimi se focalise sur les asymétries entre les COD, spécifique et non spécifique, pour postuler deux positions sous-jacentes distinctes pour le COD, schématisées en (9). Comme ce schéma le montre, Karimi (2003a) présume que le COD spécifique est généré à la base dans une position plus élevée que le COD non spécifique.

(9) *Two Object Positions Hypothesis* (TOPH)

- a. [VP DP<sub>[+Specific]</sub> [V' PP V]]  
b. [VP [V' PP [V' DP<sub>[-Specific]</sub> V]]]

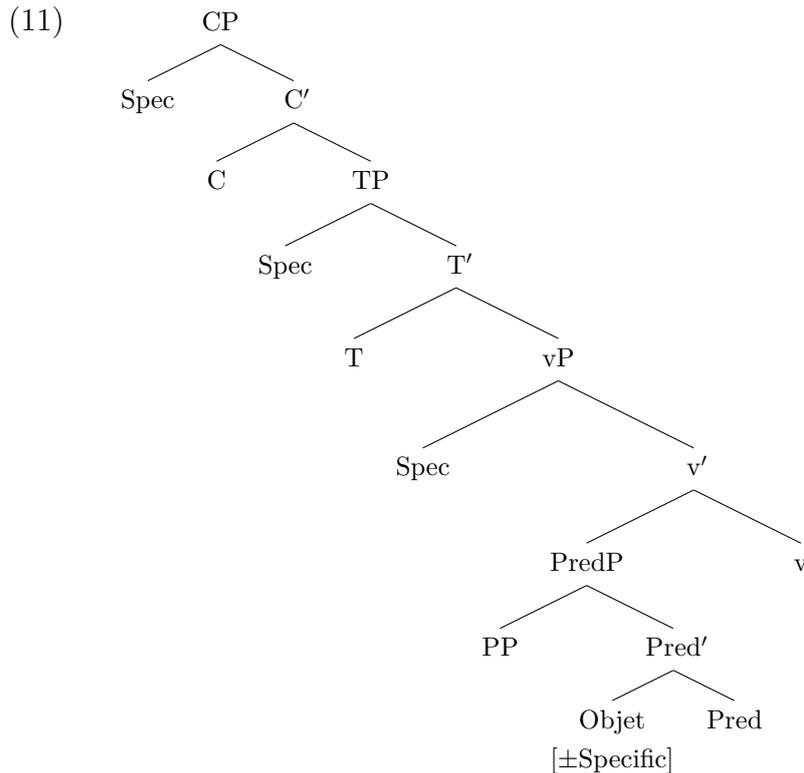
Karimi (2003a, p. 105)

En outre, l’auteur suggère l’existence d’une troisième position due au *scrambling* déclenché, selon certaines conditions sémantiques, par des traits fonctionnels discursifs tels que *Top(ic)* ou *Foc(us)*. Ces derniers, comme cela est illustré par le schéma (10), sont représentés par des projections/têtes fonctionnelles (FP en l’occurrence), dont un XP se déplacerait vers la position du spécifieur lorsqu’il satisfait les conditions sémantiques correspondantes. Karimi (2003a) souligne qu’un COD spécifique (marqué) peut rester *in situ* en persan et que le déplacement du COD n’est pas motivé par le filtre du Cas, mais est le résultat d’une règle de *scrambling*.

(10) [FP XP<sub>i</sub> [ F [YP ... t<sub>i</sub> ... ]]]

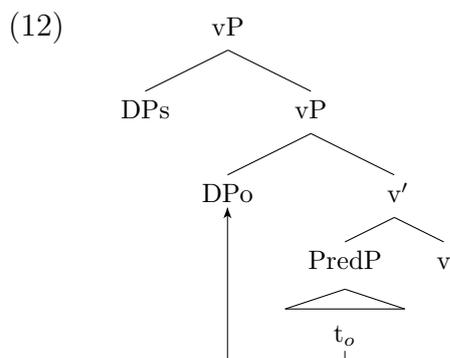
Karimi (2003a, p. 111)

Notons que Karimi reviendra sur cette analyse par la suite (Karimi, 2005, pp. 108-109) pour en proposer une version modifiée (*TOPH revisited*), schématisée en (11), dans laquelle les deux objets occupent la même position sous-jacente, la sœur du V.



Karimi (2005, p. 108)

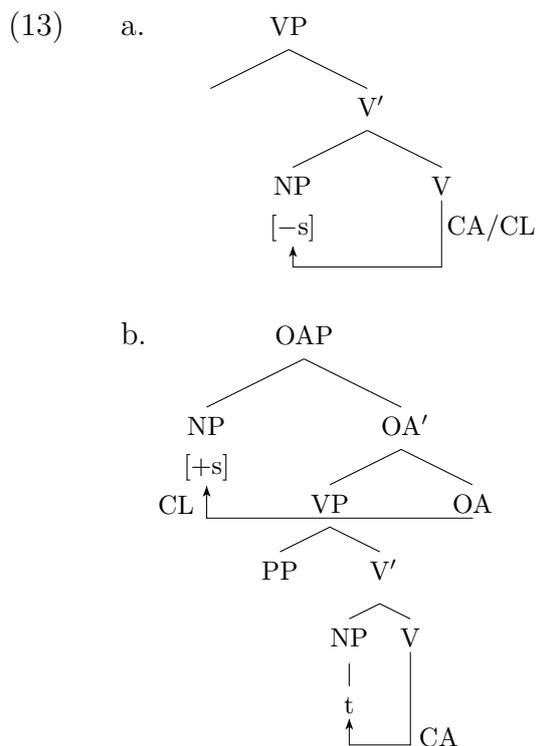
L'auteur fait donc appel au déplacement obligatoire “*object shift*”, voir schéma (12), selon lequel le COD<sub>[+Spécifique]</sub> se déplace vers la position de spécifieur du v'' (le plus bas) afin de recevoir son interprétation, autrement dit afin d'échapper au domaine de la clôture existentielle (*domain of existential closure*) au sein du Pred'' (cf. Diesing, 1992; Diesing & Jelinek, 1995; Kratzer, 1995). Karimi (2005, p. 108) affirme que seuls les COD spécifiques reçoivent le Cas accusatif, dans la mesure où un SN non spécifique, se trouvant dans le domaine de la clôture existentielle, fait partie du prédicat et n'est par conséquent pas concerné par le filtre du Cas. Rappelons que Karimi propose la même analyse pour le sujet (voir §1.4.1) en postulant une distinction similaire entre les sujets spécifiques et non spécifiques (2005, pp. 94–104).



Karimi (2005, p. 109)

### 6.1.2 Objets générés dans la même position

Browning & Karimi (1994) postulent que l'ordre sous-jacent de la phrase en persan est S-COI-COD-V et que *=rā* marque les COD spécifiques. Ils rendent compte de l'ordre S-COD<sub>[+Spécifique]</sub>-COI-V par la dérivation “*object shift*” motivée par le filtre du Cas, schématisée en (13). Pour ce faire, ils adoptent une variante de la théorie du Cas qui admet deux composantes : “*Case assignment*” (CA) et “*Case licensing*” (CL)<sup>6</sup>. Ce qui signifie que la vérification du Cas requiert distinctement l'assignation et l'autorisation de celui-ci.



Adaptés de Browning & Karimi (1994, p. 69)

Les auteurs suggèrent qu'en persan, le Cas assigné par un verbe transitif à son complément nominal, autrement dit le cas accusatif, peut être autorisé soit par le verbe lui-même, soit par une tête flexionnelle. Empruntant une approche suggérée par Belletti (1988)<sup>7</sup>, ils présument que le Cas autorisé directement par le verbe n'est

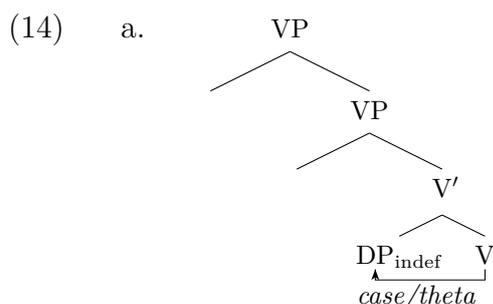
6. Les auteurs ajoutent que cela reviendrait à poser l'assignation de deux différents types de Cas : le Cas lexical (ou inhérent), assigné par une tête lexicale, en l'occurrence le V'', et le Cas structurel, assigné par une tête fonctionnelle, en l'occurrence la tête flexionnelle ObjectAGR (OA''). La première solution a été suggérée à titre d'exemple par Freidin & Sprouse (1991) pour l'islandais et la seconde par Mahajan (1990) pour le hindi.

7. Il s'agit vraisemblablement de l'observation de Belletti (1988) à partir des données du finnois, une langue dotée d'une riche morphologie casuelle, selon laquelle le Cas accusatif entraîne une lecture définie du COD et le COD indéfini requiert le Cas partitif. L'auteur soutient par ailleurs que

pas compatible avec une interprétation spécifique. Autrement dit, lorsqu'un SN se trouve dans la position du complément verbal, c'est-à-dire la sœur du V, le Cas qui lui est assigné n'est compatible qu'avec une interprétation non spécifique. Ainsi, selon Browning & Karimi (1994), un  $COD_{[-Spécifique]}$  peut rester dans sa position initiale, tandis qu'un  $COD_{[+Spécifique]}$  doit être déplacé afin d'être gouverné par une tête flexionnelle, ObjectAgr" en l'occurrence, et de recevoir son Cas. Les auteurs supposent que  $=r\bar{a}$  est la manifestation de ce Cas.

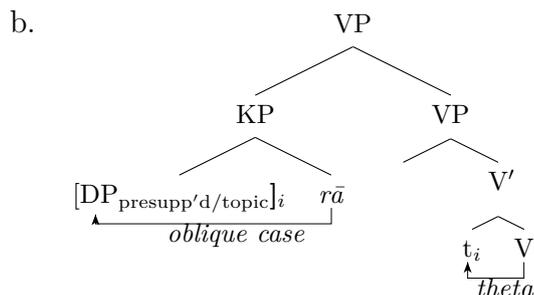
L'analyse proposée par Ghomeshi (1997) rejoint l'analyse susmentionnée en ce qu'elle postule également que le COD indéfini<sup>8</sup> (non marqué) peut demeurer dans sa position de base, qui est la position canonique du complément, et recevoir son cas et son rôle thématique directement du verbe, comme illustré dans le schéma (14-a). Le COD défini/présumé (marqué de  $=r\bar{a}$ ), en revanche, doit se déplacer, selon le schéma (14-b), car le cas assigné directement par le verbe n'est pas compatible avec ses propriétés sémantiques.

Toutefois, Ghomeshi rejoint Karimi (1990, 2003a) pour écarter l'analyse de  $=r\bar{a}$  comme la réalisation du Cas (accusatif) assigné par une tête fonctionnelle (Ghomeshi, 1997, pp. 163-164). Elle avance que  $=r\bar{a}$  marque les SN qui se trouvent en position de l'adjoint du  $V''$ . Plus précisément, elle analyse ce dernier comme un affixe syntagmatique tête du  $K''$  (*Kase Phrase*) placé dans cette position.



le Cas partitif est un Cas inhérent (dans le sens de Chomsky, 1986) assigné par le verbe. Rappelons que, dans l'approche adoptée par Browning & Karimi (1994), la division entre Cas inhérent et Cas structurel se traduit par la distinction entre "l'assignation du Cas" et "l'autorisation du Cas".

8. Rappelons que Ghomeshi (1997) réfute l'analyse sémantique de  $=r\bar{a}$  par le seul trait binaire  $[\pm spécifique]$  et propose un ensemble de traits, notamment définitude, présupposition et topicalité. Par convention, elle utilise, d'un côté, COD indéfini et, de l'autre, COD défini/présumé/topic.



Ghameshi (1997, p. 149)

Pour apparaître sous  $K''$ , un syntagme nominal doit vérifier le critère thématique. L'auteur définit celui-ci de manière à ce qu'il permet également de rendre compte de la fonction discursive de  $=rā$ . Ainsi, ce critère comprend, en plus des COD, les circonstanciels nominaux topicalisés, cf. ex. (3) ci-dessus (p. 215)<sup>9</sup>, ainsi que les SN topicalisés extraits d'une position thématique, c'est-à-dire co-indexés avec un élément  $\theta$ , cf. ex. (7-a) ci-dessus<sup>10</sup>. Notons qu'à la différence des COD marqués qui se déplaceraient pour se trouver en position de l'adjoint au  $V''$ , dans les autres cas de figure, Ghameshi soutient la génération à la base.

Nous venons de passer en revue les analyses les plus répandues en syntaxe formelle qui, leurs différences mises à part, défendent l'hypothèse de la PSDO, c'est-à-dire une vision du  $V''$  persan dans laquelle le COD marqué de  $=rā$  occupe une position syntaxique plus élevée que sa contrepartie non marquée. Dans la section suivante, nous nous pencherons sur les arguments mis en avant pour soutenir cette hypothèse.

## 6.2 Examen des arguments en faveur de l'hypothèse de la PSDO

La divergence entre les COD marqués, d'un côté, et les COD non marqués, de l'autre, est incontestable, étant donné qu'ils ne relèvent pas du même niveau de détermination, c'est-à-dire, ne partagent pas les mêmes propriétés sémantico-

9. Notons que Ghameshi se distingue de Karimi (1990) dans son analyse des circonstanciels nominaux topicalisés par  $=rā$  qu'elle propose de traiter comme "pseudo"-COD, dans la mesure où, selon l'auteur, ils manifestent certaines propriétés prototypiques des COD (Ghameshi, 1997, p. 151).

10. Ici encore l'analyse de Ghameshi diverge de celle proposée par Karimi (1990). Rappelons que cette dernière suggère la position de spécifieur du  $C''$  dans ces cas de figure, tandis que Ghameshi (1997, pp. 153-158) soutient une analyse par adjonction multiple au  $V''$ .

référentielles. La question qui se pose est de savoir si cette divergence se rencontre au niveau structurel et doit se traduire dans une analyse en syntaxe des positions.

Les travaux détaillés dans la section précédente avancent différents arguments dans ce sens. En particulier, Karimi (2003a) énumère différentes asymétries sémantiques et syntaxiques entre les deux types du COD, qui constituent de loin l'argumentaire le plus exhaustif. Sur un plan sémantique, l'auteur évoque l'autonomie sémantique par rapport au verbe, la compatibilité avec des adverbes délimitatifs et la variabilité de la portée et, sur le plan syntaxique, outre l'ordre canonique, elle énumère des asymétries en lien avec le liage, les vides parasites, ainsi qu'un argument portant sur la coordination.

L'étude empirique, exposée dans la deuxième partie de cette thèse, réfute l'asymétrie portant sur l'ordre canonique, qui constitue l'épine dorsale de cette ligne d'argumentation. Par la suite, nous examinerons les autres asymétries et démontrons qu'aucune n'est exempte d'objections. Nous verrons, en fin de compte, qu'il n'y a aucun argument probant pour maintenir l'hypothèse de la PSDO. Dans l'ordre, seront évoqués la fusion (sémantique) avec le verbe (§6.2.1), la variabilité de la portée (§6.2.2), les relations de liage (§6.2.3), les vides parasites (§6.2.4) et la coordination (§6.2.5).

Avant de poursuivre, rappelons que la dichotomie marqué vs. non marqué, présumée dans les analyses susmentionnées, est discutable en raison du fait qu'elle implique que les COD marqués, d'une part, et les COD non marqués, d'autre part, constituent des classes homogènes. Ainsi, ces analyses négligent le fait que les COD non marqués varient en fonction de leur niveau de détermination et peuvent avoir une réalisation nue, c'est-à-dire sans aucune détermination, ni quantification, aussi bien qu'une réalisation déterminée, indéfinie ou quantifiée (cf. §1.1.2 et §3.2.2.1). Le COD nu est considéré de manière consensuelle comme ayant une relation étroite avec le verbe, qui rappelle celle entretenue entre les composantes d'un prédicat complexe<sup>11</sup>, en faisant partie en quelque sorte du prédicat sémantique exprimé par le verbe plutôt qu'en constituant une entité indépendante<sup>12</sup> (Lazard, 1982, 1994; Ghomeshi & Massam, 1994; Samvelian, 2001, 2012). En cela, il se distingue non seulement du COD marqué, mais aussi du COD non marqué indéfini et/ou quantifié, qui, dans son emploi prototypique, est incontestablement référentiel (Ghomeshi & Massam,

---

11. Notons que traduire cette différence en termes d'une syntaxe des positions n'est pas toujours endossée par ces travaux. Tout au contraire, Samvelian (2001, 2012) conteste le statut syntaxique singulier attribué aux objets nus, aussi bien que celui attribué aux éléments nominaux des prédicats complexes (cf. §1.1.4).

12. À ce propos, rappelons que ces faits ne se limitent pas au persan et qu'en syntaxe générale, l'on considère que les objets non référentiels, nom nus en l'occurrence, mettent l'accent sur l'évènement plutôt que de souligner l'entité à laquelle ils réfèrent (Creissels, 2006, p. 128).

1994; Ghomeshi, 2003).

### 6.2.1 Fusion sémantique avec le verbe

Karimi (2003a) affirme qu'un COD non marqué, qu'il ait une réalisation nue ou déterminée, fait partie intégrante de l'événement décrit par le prédicat, tandis que le COD marqué constitue une entité indépendante qui participe à l'événement décrit par le verbe (ou le subit) :

The nonspecific object (kind-level or existential) is part of the *event*, rather than the *participant in the event*.

Karimi (2003a, p. 99)

L'exemple ci-dessous, en réponse à la question *Que fait Kimea chaque soir ?*, illustre bien l'assimilation par l'auteur des COD indéfinis aux COD nus.

- (15) Kimea har šab (ye) sib mi-xor-e  
 Kimea chaque soir (un) pomme IPFV-manger.PRS-3SG  
 'Kimea mange des pommes chaque soir.'

Plus exactement '*Kimea does '(an) apple eating' every night.*'

Karimi (2003a, p. 100)

De plus, l'auteur évoque un argument, avancé initialement par Ghomeshi & Massam (1994), en vue de suggérer une analyse syntaxique des COD nus ( $N^0$ ) différente des autres COD ( $N''$ ) en termes d'incorporation syntaxique<sup>13</sup>.

Il s'agit du test de l'emploi des adverbes duratifs, communément utilisés pour déterminer les propriétés aspectuelles (*aktionsart*), notamment télélique (accomplissement) vs. atélique (processus) de l'événement exprimé par le verbe (cf. Vendler, 1967; Dowty, 1979), ex. (16). Les auteurs soutiennent que les COD marqués sont compatibles avec une lecture télélique, ex. (16-a), alors que les COD nus impliquent une lecture atélique, ex. (16-b)<sup>14</sup>.

13. Ghomeshi & Massam (1994, p. 190) :

- (i) a. 
$$\begin{array}{c} V' \\ \widehat{N'' \ V^0} \end{array}$$
 b. 
$$\begin{array}{c} V' \\ | \\ V^0 \\ \widehat{N^0 \ V^0} \end{array}$$

14. Néanmoins, il faut noter qu'il existe des contextes où un COD nu est compatible avec un adverbe délimitatif, dénotant un accomplissement, et que, à l'inverse, un COD marqué est compatible avec un adverbe duratif, dénotant un processus. Par exemple, en remplaçant le verbe 'manger' par un autre verbe d'accomplissement comme 'acheter' dans l'exemple (16), la phrase en (a) devient, selon notre jugement, acceptable. De même que la phrase en (b) avec un verbe d'activité comme 'mâcher' peut dénoter un événement atélique. Notons que Modarresi (2014), mention-

- (16) a. (man) (do-daqiqe-i / \*do daqiqe) sib=rā xord-am  
 moi en deux minutes / deux minute pomme=RĀ manger.PST-1SG  
 ‘J’ai mangé la pomme en/\*pendant deux minutes.’
- b. (man) (\*do-daqiqe-i / do daqiqe) sib xord-am  
 moi en deux minutes / deux minute pomme manger.PST-1SG  
 ‘J’ai mangé des pommes \*en/pendant deux minutes.’

Adaptés de Ghomeshi & Massam (1994, pp. 190-191)

Insistons sur le fait que la compatibilité avec des adverbes duratifs et l’incompatibilité avec des adverbes délimitatifs concernent seulement les COD nus. Le COD non marqué déterminé est, en revanche, tout à fait compatible avec une lecture télique, comme le montre l’exemple (17). Or, l’incompatibilité des COD nus avec une lecture télique est liée à leur lecture massive ou “cumulative” (voir par ex. Krifka, 1989, 1992)<sup>15</sup>, qui est en contradiction avec le caractère fini/délimité (ou *bound* en anglais) requis dans une lecture télique (voir également Modarresi, 2014).

- (17) Maryam do-daqiqe-i čand sib=e tāze xarid  
 Maryam en deux minutes quelques pomme=EZ frais acheter.PST.3SG  
 ‘Maryam a acheté quelques pommes fraîches en deux minutes.’

Bien que Karimi (2003a) prenne le soin dans sa démonstration de bien préciser que l’argument de Ghomeshi & Massam (1994) concerne, d’un côté, les COD nus et, de l’autre, les COD marqués de =rā, dans sa synthèse, elle généralise cet argument pour y inclure toute la classe des COD non marqués, dont, rappelons-le, elle tient

---

nant également ces possibilités, avance que les COD nus sont compatibles avec une lecture télique lorsque la connaissance encyclopédique/épistémologique (*world knowledge*) permet de déterminer le nombre du référent (sous-spécifié chez les COD nus en persan). Dans l’exemple donné par l’auteur (*āpārtēmān xaridan* ‘acheter un appartement’), le COD nu ne peut référer vraisemblablement qu’à une entité unique (Modarresi, 2014, p. 125). De même que dans la situation décrite par l’exemple ci-dessous, *sib* réfère forcément à une entité plurielle (par exemple, un sachet de pommes). Voir également la note de bas de page 15.

- (i) a. (man) do-daqiqe-i sib xarid-am (va raft-am)  
 moi en deux minutes pomme acheter.PST-1SG et aller.PST-1SG  
 ‘J’ai acheté des pommes en deux minutes (et je suis parti).’
- b. (man) do daqiqe sib=rā javid-am  
 moi deux minute pomme=RĀ mâcher.PST-1SG  
 ‘J’ai mâché la pomme pendant deux minutes.’

15. Comme cela est souligné dans de nombreux travaux traitant des propriétés aspectuelles des expressions verbales, il existe une correspondance directe entre les propriétés référentielles du nom (le COD en l’occurrence) et les propriétés événementielles du prédicat. Plus particulièrement, en établissant une distinction entre les objets dits *quantized* et *cumulative (non quantized)*, Krifka (1989, 1992) soutient, pour simplifier, que le premier est compatible avec une lecture télique – définie comme dénotant une activité délimitée dans le temps et étant, par conséquent, incompatible avec les adverbes duratifs –, tandis que le second ne l’est qu’avec une lecture atélique – définie à l’inverse comme dénotant une activité non délimitée dans le temps.

compte par le seul trait [–spécifique]<sup>16</sup>.

Non-specific objects are compatible with adverbs denoting process; specific objects are compatible with adverbs denoting accomplishment.

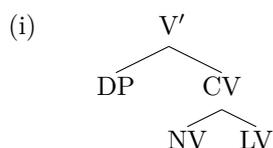
Karimi (2003a, p. 104)

Dans la même ligne d'argumentation, Karimi (2003a, pp. 104-105) rapproche les COD non marqués des éléments non verbaux des prédicats complexes, bien qu'elle admette des différences formelles entre les deux types de combinaison et en propose une analyse sémantique identique dans le cadre de la morphologie distribuée (cf. *Distributed Morphology*, Halle & Marantz, 1993), que l'on peut résumer, sans entrer dans les détails, par le schéma [N+Ving]<sup>17</sup>.

Cette analyse revient, en effet, à supposer que l'analyse sémantique d'une combinaison telle que *sib xaridan*, que l'on peut traduire par “*apple-buying*” dans une majorité de contextes, ne diffère pas d'une combinaison dans laquelle le COD est déterminé par un quantifieur ou par un déterminant indéfini, tout en demeurant non marqué. Ainsi, les deux variantes s'opposeraient, de manière analogue, à une combinaison faisant intervenir un COD marqué de =*rā*, comme en témoigne les phrases en (15) ci-dessus, en réponse à la question ‘Que fait Kimea chaque soir?’<sup>18</sup>.

16. Il est opportun de rappeler ici que, non seulement la façon dont l'auteur définit la spécificité est contestable (cf. Ghomeshi, 1997, pp. 138-139), mais aussi que ce trait, quelle qu'en soit la définition, ne suffit pas à lui seul pour rendre compte du marquage par =*rā* (cf. §1.1.3).

17. Cette analyse sémantique a, selon l'auteur, une contrepartie syntaxique, corroborée par les autres arguments avancés, qui consiste à postuler la position sœur du V (sous V') pour les COD non spécifiques, à l'instar de la configuration proposée pour la combinaison d'un élément non verbal (NV) et d'un verbe support (LV) dans un prédicat complexe (*Complex Verb* (CV) dans la terminologie utilisée par l'auteur). Cette analyse est schématisée ci-dessous (Karimi, 2003a, p. 105) :



18. De plus, les phrases en (i), en réponse à la question “Que fait Maryam au marché?”, indiquent que les questions appelant comme réponse une activité ne permettent pas toujours de distinguer entre les différents types de COD. Remarquons que toutes ces phrases conviennent également comme réponses à la question “Qu'achète Maryam en général au marché?”. Il en va de même pour les phrases en (ii), en réponse à la question “Quelle(s) pomme(s) a achetée(s) Maryam?”, qui montrent que les COD non marqués déterminés peuvent également convenir dans les réponses aux questions qui soulignent l'objet plutôt que l'activité.

(i) Réponses à la question “Que fait Maryam au marché?” :

- a. Maryam sib mi-xar-ad  
 Maryam pomme IPFV-acheter.PRS-3SG  
 ‘Maryam achète des pommes’.
- b. Maryam čand sib=e tāze mi-xar-ad  
 Maryam quelques pomme=EZ frais IPFV-acheter.PRS-3SG

Comme le montre la discussion amorcée plus haut, un tel regroupement est contestable. En effet, sauf dans les cas marginaux, ces COD non marqués sont référentiels et se différencient en cela clairement des COD nus. Notons que certains auteurs, notamment Ghomeshi (1997, 2003), admettent cette généralisation de manière absolue. Quoi qu'il en soit, la divergence entre les deux types de COD non marqués, quant à leurs propriétés référentielles et au rapport qu'ils entretiennent avec le verbe et l'événement décrit par celui-ci, remet en question la généralisation faite par Karimi.

## 6.2.2 Portée des quantifieurs

Karimi (2003a) affirme que seul un COD marqué est susceptible de provoquer une ambiguïté de portée lorsqu'il est déplacé vers la position initiale de la phrase. Afin d'évaluer la crédibilité empirique d'une telle généralisation, quelques mises au point théoriques et méthodologiques s'imposent.

### 6.2.2.1 Concepts et définitions

La portée (*scope*, voir Beyssade, 2006) désigne le domaine de l'effet des opérateurs tels que la négation, les quantifieurs, les mots interrogatifs, etc.

L'ambiguïté de portée se manifeste notamment quand un opérateur, par exemple la négation, ex. (18), est susceptible d'avoir une portée large ou étroite, c'est-à-dire respectivement que toute la phrase, ex. (18-a), ou seulement un de ces constituants, ex. (18-b) et ex. (18-c), peuvent se trouver sous sa portée.

- (18) Je ne suis pas allée en Iran cette année.
- a. Il est faux que je suis allée en Iran cette année.
  - b. Ce n'est pas cette année que je suis allée en Iran.

---

'Maryam achète quelques pommes fraîches.'

- c. Maryam hame=ye sib-hā=ye kāl=rā mi-xar-ad  
 Maryam tout=EZ sib-PL=EZ vert=RĀ IPFV-acheter.PRS-3SG  
 'Maryam achète toutes les pommes vertes.'

(ii) Réponses à la question "Quelle(s) pomme(s) a achetée(s) Maryam?" :

- a. \*Maryam sib-e           sabz xarid                   (ne-mi-dun-am                   kodum-hā)  
 Maryam pomme=EZ vert acheter.PRS.3SG NEG-IPFV-savoir.PRS-1SG quel-PL  
 'Maryam a acheté des pommes vertes (je ne sais pas lesquelles).'
- b. Maryam čand=tā           sib=e           sabz xarid  
 Maryam quelques=CLF pomme=EZ vert IPFV-acheter.PRS-3SG  
 (ne-mi-dun-am                   kodum-hā)  
 NEG-IPFV-savoir.PRS-1SG quel-PL  
 'Maryam a acheté quelques pommes vertes (je ne sais pas lesquelles).'

- c. Ce n'est pas en Iran que je suis allée cette année.

L'ambiguïté de portée se manifeste également quand deux ou plusieurs opérateurs interagissent. En (19), la négation peut avoir une portée large ( $\neg > \forall$ ) ou étroite ( $\forall > \neg$ ) sur le quantifieur universel. En (20) et (21), les portées des deux quantifieurs universel et existentiel interagissent ( $\exists > \forall$  et  $\forall > \exists$ ). La portée large d'un quantifieur existentiel sur un quantifieur universel résulte dans une lecture collective et sa portée étroite dans une lecture distributive. Ainsi, chacune de ces phrases peut recevoir deux interprétations vériconditionnelles différentes. Il va sans dire que, dans chaque cas de figure, les deux lectures ne sont pas forcément également accessibles.

- (19) Tous les politiciens ne sont pas corrompus.
- a.  $\neg\forall$  : Il est faux que tous les politiciens sont corrompus.  
 $\equiv$  Il existe des politiciens qui ne sont pas corrompus.
- b.  $\forall\neg$  : Tous les politiciens, ne sont pas corrompus.  
 $\equiv$  Il n'existe pas de politicien qui soit corrompu.
- (20) Le professeur a donné **un** livre à **tous** les étudiants.
- a.  $\exists > \forall$  : Le même livre est donné à tous les étudiants.
- b.  $\forall > \exists$  : Différents livres sont donnés aux étudiants.
- (21) **Tous** les étudiants ont reçu **un** livre.
- a.  $\exists > \forall$  : Tous les étudiants ont reçu le même livre.
- b.  $\forall > \exists$  : Les étudiants ont reçu différents livres.

De prime abord, l'ambiguïté de portée, plus particulièrement la portée dite inverse, ex. (20-b) et ex. (21-a), pose problème pour toute analyse syntaxique de la phrase en structure syntagmatique hiérarchisée – c'est-à-dire qui admet une relation de c-commande asymétrique entre les constituants immédiats, dans la mesure où la portée, large ou étroite, est considérée comme dépendant de la position syntaxique respective des quantifieurs, c'est-à-dire les relations de c-commande. En effet, dans ces cas de figure, un constituant, dont la position syntaxique est plus basse, domine sur le plan sémantique un constituant dont la position est plus haute.

En syntaxe générative, un mouvement obligatoire, dénommé "montée des quantifieurs" (*Quantifier Raising, QR*), qui a lieu au niveau de la composante LF (*Logical Form*) où chaque phrase reçoit une interprétation unique, a été proposé pour résoudre ce paradoxe (May, 1985). La montée des quantifieurs implique que toute expression quantifiée soit déplacée de sa position de surface vers une position à la périphérie gauche de la phrase. Suite à ce déplacement, l'ordre des quantifieurs, plus

précisément l'ordre dans lequel les déplacements ont été opérés, détermine leur portée. À l'instar des mouvements qui ont lieu au niveau LF, la montée des quantifieurs est un mouvement implicite (ou caché, *covert* en anglais), c'est-à-dire sans effet sur la structure de surface ou l'ordre linéaire de la phrase. De plus, comme tout mouvement, elle est soumise à des contraintes, qui, le cas échéant, expliqueraient l'absence d'ambiguïté.

Cette règle syntaxique présumée universelle fait depuis l'objet de débats controversés, notamment par rapport au comportement des SN indéfinis, en raison de l'existence d'une multitude de faits empiriques que cette règle laisse au bord de la route<sup>19</sup> (Fodor & Sag, 1982; Diesing, 1992; Farkas, 1994; Dobrovie-Sorin, 1995; Beck, 1996; Szabolcsi, 1997b; Reinhart, 1997; Kratzer, 1998; De Swart, 1998; Barker, 2002). Toutefois, la revue de ces débats dépasse le cadre de cette thèse.

En bref, ces travaux démontrent la complexité, c'est-à-dire la multidimensionalité, des phénomènes qui relèvent de l'ambiguïté de portée, et s'accordent pour écarter toute approche en termes exclusivement de contraintes structurelles<sup>20</sup>.

La portée des quantifieurs dans les langues à ordre flexible, telles que l'allemand, le japonais ou le hongrois, constitue un volet important de ces débats, dans la mesure où l'on observe que les relations de portée sont largement déterminées dans ces langues par la structure de surface, c'est-à-dire plus précisément par l'ordre linéaire respectif des quantifieurs. Pour résumer, au sein des travaux en syntaxe générative chomskyenne, deux approches théoriques s'opposent pour définir les règles de détermination de la portée : l'une se base sur la structure au niveau du LF et l'autre sur la structure de surface.

En effet, un lien direct entre la précédence (au niveau de la structure de surface) et la portée se dessine à travers l'ensemble des travaux sur les langues à ordre flexible, ceci de manière telle que, dans la littérature, ces langues sont considérées comme rigides quant à la variabilité de portée (*scope rigid* en anglais). À l'opposé, les langues présentant la variabilité de portée, comme le français, connaissent un ordre des mots rigide. Cela revient à dire que seule une des deux lectures disponibles pour

---

19. Pour en donner un exemple, l'on note qu'un SN indéfini ne permet pas la variabilité de portée dans un contexte où un autre type de quantifieur le permet. Comparons : "j'ai écrit à chaque reviewer qui a commenté un article refusé" et "j'ai écrit à un reviewer qui a commenté chaque article refusé". Tandis que la première phrase est ambiguë entre une portée large ou étroite du quantifieur universel sur le quantifieur existentiel, dans la seconde phrase le SN indéfini ne permet que l'interprétation collective, c'est-à-dire la portée large.

20. En simplifiant grandement : différentes solutions sont préconisées, dont la prise en compte plus importante de l'appareil sémantique dans la répartition des tâches et l'intégration des facteurs lexicaux et discursifs. Certains auteurs suggèrent de déplacer l'analyse de la portée dans le domaine de la sémantique (et/ou la pragmatique) et d'autres proposent d'accommoder les faits en ajoutant des contraintes de type lexical à l'analyse syntaxique initiale en affinant le sémantisme des SN quantifiés ou en réduisant leur catégorie.

les phrases (20) et (21) du français (celle qui correspond à l'ordre respectif des deux quantifieurs) est disponible pour des exemples équivalents dans une langue à ordre flexible, comme l'allemand, le japonais ou le hongrois (ex. Hoji, 1986; Beck, 1996; Szabolcsi, 1997a). Or, cette généralisation est trop forte puisque la portée inverse n'est pas exclue dans ces langues et l'ambiguïté de portée y est également présente dans certaines configurations (cf. par ex. Kuno *et al.*, 1999; Bobaljik & Wurmbrand, 2012).

### 6.2.2.2 Lien controversé de la portée avec la structure syntaxique

Tout d'abord, les travaux ne s'accordent pas toujours sur les lectures disponibles. À titre d'exemple, Saito (2005, p. 343) affirme que la portée rigide, attribuée de manière consensuelle aux phrases telle que (22-a)<sup>21</sup> du japonais, est certes une préférence forte mais pas une condition absolue<sup>22</sup>. De même, Kuno *et al.* (1999, p. 105) avancent que cette phrase est ambiguë pour une majorité des locuteurs japonais, tandis que la phrase (22-b), dans laquelle seuls les quantifieurs sont permutés, ne l'est pas.

- (22) a. Dareka-ga daremo-o aisite-iru  
 quelqu'un-NOM tout le monde-ACC aime  
 'Quelqu'un aime tout le monde.' ( $\exists > \forall$  et  $?\forall > \exists$ )
- b. Daremo-ga dareka-o aisite-iru  
 tout le monde-NOM quelqu'un-ACC aime  
 'Tout le monde aime quelqu'un.' ( $\forall > \exists$ )

Deuxièmement, l'intonation et la prosodie, ou plus précisément la configuration discursive de la phrase réalisée par la prosodie, peut modifier les relations de portée. Par exemple, en allemand, une phrase, dont la lecture correspond normalement à la portée rigide, ex. (23-a), peut recevoir une lecture à portée inverse, ex. (23-b), sous l'intonation "*rise-fall*" correspondant à la configuration discursive "*topic-focus*" (Büring, 1997; Krifka, 1998).

- (23) a. Mindestens ein Student hat jeden Roman gelesen  
 au-moins un.NOM étudiant a tout.ACC roman lu  
 'Au moins un étudiant a lu tous les romans.' ( $\exists > \forall$ )

Adapté de Krifka (1998, p. 77)

21. Dans tous les exemples glosés que nous reproduisons ici, nous conservons la glose utilisée par les auteurs en la traduisant en français.

22. L'auteur affirme, en tant que locuteur natif de cette langue, que dans cette phrase la portée inverse, bien que peu préférable, est cependant possible.

- b. Mindestens /EIN Student hat \JEDen Roman gelesen  
 au-moins un.NOM étudiant a tout.ACC roman lu  
 'Au moins un étudiant a lu tous les romans.' ( $\exists > \forall$  et  $\forall > \exists$ )

Adapté de Krifka (1998, p. 80)

Troisièmement, la plausibilité du contexte peut favoriser la lecture avec la portée inverse, avec ou sans une intonation particulière. Bobaljik & Wurmbrand (2012), se basant sur l'exemple (24)<sup>23</sup>, dont la structure est identique à (23-a) ci-dessus, affirment que la portée inverse ( $\forall > \exists$ ) n'est pas seulement disponible, mais qu'elle est aussi la seule interprétation possible dans un contexte où l'on sait de quels musiciens il s'agit et qu'il n'existe pas de disque fait par tous les musiciens (Bobaljik & Wurmbrand, 2012, p. 381).

- (24) Peter hat eine Platte jedes Musikers aufgelegt  
 Peter a un disque tout.GEN musicien joué  
 'Peter a joué un disque de tous les musiciens.'

En effet, en règle générale, les facteurs d'ordre lexical, pragmatique et discursif sont considérés comme jouant un rôle déterminant dans l'interprétation de la portée (Büiring, 1997; Kuno *et al.*, 1999).

Enfin, le *scrambling*<sup>24</sup> peut créer une ambiguïté de portée. Cela s'expliquerait par le fait que, suite à ce mouvement, la trace du constituant déplacé rentre en ligne de compte. Par exemple, bien que les jugements divergent sur les phrases (22) du japonais dont l'ordre linéaire respecte l'ordre canonique SOV, un consensus s'établit (Kuno *et al.* 1999, pp. 105–106 et Saito 2005, p. 346) sur le fait que les phrases dans lesquelles le COD est déplacé en position initiale sont ambiguës, ex. (25).

- (25) a. Daremo-o<sub>i</sub> dareka-ga t<sub>i</sub> aisite-iru  
 quelqu'un-NOM tout le monde-ACC aime  
 'Quelqu'un aime tout le monde.' ( $\exists > \forall$  et  $\forall > \exists$ )  
 b. Dareka-o<sub>i</sub> daremo-ga t<sub>i</sub> aisite-iru  
 tout le monde-NOM quelqu'un-ACC aime  
 'Tout le monde aime quelqu'un.' ( $\forall > \exists$  et  $\exists > \forall$ )

De ce fait, le débat sur la portée des quantifieurs dans les langues à ordre flexible rejoint celui plus général sur le *scrambling* au sein de cette théorie (cf. §2.2.1.3), selon laquelle les relations de portée seraient des diagnostics pour le mouvement et

23. Extrait des données expérimentales de Sauerland & Bott (2002).

24. Rappelons qu'il s'agit du mouvement qui en grammaire générative chomskyenne rend compte dans les langues à ordre flexible du bouleversement de l'ordre de base dans une phrase à ordre non canonique (cf. §2.2.1.3).

son type (A-position vs. A'-position). C'est dans cette même optique que Karimi (2003a, 2005) aborde les données du persan.

Deux approches s'opposent dans le débat relatif au *scrambling* chez les auteurs qui adhèrent à une analyse configurationnelle des ordres non canoniques. Certains défendent l'idée que le *scrambling* est défait au niveau LF, et ne joue par conséquent pas de rôle dans la détermination de la portée (ex. Saito & Fukui, 1998; Saito, 2005). D'autres, y compris Karimi, considèrent, à l'opposé, que la portée est déterminée selon la structure de surface<sup>25</sup> et est sensible au *scrambling*, dans la mesure où la trace, laissée suite à ce mouvement, crée une nouvelle relation c-commande. Le cas échéant, ceci explique la disponibilité de deux lectures vériconditionnelles. De même, les différentes contraintes imposées à ce mouvement expliqueront l'absence d'une telle possibilité<sup>26</sup>.

Notons que ce débat devient vite circulaire si, pour une phrase donnée, les interprétations disponibles ne sont pas posées de manière raisonnablement objective et indépendante de la théorie, car, comme nous avons déjà vu, les lectures disponibles divergent souvent par rapport à leur degré d'accessibilité et peuvent également varier d'un auteur/locuteur à un autre ou d'un contexte à un autre<sup>27</sup>. Ainsi, sans recourir aux méthodes empiriques standards pour déterminer les interprétations disponibles pour une même phrase et leur accessibilité respective, aucune analyse ne pourrait être considérée comme crédible (cf. Kuno et al., 2001, p. 142).

En conclusion, l'on constate que la variabilité de portée des quantifieurs demeure une question théorique ouverte aussi bien en syntaxe qu'en sémantique formelle. Comme en témoigne la littérature (cf. Szabolcsi, 1997b; Kuno et al., 1999; Brasoveanu & Farkas, 2011; Steedman, 2012; Dobrovie-Sorin & Beyssade, 2012, entre autres), il ne s'agit pas d'un phénomène homogène et monolithique que l'on peut prétendre étudier seulement à travers un nombre réduit d'exemples isolés, qu'ils soient attestés ou construits. À ce propos, la remarque de Kuno et al. (1999) est

25. « *Scope is determined in overt syntax.* » (Karimi, 2005, p. 164)

26. Dans une analyse non configurationnelle ou plate de l'ordre libre, les relations de c-commande sont considérées comme non asymétriques entraînant, par définition, les relations de portée symétriques et, par conséquent, l'ambiguïté (Kiss, 2008, pp. 470–472).

27. À titre d'exemple, Kuno et al. (1999) avancent qu'une analyse syntaxique en termes de relation binaire de c-commande ne peut être adéquate pour traiter de la variabilité des interprétations disponibles :

[...] syntactic analyses depend crucially on the notion of c-command, and they treat scope interpretations as an all-or-nothing phenomenon. [...] scope interpretations exhibit a continuum, and [...] even in an ambiguous sentence, one interpretation is generally dominant. Since there cannot be, for instance, a 'thirty percent c-command relationship', [these] syntactic models fail to account for this gradation.

Kuno et al. (1999, p. 109)

éloquente :

Quantifier scope is a delicate phenomenon. It is not simply a matter of ambiguity vs. nonambiguity, but a continuum, and the judgments on a given sentence often fluctuate from speaker to speaker.

Kuno et al. (1999, p. 110)

Toujours d'après les mêmes auteurs, plutôt que des modèles syntaxiques ne tenant compte que d'un ensemble réduit de données, une analyse fonctionnelle en termes d'interaction d'une série de préférences, qu'ils nomment *expert system*, pourrait rendre compte de ce phénomène (Kuno et al., 1999, p. 109).

Par ailleurs, les travaux plus récents plaident en faveur des méthodes expérimentales pour l'élicitation des relations de portée, dans la lignée des travaux en psycholinguistique étudiant l'interprétation de la portée et les différents facteurs qui y interviennent du point de vue du traitement et de la compréhension (voir Szabolcsi, 2010, ch. 6, pp. 87–90). En effet, dans ce domaine, il existe une riche littérature remontant à Kurtzman & MacDonald (1993), qui s'intéresse à la portée des quantifieurs en tant que phénomène relevant de l'ambiguïté sémantique (ex. Filik et al., 2004; Dwivedi, 2013; Urbach et al., 2015)

Notons que le cadre de cette thèse ne nous permet pas une telle entreprise, étant donné que nous n'y abordons la question de la portée que marginalement, dans le but seulement de s'interroger sur la fiabilité des interprétations mises en avant par Karimi (2003a) pour soutenir l'hypothèse de la PSDO, sans rentrer dans les débats théoriques.

### 6.2.2.3 Données du persan

Karimi (2003a) affirme que seul un COD marqué est susceptible de provoquer une ambiguïté de portée lorsqu'il est déplacé vers la position initiale de la phrase, ex. (26). Un COD non marqué, faisant partie intégrante (sémantiquement et syntaxiquement) du prédicat exprimé par le verbe, ne peut jamais avoir une portée large, ex. (27).

- (26) a. **har** dānešju=i **ye** še'r=ro bāyād be-xun-e  
chaque étudiant=INDF un poème=RĀ doit SBJV-lire.PRS-3SG  
'Chaque étudiant doit lire un poème (choisi dans un ensemble spécifique).' ( $\forall > \exists$ )
- b. **ye** še'r=ro<sub>i</sub> **har** dānešju=i t<sub>i</sub> bāyād be-xun-e ( $\forall > \exists$ ;  $\exists > \forall$ )
- (27) a. **har** dānešju=i **ye** še'r bāyād be-xun-e  
chaque étudiant=INDF un poème doit SBJV-lire.PRS-3SG  
'Chaque étudiant doit lire un certain poème.' ( $\forall > \exists$ )

- b. **ye** še'r<sub>i</sub> **har** dānešju=i t<sub>i</sub> bāyād be-xun-e ( $\forall > \exists$ )

Adaptés de Karimi (2003a, p. 103)

Nous n'adhérons pas aux interprétations formulées par l'auteur. Toutefois, en l'absence de données empiriques, et pour éviter toute spéculation de notre part, nous nous appuyerons sur les observations relatives à la portée des COD non marqués en persan faites par d'autres auteurs qui contredisent cette affirmation, à savoir Ghomeshi (1997, pp. 139–140) et Modarresi (2014, ch. 3). Ces derniers affirment, en effet, que les COD non marqués déterminés, c'est-à-dire les COD non marqués indéfinis ou quantifiés<sup>28</sup>, peuvent recevoir une portée large. Il n'existerait aussi pas de relation directe entre le marquage par  $=r\bar{a}$  et la portée large ou étroite.

À titre d'exemple, dans les phrases (28) et (29-a), les deux lectures avec la portée large ou étroite sont disponibles pour le COD indéfini non marqué. Par ailleurs, selon Ghomeshi (1997, p. 139), le COD marqué, ex.(29-b), peut également recevoir la portée large dans sa position de base.

- (28) hame film=i did-and  
tout film=INDF voir.PST-3PL  
'Tout de monde regardèrent un film.'

Adapté de Modarresi (2014, p. 30)

- (29) a. ?hame=ye mo'allem-ā ye šāgerd=i mo'arrefi kard-and  
tout=EZ enseignant-PL un élève=INDF présentation faire.PST-3PL  
'Tous les instituteurs présentèrent un élève.'
- b. hame=ye mo'allem-ā ye šāgerd=i=ro mo'arrefi  
tout=EZ instituteur-PL un élève=INDF=RĀ introduction  
kard-and  
faire.PST-3PL  
'Tous les instituteurs présentèrent un élève.'

Adaptés de Ghomeshi (1997, p. 139)

- (30) a. hame=ye mo'allem-ā ye ketāb=i entexāb kard-and  
tout=EZ instituteur-PL un livre=INDF sélection faire.PST-3PL  
'Tous les instituteurs choisirent un livre.'
- b. ?hame=ye mo'allem-ā ye ketāb=i=ro entexāb kard-and  
tout=EZ instituteur-PL un livre=INDF=RĀ sélection faire.PST-3PL  
'Tous les instituteurs choisirent un livre.'

Adaptés de Ghomeshi (1997, p. 140)

28. Notons que la question ne se pose pas pour les COD nus, qui, étant non référentiels, n'introduisent pas de variable leur permettant de rentrer dans les relations de portée.

Il importe de noter que, comme dans beaucoup d'autres langues, le déterminant *har* 'chaque' favorise en règle générale une lecture distributive qui se traduit par la portée large du quantifieur universel sur le quantifieur existentiel. Or, avec le quantifieur universel *hame* 'tous', la lecture collective est facilitée<sup>29</sup> et, par conséquent, la portée large du quantifieur existentiel devient tout autant accessible, ainsi que parfois la lecture primaire, comme cela semble être le cas dans la phrase (28) ci-dessus.

Au-delà de ces contre-exemples fragilisant les données sur lesquelles s'appuie Karimi (2003a), d'autres points viennent remettre en question une analyse catégorique syntaxique de ces données.

Tout d'abord, Ghomeshi (1997) souligne que les données qui semblent parfois suggérer un effet de la portée démontrent, en réalité, l'effet de facteurs tels que l'animéité sur le marquage par  $=r\bar{a}$ . L'auteur donne les paires d'exemples, (29) et (30), pour appuyer ce propos.

Ensuite, il est opportun de noter le lien entre le statut discursif et la portée du COD en ex. (26-b) ci-dessus. Comme cela a été signalé pour d'autres langues, le statut discursif topic facilite, voire impose, la lecture à portée large. Or, la topicalité est un facteur favorisant le marquage par  $=r\bar{a}$ .

Enfin, la portée large du quantifieur existentiel entraîne la lecture collective qui implique l'unicité du référent, une propriété sémantique qui favorise le marquage par  $=r\bar{a}$ .

L'ensemble de ces observations nous conduit à conclure qu'il n'est pas possible d'établir une asymétrie catégorique entre les deux COD marqué et non marqué sur la base des relations de portée.

### 6.2.3 Liage des anaphores

Karimi (2003a) affirme que seul un COD marqué/spécifique est en mesure de lier une anaphore (réciproque)<sup>30</sup> en fonction du COI, dans les constructions ditran-

29. Cette distinction rappelle celle entre *every* et *each* de l'anglais.

30. Notons que Karimi n'aborde pas l'anaphore réflexive sous prétexte que l'usage du pronom réflexif non fléchi en persan, *xod*, orienté vers le sujet, est limité au registre formel. Or, étant donné la rigidité de ce registre, il ne convient pas d'utiliser cette forme pour étudier le phénomène de liage. Quant aux formes réflexives fléchies, ex. *xod=am* (*xod=1SG*) 'moi-même', employées dans le langage courant, l'auteur affirme que le pronom personnel clitique les prive de leurs propriétés anaphoriques (cf. Karimi, 2005, p. 174). Cette dernière remarque est toutefois discutable. Dans beaucoup de contextes (y compris celui d'un emploi clairement réflexif et non emphatique), ces deux variantes ont une distribution sensiblement identique et leur distinction se limite à une différence de registre, cf. Moyne (1971), mais voir Ghomeshi (1996, ch. 4). Par ailleurs, la disponibilité de différentes variantes de pronoms réflexifs, en gros forme simple et forme complexe, n'est pas limitée au persan. De nombreuses langues, telles que les langues germaniques, présentent des cas similaires

sitives, ex. (31).

- (31) a. man [se=tā bačče-hā=ro]<sub>i</sub> be hamdige<sub>i</sub> mo'arrefi  
 moi trois=CLF enfant-PL=RĀ à l'un l'autre présentation  
 kard-am  
 faire.PST-1SG  
 'J'ai présenté les trois enfants les uns aux autres.'
- b. \*man se=tā bačče be hamdige mo'arrefi kard-am

Karimi (2003a, p. 102)

Avant de procéder à l'examen empirique de cet argument, quelques précisions terminologiques et théoriques s'imposent.

### 6.2.3.1 Concepts et définitions

La théorie du “liage” (*binding theory*, Chomsky, 1981), développée initialement dans le cadre de la théorie générative chomskyenne, désigne l'appareil qui rend compte des (contraintes sur les) relations d'antécédence entre des éléments co-référentiels. Dans ce cadre, anaphore<sup>31</sup> se dit des pronoms, telles que des expressions réfléchies, ex. “lui-même”, et réciproques, ex. “l'un l'autre”, dont l'interprétation dépend de la présence d'un antécédent<sup>32</sup> local – c'est-à-dire, techniquement parlant, qui se trouve dans son “domaine de liage” –, faute de valeur référentielle intrinsèque<sup>33</sup>. Le “domaine de liage” désigne grossièrement l'unité phrastique minimale à laquelle appartient l'élément pronominal. Dans la définition configurationnelle du liage, un élément pronominal est “lié” (*bound*) lorsque son antécédent se trouve dans la première projection maximale qui l'englobe – dans cette configuration, l'antécédent est appelé “lieur” (*binder*) –, et “libre” dans le cas contraire. En termes plus techniques, il s'agit de la relation de c-commande – occasionnellement appelée “m-

---

(cf. par ex. Buring, 2005, pp. 75–77). Force est de constater que ces systèmes pronominaux sont plus complexes que des systèmes binaires comme celui de l'anglais, ex. *him* vs. *himself*.

31. Rappelons que, en dehors de ce cadre théorique, l'anaphore désigne plus généralement « le phénomène par lequel un élément dépend pour son interprétation d'une entité établie dans la phrase (anaphore phrastique) ou hors de la phrase (anaphore discursive) » (Creissels, 2006, p. 82).

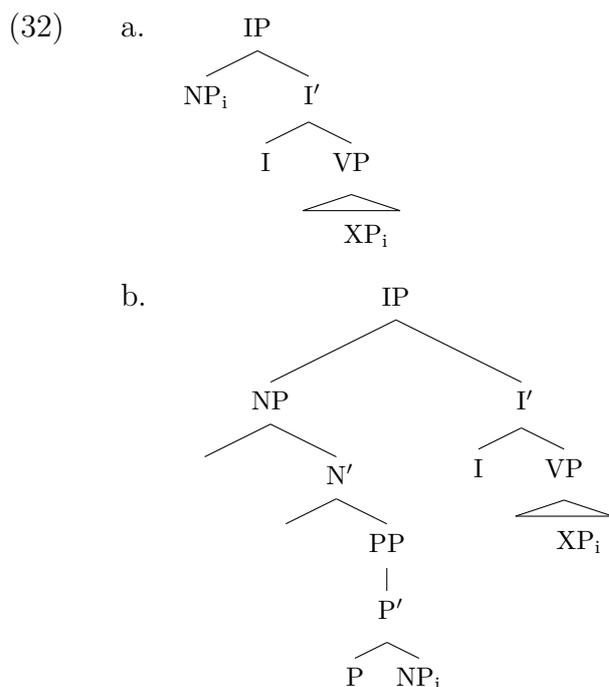
32. Dans une relation anaphorique, l'antécédent désigne l'expression dont l'élément pronominal dépend pour son interprétation. Autrement dit, l'antécédent est l'élément non pronominal avec lequel ce dernier est co-référentiel ou, techniquement parlant, co-indexé. Naturellement, un élément pronominal doit être sémantiquement compatible avec son antécédent.

33. Dans cette terminologie, le pronom (ordinaire), ex. “lui”, désigne les éléments pronominaux qui ne sont pas des anaphores. Les pronoms, à l'inverse des anaphores, ne toléreraient pas un antécédent local. Comparez :

- (i) a. Jean<sub>i</sub> parle à lui-même<sub>i/\*j</sub>  
 b. Jean<sub>i</sub> parle à lui<sub>j/\*i</sub>

commande” dans les ouvrages ultérieurs –, qui prend tout son sens dans la structure syntagmatique hiérarchisée postulée dans ce cadre théorique (cf. §2.2.1.3).

À titre d'exemple, le sujet est en position de c-commander le constituant occupant la position de l'objet, tandis que le complément adnominal du sujet ne l'est pas, comme l'illustrent les deux schémas en (32). Selon le “Principe A” de la théorie du liage, une anaphore doit être liée dans son domaine de liage<sup>34</sup>.



Ainsi, dans la série de phrases en (33), illustrant l'anaphore réciproque “l'un ... l'autre” du français, la (a)grammaticalité des phrases s'expliquerait par le (non) respect du principe A. Rappelons que, dans cette théorie, *PRO*, c'est-à-dire le sujet nul implicite dans les propositions infinies, peut faire office de lieu.

- (33) a. Ces jeunes gens<sub>i</sub> sont amoureux l'un de l'autre<sub>i</sub>  
 b. Les enfants<sub>i</sub> se<sub>i</sub> photographient l'un l'autre<sub>i</sub> / les uns les autres<sub>i</sub>  
 c. Pierre croit [ces jeunes gens<sub>i</sub> amoureux l'un de l'autre<sub>i</sub> / les uns des autres<sub>i</sub> ]  
 d. Marie a obligé les enfants de [*PRO*<sub>i</sub> être solidaires l'un de l'autre<sub>i</sub> / les uns des autres<sub>i</sub> ]  
 e. \*La voisine des enfants<sub>i</sub> se<sub>i</sub> photographiera l'un l'autre<sub>i</sub> / les uns les autres<sub>i</sub>  
 f. \*Les enfants<sub>i</sub> sont rentrés. Marie est jalouse l'un de l'autre<sub>i</sub> / les uns des autres<sub>i</sub>

34. *A contrario*, un pronom doit y être libre (“Principe B”).

- g. \*Marie<sub>i</sub> a proposé aux enfants<sub>z</sub> de [*PRO*<sub>i</sub> être solidaires l'un l'autre<sub>i</sub> / les uns des autres<sub>i</sub> ]  
 h. \*Ces jeunes gens<sub>z</sub> croient [ Pierre<sub>i</sub> amoureux l'un de l'autre<sub>z</sub> ]

Adaptés de Zribi-Hertz (1996, pp. 104-114)

Nous aurons l'occasion de revenir sur le phénomène du liage dans la section 6.3, où nous verrons que cette théorie, du fait de sa couverture empirique controversée, a été concurrencée par des approches alternatives n'associant pas le liage avec la structure syntagmatique de la phrase. Celles-ci rendent compte des phénomènes présentés ci-dessus sur la base de la hiérarchie des rôles grammaticaux sans faire intervenir la structure syntagmatique.

Notons enfin que les travaux étudiant les anaphores du point de vue de la théorie du liage se concentrent souvent principalement sur l'anaphore réflexive, arguant que la réciproque dispose sensiblement des mêmes propriétés distributionnelles tout en présentant des complexités sémantiques (par ex. Truswell, 2014). En effet, l'anaphore réciproque, quant à elle, a fait couler beaucoup d'encre de par de son sémantisme (cf. Fiengo & Lasnik, 1973; Langendoen, 1978; Heim et al., 1991; Dalrymple et al., 1998; Sauerland, 1998; Sternefeld, 1998; Beck, 2001, parmi beaucoup d'autres).

### 6.2.3.2 Impact des facteurs fonctionnels

Quelle que soit la théorie, il semble qu'il y ait un consensus pour dire que le liage dans son ensemble est un phénomène complexe, tant du point de vue syntaxique que sémantique (Büring, 2005). D'un côté, les faits rapportés pour une même langue peuvent varier d'une étude à l'autre, c'est-à-dire que les jugements et interprétations des auteurs – locuteurs natifs d'une même langue – divergent pour un même cas de figure. De l'autre côté, comme en témoignent de nombreux travaux expérimentaux<sup>35</sup>, les données obtenues des locuteurs naïfs révèlent un degré certain de variabilité et, de surcroît, ne convergent pas forcément avec les prédictions de la théorie (Keller & Asudeh, 2001; Runner et al., 2002, 2006; Kaiser et al., 2009; Runner & Head, 2014, entre autres). De plus, les études expérimentales ont permis de souligner le rôle des facteurs fonctionnels (extra-syntaxiques), notamment les facteurs d'ordre pragmatique et discursif. Ces facteurs étaient antérieurement mis en avant dans les travaux, adoptant une approche fonctionnelle (cf. Kuno, 1987).

Ainsi, de même que pour la variabilité de la portée, le doute s'installe quant à

35. Notons que les travaux expérimentaux dans ce domaine se situent majoritairement dans la lignée des recherches en psycholinguistique qui s'intéressent aux mécanismes du traitement des éléments pronominaux au sens large, que l'on désigne habituellement par "résolution de pronom", *pronoun resolution* (par ex. Hemforth et al., 2010; Colonna et al., 2010, 2012).

la possibilité d'étudier des relations du liage à travers seulement quelques exemples sans recourir à des méthodes expérimentales. En effet, toute analyse approfondie de l'asymétrie discutée ici requiert impérativement une investigation empirique poussée. Or, cela dépasse largement le cadre de de cette thèse. C'est la raison pour laquelle nous nous basons sur des exemples construits ou attestés – relevés sur l'Internet – pour exprimer notre désaccord avec la généralisation binaire de Karimi (2003a) et pour formuler une hypothèse alternative – qui sera examinée ultérieurement par des méthodes empiriques adaptées.

### 6.2.3.3 Données du persan

Selon Karimi (2003a), seul un COD spécifique, c'est-à-dire marqué, est en mesure de lier une anaphore en fonction du COI, dans les constructions ditransitives, ex. (31) ci-dessus répété en (34) et (35)<sup>36</sup>. L'auteur affirme que cela s'explique par le fait que les COD non marqués font partie du prédicat et, à ce titre, ne sont pas en mesure d'entrer dans les relations de liage<sup>37</sup>.

- (34) a. man [se=tā bačče-hā=ro]<sub>i</sub> be hamdige<sub>i</sub> mo'arrefi  
 moi trois=CLF enfant-PL=RĀ à l'un l'autre présentation  
 kard-am  
 faire.PST-1SG  
 'J'ai présenté les trois enfants les uns aux autres.'  
 b. \*man se=tā bačče be hamdige mo'arrefi kard-am

Karimi (2003a, p. 102)

- (35) a. Kimea<sub>i</sub> [ye dānešju=ro]<sub>k</sub> be hamšāgerdi-hā=š<sub>i/k</sub> mo'arrefi  
 Kimea un étudiant=RĀ à camarade-PL=3SG présentation  
 kard  
 faire.PST.3SG

36. L'auteur précise que le *scrambling*, c'est-à-dire l'ordre linéaire bouleversé pour respecter la règle de précedence requise pour le liage, ne peut pas être invoqué pour expliquer l'agrammaticalité de (34-b), car, comme en témoigne ex. (i), celui-ci est parfaitement acceptable lorsque le COD est en focus contrastif (Karimi, 2003a, p. 102).

- (i) man se=tā dānešju be dust=am mo'arrefi kard-am  
 moi trois=CLF étudiant à ami=1SG présentation faire.PST-1SG  
 'J'ai présenté trois **étudiants** à mon ami (pas des professeurs).'

37. Dans le cas des COD existentiels, cette restriction se limite selon l'auteur à la proposition. Ceci pour rendre compte de cas tels que (i) où manifestement la reprise du COD existentiel ne rend pas la phrase inacceptable (Karimi, 2003a, p. 113).

- (i) Kimea tunest ye āpārtēmān peydā kon-e un xeyli qašang=e  
 Kimea pouvoir.PST.3SG un appartement trouvé faire.PRS-3SG il très beau=COP.3SG  
 'Kimea a réussi à trouver un appartement. Il est très beau.'

‘Kimea a présenté un étudiant à ses camarades.’

- b. Kimea<sub>i</sub> [ye dānešju]<sub>k</sub> be hamšāgerdi-hā=š<sub>i</sub>/\*<sub>k</sub> t mo’arrefi kard

Karimi (2003a, p. 113)

Rappelons avant tout que traiter les COD non marqués comme des entités non référentielles dépourvues d’autonomie – sous prétexte qu’ils feraient partie intégrante du prédicat exprimé par le verbe – est en soi hautement contestable. Dans un article antérieur sur les vides parasites, Karimi (1999, p. 707) évoque le contraste sémantique, *semantic mismatch*, comme un argument plausible pour expliquer l’impossibilité pour les COD non spécifiques d’autoriser les vides parasites (cf. §6.2.4). Selon cet argument, les conditions de base pour une relation de coréférence requièrent que les éléments coindexés soient sémantiquement compatibles, y compris du point de vue du trait [ $\pm$ spécifique]. Dans la même lignée, étant donné que le pronom réciproque est spécifique, il ne peut prendre un SN non spécifique comme antécédent. Ainsi, cet argument alternatif pourrait éventuellement être évoqué pour expliquer l’asymétrie liée au liage (cf. Faghiri & Samvelian, 2015), dans le cas où celle-ci serait irréfutable par les faits. Nous nous attacherons à démontrer le contraire par la suite.

En résumé, bien que nous suivions Karimi pour dire que l’acceptabilité des phrases comme (34-b) est discutable, nous rejetons l’idée que cette inacceptabilité soit liée à une contrainte grammaticale sur la base d’exemples comparables nettement plus acceptables.

Avant de passer à la présentation des données du persan, il y a lieu de s’attarder sur les différentes formes ou variantes de l’anaphore réciproque disponibles dans cette langue.

Les ouvrages sur la grammaire du persan s’accordent pour dire que le persan dispose des deux pronoms réciproques (Mahootian, 1997; Meshkatod Dini, 1990; Lazard et al., 2006; Perry, 2007; Windfuhr & Perry, 2009) : *ham-digar*<sup>38</sup>, *yek-digar*, glosés (en anglais) par *each other* et *one another* respectivement, ainsi que leurs variantes simplifiées *ham*<sup>39</sup>, qui “surtout dans la langue familière” s’utilise “au lieu de ces formes” particulièrement “après l’ezâfé ou une préposition” (Lazard et al., 2006, p. 107).

Il faut noter que dans les exemples construits discutés dans la majorité des travaux, l’on ne voit apparaître que la forme complexe *hamdigar/hamdige*. Or, l’on ne peut ignorer que, plus généralement dans les exemples attestés ou ceux se rapprochant le plus du langage courant, la forme le plus souvent utilisée après une

38. Dans le registre informel, et plus généralement à l’oral, *digar* ‘autre’ se prononce *dige*.

39. Précisons que cette forme connaît un homonyme, ‘aussi’, mais les contextes dans lesquels les deux mots risqueraient d’être confondus sont rares.

préposition est *ham*<sup>40</sup>. Par ailleurs, le cas échéant, ce dernier est glosé de la même manière que *ham-digar*, c'est-à-dire par 'l'un l'autre' ou *each other*<sup>41</sup>. De ce fait, nous n'hésitons pas à inclure les exemples avec *ham* pour appuyer nos propos<sup>42</sup>. Précisons cependant que nous prenons le soin d'écarter les combinaisons lexicalisées, telle que *bā ham* 'ensemble', dans lesquelles le statut du pronom réciproque ne va pas de soi.

Tout d'abord, il faut souligner que nous n'avons aucune difficulté à former des phrases ditransitives, parfaitement acceptables, avec l'anaphore réciproque en position du COI liée par un COD non marqué inanimé, qu'il soit quantifié, ex. (36) et (37), ou nu<sup>43</sup>, ex. (38) et ex. (39). Dans ce dernier exemple, le COD et le verbe forment un prédicat complexe, *doruq bāftan* 'mentir (Lit. mensonge tricoter)'. Nous avons également trouvé une multitude d'exemples attestés, dont ex. (40)<sup>44</sup>. Effectivement, les exemples sont encore plus nombreux avec la forme *ham*, ex. (41)<sup>45</sup>.

En revanche, il est incontestable qu'avec un COD non marqué humain, il est moins aisé de former des exemples jouissant d'une acceptabilité comparable, ainsi que de trouver des exemples attestés. Rappelons que les contextes autorisant de telles phrases ne sont pas abondants. Quoi qu'il en soit, la phrase (42), relevée sur l'Internet<sup>46</sup>, montre que, dans un contexte favorisant la lecture massive, même un

---

40. Nous n'avons pas mené d'étude quantitative à propos de cette observation, faute d'avoir accès à un corpus oral du persan. Toutefois, la comparaison entre le nombre de résultats trouvés sur Google pour chacune des trois variantes, *ham*, *hamdigar* et *hamdige*, est, à ce sujet, suffisamment éloquente. Par exemple, pour la séquence "*be ham/hamdigar/hamdige goft-and*", 'ils dirent les uns aux autres (Lit. à RÉCIP dire.PST-3PL)', susceptible de nous fournir une comparaison raisonnablement neutre et équilibrée entre les trois formes, les résultats sont respectivement 88 500, 9 210 et 93. Pour préciser, la forme standard de la conjugaison verbale explique le résultat plus élevé pour *hamdigar* par rapport à sa variante orale *hamdige*. En effet, lorsqu'on orthographie le verbe comme il est prononcé dans le registre informel, c'est-à-dire *goft-an*, la tendance s'inverse : 156 000, 373 et 67. Pour donner un autre exemple, comparer les résultats pour la séquence "*az - be*" ('de - à') : *ham* : 677 000, *hamdigar* : 13 600, *hamdige* : 23 900 (Date de consultation le 06/06/2016).

41. À titre d'exemple, la phrase suivante est donnée par Karimi (2003a, p. 5) :

(i) man emruz se=tā bačče=ro did-am ke bā ham da'vā  
 moi aujourd'hui trois=CLF enfant=RĀ voir.PST-1SG que avec l'un l'autre dispute  
 mi-kard-and  
 IPFV.faire.PST-3PL  
 'Aujourd'hui, j'ai vu trois enfants qui se disputaient les uns avec les autres (*were arguing with each other*).'

42. En effet, en registre informel et dans les syntagmes prépositionnels, cette variante paraît plus naturelle et, dans une bonne partie des cas, améliore considérablement l'acceptabilité des phrases, notamment lorsque l'antécédent est un COD nu.

43. L'on ne peut ignorer qu'avec les COD nus, le choix de la variante n'est pas neutre et que la forme *ham* permet d'obtenir des phrases nettement plus acceptables.

44. <http://www.noandishaan.com/forums/thread66514.html> [consulté le 07/06/2016].

45. <http://alef-lam-mim.blogfa.com/9210.aspx> [consulté le 07/06/2016].

46. <http://romanparisa.blogfa.com/post/10> [consulté le 07/06/2016].

COD nu humain peut lier l'anaphore réciproque, contrairement à ce que la phrase marquée d'un astérisque en (31) plus haut peut laisser croire<sup>47</sup>.

- (36) dar guše=ye otāq [čand=tā miz]<sub>i</sub> kenār/ru=ye hamdigar<sub>i</sub>  
 dans coin=EZ chambre quelque=CLF table à-côté/sur=EZ l'un l'autre  
 gozāšte bud-and  
 mettre.PP être.PST-3PL  
 'Ils avaient mis quelques tables les uns à côté des/sur les autres dans le coin  
 de la chambre.'
- (37) kar=aš faqat in ast ke [čand varaq kāqaz]<sub>i</sub> be  
 travail=3SG seulement ce être.PRS.3SG que quelque feuille papier à  
 hamdigar<sub>i</sub> mangane kon-ad  
 l'un l'autre agrafe faire.PRS-3SG  
 'Son travail consiste simplement àagrafer quelques feuilles de papier (les  
 unes aux autres).'
- (38) Maryam dust dāšt [kāqaz]<sub>i</sub> be ham<sub>i</sub> mangane  
 Maryam ami avoir.PST.3SG papier à l'un l'autre agrafe  
 kon-e  
 faire.PRS-3SG  
 'Maryam aimait agrafe des papiers les uns aux autres (agrafer du papier).'
- (39) Maryam hamiše [dorūq]<sub>i</sub> be ham<sub>i</sub> mi-bāf-e  
 Maryam toujours mensonge à l'un l'autre IPFV-tricoter.PRS-3SG  
 'Maryam accumule toujours les mensonges (Lit. tricote des mensonges les  
 uns aux autres).'
- (40) ... mi-bin-am [čand=tā sandoq]<sub>i</sub> kenār=e hamdige<sub>i</sub>  
 IPFV-voir.PRS-1SG quelque=CLF caisse à-côté=EZ l'un l'autre  
 gozāšt-an ...  
 mettre.PST-3PL  
 '... je vois qu'ils ont mis quelques caisses les unes à côté des autres...'
- (41) ... pāin=e miz=e daxl ham [čand guni berenj]<sub>i</sub> ru=ye  
 en-bas=EZ table=EZ recette aussi quelque sacs riz sur=EZ  
 ham<sub>i</sub> gozāšte bud-and ...  
 l'un l'autre mettre.PP être.PST-3PL  
 '... ils avaient également mis quelques sacs de riz en bas de la caisse'

47. De plus, les exemples comme celui illustré par (i) sont abondants. Mais, étant donné le fait qu'il sont en quelque sorte écartés de la théorie de liage, nous ne les discutons pas davantage. Cependant, ils ne sont pas sans intérêt comme cas de figure témoignant du fait qu'un COD non marqué peut servir d'antécédent pour une anaphore.

- (i) az sarmā [do=tā puluver]<sub>i</sub> ru=ye hamdigar<sub>i</sub> pušide bud-am  
 de froid deux=CLF pull-over sur=EZ l'un l'autre mettre.PP être.PST-1SG  
 'À cause du froid, j'avais mis deux pulls l'un sur l'autre.'

- (42) Lidya yek=i=ro mi-šnās-e ke [doxtar pesar]<sub>i</sub> be  
 Lidya un=INDF=RĀ IPFV-connaître.PRS-3SG que fille garçon à  
 ham<sub>i</sub> mo'arrefi mi-kon-e  
 l'un l'autre présentation IPFV-faire.PRS-3SG  
 'Lidya connaît quelqu'un qui présente des jeunes gens (Lit. filles (et) garçons) les uns aux autres.'

Ces données remettent en cause l'affirmation de Karimi et montrent qu'aucune contrainte syntaxique liée au trait [ $\pm$ spécifique] ne peut en soi expliquer l'inacceptabilité de (34-b), répété en (43) ci-dessous. Dans un contexte favorisant une lecture massive, c'est-à-dire avec un prédicat clairement atélique, même un COD nu animé est en mesure de lier une anaphore en position du COI.

- (43) \*man se=tā bačče be hamdige mo'arrefi kard-am  
 moi trois=CLF enfant à l'un l'autre présentation faire.PST-1SG  
 (Pour dire) 'J'ai présenté trois/des enfants les uns aux autres.'

Comment expliquer alors l'inacceptabilité de (43)? Nous pensons qu'elle peut s'expliquer sur la base de facteurs sémantico-conceptuels en rapport avec le marquage différentiel de l'objet. Rappelons que celui-ci n'est pas déclenché par un seul trait binaire et jouit d'un certain degré de variabilité sur laquelle une multitude de facteurs intervient, ceux-ci contribuant, d'une manière ou autre, à la saillance. Le caractère animé/humain du référent en est un, au sujet duquel aucun auteur n'en nie l'impact sur le marquage par =*rā*<sup>48</sup>. Le degré d'individuation et la distance ou la plénitude vis-à-vis du verbe sont d'autres facteurs mentionnés dans la littérature (cf. Lazard, 1982).

Du point de vue sémantique, l'anaphore réciproque implique, en règle générale, une lecture distributive de l'entité plurielle à laquelle réfère son antécédent, c'est-à-dire que les atomes constituant cette pluralité participent à l'événement de manière individuelle. Quand l'on présente deux individus l'un à l'autre, l'action de présentation a lieu deux fois (et à chaque fois l'individu présenté est différent). En effet, les analyses répandues isolent un marqueur de distributivité (caché ou manifeste, ex. *each* de *each other* de l'anglais) pour rendre compte du sémantisme des réciproques (Heim *et al.*, 1991)<sup>49</sup>.

48. À ce propos, il est opportun de rappeler la remarque de Ghomeshi (1997) quant à la variabilité de la portée des COD non marqués surlignant le rôle du caractère humain dans le marquage par =*rā* (voir §6.2.2.3, p. 233).

49. Il faut noter que l'interprétation des réciproques varie notamment en fonction du prédicat et du contexte, que la distributivité n'est pas toujours aussi forte que dans l'exemple ci-dessus et que, dans certains cas, la lecture collective peut être prépondérante (cf. Dalrymple *et al.*, 1998). Par exemple, pour que la phrase "les enfants se regardent les uns les autres" soit vraie, il n'est pas

Nous pensons que la lecture distributive impliquée par la réciproque est susceptible d'accroître la saillance du référent et, en conséquence, de favoriser son marquage différentiel. Notons que la distributivité est faible dans l'ensemble des (contre-)exemples que nous avons fournis, et que la lecture collective est privilégiée (voir la note de bas de page 49). Il est intéressant de noter que cette hypothèse est compatible avec l'observation, faite dans de nombreux travaux en sémantique formelle, selon laquelle, en position du COD, la lecture distributive est plus facilement accessible lorsqu'un SN (indéfini) est spécifique (voir par ex. Dobrovie-Sorin & Beyssade, 2012, p. 181).

Ainsi, dans les cas de figure comme (43), deux facteurs convergent et favorisent fortement le marquage différentiel par  $=r\bar{a}$ , au point que l'absence de  $=r\bar{a}$  rendrait la phrase difficilement acceptable.

Par ailleurs, l'acceptabilité limitée des phrases avec un COD nu, surtout avec un référent animé, liant l'anaphore réciproque, peut s'expliquer par le fait que la distributivité (de la réciproque) et la lecture massive (du COD nu) sont difficilement compatibles. De plus, s'il est vrai que la réciproque, dans son sens primaire ou son emploi prototypique, implique la lecture distributive, il est plausible pour la forme pleine/non simplifiée *hamdigar* de renforcer la lecture distributive, tandis que *ham*, du fait de locutions telles que *bā ham* 'ensemble', tend à la lecture collective.

#### 6.2.4 Autorisation des vides parasites

Browning & Karimi (1994) et Karimi (2003a) affirment qu'en persan, seuls les COD spécifiques (marqués de  $=r\bar{a}$ ) peuvent autoriser les vides parasites.

Dans le cadre de la théorie générative, "vides parasites" (*parasiting gaps*, Taraldsen, 1981; Engdahl, 1983) désigne la présence de deux positions de complément implicites ou vides (*empty categories*) coindexées dans une même phrase, dont la première est une trace laissée par un mouvement Qu-. Ainsi, le premier vide, la trace (*t*), est doté d'un antécédent dans la phrase, et le deuxième vide (*e*) parasite au premier, reçoit une interprétation identique à celui-ci (cf. Tellier & Valois, 2006, entre autres).

(44) Voilà les dossiers<sub>i</sub> que tu as résumés t<sub>i</sub> avant de classer e<sub>i</sub>

Adapté de Tellier & Valois (2006, p. 80)

Il faut noter que différentes analyses et contraintes sont disponibles, au sein de la

---

nécessaire que : 1) chaque enfant regarde un autre enfant et 2) chaque enfant soit regardé par un autre enfant. De plus, la phrase "les enfants jouent les uns avec les autres", hors contexte, signifie de prime abord que *les enfants jouent ensemble* et non *chaque enfant joue avec un autre*.

théorie générative, pour rendre compte des phénomènes associés à ces constructions (ex. Chomsky, 1986; Cinque, 1990; Tellier, 1991). Ne nous intéressant pas aux débats théoriques internes à ce sujet, nous nous efforcerons de discuter du fondement empirique de la contrainte (l'asymétrie) admise par les auteurs en ce qui concerne l'autorisation des vides parasites par les COD en persan, sans entrer dans les détails techniques. Par commodité, nous employons la terminologie générative pour faire référence à ces constructions.

Les exemples (45) et (46) sont donnés respectivement par Browning & Karimi (1994) et Karimi (2003a) pour illustrer cette asymétrie. Ces auteurs ne partagent ni l'analyse syntaxique de la position du COD marqué (cf. §6.1.2 et §6.1.1 respectivement) et, de surcroît, leurs analyses des phrases à vides parasites, qu'elles soient grammaticales ou non, ne sont pas identiques.

Toutefois, il n'est pas nécessaire pour nous d'entrer dans les détails des analyses esquissées dans ces travaux, dans la mesure où les auteurs s'accordent à dire que ce contraste est dû au trait  $[\pm\text{spécifique}]$  du COD. Au-delà de l'explication théorique fournie pour étayer telle ou telle analyse de ce phénomène, nous voudrions mettre en question la validité empirique de cette affirmation, dans la mesure où nous contestons les jugements de grammaticalité sur lesquels les auteurs se sont basés pour y parvenir.

- (45) a. Ali in ketāb=rā<sub>i</sub> [ba'd az in-ke —<sub>i</sub> xānd] t<sub>i</sub> be man  
 Ali ce livre=RĀ après de ce-que lire.PST.3SG à moi  
 dād  
 donner.PST.3SG  
 'Ali m'a donné ce livre après l'avoir lu.'  
 Browning & Karimi (1994, p. 82)
- b. \*Ali yek ketāb<sub>i</sub> [ba'd az inke —<sub>i</sub> xānd] be man t<sub>i</sub>  
 Ali un livre après de que lire.PST.3SG à moi  
 dād  
 donner.PST.3SG  
 'Ali m'a donné un livre après l'avoir lu.'  
 Browning & Karimi (1994, p. 85)
- (46) a. Kimea in ketāb=o<sub>i</sub> [qablaz in-ke —<sub>i</sub> be-xun-e] t<sub>i</sub> be man  
 Kimea ce livre=RĀ avant ce-que SBJV-lire.PRS-3SG à moi  
 dād  
 donner.PST.3SG  
 'Kimea m'a donné ce livre avant qu'elle le lise.'
- b. \*Kimea ketāb<sub>i</sub> [qablaz in-ke —<sub>i</sub> be-xun-e] be man t<sub>i</sub> dād  
 Karimi (2003a, p. 116)

Premièrement, en tant que locutrice native, nous acceptons la phrase marquée d'un astérisque en (45), et, deuxièmement, nous pensons que l'inacceptabilité de (46-b) vient du contraste entre les propriétés aspectuelles du verbe et les propriétés référentielles du COD nu. Notamment, l'aspect perfectif/accompli, dénotant un événement (en l'occurrence de type accomplissement) délimité dans le temps, n'est pas compatible avec la lecture cumulative requise par l'objet nu. Rappelons l'incompatibilité de ce dernier avec toute lecture finie (*bound*) abordée plus haut. En effet, comme l'illustre (47), l'acceptabilité de (46-b) s'améliore considérablement lorsque l'on élimine ce contraste – en favorisant la lecture non finie (*unbound*), par l'emploi de l'aspect imperfectif (voir par ex. Krifka, 1992, pp. 50–51), ex. (47-a), ou simplement en quantifiant le COD non marqué, ex. (47-b).

- (47) a. Kimea ketāb<sub>i</sub> [qablaz in-ke —<sub>i</sub> be-xun-e] hičvaqt (t<sub>i</sub>) be  
 Kimea livre avant ce-que SBJV-lire.PRS-3SG jamais à  
 man (t<sub>i</sub>) ne-mi-d-e  
 moi NEG-IPFV-donner.PRS-3SG  
 'Kimea ne me donne jamais de livres avant de les avoir lus.'
- b. Kimea ye ketāb<sub>i</sub> [qablaz in-ke —<sub>i</sub> be-xun-e] (t<sub>i</sub>) be man (t<sub>i</sub>)  
 Kimea un livre avant ce-que SBJV-lire.PRS-3SG à moi  
 dād  
 donner.PST.3SG  
 'Kimea m'a donné un livre avant qu'elle le lise.'

Dans (47-b) comme dans (45-b) ci-dessus, nous interprétons le COD non marqué quantifié, *ye(k) ketāb*, avec une lecture spécifique, c'est-à-dire à portée large (c'est-à-dire 'il existe un livre que Kimea m'a donné avant de le lire'), et acceptons parfaitement les phrases. À ce propos, il faut rappeler que nous n'adhérerons pas à la dichotomie [ $\pm$ spécifique] admise dans lesdits travaux quant à la présence (et l'absence) de  $=rā$ . En effet, comme discuté au chapitre 1 (§1.1.3), cette dichotomie a été mise en question dans une majorité de travaux<sup>50</sup>.

Cela va sans dire qu'en règle générale, les COD marqués, du fait de leur saillance sur le plan discursif, sont plus faciles à extraire et présentent davantage de résistance aux contraintes de l'îlot. De plus, rappelons qu'en persan la topicalisation d'un

---

50. Pour écarter toute hésitation, voici un exemple tiré de Ghomeshi (1997, p. 139), qui montre clairement la possibilité pour un COD non marqué de recevoir une interprétation spécifique :

- (i) ĵian mi-xāst                      **ye qalam=i**    peydā kon-e                      va peydā=**š**  
 jian IPFV-vouloir.PST.3SG un stylo=INDF trouvé faire.PRS-3SG et trouvé=3SG  
 kard  
 faire.PST.3SG  
 'Jian voulait trouver un stylo et il l'a trouvé.'

élément enchâssé par son extraction vers la position initiale va, en règle générale, de pair avec son marquage par l'enclitique  $=r\bar{a}$ , et cela indépendamment de sa fonction (cf. section 1.3). Cependant, dans les contextes où le COD non marqué bénéficie d'un certain niveau de saillance, l'extraction de ce dernier reste possible. Les exemples (49) à (51) ci-dessous présentent d'autres exemples de phrases que nous jugeons acceptables.

Par souci de cohérence méthodologique, afin d'évaluer cette affirmation de manière plus objective, nous avons recueilli les jugements d'acceptabilité des locuteurs natifs du persan pour ce type de phrases, via la méthode du questionnaire auto-administré en ligne, que nous avons déjà utilisée pour nos expériences précédentes sur l'ordre relatif (cf. §5.2). Nous présenterons brièvement cette expérience par la suite. Les items de cette expérience, dix phrases au total, ont été intégrés, en tant qu'items distracteurs, au questionnaire utilisé pour l'élicitation des jugements d'acceptabilité sur la coordination, que nous présenterons dans la sous-section suivante (cf. §6.2.5)<sup>51</sup>.

#### 6.2.4.1 Expérience d'élicitation de jugement d'acceptabilité

**Méthode.** La tâche de cette expérience suit le paradigme classique du jugement d'acceptabilité sur une échelle Likert de 1 (absolument inacceptable) à 7 (parfaitement acceptable). Cette échelle à nombre impair permet d'inclure le jugement neutre par le chiffre 4.



FIGURE 6.1 – Exemple d'échelle de Likert utilisée pour l'élicitation des jugements

Les chiffres sont présentés horizontalement dans l'ordre croissant de gauche à droite, dans des cases encadrées et suffisamment espacées, et le choix de la note s'effectue en cliquant dans la case correspondante, cf. Figure 6.1.

L'instruction donnée aux participants est de lire chaque phrase attentivement et d'indiquer à quel point ils la considèrent comme une phrase acceptable de la

51. Ce questionnaire est consultable à l'adresse suivante : <http://spellout.net/ibexexps/pegah/baaraan/experiment.html>.

langue persane, sachant qu'il ne s'agit pas d'évaluer les phrases selon les règles de la grammaire scolaire ou prescriptive, ni selon la validité de leur contenu, mais plutôt de décider si elles sont susceptibles d'être produites, en tant que phrases biens formées du persan, par des locuteurs natifs de cette langue<sup>52</sup>. Cette instruction comprend également des exemples de phrases pour lesquelles les notes maximum et minimum seraient attendues<sup>53</sup>. Il est également précisé que certaines phrases susciteraient plus d'avis divergents. Les participants sont alors priés de noter les phrases de manière spontanée, en faisant confiance à leur première réaction, sans y passer un temps excessif<sup>54</sup>.

Pour permettre au participant d'assimiler la tâche, une phase d'entraînement est prévue. Elle comporte 5 items couvrant différents degrés d'acceptabilité relevant aussi bien du registre formel qu'informel. Quand le participant termine la phase d'entraînement, il lui est demandé de cliquer sur "Continuer" dès qu'il est prêt pour démarrer le questionnaire. Ensuite, les phrases se présentent sur l'écran, chacune sur une page à part, qui se renouvelle dès que le participant clique pour en noter une.

**Stimuli.** Chaque item est décliné en deux versions, illustrées par la paire de phrases en (48), selon que la phrase comporte un vide parasite ou non. La version sans le vide parasite, ex. (48-a), correspond à notre condition de contrôle.

- (48) a. man **bastani sonnati**<sub>i</sub> [bā-inke xeyli -<sub>i</sub> dust dār-am] be  
 moi glace traditionnel bien que très ami avoir.PRS-1SG à  
 xāter=e diābet=am aslan na-bāyad -<sub>i</sub>  
 cause=EZ diabète=1SG absolument NEG-il-faut  
 bo-xor-am  
 SBJV-manger.PRS-1SG
- b. man [bā-inke **bastani sonnati**<sub>i</sub> xeyli dust dār-am] be  
 moi bien que glace traditionnel très ami avoir.PRS-1SG à  
 xāter=e diābet=am aslan na-bāyad -<sub>i</sub>  
 cause=EZ diabète=1SG absolument NEG-il-faut

52. Pour faciliter la compréhension des critères pour décider de cela et éviter toute confusion, nous avons ajouté qu'une phrase est mal formée, malgré le fait qu'elle soit plausiblement interprétante, lorsqu'en tant que locuteur natif du persan, on a toutes les raisons pour penser qu'elle n'est pas produite par un locuteur natif, mais plutôt par un apprenant ou un robot.

53. Il s'agit de deux exemples de phrases, l'un dans le registre formel et l'autre informel – ce dernier registre est sous entendu par l'écriture communément utilisée pour transcrire la langue orale dans son registre familier –, parfaitement bien formées malgré leur contenu contestable. Il y a également un exemple de phrase absolument mal formée, indépendamment du registre et de son contenu.

54. Comme dans nos questionnaires précédents (cf. chapitre 5), les participants sont informés que le temps qu'ils consacrent à chaque phrase est enregistré. Cela non seulement pour les amener à juger les phrases de manière la plus spontanée possible, mais aussi pour les encourager à se concentrer sur la tâche.

bo-xor-am

SBJV-manger.PRS-1SG

‘Il ne faut absolument pas que je mange de la glace traditionnelle à cause de mon diabète, bien que je l’aime beaucoup.’

Cette expérience comporte 10 items qui soit sont neutres au niveau du registre, comme l'exemple ci-dessus, soit relèvent du registre familier, comme les exemples donnés plus bas<sup>55</sup>. Ces 10 items sont combinés avec 50 autres items, dont 20 items expérimentaux de notre expérience sur la coordination et 30 autres items distracteurs (comprenant 15 phrases bien formées et 15 mal formées). La liste de l'ensemble de ces items se trouve dans les annexes (cf. Annexe C).

**Participants.** Au total 74 participants, locuteurs natifs du persan, ont rempli le questionnaire<sup>56</sup>. Toutefois, au final, pour avoir un groupe homogène, nous avons écarté les résultats de trois participants bilingues, ayant l'azéri comme langue maternelle, ainsi qu'un participant polyglotte déclarant trois langues comme ses langues dominantes.

Ainsi, nous présentons les données de 70 locuteurs natifs du persan, dont 49 femmes et 21 hommes, âgés de 19 à 64 ans (âge moyen de 37,5 ans et âge médian de 35 ans). En ce qui concerne la distribution géographique des participants, 61 d'entre eux vivent en Iran, dont 49 à Téhéran, et les 9 autres habitent depuis plus de deux ans à l'étranger, mais déclarent avoir le persan comme langue dominante.

**Résultats et discussions.** L'analyse statistique des données obtenues s'est faite par un modèle de régression mixte avec la note comme variable dépendante, la condition (contrôle, ex. (48-a), vs. cible, ex. (48-b)) comme variable indépendante, et les variables item et participant comme variables à effet aléatoire. À la différence des modèles précédemment utilisés, ici la variable dépendante n'est pas catégorielle et nous avons pu calculer un modèle de régression linéaire.

Les résultats du modèle<sup>57</sup> montrent, comme cela était attendu, que nos phrases de contrôle ont été jugées significativement meilleures que nos phrases cibles de 1,42 points (Erreur-Type=0,272,  $t=-5,237$ ). Ceci étant, les phrases cibles ont reçu

---

55. Pour rappel, ces items sont présentés à chaque participant dans un ordre aléatoire et suivant la procédure du carré latin, c'est-à-dire que chaque participant juge tous les 10 items expérimentaux, dont 5 dans la condition de contrôle (a) et 5 dans la condition cible (b), mais jamais les deux versions d'un même item.

56. Comme pour nos expériences précédentes, les participants ont été contactés via les réseaux sociaux et les listes de diffusion. Leur participation au questionnaire s'est faite sur la base de volontariat. Notre seul critère de sélection a été qu'ils soient locuteurs natifs du persan et d'avoir le persan comme langue dominante.

57. Voir Annexe D.2.1 pour la présentation des résultats complets de ce modèle.

une note moyenne de 4,477 (IC<sub>95%</sub> :  $\pm 0,207$ ) qui, bien que relativement basse, est numériquement et significativement supérieure au jugement neutre, c'est-à-dire la note de 4 ( $t=4,843$ ,  $df=369$ ,  $p<0,0001$ ). Par ailleurs, la note médiane obtenue pour ces phrases est de 6<sup>58</sup>.

Cette note moyenne est, en outre, significativement supérieure à la note moyenne des phrases indiscutablement mal formées de notre questionnaire<sup>59</sup> ( $t=21,0764$ ,  $df=369$ ,  $p<0,0001$ ), qui s'élève à 2,38 (IC<sub>95%</sub> :  $\pm 0,109$ ). Cela témoigne du fait que les locuteurs ne considèrent pas ces phrases cibles comme des phrases mal formées.

Sans surprise, nous observons un degré important de variation parmi nos items expérimentaux, absente dans la condition de contrôle. Cela est particulièrement probant quand on considère la distribution de la fréquence des notes dans les deux conditions pour chaque item, cf. Figure 6.2.

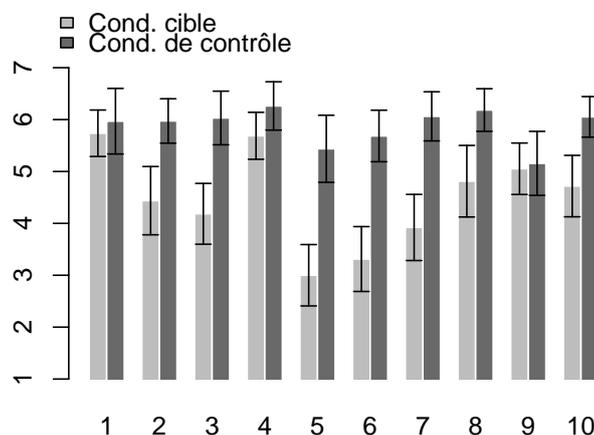


FIGURE 6.2 – Exp. Vide parasite : Moyennes et intervalles de confiance à 95% par item

Les phrases ci-dessous, correspondant respectivement aux items 1, 4 et 9 dans

58. La note moyenne des phrases de contrôle est de (IC<sub>95%</sub> :  $\pm 1,65$ ) avec la note médiane de 7.

59. Nous avons 15 phrases mal formées parmi nos distracteurs (voir Annexe C pour la liste), dont les deux exemples ci-dessous. La première phrase, qui viole une contrainte d'accord en nombre entre le verbe et le sujet et ne manifeste pas l'ordre canonique, a reçu une note moyenne de 2,284 (IC<sub>95%</sub> :  $\pm 0,308$ ). La seconde, qui souffre de plusieurs anomalies, a reçu une note moyenne de 1,189 (IC<sub>95%</sub> :  $\pm 0,129$ ).

- (i) \*parking agar xāli šod-**and** mi-tavān-id vāred šav-id  
 parking si vide devenir.PST-**3pl** IPFV-pouvoir.PRS-2PL entrée devenir.PRS-2PL  
 (vs. agar parking xāli šod mi-tavān-id vāred šav-id)
- (ii) \*Mohsen dār-ad lāzem bā-inke ziād na-gereft  
 Mohsen avoir.PRS-3SG nécessaire bien-que beaucoup NEG-prendre.PST.3SG  
 (vs. Mohsen bā-inke (?) lāzem dār-ad ziād na-gereft)

la Figure 6.2, ont toutes reçu la note médiane de 7, et une note moyenne supérieure à 5, qui est numériquement, mais pas significativement, inférieure à la note reçue dans la condition de contrôle.

- (49) ... man tā be hāl na-šode šalvār<sub>i</sub> [bedun=e inke -<sub>i</sub>  
 moi jusque à présent NEG-devenir.PP pantalon sans=EZ que  
 porov kon-am] -<sub>i</sub> be-xar-am  
 essai faire.PRS-1SG SBJV-acheter.PRS-1SG  
 ‘... il n’est jamais arrivé que j’achète un pantalon sans l’avoir essayé.’  
Note moyenne : 5,737 (IC<sub>95%</sub> : ± 0,448)
- (50) xob pālto<sub>i</sub> [age -<sub>i</sub> ne-mi-puš-e] barā=š -<sub>i</sub>  
 alors manteau si NEG-IPFV-mettre.PRS-3SG pour=3SG  
 na-xar-im  
 NEG-acheter.PRS-1PL  
 ‘... alors, s’il ne met pas de manteau, ne lui en achetons pas un.’  
Note moyenne : 5,688 (IC<sub>95%</sub> : ± 0,436)
- (51) **film=e tarsnāk<sub>i</sub>** [bā-inke aslan -<sub>i</sub> dust na-dār-am]  
 film=EZ effrayant bien que absolument ami NEG-avoir.PRS-1SG  
 daste-jam’i bad=am ne-mi-ā-d -<sub>i</sub> negāh  
 ensemble mauvais=1SG NEG-IPFV-venir.PRS-3SG regard  
 kon-am  
 faire.PRS-1SG  
 ‘Bien que je n’aime pas les films d’horreur, cela ne me dérange pas d’en  
 regarder en compagnie.’  
Note moyenne : 5,053 (IC<sub>95%</sub> : ± 0,496)

À ce stade, nous ne pouvons formuler d’hypothèse pour expliquer cette divergence. Mais, il nous semble que les notes reflètent la facilité avec laquelle, dans chaque phrase, le COD déplacé peut recevoir une lecture saillante (du point de vue discursif), par exemple comme focus contrastif ou comme topic. À ce sujet, la discussion abordée par Goldberg (2006, ch. 7) sur les contraintes de l’îlot est également très éclairante. Par ailleurs, comme nous l’avons rappelé plus haut, le marquage par =*rā* est un moyen dont dispose les locuteurs du persan pour augmenter la saillance d’un élément lors de son déplacement vers la position initiale.

Un autre point qui mérite réflexion est le fait que le persan permet la non réalisation des arguments (cf. §1.1.5.1). Considérons les exemples (52) et (53) ci-dessous, qui sont tout à fait acceptables<sup>60</sup> :

60. Les jugements d’acceptabilité d’un ensemble de 25 locuteurs natifs du persan ont été recueillis pour ces deux phrases, dans une étude pilote, selon les mêmes méthodes que pour l’expérience présentée ici. Elles ont reçu respectivement des notes moyennes de 5,84 et de 6,64.

- (52) [s<sub>1</sub> Omid ahl=e sigār hast] vali [s<sub>2</sub> jolo=ye  
 Omid habitué=EZ cigarette être.PRS.3SG mais devant=EZ  
 bābā=š [hattā age xeyli ham (~~sigār~~) havas karde  
 père=3SG même si beaucoup aussi cigarette envie faire.PP  
 bāš-e] (~~sigār~~) ne-mi-keš-e]  
 être.SBJV-3SG cigarette NEG.IPFV-tirer.PRS-3SG  
 'Omid fume/aime fumer mais, il ne fume pas (de cigarettes) devant son père  
 même s'il en a beaucoup envie.'
- (53) [s<sub>1</sub> man qalyun dust dār-am] vali [s<sub>2</sub> [vaqti tu jam'  
 moi narguilé ami avoir.PRS-1SG mais quand dans collectivité  
 (~~qalyun~~) dast-be-dast mi-š-e] [xoš-am  
 narguilé main-à-main IPFV-devenir.SBJV-3SG plaisir=1SG  
 ne-mi-ā-d [(~~qalyun~~) be-keš-am]]]  
 NEG-IPFV-venir-PRS-3SG narguilé SBJV-tirer.PRS-1SG  
 'J'aime le narguilé mais je n'aime pas le fumer quand (l'on fume collective-  
 ment et que) l'on le passe de main en main.'

Ces phrases ne peuvent être analysées en termes de vides parasites, étant donné qu'aucun mouvement ne peut y être postulé. Comparons à ce sujet (54-a) et (54-b). La différence entre ces derniers exemples semble porter sur leur structure informationnelle. Cette comparaison renforce l'idée que l'acceptabilité relative et graduée de nos phrases (cibles) soit liée à des préférences purement discursives, c'est-à-dire le fait que l'extraction d'un élément enchâssé vers la position initiale doit être justifiée sur le plan discursif. Quoi qu'il en soit, la question demeure ouverte et requiert davantage d'investigation.

- (54) a. [s<sub>1</sub> Omid ahl=e sigār hast] vali [s<sub>2</sub> (~~Omid~~) jolo=ye bābā=š [hattā age  
 xeyli ham (~~sigār~~) havas karde bāš-e] (~~sigār~~) ne-mi-keš-e]  
 b. [s<sub>1</sub> Omid [sigār]<sub>i</sub> jolo=ye bābā=š [hattā age xeyli ham -<sub>i</sub> havas karde  
 bāš-e] -<sub>i</sub> ne-mi-keš-e]

En résumé, les résultats de ce recueil de jugements d'acceptabilité appuient notre désaccord avec l'affirmation de Karimi. Toutefois, il ne s'agit que d'une étape préliminaire. Entre autres choses, il ne permet pas d'étudier dans quelle mesure l'acceptabilité des phrases avec un COD non marqué diffère de celles avec un COD marqué ou de déterminer quels facteurs sont susceptibles d'intervenir dans leur acceptabilité. En effet, notre objectif, à ce stade, n'était que de garantir l'objectivité des jugements d'acceptabilité rapportés. Une étude expérimentale approfondie s'impose

pour répondre à l'ensemble des questions soulevées par ces constructions.

### 6.2.5 Coordination

La coordination constitue généralement un test fiable pour déterminer l'identité de la position syntaxique des constituants coordonnés. En effet, il est généralement admis que la coordination est possible à la seule condition que les éléments coordonnés soient structurellement analogues. Karimi (2003a) utilise ce test pour montrer que les COD marqué et non marqué sont syntaxiquement différents. Elle considère que leur coordination est agrammaticale, alors qu'il est possible de coordonner des COD marqués, d'un côté, et des COD non marqués, de l'autre, ex. (55).

- (55) a. man diruz in aks=ro va in ketāb=ro xarid-am  
 moi hier ce photo=RĀ et ce livre=RĀ acheter.PST-1SG  
 'Hier, j'ai acheté cette photo et ce livre.'
- b. man diruz aks va ketāb xarid-am  
 moi hier photo et livre acheter.PST-1SG  
 'Hier, j'ai acheté un/des livre(s) et un/des photo(s).'
- c. \*man diruz in aks=ro va ketāb xarid-am  
 moi hier ce photo=RĀ et livre acheter.PST-1SG  
 Karimi (2003a, p. 103)

Nous contestons cette observation (cf. également Samvelian, 2001) et considérons la coordination des différents types de COD acceptable quand le contexte s'y prête, comme cela est illustré par ex. (56).

- (56) a. barā=ye bačče-ha medād-rangi yek daftar=e naqqāši va in  
 pour enfant-PL crayon de couleur un cahier=EZ dessein et ce  
 ketāb-hā=rā xarid-am  
 livre-PL=RĀ acheter.PST-1SG  
 'J'ai acheté des crayons de couleur, un cahier à dessin et ces livres-là pour les enfants.'
- b. diruz dar bāqevahš čand šir=e darande va bozorg-tarin  
 hier dans zoo quelques lion=EZ dévorant et grand.SPRL  
 yuzpalang=e donyā=rā did-im  
 guépard=EZ monde=RĀ voir.PST-1PL  
 'Hier au zoo, nous avons vu des lions féroces et le plus grand guépard du monde.'

Par ailleurs, nous acceptons une version améliorée de (55-c) en réponse, par exemple, à la question "qu'est-ce que tu as acheté hier?", en remplaçant *va*, 'et', par sa contrepartie clitique =o, qui convient au registre familier sous-entendu par

l'emploi de =ro, et en insérant un adverbe comme *faqat* 'seulement', ex. (57).

- (57) diruz faqat in aks=ro=o ketāb xarid-am  
 hier seulement ce photo=RĀ=ET livre acheter.PST-1SG  
 'Hier, j'ai seulement acheté cette photo et des livres.'

Nul ne peut nier que la coordination des différents types de COD – nu, indéfini/quantifié et marqué –, qui, du fait de leur niveau de détermination différent ne peuvent avoir le même statut discursif, n'est pas chose aisée. Celle-ci, toutefois, comme le montrent les exemples ci-dessus, n'est pas impossible.

Toujours avec le même souci de cohérence méthodologique, nous avons aussi mené une expérience d'élicitation de jugement d'acceptabilité que nous présentons par la suite.

### 6.2.5.1 Expérience d'élicitation de jugement d'acceptabilité

Dans cette expérience, nous étudions l'acceptabilité de la coordination de deux COD. Une possibilité, en persan, pour paraphraser une telle construction est d'utiliser une coordination elliptique, comme en (58), que nous appelons par commodité postposition. Afin de pouvoir évaluer l'acceptabilité de la coordination de deux COD, nous incluons cette possibilité comme un facteur expérimental.

- (58) a. Maryam ketāb va daftar xarid  
 Maryam livre et cahier acheter.PST.3SG  
 'Maryam a acheté des livres et des cahiers.'
- b. Maryam ketāb xarid va daftar  
 Maryam livre acheter.PST.3SG et cahier  
 'Maryam a acheté des livres et (elle a acheté) des cahiers.'

Cette expérience est menée via le même questionnaire que pour notre expérience sur les vides parasites, présenté dans la sous-section précédente. Voir §6.2.4.1 pour les détails sur la méthode et les participants.

**Stimuli.** Chaque item est décliné en 5 versions.

Premièrement, une version avec un seul COD, le COD marqué de =rā, nous sert de condition de contrôle, ex. (59-a). Ensuite, nous avons quatre autres conditions qui correspondent à la manipulation de deux paramètres expérimentaux : 1) construction : coordination (ex. (59-b) et (59-c)) vs. postposition (ex. (59-d) et (59-e)), et 2) ordre : COD<sub>[+Ra]</sub><COD<sub>[-Ra]</sub> (ex. (59-b) et (59-d)) vs. COD<sub>[-Ra]</sub><COD<sub>[+Ra]</sub> (ex. (59-c) et (59-e)). Ce dernier facteur semble intervenir dans l'acceptabilité de ces coordinations, dans la mesure où les coordonnées suivant l'ordre COD<sub>[+Ra]</sub>-COD<sub>[-Ra]</sub>

nous semblent plus difficiles à accepter que les coordonnées présentées dans l'ordre inverse.

- (59) a. barāye sabtenām kāfi ast [form=e takmil-šode=rā]  
 pour inscription suffisant est formulaire=EZ complet=RĀ  
 barāye mā ersāl kon-id  
 pour nous envoyer faire.PRS-2PL  
 'Pour s'inscrire, il suffit de nous envoyer le formulaire rempli.'
- b. ... [yek qat'e aks va form=e takmil-šode=rā] ...  
 un pièce photo et formulaire=EZ complet=RĀ  
 '... de nous envoyer une photo et le formulaire rempli.'
- c. ... [form=e takmil-šode=rā va yek qat'e aks ] ...  
 formulaire=EZ complet=RĀ et un pièce photo  
 '... de nous envoyer le formulaire rempli et une photo.'
- d. ... [yek qat'e aks] barāye mā ersāl kon-id va  
 un pièce photo pour nous envoyer faire.PRS-2PL et  
 [form=e takmil-šode=rā]  
 formulaire=EZ complet=RĀ  
 '... de nous envoyer une photo et le formulaire rempli.'
- e. ... [form=e takmil-šode=rā] barāye mā ersāl kon-id  
 formulaire=EZ complet=RĀ pour nous envoyer faire.PRS-2PL  
 va [yek qat'e aks]  
 et un pièce photo  
 '... de nous envoyer le formulaire rempli et une photo.'

Cette expérience comporte 20 items, dont 4 relèvent du registre formel. Notons que, dans tous les phrases, la conjonction de coordination, *va*, est écrite comme un mot séparé de façon à ce que la coordination soit clairement visible dans toutes les phrases indépendamment du registre<sup>61</sup>.

Par ailleurs, dans chaque item nous avons choisi les deux COD de manière à pouvoir éviter, dans la condition 3, toute ambiguïté sur la portée de =*rā*, plus précisément, de telle manière que =*rā* ne peut être interprété que comme portant seulement sur le dernier COD.

Les 20 items expérimentaux sont combinés avec 40 items distracteurs, dont 10 items expérimentaux de notre expérience d'élicitation de jugement d'acceptabilité précédente (cf. §6.2.4.1).

**Résultats et discussions.** Les moyennes générales de la note obtenue dans la condition de contrôle, la coordination et la postposition sont les suivantes, respecti-

61. Dans les phrases relevant du registre familier, les COD sont choisis de telle manière que la conjonction de coordination ne puisse s'attacher à l'élément la précédant. Rappelons que, dans le registre familier, ce dernier se comporte comme un enclitique et se prononce *o*.

vement : 5,996 (IC<sub>95%</sub> :  $\pm 0,185$ ), 5,414 (IC<sub>95%</sub> :  $\pm 0,147$ ) et 5,386 (IC<sub>95%</sub> :  $\pm 0,139$ ), avec les notes médianes de 7 pour la condition de contrôle et 6 pour les deux autres.

L'analyse des données par la modélisation linéaire mixte<sup>62</sup>, avec la note comme variable dépendante, les variables item et participant comme variables à effet aléatoire, et, dans un premier temps, une seule variable indépendante avec les trois conditions ci-dessus montre que la note moyenne pour la condition de contrôle, 5,983, est significativement meilleure que pour les deux autres, dans l'ordre, de 0,57 (Erreur-Type=0,123,  $t=-4,675$ ) et 0,59 (Erreur-Type=0,113,  $t=-5,213$ ) points<sup>63</sup>. Or, l'écart entre les deux demeure peu élevé et, étant donné que les notes moyennes de ces deux conditions sont (significativement) supérieures à 5 ( $t_s=5,70$   $df=591$ ,  $p_s<0,0001$ ), nous pouvons les considérer comme acceptables.

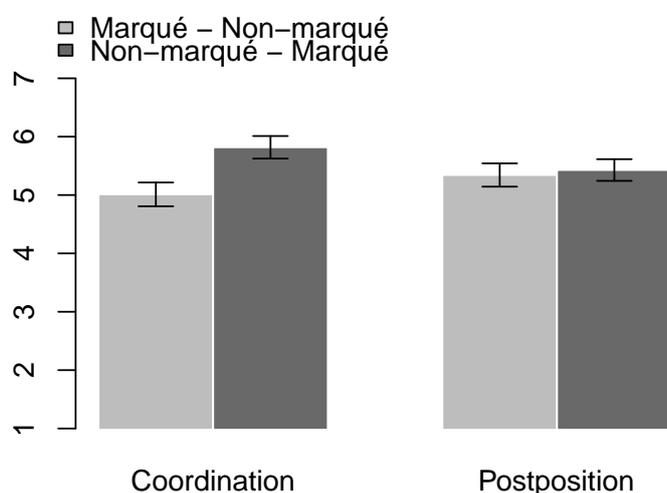


FIGURE 6.3 – Exp. Coordination : Moyennes et intervalles de confiance à 95%

Pour étudier l'effet de l'ordre sur l'acceptabilité de ces phrases, nous avons calculé un modèle linéaire mixte, sur un sous-ensemble des données, avec nos deux facteurs expérimentaux comme variables indépendantes en excluant la condition de contrôle. Les résultats de cette analyse montrent que l'ordre des COD dans ces constructions a un effet significatif sur l'acceptabilité de la phrase ( $t=-6,00$ ), mais seulement dans le cas d'une coordination, cf. Figure 6.3. En effet, l'interaction entre les deux facteurs est significative ( $t=-5,157$ ). Les coordonnées qui suivent l'ordre  $COD_{[-Ra]}-COD_{[+Ra]}$  ont une meilleure acceptabilité (+0,40 points) que leurs contreparties dans l'ordre inverse. Toutefois, l'estimation de la note moyenne des coordonnées dans ce dernier

62. Voir Annexe D.2.2 pour la présentation des résultats complets.

63. Mais, comme ces chiffres le montrent, il n'y a qu'une différence numérique entre les notes moyennes obtenues dans ces deux conditions.

ordre demeure relativement élevée (5,0).

En résumé, ces résultats confirment clairement que les COD marqué et non marqué peuvent apparaître ensemble dans une coordination. Ce fait est difficilement compatible avec une différence entre ces deux types de COD sur le plan syntaxique, telle qu'affirmée par Karimi (2003a).

- (60) morāje'e-konande-gān=e mohtaram be tālār-hā serfan  
 visiteur-PL=EZ respectable à amphi-PL seulement  
 mi-tavan-and qalam čand barg kāqaz va rāyāne=ye  
 IPFV-pouvoir.PRS-3PL crayon quelques feuille papier et ordinateur=EZ  
 dasti hamrāh dāšte bāš-and  
 portable avec avoir.PP être.SBJV-3PL  
 'Les visiteurs respectables des amphes ne peuvent avoir sur eux que de quoi  
 écrire (crayons), quelques feuilles de papier et un/des ordinateur(s) por-  
 table(s).'

Il est vrai que notre expérience concerne seulement la coordination entre un COD marqué et un COD non marqué indéfini, mais cela suffit en soi pour soutenir l'affirmation mentionnée ci-dessus. Davantage de recherche apparaît nécessaire pour aborder un éventail plus complet des coordonnées mixtes et pour étudier leurs conditions d'acceptabilité. Enfin, il est important de noter que la coordination de différents types de COD non marqués est également possible, comme cela est illustré par l'exemple attesté<sup>64</sup> en (60), et, à notre connaissance, ne fait pas l'objet de controverse.

### 6.2.6 Conclusion

Nous avons passé en revue, dans les sous-sections §6.2.1 à §6.2.5, les diagnostics – outre celui relatif à l'ordre canonique – proposés dans la littérature en faveur de l'asymétrie syntaxique entre un COD marqué et un COD non marqué. Nous avons examiné la validité empirique de chacun par la prise en compte d'un éventail relativement large de faits linguistiques. Pour ce faire, nous nous sommes appuyée sur les travaux existants, des données attestées, ainsi que sur deux recueils de jugements d'acceptabilité.

Aucun de ces diagnostics ne s'est montré probant et n'a pu s'imposer à l'issue de cet examen critique. Le premier problème dont souffre une partie de ces diagnostics est de confondre les COD non marqués indéfinis et nus, alors que le fait qu'ils divergent sur des aspects sémantiques et syntaxiques non triviaux est incontestable.

64. <http://www.nlai.ir/Default.aspx?tabid=828> [consulté le 08/06/2016].

Le second problème, concernant en particulier les arguments relatifs à la portée et au liage, provient de la vision théorique et méthodologique de la théorie générative chomskyenne et de la place privilégiée qu'y occupent les contraintes (catégoriques) syntaxiques. En effet, un certain nombre de travaux sur la portée et le liage dans différentes langues, ainsi que les données du persan que nous avons apportées ici, soulignent l'insuffisance des analyses en termes exclusivement de structure syntaxique lorsque des facteurs fonctionnels rentrent indiscutablement en ligne de compte.

Enfin, s'appuyer sur son intuition pour décider de l'(a)grammaticalité à partir d'un ensemble réduit d'exemples peut constituer une méthodologie non fiable. *A contrario*, recueillir des jugements d'acceptabilité d'un échantillon de locuteurs natifs, selon les protocoles standards et reproductibles en vigueur, permet d'obtenir des données dont la fiabilité est plus assurée, sur la base desquelles il est possible de faire des analyses plus solides du point de vue empirique.

### 6.3 Structure hiérarchique vs. structure plate

Les travaux en syntaxe formelle sur le persan, notamment dans le cadre de la grammaire générative chomskyenne, postulent une structure hiérarchique pour le syntagme verbal, dans laquelle le COD marqué de  $=r\bar{a}$  occupe une position plus élevée que sa contrepartie non marquée, hypothèse que nous avons désignée par celle de la PSDO dans ce chapitre.

Cette hypothèse rappelle le *Mapping Hypothesis* de Diesing (1992), qui, à partir des données d'un certain nombre de langues, postulait une règle d'appariement sémantique (universelle), selon laquelle, en simplifiant, les COD (ainsi que les sujets) indéfinis occupant une position à l'intérieur du  $V''$ , correspondant au domaine existentiel, reçoivent une lecture existentielle non spécifique, tandis que ceux occupant une position externe reçoivent une lecture spécifique (voir également Diesing & Jelinek, 1995)<sup>65</sup>.

Or, l'examen critique exhaustif de l'ensemble des arguments avancés en faveur de l'hypothèse de la PSDO, que nous avons effectué dans cette thèse, démontre que le recours à une structure syntagmatique asymétrique, dans laquelle le COD marqué occupe une position plus élevée que sa contrepartie non marquée, ne s'impose pas comme une nécessité empirique.

En outre, l'hypothèse de la PSDO fait des prédictions erronées en ce qui concerne les préférences de l'ordonnancement linéaire des constituants dans les constructions

---

65. Rappelons que, pour une majorité de ces travaux, le marquage par  $=r\bar{a}$  est déclenché par le trait  $[\pm\text{spécifique}]$ .

ditransitives, que nous avons pu établir dans cette thèse à travers un ensemble d'études quantitatives conduites dans la lignée des travaux clés sur le sujet sur des langues telles que l'anglais, l'allemand ou le japonais.

Les résultats de ces études contredisent clairement l'hypothèse d'un ordre canonique dichotomique pour la phrase ditransitive. De surcroît, ils suggèrent que, si une analyse syntagmatique hiérarchique du syntagme verbal est à maintenir, il faudrait que l'opposition soit basée sur la nudité au lieu du marquage par  $=r\bar{a}$ , ou bien postuler trois positions au lieu de deux en croisant ces deux facteurs. Or, aucune de ces solutions ne nous semble satisfaisante, étant donné tout d'abord le fait que les différents types du COD peuvent être coordonnés. Par ailleurs, les COD nus "lourds", c'est-à-dire comportant des ajouts, qui, sur le plan purement syntaxique ont strictement la même nature que les COD nus "monolexémiques", se distinguent de ces derniers par leur autonomie relative par rapport au verbe, qui se reflète clairement dans leur ordonnancement linéaire. En effet, le poids grammatical, en termes de nombre de mots, ainsi que d'autres facteurs extra-syntaxiques, sont sources de variation dans l'ordonnancement relatif des objets.

Par conséquent, en dépit de son intérêt théorique, nous rejetons l'hypothèse de la PSDO au profit d'une structure syntagmatique plate du syntagme verbal, qui permet une meilleure couverture empirique, notamment en rapport avec l'ordre linéaire dans les constructions ditransitives. Une structure plate du  $V''$  n'impose aucune contrainte syntaxique sur la précérence des compléments verbaux et permet de rendre compte des préférences dans l'ordonnancement linéaire par des contraintes, qu'il est désormais coutumier en français d'appeler préférentielles (*soft constraints*), dont nous avons discuté dans la deuxième partie de la thèse (cf. section 5.5).

La question reste de savoir si, au-delà du syntagme verbal, les données du persan sont compatibles avec une structure syntagmatique plate, ex. (61-b), ou hiérarchisée de la phrase, supposant une asymétrie entre le sujet et les autres compléments du verbe, ex. (61-a).



Cette question revient à nous demander si le persan est une langue configurationnelle ou non<sup>66</sup>. Nous tentons d'y apporter quelques éléments de réponse par la

66. Une précision s'impose. Il va sans dire que le persan n'est pas une langue strictement non configurationnelle, dans le sens d'avoir à montrer l'absence de toute structure syntagmatique hiérarchisée. En effet, bien que l'ordre des constituants immédiats d'une phrase soit flexible, la structure

suite.

La littérature sur l'ordre des mots distingue deux types de langue vis-à-vis de la structure de la phrase : langue configurationnelle vs. langue non configurationnelle (cf. §2.2.1.3 au chapitre 2). Le premier type se réfère aux langues manifestant un ordonnancement relativement fixe des mots, dans lesquelles les relations syntaxiques s'expriment par la syntaxe des positions, c'est-à-dire par une relation structurale asymétrique. Le second se réfère aux langues qui manifestent un ordonnancement relativement libre des mots, reflétant souvent une organisation pragmatique de la phrase, et s'appuient principalement sur "l'accord" pour exprimer les relations syntaxiques (Austin, 2001)<sup>67</sup>.

En persan, comme nous l'avons observé dans le premier chapitre de cette thèse (cf. section 1.3), l'ordre canonique SOV peut être bouleversé et tous les cas de figure possibles (SVO, OSV, etc.) peuvent se produire. Par ailleurs, contrairement à une langue comme le français, les relations syntaxiques en persan ne se définissent pas par la syntaxe des positions. Comparer (63) et (62). Le changement de l'ordre, toutes choses égales par ailleurs, entraîne, en français, le bouleversement des rôles grammaticaux, car celui-ci en est l'unique manifestation. En revanche, les deux phrases (63-a) et (63-b) du persan ont le même sens vériconditionnel, de même que les phrases (63-c) et (63-d). Ceci résulte du fait que le marquage du COD par =*rā*, et non l'ordre linéaire, indique les relations syntaxiques dans ces phrases.

- (62) a. Maryam aime Amir.  $\equiv$  Aimer(M,A)  
 b. Amir aime Maryam.  $\equiv$  Aimer(A,M)
- (63) a. Maryam Amir=*rā* dust dār-ad  
 Maryam Amir=*RĀ* ami avoir.PRS-3SG  
 $\equiv$  Aimer(M,A)  
 b. Amir=*rā* Maryam dust dār-ad  
 Amir=*RĀ* Maryam ami avoir.PRS-3SG  
 $\equiv$  Aimer(M,A)  
 c. Amir Maryam=*rā* dust dār-ad  
 Amir Maryam=*RĀ* ami avoir.PRS-3SG  
 $\equiv$  Aimer(A,M)

---

des syntagmes tels que N'', P'' et C'' est incontestablement configurationnelle dans cette langue. De plus, même au sein du syntagme verbal, certains types de compléments, notamment les compléments phrastiques, ont un placement fixe.

67. Notons que l'accord est employé ici comme un terme général, incluant tout marquage qui rappelle une relation de dépendance, y compris le marquage différentiel de l'objet. Il est plus judicieux de parler d'alignement par *flagging*, c'est-à-dire le marquage (morphologique ou adpositionnel) du dépendant (en l'occurrence des dépendants verbaux), et/ou par indexation (*indexing*), c'est-à-dire la manifestation de l'accord sur la tête (en l'occurrence le verbe) – Creissels (2006) emploie les termes marquage vs. indexation –, par opposition à un alignement par ordre des mots.

- d. Maryam=rā Amir dust dār-ad  
 Maryam=RĀ Amir ami avoir.PRS-3SG  
 ≡ Aimer(A,M)

Une autre observation faite au sujet des langues non configurationnelles est l'usage extensif des arguments nuls ou implicites (cf. Hale, 1983), ce qui est également le cas en persan, comme nous l'avons vu auparavant (cf. §1.1.5.1).

Malgré ces observations, comme nous l'avons vu au premier chapitre, la majorité des études en syntaxe formelle adhère à la vision hiérarchisée de la phrase en persan. La justification la plus notable pour une telle vision est proposée par Karimi, dans sa monographie, intitulée *A Minimalist Approach to Scrambling : Evidence from Persian* (2005), dédiée à l'étude du *scrambling* dans le cadre de la théorie minimaliste. Or, comme nous l'avons vu auparavant, l'argumentation qui y est articulée est sensiblement ancrée de point de vue théorique et les faits sont abordés dans le seul objectif d'argumenter en faveur d'une certaine vision du *scrambling*.

L'argument essentiel avancé par Karimi s'appuie sur l'asymétrie des relations du liage et de la portée :

Persian is configurationnal since the c-command relation of the phrasal elements have an impact on binding and scope relations.

Karimi (2005, p. 39)

Nous ne reviendrons pas sur la question de la portée et l'analyse qu'en fait Karimi (2005), que nous avons largement abordées dans la section précédente (cf. §6.2.2). Nous nous contentons de rappeler le lien controversé entre la portée et la structure syntaxique ainsi que le fait que l'interprétation de portée est une question complexe et toute analyse fiable de cette question nécessite une investigation approfondie par les méthodes expérimentales.

Quoi qu'il en soit, nous sommes d'avis que les données existantes sur l'interprétation de portée en persan ne permettent pas d'établir une asymétrie structurale entre les dépendants verbaux. Pour corroborer ce point de vue, nous répétons en (64), l'exemple, (29-b) ci-dessus, donné par Ghomeshi (1997), qui, selon l'auteur, présente l'ambiguïté de portée des quantifieurs, contrairement aux affirmations de Karimi (2003a, 2005). Dans cet exemple, le COD peut recevoir une portée large ainsi qu'une portée étroite sur le sujet. Cette ambiguïté ne peut, en effet, recevoir une explication en termes de c-commande asymétrique.

- (64) hame=ye mo'alle-m-ā ye šāgerd=i=ro mo'arrefi kard-and  
 tout=EZ instituteur-PL un élève=INDF=RĀ introduction faire.PST-3PL  
 'Tous les instituteurs présentèrent un élève.'

Adapté de Ghomeshi (1997, p. 139)

En ce qui concerne les relations de liage, il importe, tout d'abord, de noter qu'elles n'impliquent pas forcément une relation structurelle asymétrique, car elles peuvent s'expliquer sur la base d'autres considérations, notamment en termes de hiérarchie des rôles syntaxiques, observée de manière générale dans les langues du monde (Creissels, 2006, p. 83).

Depuis son émergence, la théorie du liage a dû évoluer du fait de la remise en cause de sa couverture empirique (voir Jackendoff, 1992; Pollard & Sag, 1992; Reinhart & Reuland, 1993). En effet, des exemples contredisant ces principes abondent. À titre d'exemple, la définition des principes A et B implique que les anaphores et les pronoms soient mutuellement exclusifs, ex. (65-a), cf. §6.2.3.1. Or, il existe des configurations dans lesquelles l'emploi de l'une ou l'autre forme est également grammatical, ex. (65-b) à (65-e)<sup>68</sup>.

- (65)
- a. Marie parle à \*elle/elle-même.
  - b. Marie a vu une vidéo d'elle/elle-même sur l'Internet.
  - c. Marie aime les blagues sur elle/elle-même.
  - d. Marie a tiré la couverture sur elle/elle-même.
  - e. Marie a trouvé une place pour Jean près d'elle/elle-même.

Pour accommoder ces contre-exemples, plusieurs définitions non structurelles de la notion du liage – basée grossièrement sur la hiérarchie des rôles grammaticaux, appelée également la hiérarchie de l'oblicité (Keenan & Comrie, 1977) – ont vu le jour non seulement dans le cadre des théories alternatives non transformationnelles, comme le LFG ou le HPSG, mais aussi dans le cadre du modèle chomskyen (ex. Reinhart & Reuland, 1993). Ces approches se sont renforcées notamment par les données qui montrent que, même dans le contexte des structures plates, l'on peut observer une asymétrie dans les relations du liage. À titre d'exemple, Kiss (1994, pp. 236–237) affirme qu'en hongrois, il existe des cas de figure où, indépendamment de l'ordre respectif, le sujet peut lier un objet, mais pas l'inverse. Or, cela ne s'explique pas par la structure syntagmatique de la phrase, car, selon l'auteur, celle-ci est plate, et, par conséquent, l'on ne peut établir une relation c-commande asymétrique entre ces constituants.

Force est de constater que la vision configurationnelle du liage ne fait pas l'una-

---

68. Ces données sont à l'origine de l'hypothèse des "anaphores exemptées" (*exempte anaphors*) (Pollard & Sag, 1992), que l'on traite souvent comme logophoriques (Reinhart & Reuland, 1993) et non anaphoriques, dans la mesure où l'on considère qu'elles n'obéissent pas à la théorie du liage mais aux principes discursifs et pragmatiques, notamment le point de vue du locuteur.

nimité. Nous adhérons également à la vision fonctionnelle, qui permet de rendre compte des données des langues typologiquement différentes par le même appareil. Enfin, il est à noter que les relations asymétriques du liage servent de moins en moins d'argument pour déterminer la configuration syntagmatique d'une langue. Ainsi, bien que nous ne remettons pas en cause les données présentées par Karimi (2005, ch. 5.3) en faveur des relations asymétriques du liage, nous ne sommes pas pour autant encline à en rendre compte par le biais d'une relation de c-commande asymétrique.

En résumé, ni les relations du liage, ni les ambiguïtés de portée ne permettent de trancher en faveur d'une vision hiérarchique de la structure de la phrase en persan. Par la suite, nous apportons quelques arguments supplémentaires en faveur de la structure plate :

- L'existence de locutions idiomatiques impliquant le sujet (*subject idioms*), considéré comme un reliquat de la structure plate de la phrase (cf. par ex. Kiss, 1994). De telles locutions ne sont pas rares en persan. En effet, dans la formation des prédicats complexes, notamment les combinaisons nom-verbe, le nom peut aussi occuper la fonction du sujet par rapport au verbe. Voici quelques exemples d'une telle combinaison : *sarmā zadan* 'être attaqué par le froid (Lit. froid frapper)', *barq gereftan* 'recevoir une décharge électrique, être électricisé (Lit. électricité prendre)', ex. (66)<sup>69</sup>.

(66) Ali=rā barq gereft  
 Ali=RA électricité prendre.PST.3SG  
 'Ali a reçu une décharge électrique.'

- L'effet dit de supériorité, *superiority effect*, dans les interrogatives à mots Qu- multiples, un diagnostic pour la présence de l'asymétrie sujet/objet, est absent en persan, comme illustré par la paire d'exemples en (67) empruntée à Lotfi (2003), citée par Adli (2010).

(67) a. ki čī=ro xord ?  
 qui quoi=RA manger.PST.3SG  
 b. čī=ro ki xord ?  
 quoi=RA qui manger.PST.3SG  
 'Qui a mangé quoi?'

Pour conclure, l'ensemble des observations et faits linguistiques abordés ici nous

---

69. L'on peut trouver plus de combinaisons de ce type, ainsi que des exemples attestés pour chacune dans la base de données PersPred (cf. Samvelian & Faghiri, 2013) à l'adresse suivante : [www.perspred.cnrs.fr](http://www.perspred.cnrs.fr).

conduit à nous joindre à Bonami & Samvelian (2015) pour préférer une structure syntagmatique plate pour la phrase en persan.

# Chapitre 7

## Considérations sur les préférences de l'ordre des mots

Ce dernier chapitre est consacré à l'ancrage de nos conclusions relatives à l'étude des préférences de l'ordre linéaire, au sein de la recherche à travers les langues du monde présentée au chapitre 2 en première partie de la thèse.

Dans la première section, nous synthétiserons les conclusions relatives aux différents facteurs cognitivo-fonctionnels susceptibles d'intervenir dans l'ordonnement linéaire des constituants que nous nous sommes proposée d'investiguer dans ce travail. Dans la seconde section, nous nous pencherons plus particulièrement sur l'effet du poids grammatical, c'est-à-dire la préférence "long-avant-court" en persan, et procéderons à une évaluation des différentes hypothèses évoquées pour en rendre compte.

### 7.1 Les facteurs influant sur l'ordre linéaire

Dans le deuxième chapitre de cette thèse, nous avons détaillé les approches existantes de l'étude de la variation de l'ordre des mots, en particulier les approches cognitivo-fonctionnelles, qui se démarquent par leur rigueur méthodologique et leur couverture empirique poussée (cf. section 2.2). Ces approches rejoignent les travaux en psycholinguistique et en sciences cognitives portant sur les mécanismes de la production et du traitement de la parole en temps réel, au sein desquels la question de la variation de l'ordre linéaire reçoit une attention particulière.

Après avoir passé en revue les différents facteurs mis en avant dans l'ensemble de ces travaux (cf. §2.3.1), nous nous sommes proposée d'en étudier un certain nombre, énumérés dans le chapitre 3 (cf. §3.2.2.3). Par la suite, nous présentons nos observations et conclusions pour chacun d'entre eux en vue d'apprécier les prédictions

disponibles dans la littérature. Celles-ci ont été détaillées dans le chapitre 2.

### 7.1.1 Définitude

Nos études successives soulignent le rôle prépondérant du degré de détermination (du COD), que l'on peut représenter, pour simplifier, par le trait de la définitude sur l'ordre relatif. La définitude constitue une des dimensions de l'accessibilité conceptuelle mise en avant pour favoriser le placement plus en amont de la phrase.

Dans notre étude, nous nous sommes penchée particulièrement sur la définitude du COD, pour lequel celle-ci constitue le seul trait à déclencher systématiquement le marquage différentiel, c'est-à-dire la présence de l'enclitique  $=r\bar{a}$  (cf. §1.1.3). Le COD marqué de  $=r\bar{a}$  correspond ainsi au degré le plus élevé de la définitude pour un SN dans cette position en persan, le degré le plus bas étant occupé par les COD nus, ne portant aucune marque de détermination. Les COD (non marqués) indéfinis se trouvent dans une position intermédiaire.

Nous avons pu constater dans nos données du corpus que, dans les phrases transitives pour lesquelles l'ordre SOV s'impose comme ordre canonique, les rares cas de l'ordre OSV se produisent avec des COD marqués de  $=r\bar{a}$ .

Le rôle de ce facteur est encore plus prépondérant dans les phrases ditransitives, qui manifestent davantage de variation et pour lesquelles la définitude se démarque comme la contrainte la plus forte pour déterminer l'ordre relatif entre les deux objets. Conformément à ce qui est attendu, plus le COD est défini, plus il tend à précéder le COI<sup>1</sup>.

Notons néanmoins, que, dans nos études, la définitude ni du COI, ni du sujet n'a pas été prise en compte en tant que telle. Or, il sera intéressant de voir dans quelle mesure celle-ci est susceptible d'intervenir sur l'ordre. Cela est d'autant plus intéressant que nous savons déjà que l'animéité de ces deux constituants intervient activement sur leur ordre relatif (voir §7.1.4). Cela pourra faire l'objet de recherches à venir.

### 7.1.2 Statut informationnel

En règle générale, l'importance de la structure informationnelle ou discursive de la phrase sur l'ordre linéaire est incontestable en persan (cf. §1.3.1).

Nous venons d'aborder dans la sous-section précédente le rôle du degré de la détermination (du COD) sur l'ordre linéaire. Or, ce dernier, comme nous l'avons vu dans le chapitre 2 (cf. §2.3.1.4), est étroitement lié avec la structure informationnelle

---

1. Voir la Figure 5.13 du ch. 5 (p. 196).

de la phrase, notamment en ce qui concerne la dimension de l'accessibilité du référent dans le discours (*Referential Givenness* en anglais), voir §2.3.1.2. Effectivement, les COD de type nu et indéfini ne peuvent pas correspondre au statut "donné", tandis qu'un COD marqué de  $=r\bar{a}$  le peut. Aussi, la configuration discursive de la phrase est également sensible à cet enclitique, dans la mesure où il peut également marquer la topicalité, ou plus généralement la saillance discursive.

En dehors de ces considérations, nous avons tenté de nous pencher sur ce facteur dans nos deux premières expériences (D1 et D2), qui, excepté la catégorie du COD, ont une conception et des stimuli (presque) identiques. Nous avons manipulé le statut discursif du COI (c'est-à-dire son accessibilité discursive, *givenness* en anglais) selon que son référent était présent dans le contexte précédant la phrase cible ou non (donné vs. nouveau), sachant que le COD, du fait de sa catégorie, était nouveau dans le discours. Nous n'avons pas trouvé d'effet clair nous permettant de nous prononcer sur ce facteur. Ces résultats non concluants peuvent toutefois être dus à notre paradigme expérimental. Rappelons également qu'il est plausible qu'en raison de la position basse du COI (par rapport au COD) sur la hiérarchie des rôles grammaticaux, son statut discursif joue un rôle moindre sur cet ordre relatif. Quoi qu'il en soit, des expériences relevant des paradigmes mieux adaptés pour détecter l'effet du statut informationnel dans les recherches ultérieurement permettront d'approfondir ce point.

### 7.1.3 Poids grammatical

Une des conclusions fondamentales de notre étude est de valider l'effet du poids grammatical sur l'ordre relatif entre les constituants, en termes de longueur relative en nombre de mots – ou plus précisément d'unités syntaxiques. Cet effet, quand il apparaît, correspond sans exception à la préférence "long-avant-court", mais il n'est pas systématique dans tous les cas de figure que nous avons examinés.

Tout d'abord, nous n'avons trouvé aucun effet de la longueur relative entre le sujet et le COD<sub>[marqué][humain]</sub> dans les phrases transitives dans nos données expérimentales (Exp. T1), sachant que, dans notre étude de corpus, les occurrences dans l'ordre OSV sont trop rares pour permettre d'étudier l'effet d'un quelconque facteur.

De plus, dans les phrases ditransitives, nous n'avons trouvé aucun effet (ni numérique, ni significatif) de la longueur dans nos données de corpus quant à l'ordre relatif entre le COI et un COD marqué de  $=r\bar{a}$ , pour lesquels l'ordre OD-OI-V se produit à 95%. Mais, il reste encore à tester cela par une expérience contrôlée<sup>2</sup>.

---

2. Il est intéressant de noter que dans notre expérience D3, dans laquelle la longueur (COD = COI) et l'animéité (COD<sub>[inanimé]</sub> et COI<sub>[humain]</sub>) sont contrôlées, le taux de l'ordre OI-OD-V

Dans ce même type de phrases, nous trouvons pour les COD non marqués, et de surcroît inanimés, un effet de la longueur identique dans nos données expérimentales et nos données du corpus.

Pour conclure, bien que nos données confirment l'effet de la longueur relative sur l'ordre linéaire, elles soulignent le fait que cet effet n'est pas systématique et dépend d'autres facteurs, tels que le rôle grammatical des constituants, le degré de détermination et l'animéité. Des expériences complémentaires permettraient de tester cette hypothèse afin de déterminer la façon exacte dont la longueur interagit avec les autres facteurs.

Nous reviendrons plus longuement sur cet effet dans la section suivante, où nous débattrons des différentes analyses proposées pour rendre compte de la préférence "long-avant-court" dans les langues (S)OV.

#### 7.1.4 Hiérarchie de la personne

Une autre observation essentielle, que nous avons pu relever à partir de nos études expérimentales, se rapporte à la hiérarchie de la personne. Le persan est une langue à ordre flexible qui n'impose pas de contrainte syntaxique sur l'ordre et, de ce fait, permet de contribuer aux débats sur la façon dont l'animéité est susceptible d'influencer l'ordre linéaire (cf. §2.3.1.3). Nos résultats mettent en évidence l'effet direct de l'animéité sur l'ordre linéaire correspondant à la préférence "animé-avant-inanimé".

Les données de notre expérience sur les phrases transitives (Exp. T1) permettent de confirmer l'effet de l'animéité sur l'ordre linéaire correspondant, comme attendu, à la préférence "animé-avant-inanimé". Dans cette expérience, où le COD est toujours marqué de  $=r\bar{a}$  et animé, c'est-à-dire occupant une position haute sur la hiérarchie d'accessibilité conceptuelle – par l'accumulation de deux traits convergents de la définitude et de l'animéité –, le taux de l'ordre OSV augmente significativement lorsque le sujet est inanimé. Cela confirme le fait que l'animéité peut influencer sur l'ordre linéaire de manière directe et indépendamment de son rôle sur l'appariement des rôles grammaticaux.

Par ailleurs, dans nos expériences sur les phrases ditransitives, bien que les variables indépendantes n'incluent jamais l'animéité, nous avons pu constater des variations dans les résultats que l'on peut attribuer à ce facteur. La comparaison entre

---

pour ces COD est plus élevé par rapport aux données du corpus (16,3% vs. 5%). Cette différence peut laisser présager la possibilité d'une certaine variation liée à ces deux derniers facteurs, mais la part de la variation liée à l'animéité est plausiblement plus prépondérante que celle liée à la longueur, car la préférence "animé-avant-inanimé" pourrait avoir favorisé l'ordre OI-OD-V dans cette configuration.

nos deux expériences sur les COD nus (D2 et D4) est particulièrement éclairante à ce sujet.

Nos données montrent que, pour un COD nu correspondant à un argument thème inanimé, lorsque le COI est un argument locatif inanimé, le taux de l'ordre OD-OI-V est de 28,2% (et peut atteindre un taux quasiment double si l'on y ajoute deux modificateurs adjectivaux). Or, avec le même type de COD nu, dans une autre configuration où le COI est un argument bénéficiaire humain, le taux de l'ordre OD-OI-V ne va pas au-delà de 14,5%, sans effet de la longueur, et descend jusqu'à 4,5% lorsque le COI est modifié par une subordonnée relative.

De plus, le taux relativement élevé de l'ordre OI-OD-V pour les COD marqués dans notre expérience D3, dans laquelle le COD correspond toujours à un argument thème inanimé et le COI à un argument bénéficiaire humain, par rapport aux données du corpus pour la même catégorie (16,3% vs. 5%), peut laisser présager une préférence "animé-avant-inanimé".

Toutefois, le caractère animé dans ces expériences est naturellement lié à la structure argumentale du verbe et, en conséquence, la question de savoir si ces variations sont liées à l'animéité en tant que telle ou à la hiérarchie des rôles sémantiques reste à étudier.

### 7.1.5 Dépendance sémantique

Nous avons étudié la dépendance sémantique dans notre étude de corpus sur les phrases ditransitives par le biais du lien collocationnel que le verbe peut entretenir avec ses compléments. Nous constatons, conformément à ce qui est attendu, que le lien collocationnel favorise l'adjacence. Toutefois, nous n'avons pas pu isoler l'effet de ce facteur, qui, du fait notamment de la façon dont nous étions amenée à l'opérationnaliser, est en forte corrélation avec le type de COD.

La dépendance sémantique entre le verbe et ses compléments est en soi un sujet stimulant en persan, qui, rappelons-le, disposant d'un nombre limité de verbes simples, s'appuie essentiellement sur les prédicats complexes pour exprimer les concepts verbaux. Les composantes d'un prédicat complexe, entretenant un lien syntaxique similaire à celui entre un verbe et son complément, n'apparaissent pas systématiquement dans des positions adjacentes ; mais la question de la séparabilité de celles-ci demeure sujet à controverse dans la littérature.

Certains auteurs affirment qu'en dehors de la classe d'éléments susmentionnée, l'insertion d'un constituant syntaxique entre les deux composantes d'un prédicat complexe est impossible. D'autres proposent une version plus faible de cette affirmation en soumettant l'éloignement de l'élément nominal d'un prédicat complexe

du verbe à des contraintes syntaxiques et/ou discursives, non requises pour un complément nominal ordinaire. Bien que ces affirmations soient remises en cause par d'autres auteurs, contre-exemples attestés à l'appui, cette controverse perdure dans la littérature (cf. Samvelian, 2012).

Il serait intéressant dans le cadre de futures recherches de tenter d'apporter des approfondissements empiriques à ce débat. Par exemple, dans la continuité de notre travail, il s'agirait de tester dans quelle mesure l'élément nominal d'un prédicat complexe se différencie d'un complément verbal nominal ordinaire en ce qui concerne son ordre relatif par rapport au COD.

### 7.1.6 Biais lexical du verbe

Nous avons pu détecter, à travers l'ensemble de nos études (expérimentales en particulier), un biais exercé par le lemme verbal sur l'ordre linéaire. Nous n'avons pas étudié ce facteur en tant que tel, mais nos expériences, dans l'ensemble, ont des conceptions équilibrées pour le lemme verbal. Nous avons pu, en conséquence, évaluer le biais lexical, par la comparaison entre le taux moyen et les taux individuels pour chaque verbe. Non seulement nous relevons des écarts significatifs de certains verbes par rapport à la moyenne, mais constatons aussi que le penchant que peut manifester un verbe pour un ordre donné est constant tout au long des expériences. Cela nous permet d'affirmer que les verbes peuvent avoir une prédisposition pour un ordre donné de leur compléments, même dans une langue à verbe final.

## 7.2 La préférence “long-avant-court” en persan et son impact théorique

Une des conclusions les plus importantes de nos études est de confirmer la préférence “long-avant-court” dans le domaine préverbal en persan.

Parmi les facteurs influant sur l'ordre linéaire, le débat sur l'effet de la longueur ou le poids grammatical reste vif et suscite le plus de controverses. En premier lieu parce que la préférence observée pour différentes langues SOV contrevient à une hypothèse bien ancrée et largement répandue sur le système de production de la parole, selon laquelle, rappelons-le, les éléments apparaissent dans l'ordre décroissant de leur accessibilité ; les éléments les plus courts étant plus simples (moins complexes et plus accessibles), l'on s'attend à ce qu'ils apparaissent avant les éléments les plus longs lorsque l'ordre n'est pas contraint par la grammaire (cf. §2.2.3, p. 80). Pour résumer, les théories de la production basée sur l'accessibilité prédisent une

préférence “court-avant-long” universelle.

La préférence “long-avant-court” observée pour le persan, comme pour d’autres langues SOV comme le japonais, le coréen ou le basque, contredit les prédictions de cette théorie. Comme nous l’avons vu au chapitre 2, cette préférence est prédite par des modèles orientés vers le traitement (*parsing* ou *processing*) et la compréhension. Ces modèles se concentrent également sur la complexité, mais cela du point de vue du nombre de mots à analyser avant d’identifier l’ensemble des constituants immédiats (Hawkins, 2004) ou la distance entre le verbe et ses dépendants (Gibson, 2000), et prédisent que l’ordre, qui permet de minimiser cette distance, obtient une meilleure efficacité dans le traitement et sera donc préféré aux ordres alternatifs (cf. §2.3.2.1 et §2.3.2.2 respectivement).

Le persan n’est pas une langue (uniformément) tête finale et nous avons vu que le principe *EIC* (et sa version plus récente *MiD*) de Hawkins (2004), à l’inverse du japonais, ne prédit pas la préférence “long-avant-court” pour le persan, mais une légère préférence “court-avant-long” (voir §3.2.3.1). Le modèle *DLT* (Gibson, 2000), en revanche, prédit une préférence “long-avant-court” pour le persan comme pour le japonais. Il en va de même pour l’hypothèse de Wasow (2013), qui explique la préférence “long-avant-court” dans les langues à verbe final par une tendance à éviter les constituants complexes au milieu de la phrase (*center-embedded structures* en anglais), plus difficiles à analyser.

Une seconde approche disponible pour expliquer la préférence “long-avant-court” dans une langue SOV est l’hypothèse de l’accessibilité conceptuelle de Yamashita & Chang (2001) (cf. §2.3.2.3), se situant du point de vue de la production. Pour rappel, selon cette hypothèse, les constituants plus longs, tout en étant plus complexes (au niveau formel), sont plus accessibles sur le plan conceptuel. Cela conduit à ce que, dans une langue flexible comme le japonais, dans laquelle, contrairement à une langue rigide comme l’anglais, la production de la parole est plus sensible aux facteurs conceptuels qu’aux facteurs formels, les constituants plus longs tendent à précéder les constituants plus courts. Cette analyse explique également la préférence “long-avant-court” en persan qui a en commun avec le japonais tous les paramètres sur lesquels les auteurs fondent leur argumentaires (cf. §3.2.3.3).

Cette conclusion est d’autant plus importante que les travaux plus récents sur le japonais (Tanaka et al., 2011b, pp. 125–126) et sur le basque (Ros et al., 2015, p. 1168), toutes deux uniformément à tête finale, écartent l’hypothèse d’accessibilité conceptuelle au profit de la théorie de Hawkins (1994, 2004). Par la suite, nous mentionnons d’autres conclusions de notre étude qui permettent de corroborer l’hypothèse de l’accessibilité conceptuelle pour le persan.

Comme nous l'avons vu dans la première section de ce chapitre, nos données montrent que l'effet de la longueur n'est pas systématique en persan et interagit avec d'autres facteurs, tels que le rôle grammatical des constituants, la définitude, ou encore l'animéité (cf. §7.1.3). Les études susmentionnées, mettant en cause l'hypothèse de l'accessibilité conceptuelle, n'abordent pas l'effet croisé de la longueur avec des facteurs tels que la définitude et l'animéité. Du point de vue de l'efficacité du traitement et de la compréhension, les préférences de l'ordre linéaire liées à la longueur sont censées s'appliquer indépendamment de ces facteurs, dans la mesure où ces modèles prennent en compte seulement la distance entre les constituants<sup>3</sup>. Ainsi, l'on s'attendrait à observer une tendance (générale, indépendante de l'animéité ou la définitude) à déplacer les constituants longs vers la gauche – dans les langues OV – ou vers la droite – dans les langue VO.

L'accessibilité conceptuelle connaît différentes dimensions dont le rôle sémantique, la fonction grammaticale, l'animéité, l'accessibilité du référent (dans le discours), la définitude, etc. Si l'on admet que la longueur (en nombre d'unités lexicales) en fait également partie, il est judicieux de considérer que cette dimension intervient avec une ampleur moindre que les autres dimensions. Ainsi, l'on s'attendrait à ce que l'effet de la longueur relative sur l'ordre linéaire soit minoré par d'autres dimensions de l'accessibilité conceptuelle plus prépondérantes.

Les données du persan sont compatibles avec cette seconde observation. Tout d'abord, il importe de noter que, dans notre étude, les effets de ces différentes dimensions de l'accessibilité conceptuelle convergent et correspondent tous à une préférence de placer les constituants les plus accessibles plus en amont de la phrase.

Par ailleurs, nos résultats montrent que la longueur relative a un effet significatif sur l'ordre dans les configurations où les constituants concernés occupent une position (relativement) basse sur les autres échelles de l'accessibilité conceptuelle. À l'inverse, dans le cas où les autres dimensions jouent déjà en faveur d'un positionnement plus en amont de la phrase, la longueur relative ne manifeste pas ou peu d'effet.

Plus précisément, nos données montrent que l'ordre relatif entre le COD et le COI n'est pas affecté par la longueur lorsque le COD est marqué de  $=r\bar{a}$ , c'est-à-dire lorsqu'il occupe la position la plus haute sur l'échelle de la définitude (et donc celle de l'accessibilité du référent également). Tandis que dans les cas de figure, où le COD occupe une position basse ou intermédiaire sur ces échelles, la longueur manifeste un effet significatif. C'est également le cas en ce qui concerne l'ordre relatif entre le

---

3. N'omettons pas de mentionner que ces approches n'affirment pas non plus que la préférence liée à la longueur soit l'unique force explicative des variations de l'ordre linéaire. Toutefois, dans ces approches, l'effet de la longueur est formulé indépendamment des autres facteurs.

sujet et le COD, lorsque le COD est à la fois marqué de  $=r\bar{a}$  et animé.

Ces observations soulignent la pertinence de l’hypothèse de l’accessibilité conceptuelle, au moins en ce qui concerne les données du persan. Nous nous joignons ainsi à Yamashita & Chang (2001) pour affirmer qu’en persan, l’ordre linéaire dans lequel les constituants apparaissent est sensible à l’accessibilité conceptuelle des constituants. Celle-ci inclut, à côté des autres facteurs reconnus consensuellement dans la littérature, la longueur en termes d’unités lexicales, laquelle, en augmentant la richesse lexicale et sémantique du constituant, contribue à sa saillance<sup>4</sup>.

Pour finir, précisons que nos conclusions ne tendent pas à écarter les explications orientées vers la compréhension et le traitement, en dehors des prédictions du principe de *EIC/MiD* clairement contestées par nos données. Bien au contraire, nous sommes d’avis que, comme pour les autres langues, les deux explications sont complémentaires pour le persan. Cette complémentarité est particulièrement appréciable quand l’on compare les préférences de l’ordre linéaire dans le domaine préverbal et dans le domaine postverbal (que nous n’avons pas pu aborder dans le cadre de cette thèse). En effet, il semble que la postposition des constituants suit une préférence selon laquelle les constituants les plus longs sont à placer après le verbe (cf. Rasekhmahand & Ghiasvand, 2015). Cela constitue une possibilité pour éviter un constituant complexe au milieu de la phrase, et, par conséquent, cette préférence converge avec l’hypothèse de Wasow (2013). Elle converge également avec les prédictions des modèles *DLT* (Gibson, 2000) et *EIC/MiD* (Hawkins, 2004).

Les recherches ultérieures devront approfondir les divergences des préférences liées à la longueur dans les deux domaines. Le persan offre, en effet, un terrain propice pour tester les prédictions de ces hypothèses dans les deux domaines de manière simultanée, que des langues comme le japonais ou comme l’anglais ne permettent pas.

Par ailleurs, le paradigme expérimental sur lequel nous nous sommes basée dans nos expériences se situe du point de vue de la production. Il serait intéressant de vérifier, dans les recherches futures, par des paradigmes se situant de côté de la compréhension et du traitement, tels que la lecture à vitesse autogérée (*self-paced reading*) ou le mouvement oculaire (*eye-tracking*), dans quelle mesure la longueur ou plus globalement la complexité formelle des constituants peut affecter le traitement en persan. Cela devrait également permettre d’étudier la façon dont ces deux forces,

---

4. À ce sujet, il est intéressant de mentionner que d’autres travaux plus récents en psycholinguistique, portant notamment sur la production des expressions référentielles, insistent également sur l’effet de l’accessibilité liée à la longueur, qu’ils expliquent par le fait que les informations supplémentaires de ces constituants résultent en une meilleure représentation de ceux-ci dans la mémoire du sujet parlant (Karimi *et al.*, 2014; Karimi & Ferreira, 2016).

à savoir, “optimiser la production” et “faciliter le traitement”, interagissent pour façonner les préférences de l'ordonnancement linéaire.

# Conclusion

La question de l'ordre des mots est fondamentale en syntaxe du fait de son lien avec la question de la structure et celle de la constituance. L'ordre des constituants dans une phrase et ses variations constituent une des données primaires sur laquelle se base la syntaxe formelle pour déterminer la structure syntagmatique de la phrase dans une langue donnée. Cet enjeu est d'autant plus crucial dans une langue à ordre flexible qui permet un degré certain de variation dans l'ordonnement linéaire de ses constituants de base.

Cette question se trouve ainsi au cœur des travaux qui se sont penchés sur la structure syntagmatique du persan que nous avons détaillés dans le chapitre 6, ainsi que dans la section 1.4 du chapitre 1. Ceux-ci portent une attention particulière à l'ordre relatif entre le COD et le COI dans les constructions ditransitives, du fait du marquage différentiel de l'objet, présenté dans le premier chapitre (§1.1.3), qui est une des propriétés les plus étudiées du persan et qui continue encore de susciter des débats controversés. Les analyses les plus répandues sur la structure du syntagme verbal persan affirment que le COD marqué de  $=r\bar{a}$  n'occupe pas la même position syntaxique que sa contrepartie non marquée. L'argument essentiel sur lequel reposent ces analyses est l'hypothèse (MDO) largement admise de l'existence d'un ordre canonique dichotomique dans la phrase ditransitive dépendant du marquage différentiel de l'objet : les COD marqués précéderaient, tandis que les COD non marqués suivraient le COI.

Or, en dépit du caractère crucial de cette hypothèse, elle n'avait jusqu'à présent pas fait l'objet d'études empiriques approfondies permettant d'en évaluer la validité empirique de façon la plus objective possible. Dans cette thèse, nous avons pris cette question à bras-le-corps et avons mené une étude quantitative extensive portant aussi bien sur les phrases ditransitives que sur les phrases transitives, dans la lignée des travaux fondateurs en syntaxe formelle portant sur l'étude de l'alternance, à propos de laquelle nous avons proposé une revue dans le chapitre 2. Cette étude quantitative a été réalisée en mobilisant conjointement deux sources de collecte de données, à savoir le corpus et les expériences psycholinguistiques, détaillées respectivement dans

les chapitres 4 et 5. Cette étude a montré clairement que l'hypothèse MDO, malgré le consensus dont elle fait l'objet, n'est pas empiriquement valide.

Un des apports majeurs de ce travail, basé sur un ensemble d'études successives, est que le marquage par  $=r\bar{a}$  n'est pas le critère adéquat pour décrire l'ordre relatif entre les deux compléments verbaux. Les COD non marqués n'ont pas un comportement homogène quant à l'ordre relatif "canonique" vis-à-vis du COI. Parmi ces derniers, la présence (ou l'absence) de détermination constitue le critère décisif : les COD non marqués qui portent une détermination ou une quantification ont un comportement plus proche des COD marqués de  $=r\bar{a}$  que des COD non marqués nus ne portant aucune détermination. Ainsi, il est d'ores et déjà clair que si un critère binaire doit être maintenu pour définir l'ordre relatif par défaut, il doit porter sur la "nudité" (c'est-à-dire l'absence de détermination) et non sur le marquage par  $=r\bar{a}$ . Néanmoins, nous ne préconisons pas un critère binaire pour décrire l'ordre relatif dans les phrases ditransitives. Les conclusions de nos différentes études convergent pour affirmer que la meilleure description de l'ordre relatif entre les deux compléments verbaux s'obtient par le croisement des traits du marquage et de la nudité, que nous représentons par le degré de détermination ou la définitude du COD. Nous pouvons résumer cette description comme suit : plus le COD est défini et/ou déterminé, plus il tend à s'éloigner du verbe. Ce trait étant en lien direct avec sa saillance dans le discours et avec son autonomie sémantique par rapport au verbe, cette description de l'ordre peut recevoir une explication fonctionnelle, sachant que ces deux propriétés convergent en persan pour favoriser le placement le plus en amont de la phrase.

Les COD marqués sont des entités hautement saillantes et disposent d'une autonomie pleine par rapport au verbe, tandis que les COD nus correspondent, en règle générale, au statut discursif le moins saillant et entretiennent un lien sémantique étroit avec le verbe<sup>5</sup>. Quant aux COD (non marqués) indéfinis, ils disposent d'un niveau de saillance intermédiaire, mais, étant construits comme entités autonomes (par rapport au verbe), ils tendent à se séparer de celui-ci. À ce sujet, il est intéressant de rappeler que la possibilité pour un COD de se séparer du verbe a été généralement associée à son statut discursif dans la littérature sur le sujet. Par exemple, un COD nu peut précéder le COI (c'est-à-dire s'éloigner du verbe) lorsqu'il reçoit une lecture de focus contrastif (cf. Karimi, 2003a)<sup>6</sup>. Il est également pertinent de noter le fait que les COD nus comportant des modifieurs tendent à

5. La discussion de Lazard (1982) sur la "polarisation" attribuée à l'enclitique  $=r\bar{a}$  est à ce sujet éclairante.

6. Rappelons que, selon cet auteur, cette condition est le seul cas de figure permettant le déplacement d'un COD non marqué, et non seulement un COD nu, de sa position de base.

se séparer du verbe, ce qui peut s'expliquer, suivant le même raisonnement : d'une part, leur richesse lexicale peut rehausser leur saillance discursive et, d'autre part, leur "lourdeur" relative peut leur procurer une certaine autonomie par rapport au verbe.

À la lumière de cette découverte, dont l'impact est conséquent pour l'étude de la structure du syntagme verbal en persan, nous avons dans le chapitre 6 (section 6.2) mené un examen critique approfondi de l'ensemble des arguments mis en avant pour soutenir l'hypothèse d'une position syntaxique double de l'objet (PSDO). Aucun de ces arguments ne s'est montré empiriquement probant. Dès lors, nous avons rejeté cette hypothèse au profit d'une structure plate du syntagme verbal en persan. Cela nous a conduit à aller plus loin et à poser la même question au niveau de la phrase. Nous nous sommes ainsi penchée brièvement sur la configurationnalité du persan. Après avoir passé en revue les propriétés liées à la distinction "configurationnel" vs. "non configurationnel", évoquées dans la littérature, nous avons opté pour l'extension de cette analyse en structure syntagmatique plate à la phrase.

En outre, la comparaison entre la variation de l'ordre entre le COD et le COI dans les phrases ditransitives, d'une part, et celle entre le sujet et le COD dans les phrases transitives, d'autre part, a mis en évidence qu'il n'est pas judicieux d'envisager un ordre "canonique" entre les deux objets pour la phrase ditransitive. En effet, dans nos données, l'ordre SOV s'impose comme ordre relatif par défaut entre le sujet et le(s) objet(s). *A contrario*, l'ordre relatif entre les deux objets manifeste un degré certain de variation qui fragilise l'hypothèse de l'existence d'un ordre relatif par défaut, le critère choisi pour le déterminer important peu. De plus, nos études montrent que les facteurs extra-syntaxiques interviendraient plus activement pour déterminer l'ordre relatif entre le COD et le COI qu'entre le sujet et le COD.

Cela nous ramène à la seconde problématique à laquelle nous nous sommes attelée dans cette thèse, à savoir la question des préférences dans l'ordonnement linéaire des constituants. Celle-ci, délaissée dans les travaux antérieurs sur le persan, continue de présenter un enjeu manifeste pour les travaux qui se penchent sur les mécanismes de la production, d'une part, et du traitement et de la compréhension de la parole, d'autre part, tant du point de vue cognitif que typologique. Ces travaux, dont nous avons discuté dans le chapitre 2, s'intéressent aux facteurs cognitivo-fonctionnels susceptibles d'intervenir sur l'ordre linéaire au-delà des contraintes grammaticales. Il s'agit notamment du poids grammatical, du statut informationnel, de la définitude, de l'animéité et de la dépendance sémantique, dont nous avons pu confirmer l'effet pour le persan. La littérature sur le fonctionnement du système langagier, dont l'étude des préférences dans l'ordonnement linéaire constitue un volet essentiel,

ne manque pas de rappeler l'importance des études quantitatives sur des langues diverses et variées pour affiner les théories et hypothèses existantes, notamment en ce qui concerne l'effet des différents facteurs cognitivo-fonctionnels et la façon dont les théories en rendent compte. Les conclusions de notre travail sont à apprécier à cet égard.

Plus particulièrement, partageant des propriétés avec des langues typologiquement opposées comme l'anglais et le japonais, le persan s'est révélé être un terrain propice pour évaluer les différentes hypothèses disponibles pour expliquer l'effet du poids grammatical (la longueur, pour simplifier) correspondant à la préférence "long-avant-court". Nos données appuient l'hypothèse de l'accessibilité conceptuelle mise en avant par Yamashita & Chang (2001) pour expliquer cette préférence en japonais par un modèle de production, au même titre que la préférence inverse dans une langue comme l'anglais<sup>7</sup>. L'enjeu de ce constat est d'autant plus significatif que cette hypothèse avait été minorée dans les travaux ultérieurs sur le japonais (Tanaka et al., 2011b) et sur le basque (Ros et al., 2015) au profit du modèle de Hawkins (1994). En effet, ce modèle prédit déjà la préférence "long-avant-court" dans ces langues.

Non seulement l'hypothèse de l'accessibilité conceptuelle prédit correctement la préférence "long-avant-court" pour le persan, mais elle se renforce aussi par d'autres aspects de notre étude, notamment par le fait que nous avons pu y observer l'interaction entre l'effet de la longueur et les effets d'autres facteurs tels que l'animéité et la définitude et/ou la saillance discursive, relevant de l'accessibilité conceptuelle qui n'avait pas été abordée par les études susmentionnées. Nos données indiquent qu'en persan, l'effet de la longueur n'est pas systématique et se voit moduler en fonction des autres dimensions de l'accessibilité conceptuelle. Pour être plus précise, nous constatons que la longueur relative a un effet significatif sur l'ordre lorsque les constituants concernés se positionnent en bas des autres échelles de l'accessibilité conceptuelle, par exemple les COD non marqués inanimés. À l'inverse, cette longueur ne manifeste pas ou peu d'effet lorsque les autres dimensions favorisent déjà un positionnement vers la gauche, par exemple dans le cas du COD marqué humain.

Terminons la présentation de l'apport de notre étude en soulignant que le paradigme expérimental, que nous avons mis en place dans nos expériences à l'aide

---

7. Rappelons que cette hypothèse se fonde sur deux observations : d'un côté, un constituant long tout en étant plus complexe est plus saillant et, par conséquent, plus accessible sur le plan conceptuel (que sa contrepartie courte), et, de l'autre, une langue comme le japonais est plus sensible aux facteurs conceptuels plutôt que formels, contrairement à une langue comme l'anglais. La divergence entre ces deux langues s'explique par un ensemble de spécificités qui les opposent, y compris la position du verbe dans les constructions étudiées, que l'on peut résumer, en simplifiant, par la rigidité de l'anglais et la flexibilité du japonais.

des questionnaires auto-administrés sur l'Internet, se révèle pratique et fiable pour se rapprocher des paradigmes présentsiels. Ce paradigme propose, en effet, un outil expérimental efficace pour les chercheurs qui ne peuvent avoir un accès local à une communauté notoire de locuteurs natifs de la langue faisant l'objet de l'investigation.

Ce travail offre de nombreuses perspectives de recherche, qui ont été évoquées tout au long de la thèse. Nous en présentons ici trois pistes majeures que nous souhaitons poursuivre dans nos futures recherches, outre les quelques études complémentaires que nous avons proposées d'effectuer pour affiner nos résultats (cf. section 5.5)<sup>8</sup>.

La première piste concerne les prédicats complexes et la question de leur séparabilité. La possibilité pour l'élément non verbal d'être séparé du verbe par des éléments morphologiques (ex. affixes ou clitiques) ou des éléments syntaxiques grammaticalisés (ex. auxiliaires) est incontestable et admise dans la littérature. Toutefois, la question de la séparabilité à proprement parler, à savoir si l'élément non verbal peut être séparé du verbe comme peut l'être un complément verbal ordinaire, fait l'objet de controverses. Ainsi, la question de la séparabilité entre les composantes d'un prédicat complexe rejoint la question de l'ordre linéaire des constituants investie dans cette thèse. La démarche que nous avons adoptée dans cette étude peut être élargie : par exemple, il serait intéressant de tester dans quelle mesure l'élément nominal d'un prédicat complexe se différencie d'un complément verbal nominal ordinaire en ce qui concerne son ordonnancement linéaire par rapport à un complément prépositionnel ou à un ajout. À ce sujet, nous pouvons nous appuyer sur l'étude de Safavi *et al.* (2016), qui se saisissent de cette question pour expliquer le lien entre la distance et la résolution des dépendances du point de vue du traitement et de la compréhension.

Cela nous conduit à la deuxième question de recherche sur laquelle il est indispensable de se pencher pour saisir les enjeux théoriques de cette étude concernant la préférence "long-avant-court", c'est-à-dire l'étude des préférences de l'ordonnancement linéaire du point de vue du traitement et de la compréhension. Un vif débat a lieu dans la littérature sur les hypothèses concernant le traitement de la complexité et les dépendances à distance à travers les langues typologiquement différentes (par ex. Gibson, 2000; Konieczny, 2000; Hale, 2003; Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Grodner & Gibson, 2005; Levy, 2008; Tily, 2010; Vasishth, 2011; Hale, 2011; Levy & Keller, 2013; Hawkins, 2014), qui rejoint celui sur l'ordre linéaire. Ces

---

8. Il s'agit notamment de déterminer l'interaction entre les effets de la longueur, de l'animéité et de la définitude sur l'ordre relatif entre le COD et le COI, afin de pouvoir proposer un modèle de prédiction global en fonction de ces trois préférences.

travaux s'appuient sur des expériences se situant dans des paradigmes tels que la lecture à vitesse autogérée (*self-paced reading*) et le mouvement oculaire (*eye-tracking*). Répliquer nos plans expérimentaux à l'aide de tels paradigmes devrait permettre de savoir à quel point nos conclusions sont dépendantes de notre paradigme expérimental, qui reflète le point de vue de la production. Cela peut, en outre, amorcer une étude des préférences de l'ordonnancement linéaire liées à la longueur et/ou la complexité plus large, englobant le domaine préverbal et postverbal, sachant que la position du verbe est essentielle dans ces hypothèses.

Enfin, l'examen des arguments en faveur d'une analyse hiérarchisée de la structure syntagmatique en persan, que nous avons effectuée dans la troisième partie de la thèse (section 6.2), a mis en évidence le besoin impératif d'entreprendre des études expérimentales portant sur les questions telles que la variabilité de portée, le liage des anaphores et les "vides parasites". Les études existantes sur le persan, en syntaxe aussi bien qu'en sémantique formelle, demeurent essentiellement théoriques se basant sur les jugements introspectifs des auteurs. Parmi ces sujets, nous souhaitons prolonger notre étude sur les "vides parasites", qui fait intervenir les contraintes de l'îlot et rejoint également la question de l'ordre linéaire. À ce sujet, nous souhaitons explorer les analyses éclairantes de Goldberg (2006, ch. 7) relatives au rôle essentiel de la structure discursive sur l'(in)acceptabilité de telles constructions.

Pour conclure, nous pouvons dire que cette thèse, de par sa méthodologie, apporte une contribution importante aux questions de la syntaxe du persan, qui, jusqu'à présent, n'avaient pas été étudiées sous un angle quantitatif. Nous espérons qu'elle ouvrira la voie à l'application fructueuse des mêmes méthodes aux questions largement débattues dans la littérature, mais sous-exploitées d'un point de vue empirique.

# Bibliographie

- Abeillé A. 2007. Les grammaires d'unification. Hermès sciences–Lavoisier.
- Abeillé A., Clément L. & Toussnel F. 2003. Building a treebank for French. In Treebanks, p. 165–187. Springer.
- Adli A. 2010. Constraint cumulativity and gradience : Wh-scrambling in Persian. Lingua, **120**(9), 2259–2294.
- Agresti A. 2007. An Introduction to Categorical Data Analysis. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley-Interscience.
- Aissen J. 2003. Differential object marking : Iconicity vs. economy. Natural Language & Linguistic Theory, **21**(3), 435–483.
- Andrews E. 1990. Markedness Theory. Sound and Meaning : The Roman Jakobson Series in Linguistics and Poetics. Duke University Press.
- Ariel M. 1988. Referring and accessibility. Journal of linguistics, **24**(01), 65–87.
- Arnold J. E. 1998. Reference form and discourse patterns. PhD thesis, Stanford University.
- Arnold J. E., Wasow T., Asudeh A. & Alrenga P. 2004. Avoiding attachment ambiguities : The role of constituent ordering. Journal of Memory and Language, **51**(1), 55–70.
- Arnold J. E., Wasow T., Losongco A. & Ginstrom R. 2000. Heaviness vs. newness : The effects of complexity and information structure on constituent ordering. Language, **76**(1), 28–55.
- Austin P. 2001. Word order in a free word order language : the case of Jiwari. In Forty years on : Ken Hale and Australian languages, p. 205–323.

- Baayen R. H., Davidson D. J. & Bates D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, **59**(4), 390–412.
- Bader M. & Häussler J. 2010. Word order in German : A corpus study. Lingua, **120**(3), 717–762.
- Barker C. 2002. Continuations and the nature of quantification. Natural language semantics, **10**(3), 211–242.
- Baroni M. & Evert S. 2008. Statistical methods for corpus exploitation. Corpus Linguistics. An International Handbook, Mouton de Gruyter, Berlin, p. 777–803.
- Bates D., Kliegl R., Vasishth S. & Baayen H. 2015. Parsimonious mixed models. arXiv preprint arXiv :1506.04967.
- Bates D., Maechler M. & Bolker B. 2012. lme4 : Linear mixed-effects models using s4 classes.
- Battistella E. L. 1990. Markedness : The evaluative superstructure of language. SUNY Press.
- Beck S. 1996. Quantified structures as barriers for If movement. Natural language semantics, **4**(1), 1–56.
- Beck S. 2001. Reciprocals are definites. Natural language semantics, **9**(1), 69–138.
- Behaghel O. 1909. Beziehungen zwischen umfang und reihenfolge von satzgliedern. Indogermanische Forschungen, **25**(110-142).
- Behaghel O. 1932. Deutsche Syntax, Band IV. Heidelberg : Carl Winter.
- Belletti A. 1988. The case of unaccusatives. Linguistic inquiry, p. 1–34.
- Bever T. G. 1970. The cognitive basis for linguistic structures. In R. Hayes, Ed., Cognition and language development, p. 277–360. New York : Wiley & Sons, Inc.
- Beyssade C. 2006. Portée. In D. Godard, L. Roussarie & F. Corblin, Eds., Sémanticopédie : dictionnaire de sémantique. [http ://www.semantique-gdr.net/dico/](http://www.semantique-gdr.net/dico/) : GDR Sémantique & Modélisation, CNRS.
- Bijankhan M. 2004. Naqsh-e peykare-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.

- Bijankhan M., Sheykhzadegan J., Bahrani M. & Ghayoomi M. 2011. Lessons from building a Persian written corpus : Peykare. Language resources and evaluation, **45**(2), 143–164.
- Blache P., Hemforth B. & Rauzy S. 2006. Acceptability prediction by means of grammaticality quantification. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, p. 57–64 : Association for Computational Linguistics.
- Bloomfield L. 1926. A set of postulates for the science of language. Language, **2**(3), 153–164.
- Bloomfield L. 1984. Language. Chicago : University of Chicago Press. Première édition : 1933.
- Bobaljik J. D. & Wurmbrand S. 2012. Word order and scope : Transparent interfaces and the 3/4 signature. Linguistic Inquiry, **43**(3), 371–421.
- Bock J. K. 1982. Toward a cognitive psychology of syntax : Information processing contributions to sentence formulation. Psychological Review, **89**(1), 1.
- Bock J. K. 1986. Meaning, sound, and syntax : Lexical priming in sentence production. Journal of Experimental Psychology : Learning, Memory, and Cognition, **12**(4), 575.
- Bock J. K. & Irwin D. E. 1980. Syntactic effects of information availability in sentence production. Journal of verbal learning and verbal behavior, **19**(4), 467–484.
- Bock J. K. & Levelt W. 1994. Language production : Grammatical encoding. In M. Gernsbacher, Ed., Handbook of psycholinguistics, p. 945—984. New York : Academic Press.
- Bock J. K., Loebell H. & Morey R. 1992. From conceptual roles to structural relations : bridging the syntactic cleft. Psychological review, **99**(1), 150.
- Bock J. K. & Warren R. K. 1985. Conceptual accessibility and syntactic structure in sentence formulation. Cognition, **21**(1), 47–67.
- Boersma P. 1997. How we learn variation, optionality, and probability. Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, **21**, 43–58.

- Boersma P. & Hayes B. 2001. Empirical tests of the gradual learning algorithm. Linguistic inquiry, **32**(1), 45–86.
- Bonami O. & Samvelian P. 2009. Inflectional periphrasis in Persian. In Proceedings of HPSG 2009, p. 26–46 : CSLI Publications.
- Bonami O. & Samvelian P. 2010. Persian complex predicates : Lexeme formation by itself. Paper presented at Septièmes Décembrettes Morphology Conference, Toulouse.
- Bonami O. & Samvelian P. 2015. The diversity of inflectional periphrasis in Persian. Journal of Linguistics, **51**(02), 327–382.
- Bossong G. 1985. Empirische Universalienforschung, volume 14. Narr.
- Branigan H. P. & Feleki E. 1999. Conceptual accessibility and serial order in greek language production. In Proceedings of the 21st conference of the cognitive science society, p. 96–101.
- Branigan H. P., Pickering M. J. & Tanaka M. 2008. Contributions of animacy to grammatical function assignment and word order during production. Lingua, **118**(2), 172 – 189.
- Brasoveanu A. & Farkas D. F. 2011. How indefinites choose their scope. Linguistics and philosophy, **34**(1), 1–55.
- Bresnan J. 2007. Is syntactic knowledge probabilistic ? experiments with the English dative alternation. Roots : Linguistics in search of its evidential base, p. 75–96.
- Bresnan J., Cueni A., Nikitina T. & Baayen H. 2007. Predicting the dative alternation. In I. K. Gerlof Bouma & J. Zwarts, Eds., Cognitive Foundations of Interpretation, p. 69–94. Royal Netherlands Academy of Science.
- Bresnan J., Dingare S. & Manning C. D. 2001. Soft constraints mirror hard constraints : Voice and person in English and Lummi. In Proceedings of the LFG01 Conference, p. 13–32.
- Bresnan J. & Ford M. 2010. Predicting syntax : Processing dative constructions in american and australian varieties of English. Language, **86**(1), 168–213.
- Brody J. 1984. Some problems with the concept of basic word order. Linguistics, **22**(5), 711–736.

- Browne W. 1970. More on definiteness markers : interrogatives in Persian. Linguistic Inquiry, **1**(3), 359–363.
- Browning M. & Karimi E. 1994. Scrambling to object position in Persian. In N. Corver & H. van Riemsdijk, Eds., Studies on scrambling : Movement and non-movement approaches to free word-order phenomena, p. 61–100. Mouton de Gruyter.
- Büring D. 1997. The great scope inversion conspiracy. Linguistics & Philosophy, **20**, 175–194.
- Büring D. 2005. Binding Theory. Cambridge Textbooks in Linguistics. Cambridge : Cambridge University Press.
- Bybee J. L. 2006. From usage to grammar : The mind's response to repetition. Language, p. 711–733.
- Bybee J. L. & Hopper P. J. 2001. Frequency and the emergence of linguistic structure, volume 45. John Benjamins Publishing.
- Chafe W. 1994. Discourse, consciousness, and time. Chicago University Press.
- Chang F. 2009. Learning to order words : A connectionist model of heavy NP shift and accessibility effects in Japanese and English. Journal of Memory and Language, **61**(3), 374 – 397.
- Choi H.-W. 1996. Optimizing structure in context : Scrambling and information structure. PhD thesis, Stanford University.
- Choi H.-W. 2007. Length and order : A corpus study of korean dative-accusative construction. Discourse and Cognition, **14**(3), 207–227.
- Chomsky N. 1957. Syntactic structures. Trad. fr. Seuil, 1969.
- Chomsky N. 1965. Aspects of the Theory of Syntax. Cambridge, MA : MIT Press.
- Chomsky N. 1969. La linguistique cartésienne : un chapitre de l'histoire de la pensée rationaliste. Seuil.
- Chomsky N. 1975. The logical structure of linguistic theory. Plenum press New York.
- Chomsky N. 1981. Lectures on government and binding : The Pisa lectures, volume 9. Walter de Gruyter.

- Chomsky N. 1986. Knowledge of language : Its nature, origin, and use. Greenwood Publishing Group.
- Chomsky N. 1995. The minimalist program, volume Current Studies in Linguistics. Cambridge : MIT Press.
- Chomsky N. & Lasnik H. 1993. The theory of principles and parameters. Syntax : An international handbook of contemporary research, **1**, 506–569.
- Christianson K. & Ferreira F. 2005. Conceptual accessibility and sentence production in a free word order language (odawa). Cognition, **98**(2), 105–135.
- Cinque G. 1990. Types of  $\bar{A}$ -dependencies. MIT press.
- Clark E. V. & Clark H. H. 1979. When nouns surface as verbs. Language, **55**(4), 767–811.
- Clark H. H. 1994. Managing problems in speaking. Speech communication, **15**(3), 243–250.
- Clark H. H. 1996. Using language. Cambridge university press.
- Clark H. H. & Haviland S. E. 1977. Comprehension and the given-new contract. Discourse production and comprehension. Discourse processes : Advances in research and theory, **1**, 1–40.
- Collins P. 1995. The indirect object construction in English : An informational approach. Linguistics, **33**(1), 35–50.
- Colonna S., Schimke S. & Hemforth B. 2010. Le rôle de la structure informationnelle dans l'interprétation d'une anaphore pronominale inter-phrastique. In Congrès Mondial de Linguistique Française, p. 1489–1499.
- Colonna S., Schimke S. & Hemforth B. 2012. Information structure effects on anaphora resolution in German and French : A crosslinguistic study of pronoun resolution. Linguistics, **50**(5), 901–1073.
- Comrie B. 1981. Language universals and linguistic typology : Syntax and morphology. University of Chicago press.
- Corbett G. G. 2000. Number. Cambridge University Press.
- N. Corver & H. van Riemsdijk, Eds. 1994. Studies on scrambling : Movement and non-movement approaches to free word-order phenomena, volume 41 of Studies in Generative Grammar. Mouton de Gruyter.

- Cowart W. 1997. Experimental syntax : Applying objective methods to sentence judgments. Sage Publications.
- Creissels D. 2006. Syntaxe générale : une introduction typologique. Hermès sciences–Lavoisier.
- Croft W. 2001. Radical construction grammar. Oxford University Press.
- Croft W. 2002. Typology and Universals. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Dabir-Moghaddam M. 1990. Pirāmune rā dar zabāne fārsi (sur rā dans la langue persane). Majalleye Zabānshenāsi, **13**(1), 2–60.
- Dabir-Moghaddam M. 1992. On the (in) dependence of syntax and pragmatics : Evidence from the postposition-ra in Persian. Cooperating with written texts, p. 549–573.
- Dabir-Moghaddam M. 1995. Fe'le morakkab dar zabāne fārsi [les verbes composés en persan]. Majalle-ye zabānshenāsi, (2), 2–46.
- Dabir-Moghaddam M. 2001. Word order typology of iranian languages. The Journal of Humanities, **2**, 17–23.
- Dabir-Moghaddam M. 2005. Zabāne fārsi va nazariyehāye zabāni : dar jostjuye charchubi barāye tadvin dasture jāme' zabāne fārsi [la langue persane et les théories linguistiques]. In Studies in Persian linguistics : selected articles, p. 230 – 267. Markaze nashere daneshgahi.
- Dabir-Moghaddam M. 2006. Internal and external forces in typology : evidence from iranian languages. Journal of Universal Language, **7**(2), 29–47.
- Dąbrowska E. 2010. Naive v. expert intuitions : an empirical study of acceptability judgments. The Linguistic Review, **27**(1), 1–23.
- Dalrymple M., Kanazawa M., Kim Y., Mchombo S. & Peters S. 1998. Reciprocal expressions and the concept of reciprocity. Linguistics and Philosophy, **21**(2), 159–210.
- Dalrymple M. & Nikolaeva I. 2011. Objects and information structure, volume 131. Cambridge University Press.
- Darzi A. 1996. Word order, NP movements, and opacity conditions in Persian. University of Illinois at Urbana-Champaign.

- Darzi A. 2008. On the vP analysis of Persian finite control constructions. Linguistic Inquiry, **39**(1), 103–116.
- De Smedt K. 1994. Parallelism in incremental sentence generation. In G. Adriaens & U. Hahn, Eds., Parallel natural language processing, p. 421–447. Norwood, NJ : Ablex.
- De Smedt K. & Kempen G. 1987. Incremental sentence production, self-correction and coordination. Natural Language Generation, **135**, 365–376.
- De Swart H. 1998. Licensing of negative polarity items under inverse scope. Lingua, **105**(3), 175–200.
- Déprez V. M. 1989. On the typology of syntactic positions and the nature of chains : move [alpha] to the specifier of functional projections. PhD thesis, Massachusetts Institute of Technology.
- Derbyshire D. C. & Pullum G. K. 1981. Object-initial languages. International Journal of American Linguistics, p. 192–214.
- Diesing M. 1992. Indefinites. Linguistic Inquiry Monographs. Cambridge, Mass : MIT Press.
- Diesing M. & Jelinek E. 1995. Distributing arguments. Natural Language Semantics, **3**(2), 123–176.
- Dixon R. M. 1979. Ergativity. Language, **55**(1), 59–138.
- Dobrovie-Sorin C. 1995. On the denotation and scope of indefinites. In L. Bruge, Ed., University of Venice Working Papers in Linguistics, volume 5, p. 67 – 114.
- Dobrovie-Sorin C. & Beyssade C. 2012. Redefining Indefinites. Studies in Natural Language and Linguistic Theory. Springer Netherlands.
- Dowty D. 1979. Word meaning and Montague grammar. Reidel, Dordrecht.
- Drummond A. 2013. Ibex farm. Version 0.3.
- Dryer M. S. 1980. The positional tendencies of sentential noun phrases in universal grammar. The Canadian Journal of Linguistics, **25**(2), 123–195.
- Dryer M. S. 1992. The greenbergian word order correlations. Language, **68**(1), 81–138.

- Dryer M. S. 2007. Word order. In T. Shopen, Ed., Clause Structure, Language Typology and Syntactic Description, volume 1, p. 61–131. Cambridge : Cambridge University Press.
- Dwivedi V. D. 2013. Interpreting quantifier scope ambiguity : Evidence of heuristic first, algorithmic second processing. PloS one, **8**(11), e81461.
- Engdahl E. 1983. Parasitic gaps. Linguistics and philosophy, **6**(1), 5–34.
- Evert S. 2006. How random is a corpus? the library metaphor. Zeitschrift für Anglistik und Amerikanistik, **54**(2), 177–190.
- Faghiri P. 2010. La morphologie du pluriel nominal du persan d'après la théorie Whole Word Morphology. Master's thesis, Université de Montréal.
- Faghiri P. & Samvelian P. 2014. Constituent ordering in Persian and the weight factor. In C. Pinon, Ed., Empirical Issues in Syntax and Semantics 10 (EISS 10), p. 215–232. CNRS.
- Faghiri P. & Samvelian P. 2015. How much is determined by the syntax? an empirical approach to the position of the direct object in Persian. In Proceedings of the 9th Iranian Conference on Linguistics, Allameh Tabataba'i University, Tehran, Feb. 24-25 2015, p. 1419–1435 : Allameh Tabataba'i University.
- Falk Y. 1983. Constituency, word order, and phrase structure rules. Linguistic Analysis, **11**(331), 60.
- Farkas D. F. 1994. Specificity and scope. In in L. Nash and G. Tsoulas (eds), Langues et Grammaire 1 : Citeseer.
- Farmer A. K. 1980. On the interaction of morphology and syntax. PhD thesis, Massachusetts Institute of Technology.
- Ferreira F. 1994. Choice of passive voice is affected by verb type and animacy. Journal of Memory and Language, **33**(6), 715–736.
- Ferreira V. S. 1996. Is it better to give than to donate? syntactic flexibility in language production. Journal of memory and language, **35**(5), 724–755.
- Ferreira V. S. & Yoshita H. 2003. Given-new ordering effects on the production of scrambled sentences in Japanese. Journal of psycholinguistic research, **32**(6), 669–692.

- Fiengo R. & Lasnik H. 1973. The logical structure of reciprocal sentences in English. Foundations of language, **9**(4), 447–468.
- Filik R., Paterson K. B. & Liversedge S. P. 2004. Processing doubly quantified sentences : Evidence from eye movements. Psychonomic Bulletin & Review, **11**(5), 953–959.
- Fodor J. D. & Sag I. A. 1982. Referential and quantificational indefinites. Linguistics and philosophy, **5**(3), 355–398.
- Frazier L. 1987. Sentence processing : A tutorial review. In M. Coltheart, Ed., Attention and Performance XII, p. 559–586. Erlbaum.
- Frazier L. & Fodor J. D. 1978. The sausage machine : A new two-stage parsing model. Cognition, **6**(4), 291–325.
- Frazier L. & Rayner K. 1982. Making and correcting errors during sentence comprehension : Eye movements in the analysis of structurally ambiguous sentences. Cognitive psychology, **14**(2), 178–210.
- Freidin R. & Sprouse R. A. 1991. Lexical case phenomena. In R. Freidin, Ed., Principles and parameters in comparative grammar, p. 392–416. MIT Press Cambridge, MA.
- Ganjavi S. 2007. Direct Objects in Persian. University of Southern California.
- Garrett M. F. 1980. Levels of processing in sentence production. Language production, **1**, 177–220.
- Gazdar G. 1985. Generalized phrase structure grammar. Harvard University Press.
- Gazdar G. & Pullum G. K. 1981. Subcategorization, constituent order, and the notion of 'head'. The scope of lexical rules, **107**, 123.
- Gelman A. & Hill J. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.
- Gharib A., Homāi, J. & Yāssemi R. 1994. Dastur zabān fārsi 'La grammaire du persan' (Panj Ostād). Enteshārāt Nāhid.
- Ghomeshi J. 1996. Projection and inflection : A study of Persian phrase structure. PhD thesis, University of Toronto.
- Ghomeshi J. 1997. Topics in Persian VPs. Lingua, **102**(2), 133 – 167.

- Ghameshi J. 2003. Plural marking, indefiniteness, and the noun phrase. Studia linguistica, **57**(2), 47–74.
- Ghameshi J. & Massam D. 1994. Lexical/syntactic relations without projection. Linguistic Analysis, **24**(3-4), 175–217.
- Gibson E. 1998. Linguistic complexity : locality of syntactic dependencies. Cognition, **68**(1), 1 – 76.
- Gibson E. 2000. The dependency locality theory : A distance-based theory of linguistic complexity. Image, language, brain, p. 95–126.
- Gibson E. & Fedorenko E. 2010. Weak quantitative standards in linguistics research. Trends in cognitive sciences, **14**(6), 233.
- Gibson E. & Fedorenko E. 2013. The need for quantitative methods in syntax and semantics research. Language and Cognitive Processes, **28**(1-2), 88–124.
- Givi Ahmadi H. & Anvari H. 1995. Dastur zabāne fārsi ‘La grammaire du persan’. Mo’assese farhangi Fātemi.
- Givón T. 1983. Topic Continuity in Discourse : A Quantitative Cross-Language Study. Typological Studies in Language. J. Benjamins.
- Givón T. 1988. The pragmatics of word order : predictability, importance and attention. In M. E. Hammond M. & W. J., Eds., Studies in Syntactic Typology. John Benjamins, Amsterdam, p. 243–284.
- Givón T. 1995. Functionalism and grammar. John Benjamins Publishing.
- Godfrey J. J., Holliman E. C. & McDaniel J. 1992. Switchboard : Telephone speech corpus for research and development. In Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, volume 1, p. 517–520 : IEEE.
- Goldberg A. E. 1996. Words by default : Optimizing constraints and the Persian complex predicate. In Annual Proceedings of the Berkeley Linguistic Society 22, p. 132–146. Berkeley.
- Goldberg A. E. 2006. Constructions at work : The nature of generalization in language. Oxford University Press.
- Greenberg J. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. Universals of language, **2**, 73–113.

- Greenberg J. H. 1966. Universals of language. MIT Press.
- Grewendorf G. & Sternefeld W. 1990. Scrambling and Barriers, volume 5 of Linguistics today. John Benjamins Publishing.
- Gries S. T. 2003. Towards a corpus-based identification of prototypical instances of constructions. Annual Review of Cognitive Linguistics, **1**(1), 1–27.
- Gries S. T. 2005. Syntactic priming : A corpus-based approach. Journal of psycholinguistic research, **34**(4), 365–399.
- Grieve R. & Wales R. J. 1973. Passives and topicalization. British Journal of Psychology, **64**(2), 173–182.
- Grodner D. & Gibson E. 2005. Consequences of the serial nature of linguistic input for sentential complexity. Cognitive science, **29**(2), 261–290.
- Gundel J. K. 1988. Universals of topic-comment structure. In M. Hammond, E. Moravcsik & J. Wirth, Eds., Studies in syntactic typology, Typological studies in language, p. 209–239. John Benjamins.
- Gundel J. K., Hedberg N. & Zacharski R. 1993. Cognitive status and the form of referring expressions in discourse. Language, **69**(2), 274–307.
- Hale J. T. 2003. The information conveyed by words in sentences. Journal of Psycholinguistic Research, **32**(2), 101–123.
- Hale J. T. 2011. What a rational parser would do. Cognitive Science, **35**(3), 399–443.
- Hale K. L. 1981. On the position of Walbiri in a typology of the base. Indiana University Linguistics Club.
- Hale K. L. 1983. Warlpiri and the grammar of non-configurational languages. Natural Language & Linguistic Theory, **1**(1), 5–47.
- Halle M. & Marantz A. 1993. Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser, Eds., The View from Building 20, p. 111–176. MIT Press.
- Halliday M. A. 1967. Notes on transitivity and theme in English : Part 2. Journal of linguistics, **3**(02), 199–244.
- Halliday M. A. 1985. An introduction to functional grammar. London : Edward Arnold.

- Hartsuiker R. J., Kolk H. H. J. & Huiskamp P. 1999. Priming word order in sentence production. The Quarterly Journal of Experimental Psychology, **52**(1), 129–147.
- Hartsuiker R. J. & Westenberg C. 2000. Word order priming in written and spoken sentence production. Cognition, **75**(2), B27–B39.
- Hawkins J. A. 1983. Word Order Universals. Academic Press.
- Hawkins J. A. 1990. A parsing theory of word order universals. Linguistics Inquiry, **21**(2), 223–261.
- Hawkins J. A. 1994. A performance theory of order and constituency. Cambridge studies in linguistics. Cambridge University Press.
- Hawkins J. A. 2000. The relative order of prepositional phrases in English : Going beyond manner–place–time. Language variation and change, **11**, 231–266.
- Hawkins J. A. 2001. Why are categories adjacent? Journal of Linguistics, **37**(1), pp. 1–34.
- Hawkins J. A. 2004. Efficiency and Complexity in Grammars. Oxford Linguistics. Oxford University Press.
- Hawkins J. A. 2008. An asymmetry between VO and OV languages. In M. N. Greville G. Corbett, Ed., Case and Grammatical Relations : Studies in honor of Bernard Comrie, volume 81, p. 167–190. John Benjamins Publishing.
- Hawkins J. A. 2014. Cross-linguistic variation and efficiency. OUP Oxford.
- Heim I., Lasnik H. & May R. 1991. Reciprocity and plurality. Linguistic Inquiry, **22**(1), 63–101.
- Hemforth B. 1993. Kognitives parsing : Repräsentation und verarbeitung sprachlichen Wissens. Sankt Augustin : Infix.
- Hemforth B., Colonna S., Petrone C. & D’Imperio M. 2013. Length matters : Informational load in ambiguity resolution. Discours. Revue de linguistique, psycholinguistique et informatique., (12).
- Hemforth B., Konieczny L., Scheepers C., Colonna S., Schimke S., Baumann P. & Pynte J. 2010. Language specific preferences in anaphor resolution : Exposure or gricean maxims? In 32nd Annual Conference of the Cognitive Science Society, p. 2218–2223.

- Hemforth B., Konieczny L. & Strube G. 1993. Incremental syntax processing and parsing strategies. In Proceedings of the 15th annual conference of the Cognitive Science Society, p. 539–545.
- Hemforth B., Mertins B. & Fabricius-Hansen C. 2014. Psycholinguistic Approaches to Meaning and Understanding Across Languages. Studies in Theoretical Psycholinguistics. Springer.
- Hofmeister P. & Sag I. A. 2010. Cognitive constraints and island effects. Language, **86**(2), 366.
- Hoji H. 1986. Scope interpretation in Japanese and its theoretical implications. In Proceedings of the West Coast Conference on Formal Linguistics, volume 5, p. 87–101.
- Hopper P. J. & Thompson S. A. 1980. Transitivity in grammar and discourse. Language, **56**(2), 251–299.
- Howell, David, Yves B., Vincent Y. & Marylène R. 2008. Méthodes statistiques en sciences humaines. De Boeck Université : Bruxelles.
- Ilkhanipour N. 2014. On the cp analysis of Persian finite control constructions. Linguistic Inquiry, **45**(2), 323–331.
- Jackendoff R. S. 1992. Semantic structures, volume 18. MIT press.
- Jaeger T. F. 2008. Categorical data analysis : Away from anovas (transformation or not) and towards logit mixed models. Journal of memory and language, **59**(4), 434–446.
- Jaeger T. F. & Norcliffe E. J. 2009. The cross-linguistic study of sentence production. Language and Linguistics Compass, **3**(4), 866–887.
- Kahnemuyipour A. 2001. On wh-questions in Persian. The Canadian journal of linguistics, **46**(1-2), 41–61.
- Kahnemuyipour A. 2003. Syntactic categories and Persian stress. Natural Language & Linguistic Theory, **21**(2), 333–379.
- Kahnemuyipour A. 2004. The syntax of sentential stress. PhD thesis, University of Toronto.

- Kaiser E., Runner J. T., Sussman R. S. & Tanenhaus M. K. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. Cognition, **112**(1), 55–80.
- Kaplan R. M. & Bresnan J. 1982. Lexical-functional grammar : A formal system for grammatical representation. Formal Issues in Lexical-Functional Grammar, p. 29–130.
- Karimi H. & Ferreira F. 2016. Informativity renders a referent more accessible : Evidence from eyetracking. Psychonomic bulletin & review, **23**(2), 507–525.
- Karimi H., Fukumura K., Ferreira F. & Pickering M. J. 2014. The effect of noun phrase length on the form of referring expressions. Memory & cognition, **42**(6), 993–1009.
- Karimi S. 1989. Aspects of Persian syntax, specificity, and the theory of grammar. PhD thesis, University of Washington.
- Karimi S. 1990. Obliqueness, specificity, and discourse functions : Ra in Persian. Linguistic Analysis, **20**, 139–191.
- Karimi S. 1994. Word-order variations in contemporary spoken Persian. Persian Studies in North America, p. 43–73.
- Karimi S. 1997. Persian complex verbs : Idiomatic or compositional. Lexicology, **3**, 273–318.
- Karimi S. 1999. A note on parasitic gaps and specificity. Linguistic Inquiry, **30**(4), 704–713.
- Karimi S. 2003a. On object positions, specificity, and scrambling in Persian. In S. Karimi, Ed., Word Order and Scrambling, p. 91–124. Blackwell Publishing Ltd.
- S. Karimi, Ed. 2003b. Word order and scrambling. Explaining linguistics. Oxford : Blackwell Publishing Ltd.
- Karimi S. 2005. A Minimalist Approach to Scrambling : Evidence from Persian. Studies in Generative Grammar. Mouton De Gruyter.
- Karimi-Doostan G. 1997. Light Verb Constructions in Persian. PhD thesis, University of Essex.
- Kayne R. S. 1994. The antisymmetry of syntax. Number 25 in Monographs, Linguistic inquiry. MIT Press.

- Keenan E. L. & Comrie B. 1977. Noun phrase accessibility and universal grammar. Linguistic inquiry, **8**(1), 63–99.
- Keller F. 2000. Gradience in grammar : Experimental and computational aspects of degrees of grammaticality. PhD thesis, University of Edinburgh.
- Keller F. & Asudeh A. 2001. Constraints on linguistic coreference : Structural vs. pragmatic factors. In Proceedings of the 23rd annual conference of the cognitive science society, p. 483–488.
- Kempen G. & Harbusch K. 2003. Word order scrambling as a consequence of incremental sentence production. In H. Härtl & H. Tappe, Eds., Mediating Between Concepts and Grammar, p. 141–64. Berlin, Germany : Mouton De Gruyter.
- 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 H. C. Pechmann, T., Ed., Multidisciplinary Approaches to Language Production, p. 173–181. Mouton de Gruyter.
- Kempen G. & Harbusch K. 2005. The relationship between grammaticality ratings and corpus frequencies : A case study into word order variability in the midfield of German clauses. In Linguistic evidence : Empirical, theoretical, and computational perspectives, p. 329–349. Walter de Gruyter.
- Kempen G. & Hoenkamp E. 1987. An incremental procedural grammar for sentence formulation. Cognitive science, **11**(2), 201–258.
- Khanlari P. 1972. Dastur-e zabān-e fārsi [La grammaire du persan]. Bonyād farhang Irān.
- Khanlari P. 1986. Tarix-e zabān-e farsi [Une histoire de la langue persane]. Editions Nashr-e Now.
- Khanlari P. 1994. Dastur-e tārixi-ye zabān-e fārsi [La grammaire historique du persan]. Enteshārāt tus.
- Kimball J. 1973. Seven principles of surface structure parsing in natural language. Cognition, **2**(1), 15–47.
- Kiss K. É. 1994. Scrambling as the base generation of random complement order. In N. Corver & H. van Riemsdijk, Eds., Studies on Scrambling : Movement and non-movement approaches to freeword-order phenomena, volume 41 of Studies in generative grammar, p. 221–256. Berlin : Mouton de Gruyter.

- Kiss K. É. 1998. Identificational focus versus Information focus. Language, **74**(2), 245–273.
- Kiss K. É. 2008. Free word order,(non) configurationality, and phases. Linguistic Inquiry, **39**(3), 441–475.
- Konieczny L. 2000. Locality and parsing complexity. Journal of Psycholinguistic Research, **29**(6), 627–645.
- Koopman H. & Sportiche D. 1991. The position of subjects. Lingua, **85**(2), 211–258.
- Kratzer A. 1995. Stage-level and individual-level predicates. In The generic book. Edited by Gregory N. Carlson and Francis J. Pelletier., p. 125–175. Chicago University Press.
- Kratzer A. 1998. Scope or pseudoscope? are there wide-scope indefinites? In S. Rothstein, Ed., Events and grammar, p. 163–196. Springer.
- Krifka M. 1989. Nominal reference, temporal constitution and quantification in event semantics. In J. v. B. Renate Bartsch & P. van Emde Boas, Eds., Nominal Reference, Temporal Constitution and Quantification Semantics and Contextual Expressions, p. 75 –115.
- Krifka M. 1992. Thematic relations as links between nominal reference and temporal constitution. In I. A. Sag & A. Szabolcsi, Eds., Lexical matters, p. 29–53. Stanford University.
- Krifka M. 1998. Scope inversion under the rise-fall contour in German. Linguistic Inquiry, **29**(1), 75–112.
- Kuno S. 1987. Functional syntax : Anaphora, discourse and empathy. Chicago : University of Chicago Press.
- Kuno S., Takami K.-i. & Wu Y. 1999. Quantifier scope in English, Chinese, and Japanese. Language, **75**(1), 63–111.
- Kuno S., Takami K.-I. & Wu Y. 2001. Response to Aoun and Li. Language, **77**(1), 134–143.
- Kurtzman H. S. & MacDonald M. C. 1993. Resolution of quantifier scope ambiguities. Cognition, **48**(3), 243–279.

- Lambrecht K. 1994. Information Structure and Sentence Form : Topic, Focus, and the Mental Representations of Discourse Referents. Cambridge Studies in Linguistics. Cambridge University Press.
- Langacker R. W. 1987. Foundations of cognitive grammar : Theoretical prerequisites. Stanford university press.
- Langendoen D. T. 1978. The logic of reciprocity. Linguistic inquiry, **9**(2), 177–197.
- Lazard G. 1966. L'enclitique nominale -i en persan : un ou deux morphèmes ? Bulletin de la Société de Linguistique de Paris, **61**, 249–264.
- Lazard G. 1982. Le morphème rā en persan et les relations actanciennes. Bulletin de la Société de Linguistique de Paris, **77**(1), 177–208.
- Lazard G. 1994. Le rā persan et le ba chinois. Cahiers de linguistique – Asie Orientale, **23**(1), 169–176.
- Lazard G. & Ghavam-Nejad M. 1990. Dictionnaire persan-français. E.J. Brill.
- Lazard G., Richard Y., Hechmati R. & Samvelian P. 2006. Grammaire du persan contemporain. Bibliothèque iranienne. Institut français de recherche en Iran.
- Lehmann W. P. 1973. A structural principle of language and its implications. Language, **84**(3), 47–66.
- Lemaréchal A. 1993. Pluriel, animéité et référentialité en palau. Faits de langues, **1**(2), 169–177.
- Levelt W. J. 1989. Speaking : From Intention to Articulation. ACL-MIT Press series in natural-language processing. Bradford Books, U.S.
- Levin B. & Rappaport Hovav M. 2005. Argument Realization. Research Surveys in Linguistics. Cambridge University Press.
- Levy R. 2008. Expectation-based syntactic comprehension. Cognition, **106**(3), 1126–1177.
- Levy R. P. & Keller F. 2013. Expectation and locality effects in German verb-final structures. Journal of Memory and Language, **68**(2), 199–222.
- Lewis R. L. & Vasishth S. 2005. An activation-based model of sentence processing as skilled memory retrieval. Cognitive science, **29**(3), 375–419.

- Lohse B., Hawkins J. A. & Wasow T. 2004. Domain minimization in English verb-particle constructions. Language, **80**(2), 238–261.
- Long C. 2010. Constituent ordering preferences in Japanese : implications for models of language production. Journal of Institute for Research in English Language and Literature, (35), 1–23.
- Lotfi A. R. 2003. Persian wh-riddles. In C. Boeckx & K. K. Grohmann, Eds., Multiple Wh-Fronting, p. 161–186. John Benjamins Publishing.
- MacDonald M. C. 1994. Probabilistic constraints and syntactic ambiguity resolution. Language and cognitive processes, **9**(2), 157–201.
- MacDonald M. C. 2013. How language production shapes language form and comprehension. Frontiers in psychology, **4**.
- Mace J. 2003. Persian grammar. London : Routledge.
- Mahajan A. K. 1990. The A/A-bar distinction and movement theory. PhD thesis, MIT.
- Mahootian S. 1997. Persian. New York : Routledge.
- Manning C. D. 2003. Probabilistic syntax. Probabilistic linguistics, p. 289–341.
- Marandin J.-M. 2006. Structure informationnelle. In D. Godard, L. Rousarie & F. Corblin, Eds., Sémanticlopédie : dictionnaire de sémantique. [http ://www.semantique-gdr.net/dico/](http://www.semantique-gdr.net/dico/) : GDR Sémantique & Modélisation, CNRS.
- Marashi M. 1970. The Persian verb : A partial description for pedagogical purposes. PhD thesis, University of Texas, Austin.
- May R. 1985. Logical Form : Its structure and derivation, volume 12. MIT press.
- McDonald J. L., Bock J. K. & Kelly M. H. 1993. Word and world order : Semantic, phonological, and metrical determinants of serial position. Cognitive Psychology, **25**(2), 188–230.
- Megerdooian K. 2002. Beyond Words and Phrases : A Unified Theory of Predicate Composition. PhD thesis, University of Southern California.
- Meshkatod Dini M. 1990. Introduction to Persian Transformational Syntax,. Mashhad : Ferdowsi University Editions.

- Meunier A. & Samvelian P. 1997. La postposition *rā* en persan : son rôle dans la détermination et sa fonction discursive. Cahiers de Grammaire, **25**, 187–232.
- Meurers W. D. 2005. On the use of electronic corpora for theoretical linguistics : Case studies from the syntax of German. Lingua, **115**(11), 1619–1639.
- Miller P. & Hemforth B. 2014. VP ellipsis beyond syntactic identity : the case of nominal antecedents. Présentation (orale) à la conférence Linguistic Evidence, Tübingen, Germany.
- Mithun M. 1992. Is basic word order universal. In D. L. Payne, Ed., Pragmatics of word order flexibility, p. 15–62. John Benjamins Publishing.
- Modarresi F. 2014. Bare nouns in Persian : Interpretation, Grammar and Prosody. PhD thesis, Humboldt-Universität zu Berlin.
- Moinzadeh A. 2001. An Antisymmetric, Minimalist Approach to Persian Phrase Structure. PhD thesis, The University of Ottawa.
- Moyne J. A. 1971. Reflexive and emphatic. Language, **47**(1), 141–163.
- Moyne J. A. 1974. The so-called passive in Persian. Foundations of Language, **12**(2), 249–267.
- Nikolaeva I. 2001. Secondary topic as a relation in information structure. Linguistics, **39**(1), 1–49.
- Oroumchian F., Tasharofi S., Amiri H., Hojjat H. & Raja F. 2006. Creating a feasible corpus for Persian POS tagging. Rapport interne, Department of Electrical and Computer Engineering, University of Tehran.
- Osgood C. E. & Bock J. K. 1977. Saliency and sentencings : Some production principles. Sentence Production : Developments in Research and Theory, Rosenberg, S., Editor, p. 89–140.
- Payne D. L. 1987. Information structuring in papago narrative discourse. Language, **63**(4), 783–804.
- Perry J. R. 2007. Persian morphology. In Morphologies of Asia and Africa, volume 2, p. 975–1019. Eisenbrauns.
- Peterson D. J. 1974. Noun phrase specificity. Xerox University Microfilms.

- Pickering M. J. & Branigan H. P. 1999. Syntactic priming in language production. Trends in cognitive sciences, **3**(4), 136–141.
- Pollard C. & Sag I. A. 1992. Anaphors in English and the scope of binding theory. Linguistic inquiry, **23**(2), 261–303.
- Pollard C. & Sag I. A. 1994. Head-driven phrase structure grammar. University of Chicago Press.
- Popper K. 1963. Conjectures and refutations : The growth of scientific knowledge. Routledge.
- Prat-Sala M. & Branigan H. P. 2000. Discourse constraints on syntactic processing in language production : A cross-linguistic study in English and spanish. Journal of Memory and Language, **42**(2), 168–182.
- Prince A. & Smolensky P. 2008. Optimality Theory : Constraint interaction in generative grammar. John Wiley & Sons.
- Prince E. F. 1981. Toward a taxonomy of given-new information. In P. Cole, Ed., Radical Pragmatics, p. 223–255. New York : Academic.
- Pullum G. K. 1982. Free word order and phrase structure rules. In NELS, volume 12, p. 209–220.
- Quirk R., Greenbaum S., Leech G. N., Svartvik J. et al. 1972. A grammar of contemporary English. London : Longman.
- R Core Team 2013. R : A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rasekhmahand M. 2004. Jāygāh maf’ul mostaqim dar fārsi [the position of the direct object in Persian]. Nāme farhangestān, **6**(4), 56–66.
- Rasekhmahand M. & Ghiasvand M. 2015. Barresi-ye peykarebonyād-e ta’sir-e avāmel-e naqši dar qalb-e nahvi-ye kutāh fārsi [une étude de corpus des facteurs fonctionnels intervenant dans le scrambling court en persan]. Dastur, **10**, 163–197.
- Reape M. 1993. A formal theory of word order : a case study in West Germanic. University of Edinburgh.
- Reape M. 1996. Getting things in order. Natural Language Processing, **6**, 209–254.

- Reinhart T. 1997. Quantifier scope : How labor is divided between  $qr$  and choice functions. Linguistics and philosophy, **20**(4), 335–397.
- Reinhart T. & Reuland E. 1993. Reflexivity. Linguistic inquiry, **24**(4), 657–720.
- Rialland A., Doetjes J. & Rebuschi G. 2002. What is focused in *c'est xp qui/que* cleft sentences in French? In Speech Prosody 2002, International Conference, p. 595–598.
- Rizzi L. 1989. The new comparative syntax : principles and parameters of universal grammar. Cahiers Ferdinand de Saussure, p. 65–78.
- Roberts J. R., Barjasteh Delforooz B. & Jahani C. 2009. A study of Persian discourse structure. *Studia Iranica Upsaliensia*, 12. Acta Universitatis Upsaliensis.
- Ros I., Santesteban M., Fukumura K. & Laka I. 2015. Aiming at shorter dependencies : the role of agreement morphology. Language, Cognition and Neuroscience, **30**(9), 1156–1174.
- Rosenbach A. 2002. Genitive variation in English : Conceptual factors in synchronic and diachronic studies, volume 42. Walter de Gruyter.
- Rosenbach A. 2005. Animacy versus weight as determinants of grammatical variation in English. Language, **81**(3), 613–644.
- Ross J. R. 1967. Constraints on variables in syntax. PhD thesis, Massachusetts Institute of Technology.
- Runner J. T. & Head K. D. L. 2014. What can visual world eye-tracking tell us about the binding theory? In C. Pinon, Ed., Empirical Issues in Syntax and Semantics 10 (EISS 10), p. 269–286. CNRS.
- Runner J. T., Sussman R. S. & Tanenhaus M. K. 2002. Logophors in possessed picture noun phrases. In Proceedings of West Coast Conference on Formal Linguistics 21, p. 401–414.
- Runner J. T., Sussman R. S. & Tanenhaus M. K. 2006. Processing reflexives and pronouns in picture noun phrase. Cognitive Science, **30**(2), 193–241.
- J. Sabel & M. Saito, Eds. 2005. The Free Word Order Phenomenon : Its Syntactic Sources and Diversity, volume 69 of Studies in generative grammar. Mouton de Gruyter.

- Sadeghi A.-A. 1970. Rā dar zabān-e fārsi emrooz (râ en persan contemporain). Nashriyeye dāneshkadeye adabiyāt va olume ensāni Tabriz, (93), 9–22.
- Sadeghi A. A. 1993. On denominative verbs in Persian. In Farsi Language and the Language of Science, p. 236–246. Tehran : University Press.
- Safavi M. S., Husain S. & Vasisht S. 2016. Dependency resolution difficulty increases with distance in Persian separable complex predicates : Evidence for expectation and memory-based accounts. Frontiers in psychology, **7**.
- Saito M. 2005. Further notes on the interpretation of scrambling chains. In M. S. Joachim Sabel, Ed., The free word order phenomenon : Its syntactic sources and diversity, volume 69 of Studies in Generative Grammar, p. 335–376. Mouton de Gruyter.
- Saito M. & Fukui N. 1998. Order in phrase structure and movement. Linguistic Inquiry, **29**(3), 439–474.
- Samiian V. 1994. The ezafe construction : some implications for the theory of x-bar syntax. Persian Studies in North America, p. 17–41.
- Samvelian P. 2001. Le statut syntaxique des objets nus en persan. Bulletin de la Société de Linguistique de Paris, **96**(1), 349–388.
- Samvelian P. 2006a. Le sujet, l’objet et l’inaccusativité dans les prédicats complexes nom-verbe en persan. Cahiers de linguistique de l’INALCO, **6**, 155–190.
- Samvelian P. 2006b. L’enclitique *-i* introducteur de relative en persan : déterminant, allomorphe de l’*ezâfé* ou autre chose encore ? Studia Iranica, **35**(1), 7–34.
- Samvelian P. 2007. A (phrasal) affix analysis of the Persian ezafe. Journal of Linguistics, **43**(3), 605–645.
- Samvelian P. 2012. Grammaire des prédicats complexes : les constructions nom-verbe. Hermès sciences–Lavoisier.
- Samvelian P. & Faghiri P. 2013. Introducing perspred, a syntactic and semantic database for Persian complex predicates. In Proceedings of the Workshop on Multiword Expressions, p. 11–20, Atlanta, Georgia, USA.
- Samvelian P. & Faghiri P. 2014. Persian complex predicates : How compositional are they ? Semantics-Syntax Interface, **1**(1), 43–74.

- Samvelian P. & Faghiri P. 2016. Re-thinking compositionality in Persian complex predicates. Annual Meeting of the Berkeley Linguistics Society 39 (BLS39), **39**(1), 212–226.
- Samvelian P. & Tseng J. 2010. Persian object clitics and the syntax-morphology interface. In S. Müller, Ed., Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, p. 212–232. Stanford : CSLI Publications.
- Sauerland U. 1998. Plurals, derived predicates and reciprocals. In U. Sauerland & O. Percus, Eds., The Interpretive Tract, volume 25 of MIT Working Papers in Linguistics, p. 177–204.
- Sauerland U. & Bott O. 2002. Prosody and scope in German inverse linking constructions. In the Speech Prosody 2002 conference, Ed., the Speech Prosody 2002 conference, p. 623–628 : Aix-en-Provence, France : Laboratoire Parole et Langage.
- Schutze C. 1996. The Empirical Base of Linguistics : Grammaticality Judgments and Linguistic Methodology. University of Chicago Press.
- Seidenberg M. S. & MacDonald M. C. 1999. A probabilistic constraints approach to language acquisition and processing. Cognitive science, **23**(4), 569–588.
- Shokouhi H. & Kipka P. 2003. A discourse study of Persian *rā*. Lingua, **113**(10), 953–966.
- Siewierska A. 1988. Word order rules. Croom Helm, Limited.
- Siewierska A. 1993. On the interplay of factors in the determination of word order. Syntax : An international handbook of contemporary research, p. 826–846.
- Silverstein M. 1976. Hierarchy of features and ergativity. In R. M. Dixon, Ed., Grammatical categories in Australian languages, p. 112–171. Australian Institute of Aboriginal Studies.
- Song J. J. 2012. Word order. Cambridge University Press.
- Song J. J. 2014. Linguistic Typology : Morphology and Syntax. Longman Linguistics Library. Taylor & Francis.
- Sorace A. & Keller F. 2005. Gradience in linguistic data. Lingua, **115**(11), 1497–1524.

- Sprouse J. & Almeida D. 2012. Assessing the reliability of textbook data in syntax : Adger's core syntax. Journal of Linguistics, **48**(03), 609–652.
- Sprouse J., Schütze C. T. & Almeida D. 2013. A comparison of informal and formal acceptability judgments using a random sample from linguistic inquiry 2001–2010. Lingua, **134**, 219–248.
- 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.
- Steedman M. 2012. Taking scope : The natural semantics of quantifiers. MIT Press.
- Sternefeld W. 1998. Reciprocity and cumulative predication. Natural language semantics, **6**(3), 303–337.
- Szabolcsi A. 1997a. Strategies for scope taking. In A. Szabolcsi, Ed., Ways of scope taking, p. 109 – 154. Dordrecht : Kluwer.
- Szabolcsi A. 1997b. Ways of Scope Taking. Studies in Linguistics and Philosophy. Dordrecht : Kluwer.
- Szabolcsi A. 2010. Quantification. Research Surveys in Linguistics. Cambridge University Press.
- Tanaka M., Branigan H. P. & Pickering M. J. 2005. The role of animacy in Japanese sentence production. Paper Presented at the CUNY Sentence Processing Conference, Tucson, AZ.
- Tanaka M. N., Branigan H. P., McLean J. F. & Pickering M. J. 2011a. Conceptual influences on word order and voice in sentence production : Evidence from Japanese. Journal of Memory and Language, **65**(3), 318–330.
- Tanaka M. N., Branigan H. P. & Pickering M. J. 2011b. The production of head-initial and head-final languages. In Y. H. Hiroko Yamashita & J. Packard, Eds., Processing and producing head-final structures, volume Studies in Theoretical Psycholinguistics, p. 113–129. Springer.
- Tanenhaus M. K. & Trueswell J. C. 1995. Sentence comprehension. Academic Press.

- Taraldsen K. T. 1981. Remarks on government, thematic structure and the distribution of empty categories. Levels of syntactic representation, **10**, 253.
- Tellier C. 1991. Licensing Theory and French Parasitic Gaps. NATO ASI Series C : Mathematical and Physical Sciences. Springer Netherlands.
- Tellier C. & Valois D. 2006. Constructions méconnues du français. Paramètres (Montréal). Presses de l'Université de Montréal.
- Temperley D. 2007. Minimization of dependency length in written English. Cognition, **105**(2), 300–333.
- Thuilier J. 2012. Contraintes préférentielles et ordre des mots en français. PhD thesis, Université Paris Diderot.
- Tily H. 2010. The Role of Processing Complexity in Word Order Variation and Change. PhD thesis, Stanford University. Department of Linguistics.
- Tomasello M. 2003. Constructing a Language : A Usage-Based Theory of Language Acquisition. Harvard University Press.
- Tomlin R. 1986. Basic Word Order : Functional Principles. London : Croom Helm.
- Toosarvandani M. 2008. Wh-movement and the syntax of sluicing. Journal of Linguistics, **44**(3), 677–722.
- Trueswell J. & Tanenhaus M. 1994. Toward a lexical framework of constraint-based syntactic ambiguity resolution. Perspectives on sentence processing, p. 155–179.
- Truswell R. 2014. Binding theory. In Y. S. Andrew Carnie, Dan Siddiqi, Ed., The Routledge handbook of syntax, p. 214–238.
- Urbach T. P., DeLong K. A. & Kutas M. 2015. Quantifiers are incrementally interpreted in context, more than less. Journal of memory and language, **83**, 79–96.
- Vallduví E. 1992. The informational component. Garland Publishers.
- Vasishth S. 2003. Working Memory in Sentence Comprehension : Processing Hindi Center Embeddings. Taylor & Francis.
- Vasishth S. 2004. Discourse context and word order preferences in hindi. The Yearbook of South Asian Languages and Linguistics, p. 113–127.

- Vasishth S. 2011. Integration and prediction in head-final structures. In Y. H. Hiroko Yamashita & J. Packard, Eds., Processing and producing head-final structures, volume Studies in Theoretical Psycholinguistics, p. 349–367. Springer.
- Vasishth S. & Lewis R. L. 2006. Argument-head distance and processing complexity : Explaining both locality and antilocality effects. Language, **82**(4), 767–794.
- Vasishth S., Shaher R. F. & Srinivasan N. 2012. The role of clefting, word order and given-new ordering in sentence comprehension : Evidence from Hindi. Journal of South Asian Linguistics, **5**, 35–56.
- Vendler Z. 1967. Linguistics in philosophy. Cornell University Press.
- Vennemann T. 1974. Analogy in generative grammar : the origin of word order. In Proceedings of the eleventh international congress of linguists (1972), p. 79–83.
- Ward G. L. & Prince E. F. 1991. On the topicalization of indefinite nps. Journal of Pragmatics, **16**(2), 167–177.
- Wasow T. 1986. The wizards of ling. Natural Language and Linguistic Theory, **3**, 485–491.
- Wasow T. 1997. Remarks on grammatical weight. Language Variation and Change, **9**(1), 81–105.
- Wasow T. 2002. Postverbal behavior. Stanford : CSLI Publications.
- Wasow T. 2013. The appeal of the PDC program. Frontiers in psychology, **4**, 236.
- Wasow T. & Arnold J. E. 2003. Post-verbal constituent ordering in English. Topics in English Linguistics, **43**, 119–154.
- Waugh L. R. 1982. Marked and unmarked : A choice between unequals in semiotic structure. Semiotica, **38**(3-4), 299–318.
- Webelhuth G. 1989. Syntactic saturation phenomena and the modern Germanic languages. PhD thesis, University of Massachusetts.
- Whaley L. J. 1997. Introduction to Typology. SAGE publication.
- Windfuhr G. L. 1979. Persian grammar : History and state of its study. Walter de Gruyter.
- Windfuhr G. L. 1990. Persian. In B. Comrie, Ed., The Major Languages of South Asia, the Middle East and Africa, p. 81–100. London : Routledge.

- Windfuhr G. L. 2009. Introduction to the Iranian languages. In G. L. Windfuhr, Ed., The Iranian Languages, p. 1–3. London : Routledge.
- Windfuhr G. L. & Perry J. R. 2009. Persian and Tajik. In G. L. Windfuhr, Ed., The Iranian languages, Routledge Language Family Series, p. 416–544. London : Routledge.
- Yamashita H. 1997. The effects of word-order and case marking information on the processing of Japanese. Journal of Psycholinguistic Research, **26**(2), 163–188.
- Yamashita H. 2000. Structural computation and the role of morphological markings in the processing of Japanese. Language and speech, **43**(4), 429–455.
- Yamashita H. 2002. Scrambled sentences in Japanese : Linguistic properties and motivations for production. TEXT, **22**(4), 597–634.
- Yamashita H. & Chang F. 2001. “Long before short” preference in the production of a head-final language. Cognition, **81**(2), B45–B55.
- Yamashita H. & Chang F. 2006. Sentence production in Japanese. In . Y. S. M. Nakayama, R. Mazuka, Ed., Handbook of East Asian Psycholinguistics (Volume 2, Japanese), p. 291–297. Cambridge : Cambridge University Press.
- Zribi-Hertz A. 1996. L’anaphore et les pronoms : une introduction à la syntaxe générative. Sens et structures. Presses universitaires du Septentrion.

# Annexes



## A Algorithme utilisé pour lemmatiser le corpus Bijankhan

Dans les paragraphes suivants, nous présentons, sans entrer dans les détails, l’algorithme que nous avons utilisé pour lemmatiser les verbes fléchis du corpus Bijankhan. Pour ce faire, nous présentons brièvement le paradigme verbal du persan (voir également §1.1.5).

Le paradigme verbal du persan connaît deux sous-classes : les verbes formés à partir du radical du “présent” (radical I) et ceux formés à partir du radical du “passé” (radical II). Dans chacune de ces sous-classes, le paradigme est parfaitement régulier, dans le sens où les conjugaisons sont identiques pour tous les lemmes verbaux. Rappelons qu’il n’y pas de syncrétisme dans la conjugaison verbale du persan et que chaque désinence verbale réalise sans ambiguïté la personne et le nombre du sujet grammatical. Il existe néanmoins une certaine alternance morphophonologique, notamment lorsque le radical commence ou finit par une voyelle, qui reste régulière. Par ailleurs, il y a également une variation liée au registre, qui est systématique et qui ne dépend pas du lemme verbal.

Le radical I s’obtient à partir de la forme de l’impératif singulier par la suppression du préfixe  $b(e)$ -. Le radical II s’obtient à partir de la forme de l’infinitif par la suppression du suffixe  $-(a)n$ . Pour une partie des verbes, le radical II peut s’obtenir, par une affixation simple, à partir du radical I, ex.  $(be-)tars : tars-id(-an)$  ‘avoir peur’. Pour d’autres verbes, le lien formel entre les deux radicaux relèvent de l’idiosyncrasie, ex.  $(be-)bin : did(-an)$  ‘voir’.

Étant donné le nombre restreint des verbes simples, nous avons tout d’abord dressé une liste des verbes accompagnés leurs deux radicaux. Nous avons pour ce faire extrait la liste des infinitifs du corpus Bijankhan en nous basant sur l’annotation déjà disponible pour cette catégorie dans celui-ci (l’étiquette **N-INF**). La lemmatisation des verbes a été faite à l’aide de l’algorithme suivant exécuté via un script Python.

Deux séries principales de gabarits de conjugaison (voir ci-dessous pour un exemple) ont été définies en tant que filtres pour extraire le radical des verbes fléchis du corpus, c’est-à-dire les tokens étiquetés en tant que verbe, selon que leur conjugaison relève d’un temps verbal formé sur la base du radical I ou de celui de radical II. L’étiquette des verbes fléchis du corpus Bijankhan fournit cette information, c’est-à-dire le temps verbal.

- (1) a.  $\{my|nmy|b|n\}$ -RadI- $\{m|y|d|ym|yd|nd\}$
- b.  $\{my|nmy|b|n|\emptyset\}$ -RadII- $\{m|y|\emptyset|ym|yd|nd\}$

Ensuite, à l'aide d'un dictionnaire établi sur la liste des radicaux existants, le verbe est lemmatisé. Ainsi, son lemme est ajouté comme une annotation supplémentaire à son token. Les radicaux extraits ne figurant pas dans la liste sont répertoriés dans une seconde liste. La première liste des radicaux est mise à jour manuellement à la fin de la première phase en ajoutant les radicaux manquants. Le filtre est exécuté une deuxième fois pour lemmatiser ces verbes.

Il faut noter qu'il existe des cas d'homographie parmi les radicaux. Par exemple, les radicaux I des deux verbes *koštan* 'tuer' et *kešidan* 'tirer' s'écrivent de manière identique. C'est également le cas des verbes *bordan* 'amener' (ou 'gagner') et *boridan* 'couper'. Un autre cas similaire concerne les verbes *nešastan* 's'asseoir' et *šostan* 'laver'. Lorsque le dernier prend le préfixe de la négation (*n(e)-*), ses formes verbales se confondent à l'écrit avec celles du premier. Comme nous ne disposons pas de moyen automatique pour différencier entre les formes homographes, nous avons fait le choix d'attribuer un lemme verbal en fonction de la fréquence relative des deux lemmes concurrents (obtenue sur la base des formes non homographes). Ainsi, par exemple, dans le premier cas, le lemme *kešidan* 'tirer', et, dans le second cas, le lemme *bordan* 'amener' (ou 'gagner'), a été attribué aux formes homographes, car nous savons que la fréquence de ces lemmes, quant aux formes verbales basées sur le radical II, est plus élevée dans le corpus.

## B Les lemmes verbaux de la table de données DS1

Le tableau ci-dessous présente les lemmes verbaux de notre table de données préliminaire (DS1) avec leurs fréquences totales dans le corpus Bijankhan, ainsi que dans cette table de données. Notons que ces lemmes se retrouvent également dans notre table de données finale (DS2).

Nb	Lemme	Freq-DS1	Freq-BK
1	gereftan	66	6 849
2	dādan	27	10 494
3	āvardan	27	1 915
4	gozāštan	27	1 496
5	bordan	23	1 560
6	andāxtan	19	310
7	zadan	18	1 304
8	resāndan	17	693
9	bardāštan	14	294
10	ferestādan	14	219
11	resānidan	14	30
12	nahādan	13	254
13	kešidan	11	731
14	rixtan	11	254
15	kešāndan	11	75
16	gozidan	11	56
17	rāndan	10	89
18	darāvardan	9	177
19	afkandan	9	69
20	gozārdan	9	65
21	bargardāndan	9	42
22	goftan	8	9 418
23	pušāndan	8	79
24	talabidan	8	67
25	fešordan	8	38
26	bastan	7	247
27	bāzgardāndan	7	36
28	duxtan	7	25
29	āmuxtan	6	130
30	xaridan	6	70
31	vādāštan	6	25
32	baxšidan	5	276
33	pardāxtan	4	1 045
34	xordan	4	636
35	paziroftan	4	535
36	bāzdāštan	4	23
37	čarxāndan	4	20
38	afzudan	3	2 394
39	sepordan	3	186
40	gardāndan	3	47
41	kubidan	3	46

Nb	Lemme	Freq-DS1	Freq-BK
42	čidan	3	32
43	āmixtan	3	30
44	pāšidan	3	23
45	gostardan	3	17
46	vāgozāštan	3	10
47	āludan	3	5
48	neveštan	2	1 457
49	foruxtan	2	82
50	barangixtan	2	60
51	negahdāštan	2	27
52	negāštan	2	26
53	nešāndan	2	22
54	fahmāndan	2	13
55	tābāndan	2	10
56	rahāndan	2	9
57	āvixtan	2	8
58	šenidan	1	303
59	porsidan	1	231
60	barāvardan	1	48
61	farāxāndan	1	36
62	robudan	1	27
63	boridan	1	26
64	damidan	1	26
65	bātan	1	24
66	gonjāndan	1	24
67	barafrāštan	1	18
68	časbāndan	1	15
69	gonjidan	1	14
70	zodudan	1	14
71	tanidan	1	13
72	bargereftan	1	10
73	barkardan	1	8
74	tarsāndan	1	8
75	darāmixtan	1	7
76	gostarāndan	1	4
77	mālidan	1	4
78	ranjāndan	1	4
79	bāzgoftan	1	3
80	forubordan	1	3
81	vāgozārdan	1	2
82	forukardan	1	1



## C Les stimuli de nos expériences

Dans les pages suivantes, nous présentons l'ensemble des stimuli de nos expériences. Nous présentons tout d'abord les stimuli de nos cinq expériences "phrases à trous", abordées dans le chapitre 5, qui portent sur la variation de l'ordre et ensuite ceux de nos deux expériences de jugement d'acceptabilité, abordées dans le chapitre 6, la première portant sur la coordination des COD marqué et non marqué et la seconde sur l'autorisation des vides parasites. Notons que pour la première série nous nous limitons aux items cibles et dans la seconde nous fournissons également les items distracteurs.

La présentation suit l'organisation suivante : tout d'abord l'ensemble des stimuli en persan – tels qu'ils apparaissent dans nos questionnaires – et ensuite les équivalents français. Précisons que, dans la traduction française, nous nous sommes efforcée de respecter, dans la mesure du possible, la structure de l'item. Ainsi, il arrive que les traductions soient approximatives. Autrement dit, nous avons privilégié la forme plutôt que le sens.

Dans chaque bloc, les items expérimentaux sont présentés sous forme d'un tableau. Celui-ci comporte pour chaque item un index composé du nom de l'expérience, le numéro de l'item et la condition dans laquelle il est décliné. L'annotation des conditions reprend celle utilisée dans le corps de la thèse.

En ce qui concerne, les expériences "phrases à trous", chaque item est présenté sur deux lignes, dont la première correspond au préambule et la seconde à la phrase à compléter. Les constituants omis de la phrase à compléter sont donnés entre crochets. Les deux choix pour un même constituant sont séparés par une barre oblique.

Pour ces items, il arrive que la traduction française d'un préambule, là où en persan l'on utilise une forme verbale fléchie, requiert l'emploi de l'infinitif. Dans ces cas de figure, nous avons suivi la traduction utilisée dans l'exemple ci-dessous, afin de rester fidèle à la structure de la phrase à compléter.

- (2) a. Sepehr tasmim gereft (...) (Sepehr décision prendre.PST.3SG)  
'Sepehr a décidé de (qu'il)...
- b. [be Omid] [yek ketāb] be-dah-ad  
[à Omid] [un livre] SBJV-donner.PRS-3SG  
'... donne un livre à Omid.'

La traduction française de certaines de nos manipulations n'était pas possible, dans la mesure où les deux conditions se traduisent de la même manière en français. Dans ces cas de figure, nous avons fourni, dans la version française, seulement une des deux

conditions concernées. Il s'agit notamment des conditions B et B' de l'expérience T1 (dans le questionnaire "phrases à trous" 3), qui se distinguent, en persan, seulement par la position de l'enclitique =*rā* par rapport à la subordonnée relative. Le deuxième cas concerne notre expérience sur la coordination dans laquelle les conditions "Post" et "Coor" ne peuvent difficilement être différenciées dans les équivalents français. Enfin, pour notre expérience sur les vides parasites, nous avons fourni une seule version (sur les deux conditions) dans laquelle nous avons marqué la manipulation entre la phrase de contrôle et la phrase cible à l'aide de parenthèses.

Ajoutons enfin que, dans le cas des pronoms ou des expressions référentielles sous-spécifiés pour le genre en persan, nous avons utilisé le masculin dans la traduction française. Par ailleurs, nous avons ajouté un sujet (grammatical) pronominal, ou un sujet lexical entre parenthèses, là où dans la version persane le sujet est implicite.

## Questionnaire "phrases à trous" 1 : Exp. D1

Item	Ind
خانم سلطانی زن دست و دلبازی بود. وقتی املاک شوهرش بعد از مرگ به او رسید [به یکی از خواهرهایش که مدتها اجارمنشین بود] [یک گردنبند/یک آپارتمان] بخشید.	D1-1: INL
خانم سلطانی زن دست و دلبازی بود. وقتی املاک شوهرش بعد از مرگ به او رسید [به یکی از خواهرهایش] [یک گردنبند مروارید دانه درشت اصل/یک آپارتمان صدوپنجاهمتری دوخوابه] بخشید.	D1-1: INC
هنگامی که خواهرزاده‌ی خانم سلطانی از دواج کرد خانم سلطانی ثروت چندانی نداشت. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش که مدتها اجارمنشین بود] [یک گردنبند/یک آپارتمان] بخشید.	D1-1: IDL
هنگامی که خواهرزاده‌ی خانم سلطانی از دواج کرد خانم سلطانی ثروت چندانی نداشت. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش] [یک گردنبند مروارید دانه درشت اصل/یک آپارتمان صدوپنجاهمتری نوساز] بخشید.	D1-1: IDC
فرزندان آقای تهرانی کوچکترین توجهی به کتابخانه‌اش نداشتند. آقای تهرانی هم هرازگاهی [به برادرزاده‌اش که شاگرد زرنگی بود] [چند دست پیراهن/چند جلد کتاب] می‌بخشید.	D1-2: INL
فرزندان آقای تهرانی کوچکترین توجهی به کتابخانه‌اش نداشتند. آقای تهرانی هم هرازگاهی [به برادرزاده‌اش] [چند دست پیراهن مردانه‌ی چهارخانه/چند جلد کتاب نفیس و گران‌قیمت] می‌بخشید.	D1-2: INC
تنها کسی که به کتابخانه‌ی آقای تهرانی علاقه نشان می‌داد، امیرعلی، پسر همسایه بود. آقای تهرانی هم هرازگاهی [به امیرعلی که شاگرد زرنگی بود] [چند دست پیراهن/چند جلد کتاب] می‌بخشید.	D1-2: IDL
تنها کسی که به کتابخانه‌ی آقای تهرانی علاقه نشان می‌داد، امیرعلی، پسر همسایه بود. آقای تهرانی هم هرازگاهی [به امیرعلی] [چند دست پیراهن مردانه‌ی چهارخانه/چند جلد کتاب نفیس و گران‌قیمت] می‌بخشید.	D1-2: IDC
شهردار جدید که با شعار حل مشکل مسکن رای آورده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانانی که قصد ازدواج دارند] [یک دستگاه پژو/یک قطعه زمین] خواهد بخشید.	D1-3: INL
شهردار جدید که با شعار حل مشکل مسکن رای آورده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به هر زوج جوان] [یک دستگاه پژوی چهارصد و پنج/یک قطعه زمین در حاشیه شمال شهر] خواهد بخشید.	D1-3: INC
شهردار جدید که با رای جوانان انتخاب شده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانانی که قصد ازدواج دارند] [یک دستگاه پژو/یک قطعه زمین] خواهد بخشید.	D1-3: IDL
شهردار جدید که با رای جوانان انتخاب شده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانان مومن] [یک دستگاه پژوی چهارصد و پنج/یک قطعه زمین در حاشیه شمال شهر] خواهد بخشید.	D1-3: IDC
پرستار با نظر پزشک معالج موافق نبود. بعد از این که پزشک از اتاق خارج شد [به مریض] [چند لقمه‌ی بزرگ نان و پنیر گردو/چند قاشق غذاخوری روغن کرچک] خوردند.	D1-4: INL
پرستار با نظر پزشک معالج موافق نبود. بعد از این که پزشک از اتاق خارج شد [به مریض] [چند لقمه‌ی بزرگ نان و پنیر گردو/چند قاشق غذاخوری روغن کرچک] خوردند.	D1-4: INC
پزشک بعد از معاینه‌ی مریض دستور سرم داد، اما پرستار بعد از این که او از اتاق خارج شد [به مریض بیچاره که از درد به خود می‌پیچید] [یک تکه نان/چند قاشق روغن] خوردند.	D1-4: IDL
پزشک بعد از معاینه‌ی مریض دستور سرم داد، اما پرستار بعد از این که او از اتاق خارج شد [به مریض] [چند لقمه‌ی بزرگ نان و پنیر گردو/چند قاشق غذاخوری روغن کرچک] خوردند.	D1-4: IDC
زهره زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به پسرش که تمام روز درس می‌خواند] [یک لیوان خاک شیر/یک لیوان آب‌هویج] می‌خوراند.	D1-5: INL
زهره زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به پسرش] [یک لیوان بزرگ معجون ویتامینه/یک لیوان بزرگ آب‌هویج تازه] می‌خوراند.	D1-5: INC
مادر کبارش زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کبارش که تمام روز درس می‌خواند] [یک لیوان خاک شیر/یک لیوان آب‌هویج] می‌خوراند.	D1-5: IDL
مادر کبارش زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کبارش] [یک لیوان بزرگ معجون ویتامینه/یک لیوان بزرگ آب‌هویج تازه] می‌خوراند.	D1-5: IDC
دخترها همه جور غذایی را امتحان کرده بودند و از هر دری وارد شده بودند تا بالاخره دیشب توانستند [به مادر بزرگ که ضعیف و لاغر شده بود] [یک بشقاب عدس‌پلو/یک کاسه سوپ مرغ] بخوراند.	D1-6: INL
دخترها همه جور غذایی را امتحان کرده بودند و از هر دری وارد شده بودند تا بالاخره دیشب توانستند [به مادر بزرگ] [یک مقدار سیب‌زمینی سرخ‌کرده با سس/یک مقدار گوشت کبابی و سبزیجات] بخوراند.	D1-6: INC
مادر بزرگ بی‌اشتها بود و از کم‌غذایی ضعیف و لاغر شده بود. دخترها دیشب به زحمت توانستند [به مادر بزرگ که رنگ‌پریده و بی‌حال بود] [یک بشقاب عدس‌پلو/یک کاسه سوپ مرغ] بخوراند.	D1-6: IDL
مادر بزرگ بی‌اشتها بود و از کم‌غذایی ضعیف و لاغر شده بود. دخترها دیشب به زحمت توانستند [به مادر بزرگ] [یک مقدار سیب‌زمینی سرخ‌کرده با سس/یک مقدار گوشت کبابی و سبزیجات] بخوراند.	D1-6: IDC

D1-7:	این موقع سال خانم و آقای نیکفر مشغول خرید هستند. آنها هر سال قبل از عید [برای دخترشان که در برلین درس می‌خواند] [یک بسته لباس/یک بسته خوراکی] می‌فرستند.
D1-7:	این موقع سال خانم و آقای نیکفر مشغول خرید هستند. آنها هر سال قبل از عید [برای دخترشان] [یک بسته می‌کلوپی آجیل و شیرینی/یک بسته می‌کلوپی خوراکی و کتاب] می‌فرستند.
D1-7:	پرستو، دختر خانم و آقای نیکفر سالهاست به آلمان رفته است. آنها هر سال نزدیک عید [برای پرستو که در اشتوتگارد درس می‌خواند] [یک بسته لباس/یک بسته خوراکی] می‌فرستند.
D1-7:	پرستو، دختر خانم و آقای نیکفر سالهاست به آلمان رفته است. آنها هر سال نزدیک عید [برای پرستو] [یک بسته می‌کلوپی آجیل و شیرینی/یک بسته می‌کلوپی خوراکی و کتاب] می‌فرستند.
D1-8:	اولین بار بود که مریم ویزای اروپا گرفته بود و به ایتالیا سفر می‌کرد. او در اولین فرصت [برای آرزو خواهر کوچکش] [یک کارت تبریک/یک کارت پستال] فرستاد.
D1-8:	اولین بار بود که مریم ویزای اروپا گرفته بود و به ایتالیا سفر می‌کرد. او در اولین فرصت [برای آرزو] [چندین نقاشی از جاهای دیدنی شهر/چندین کارت پستال از جاهای دیدنی شهر] فرستاد.
D1-8:	همکار مریم نتوانسته بود ویزا بگیرد و مریم که تصمیم گرفته بود تنها به سفر برود در اولین فرصت [برای همکارش که زن زودرنجی بود] [یک کارت تبریک/یک کارت پستال] فرستاد.
D1-8:	همکار مریم نتوانسته بود ویزا بگیرد و مریم که تصمیم گرفته بود تنها به سفر برود در حین سفر [برای همکارش] [چندین نقاشی از جاهای دیدنی شهر/چندین کارت پستال از جاهای دیدنی شهر] فرستاد.
D1-9:	هاله دختر و لخرجی است، تمام حقوقش را یک هفته‌ای خرج می‌کند. دوباره همین دیروز [از یکی از دوستانش که حراج گذاشته بود] [یک کیف کوله‌پشتی/یک انگشتر عقیق] خرید.
D1-9:	هاله دختر و لخرجی است، تمام حقوقش را یک هفته‌ای خرج می‌کند. دوباره همین دیروز [از یکی از دوستانش] [یک کیف کوله‌پشتی و یک کیف پول چرم/یک جفت گوشواره و یک گردنبند مروارید] خرید.
D1-9:	هاله هم مهربان است هم ولخرج. وقتی سپیده نمایشگاه گذاشته بود، همان روز اول [از سپیده که امیدی هم به فروش نداشت] [یک کیف پول چرم/یک دستبند نقره] خرید.
D1-9:	هاله هم مهربان است هم ولخرج. وقتی سپیده نمایشگاه گذاشته بود، همان روز اول [از سپیده] [یک کیف پول چرم و یک کوله پشتی/یک انگشتر عقیق و یک زنجیر] خرید.
D1-10:	کامپیوتر شاهین با این که نسبتاً قدیمی بود هنوز خوب کار می‌کرد، ولی دیروز وسط یک کار فوری هنگ کرد و دیگر روشن نشد. شاهین هم مجبور شد [از یکی از دوستانش که در پایتخت مغازه داشت] [یک آیفون ارزان قیمت/یک لپتاپ ارزان قیمت] بخرد.
D1-10:	کامپیوتر شاهین با این که نسبتاً قدیمی بود هنوز خوب کار می‌کرد، ولی دیروز وسط یک کار فوری هنگ کرد و دیگر روشن نشد. شاهین هم مجبور شد [از یکی از دوستانش] [یک آیفون دست دوم نسبتاً ارزان/یک لپتاپ دست دوم نسبتاً ارزان] بخرد.
D1-10:	شاهین از دوستش خواست هر جور می‌شود کامپیوترش را تعمیر کند اما وقتی دید کامپیوترش تعمیر شدنی نیست مجبور شد [از دوستش که در پایتخت مغازه داشت] [یک آیفون نو/یک لپتاپ نو] بخرد.
D1-10:	شاهین از دوستش خواست هر جور می‌شود کامپیوترش را تعمیر کند اما وقتی دید کامپیوترش تعمیر شدنی نیست مجبور شد [از دوستش] [یک آیفون دست دوم ارزان/یک لپتاپ دست دوم ارزان] بخرد.
D1-11:	آقای پارسا مدتها بود برای سفرهای مکررش به مناطق کوهستانی به دنبال یک ماشین شاسی بلند بود. او قبل از بالا رفتن قیمت‌ها موفق شد [از یکی از همکارانش که پول لازم داشت] [یک مینی‌مایندر/یک نیسان پاترل] بخرد.
D1-11:	آقای پارسا مدتها بود برای سفرهای مکررش به مناطق کوهستانی به دنبال یک ماشین شاسی بلند بود. او قبل از بالا رفتن قیمت‌ها موفق شد [از یکی از همکارانش] [یک مینی‌مایندر صفر کیلومتر خاکستری/یک پاترل چهار در صفر کیلومتر سیاه] بخرد.
D1-11:	برادرزن آقای پارسا نمایندگی فروش ماشین داشت. به محض پایین آمدن قیمت‌ها آقای پارسا موفق شد [از برادرزنش که مرد ساده‌ای بود] [یک مینی‌مایندر/یک نیسان پاترل] بخرد.
D1-11:	برادرزن آقای پارسا نمایندگی فروش ماشین داشت. به محض پایین آمدن قیمت‌ها آقای پارسا موفق شد [از برادرزنش] [یک مینی‌مایندر صفر کیلومتر خاکستری/یک نیسان پاترل چهار در صفر کیلومتر] بخرد.
D1-12:	چند روز بود که حتی یک نفر هم داخل مغازه نشده بود. وقتی جمعه فروش فوق العاده گذاشتند جوان‌ترین فروشنده توانست [به اولین مشتری که قصد خرید هم نداشت] [یک یخچال ساید بای ساید/یک کاناپه‌ی چرمی] بفرشد.
D1-12:	چند روز بود که حتی یک نفر هم داخل مغازه نشده بود. وقتی جمعه فروش فوق العاده گذاشتند جوان‌ترین فروشنده توانست [به اولین مشتری] [یک یخچال ساید بای ساید مدل بالای سامسونگ/یک کاناپه‌ی تختخوابشو تمام چرم لوکس] بفرشد.
D1-12:	چند روز بود که هیچ فروشی نداشتند. وقتی پنجشنبه اولین مشتری وارد مغازه شد، مجرب‌ترین فروشنده به سراغ او رفت و توانست [به آن مشتری که قصد خرید هم نداشت] [یک جاروبرقی/یک دست میل چرم] بفرشد.
D1-12:	چند روز بود که هیچ فروشی نداشتند. وقتی پنجشنبه اولین مشتری وارد مغازه شد، مجرب‌ترین فروشنده به سراغ او رفت و توانست [به آن مشتری] [یک ماشین ظرفشویی رومیزی سامسونگ/یک کاناپه‌ی تختخوابشو تمام چرم] بفرشد.
D1-13:	صبا دختر زرنگی است. وقتی پول لازم داشت وسائل قدیمی‌اش را برای فروش گذاشت و همان روز اول [به خانم صاحبخانه که زن ولخرجی بود] [یک دست فنجان و نعلبکی/یک سرویس غذاخوری] فروخت.
D1-13:	صبا دختر زرنگی است. وقتی پول لازم داشت وسائل قدیمی‌اش را برای فروش گذاشت و همان روز اول [به همسایه‌ها] [یک دست کارد و چنگال نقره قدیمی/یک سرویس غذاخوری و چند قابلمه] فروخت.

D1-13:	صبا وقتی پول لازم داشت همه‌ی همسایه‌ها را دعوت کرد و وسائل قدیمی‌اش را برای فروش گذاشت. او که دختر زرنگی است در عرض چند ساعت [به همسایه‌ی پایینی که زن ولخرجی بود] [یک دست فنجان و نعلبکی/یک سرویس غذاخوری] فروخت.
D1-13:	صبا وقتی پول لازم داشت همه‌ی همسایه‌ها را دعوت کرد و وسائل قدیمی‌اش را برای فروش گذاشت. او که دختر زرنگی است در عرض چند ساعت [به همسایه‌ها] [یک دست کارد و چنگال نقره قدیمی/یک سرویس غذاخوری و چند قابلمه] فروخت.
D1-14:	بهادر شاید دیپلم نداشته باشد ولی کارش را خوب بلد است. چند وقت پیش در اوج کسادی بازار [به یک مهندس معمار جوان و تازمکار] [یک ساختمان تجاری/یک خانه‌ی کلنگی] فروخت.
D1-14:	بهادر شاید دیپلم نداشته باشد ولی کارش را خوب بلد است. چند وقت پیش در اوج کسادی بازار [به یک مهندس معمار] [یک ساختمان دوازدهطبقه‌ی تجاری/یک آپارتمان چهارطبقه‌ی کلنگی] فروخت.
D1-14:	چند وقت پیش در اوج کسادی بازار گذر یک مهندس معمار به بنگاه افتاده بود. بهادر هم که کارش را خیلی خوب بلد است [به آن مهندس معمار جوان و تازمکار] [یک ساختمان تجاری/یک خانه‌ی کلنگی] فروخت.
D1-14:	چند وقت پیش در اوج کسادی بازار گذر یک مهندس معمار به بنگاه افتاده بود. بهادر هم که کارش را خیلی خوب بلد است [به آن مهندس معمار] [یک ساختمان دوازدهطبقه‌ی تجاری/یک آپارتمان چهارطبقه‌ی کلنگی] فروخت.
D1-15:	نمایشگاه بسیار شلوغ و پرازدهام بود. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه‌هایی که بی‌تابی می‌کردند] [چندتا آدامس/چندتا آبنبات] داد.
D1-15:	نمایشگاه بسیار شلوغ و پرازدهام بود. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه‌ها] [یک تفنگ بادی و چندتا فشنگ/یک ورق کاغذ و چندتا مدارنگی] داد.
D1-15:	نمایشگاه بسیار شلوغ بود. بچه‌ها با پدرمادر هایشان سروکله می‌زدند. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه‌ها که بی‌تابی می‌کردند] [چندتا آدامس/چندتا آبنبات] داد.
D1-15:	نمایشگاه بسیار شلوغ بود. بچه‌ها با پدرمادر هایشان سروکله می‌زدند. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه‌ها] [یک تفنگ بادی و چندتا فشنگ/یک ورق کاغذ و چندتا مدارنگی] داد.
D1-16:	همگی برای عیددینی رفته بودند به خانه‌ی خانم و آقای افتخاری. خانم افتخاری موقع رفتن [به همه‌ی بچه‌هایی که آمده بودند] [یک مشت نخودچی/یک مشت آبنبات] داد.
D1-16:	همگی برای عیددینی رفته بودند به خانه‌ی خانم و آقای افتخاری. خانم افتخاری موقع رفتن [به همه‌ی بچه‌ها] [یک مشت برگه‌ی هلو و زردآلو/یک مشت آبنبات‌چوبی و شکلات] داد.
D1-16:	همه‌ی نوه‌ها برای عیددینی رفته بودند به خانه‌ی پدربزرگ و مادربزرگ. مادربزرگ موقع رفتن [به همه‌ی نوه‌ها که منتظر عیدی بودند] [یک مشت نخودچی/یک مشت آبنبات] داد.
D1-16:	همه‌ی نوه‌ها برای عیددینی رفته بودند به خانه‌ی پدربزرگ و مادربزرگ. مادربزرگ موقع رفتن [به همه‌ی نوه‌ها] [یک مشت برگه‌ی هلو و زردآلو/یک مشت آبنبات‌چوبی و شکلات] داد.
D1-17:	کولر از کار افتاده بود و هوا همین‌طور گرم‌تر می‌شد. آبدارچی وقتی گرما به اوج رسید [به مشتری‌ها که از گرما کلافه بودند] [یک فنجان چای/یک لیوان شربت] داد.
D1-17:	کولر از کار افتاده بود و هوا همین‌طور گرم‌تر می‌شد. آبدارچی وقتی گرما به اوج رسید [به مشتریها] [یک قاچ هندوانه‌ی بدون هسته‌ی خنک/یک لیوان شربت سکنجبین نگری] داد.
D1-17:	کولر از کار افتاده بود و مشتری‌ها ابراز نارضایتی می‌کردند. آبدارچی وقتی اعتراض‌ها به اوج رسید [به مشتریها که از گرما کلافه بودند] [یک فنجان چای/یک لیوان شربت] داد.
D1-17:	کولر از کار افتاده بود و مشتری‌ها ابراز نارضایتی می‌کردند. آبدارچی وقتی اعتراض‌ها به اوج رسید [به مشتریها] [یک قاچ هندوانه‌ی بدون هسته‌ی خنک/یک لیوان شربت سکنجبین نگری] داد.
D1-18:	ماشین سر چهارراه نگه داشت و راننده از پشت پنجره [از مرد جوانی که از ماشین پیاده شد] [یک پاکت در بسته/یک دسته هزارتومانی] گرفت.
D1-18:	ماشین سر چهارراه نگه داشت و راننده از پشت پنجره [از مرد جوانی] [یک پاکت بزرگ مهر و موم شده/یک دسته اسکناس پنج‌هزارتومانی] گرفت.
D1-18:	راننده سر چهارراه نگه داشت و مردی از پشت پنجره [از راننده که خیلی عصبانی بود] [یک پاکت در بسته/یک دسته هزارتومانی] گرفت.
D1-18:	راننده سر چهارراه نگه داشت و مردی از پشت پنجره [از راننده] [یک پاکت بزرگ مهر و موم شده/یک دسته اسکناس پنج‌هزارتومانی] گرفت.
D1-19:	ورود به سالن با کفش ممنوع بود و دانش‌آموزان برای وارد شدن به سالن [از مردی که مسوول سالن بود] [یک کفش نو/یک جفت جوراب] گرفتند.
D1-19:	ورود به سالن با کفش ممنوع بود و دانش‌آموزان برای وارد شدن به سالن [از دربان] [یک جفت کفش یکبار مصرف/یک جفت دمپایی روفرشی] گرفتند.
D1-19:	دربان گفت ورود به سالن با کفش ممنوع است. دانش‌آموزان هم برای وارد شدن به سالن [از دربان که مرد جوانی بود] [یک کفش نو/یک جفت جوراب] گرفتند.
D1-19:	دربان گفت ورود به سالن با کفش ممنوع است. دانش‌آموزان هم برای وارد شدن به سالن [از دربان] [یک جفت کفش یکبار مصرف/یک جفت دمپایی روفرشی] گرفتند.

D1-20:	دختر و پرسی که کنار ایستگاه ایستاده بودند نگران به نظر می‌رسیدند. آنها قبل از این که اتوبوس برسد [از دستفروشی که از آنجا رد می‌شد] [یک کیسه نایلون/یک تکه کاغذ] گرفتند.	INL
D1-20:	دختر و پرسی که کنار ایستگاه ایستاده بودند نگران به نظر می‌رسیدند. آنها قبل از این که اتوبوس برسد [از یک دستفروش] [یک کیسه پلاستیکی سیاه/یک تکه کاغذپاره و یک خودکار] گرفتند.	INC
D1-20:	مردی وارد ایستگاه شد و روی صندلی نشست. دختر و پرسی که کنار ایستگاه ایستاده بودند [از آن مرد که ریش بلندی داشت] [یک کیسه نایلون/یک خودکار] گرفتند.	IDL
D1-20:	مردی وارد ایستگاه شد و روی صندلی نشست. دختر و پرسی که کنار ایستگاه ایستاده بودند [از آن مرد] [یک کیسه پلاستیکی سیاه/یک تکه کاغذپاره و یک خودکار] گرفتند.	IDC

## Questionnaire "phrases à trous" 2 : Exp. D2

Item	Ind	
D2-1:	خانم سلطانی زن دست و دلبازی بود. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش که مدتها اجارمنشین بود] [گردنبند مروارید/آپارتمان نوساز] بخشید.	INL
D2-1:	خانم سلطانی زن دست و دلبازی بود. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش] [گردنبند مروارید/آپارتمان نوساز] بخشید.	INC
D2-1:	هنگامی که خواهرزاده‌ی خانم سلطانی از دواج کرد خانم سلطانی ثروت چندانی نداشت. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش که مدتها اجارمنشین بود] [گردنبند مروارید/آپارتمان نوساز] بخشید.	IDL
D2-1:	هنگامی که خواهرزاده‌ی خانم سلطانی از دواج کرد خانم سلطانی ثروت چندانی نداشت. وقتی املاک شوهرش بعد از مرگ به او رسید [به خواهرزاده‌اش] [گردنبند مروارید/آپارتمان نوساز] بخشید.	IDC
D2-2:	تمام فرزندان آقای تهرانی از ایران رفته بودند. آقای تهرانی که از تنهایی گریزان بود [به برادرزاده‌اش که مرتب به او سر می‌زد] [پیراهن مردانه/کتاب عکاسی] می‌بخشید.	INL
D2-2:	تمام فرزندان آقای تهرانی از ایران رفته بودند. آقای تهرانی که از تنهایی گریزان بود [به برادرزاده‌اش] [پیراهن مردانه/کتاب عکاسی] می‌بخشید.	INC
D2-2:	برادرزاده‌ی آقای تهرانی تنها کسی بود که به او سر می‌زد. آقای تهرانی که از فرزندانش ناامید شده بود [به برادرزاده‌اش که دانشجوی معماری بود] [پیراهن مردانه /کتاب عکاسی] می‌بخشید.	IDL
D2-2:	برادرزاده‌ی آقای تهرانی تنها کسی بود که به او سر می‌زد. آقای تهرانی هم هرازگاهی [به برادرزاده‌اش] [پیراهن مردانه /کتاب عکاسی] می‌بخشید.	IDC
D2-3:	شهردار جدید که با شعار حل مشکل بیکاری رای آورده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانانی که قصد ازدواج دارند] [جهاز کامل/سفر مکه] خواهد بخشید.	INL
D2-3:	شهردار جدید که با شعار حل مشکل بیکاری رای آورده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانان] [جهاز کامل/سفر مکه] خواهد بخشید.	INC
D2-3:	شهردار جدید که با رای جوانان انتخاب شده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانانی که قصد ازدواج دارند] [جهاز کامل/سفر مکه] خواهد بخشید.	IDL
D2-3:	شهردار جدید که با رای جوانان انتخاب شده بود، چند ماه بعد از به روی کار آمدن اعلام کرد که [به جوانان] [جهاز کامل/سفر مکه] خواهد بخشید.	IDC
D2-4:	پرستار با نظر پزشک معالج موافق نبود. بعد از این که پزشک از اتاق خارج شد [به مریض بیچاره که از درد به خود می‌پیچید] [چای ولرم/شیر ولرم] خوراند.	INL
D2-4:	پرستار با نظر پزشک معالج موافق نبود. بعد از این که پزشک از اتاق خارج شد [به مریض] [چای ولرم/شیر ولرم] خوراند.	INC
D2-4:	پزشک بعد از معاینه‌ی مریض دستور سرم داد، اما پرستار بعد از این که پزشک از اتاق خارج شد [به مریض بیچاره که از درد به خود می‌پیچید] [چای ولرم/شیر ولرم] خوراند.	IDL
D2-4:	پزشک بعد از معاینه‌ی مریض دستور سرم داد، اما پرستار بعد از این که پزشک از اتاق خارج شد [به مریض] [چای ولرم/شیر ولرم] خوراند.	IDC
D2-5:	زهره زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کپارش که تمام روز درس می‌خواند] [برگه‌ی هلو/روغن کرچک] می‌خوراند.	INL
D2-5:	زهره زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کپارش] [برگه‌ی هلو/روغن کرچک] می‌خوراند.	INC
D2-5:	مادر کپارش زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کپارش که تمام روز درس می‌خواند] [برگه‌ی هلو/روغن کرچک] می‌خوراند.	IDL
D2-5:	مادر کپارش زن وسواسی و سختگیری بود. زمان کنکور، هر روز صبح [به کپارش] [برگه‌ی هلو/روغن کرچک] می‌خوراند.	IDC

D2-6:	دخترها همه جور غذایی را امتحان کرده بودند و از هر دری وارد شده بودند تا بالاخره دیشب توانستند [به مادر بزرگ که ضعیف و بی‌حال شده بود] [حریری بادام / خوراک ماهیچه] بخوراند.
D2-6:	دخترها همه جور غذایی را امتحان کرده بودند و از هر دری وارد شده بودند تا بالاخره دیشب توانستند [به مادر بزرگ] [حریری بادام / خوراک ماهیچه] بخوراند.
D2-6:	مادر بزرگ بی‌اشتها بود و از کم‌غذایی لاغر شده بود. دخترها دیشب به زحمت توانستند [به مادر بزرگ که ضعیف و بی‌حال شده بود] [حریری بادام / خوراک ماهیچه] بخوراند.
D2-6:	مادر بزرگ بی‌اشتها بود و از کم‌غذایی لاغر شده بود. دخترها دیشب به زحمت توانستند [به مادر بزرگ] [حریری بادام / خوراک ماهیچه] بخوراند.
D2-7:	این موقع سال خانم و آقای نیکفر مشغول خرید هستند. آنها هر سال قبل از عید [برای دخترشان که در برلین درس می‌خواند] [لباس زمستانی/شیرینی خانگی] می‌فرستند.
D2-7:	این موقع سال خانم و آقای نیکفر مشغول خرید هستند. آنها هر سال قبل از عید [برای دخترشان] [لباس زمستانی/شیرینی خانگی] می‌فرستند.
D2-7:	ختر خانم و آقای نیکفر سالهاست به آلمان رفته است. آنها هر سال نزدیک عید [برای دخترشان که در برلین درس می‌خواند] [لباس زمستانی/شیرینی خانگی] می‌فرستند.
D2-7:	ختر خانم و آقای نیکفر سالهاست به آلمان رفته است. آنها هر سال نزدیک عید [برای دخترشان] [لباس زمستانی/شیرینی خانگی] می‌فرستند.
D2-8:	اولین بار بود که مریم ویزای اروپا گرفته بود و به ایتالیا سفر می‌کرد. او در اولین فرصت [برای همکارش که زن زودرنجی بود] [کارت تبریک سبعمدی/کارت پستال توریستی] فرستاد.
D2-8:	اولین بار بود که مریم ویزای اروپا گرفته بود و به ایتالیا سفر می‌کرد. او در اولین فرصت [برای همکارش] [کارت تبریک سبعمدی/کارت پستال توریستی] فرستاد.
D2-8:	همکار مریم نتوانسته بود ویزا بگیرد و مریم که تصمیم گرفته بود تنها به سفر برود در اولین فرصت [برای همکارش که زن زودرنجی بود] [کارت تبریک سبعمدی/کارت پستال توریستی] فرستاد.
D2-8:	همکار مریم نتوانسته بود ویزا بگیرد و مریم که تصمیم گرفته بود تنها به سفر برود در حین سفر [برای همکارش] [کارت تبریک سبعمدی/کارت پستال توریستی] فرستاد.
D2-9:	هاله دختر و لخرجی است که تمام حقوقش را یک هفته‌ای خرج می‌کند. دوباره همین دیروز [از سپیده که نمایشگاه گذاشته بود] [تقویم جیبی/دستبند نقره] خرید.
D2-9:	هاله دختر و لخرجی است که تمام حقوقش را یک هفته‌ای خرج می‌کند. دوباره همین دیروز [از سپیده] [تقویم جیبی/دستبند نقره] خرید.
D2-9:	هاله هم مهربان است هم و لخرج. وقتی سپیده نمایشگاه گذاشته بود، همان روز اول [از سپیده که امیدی به فروش نداشت] [تقویم جیبی/دستبند نقره] خرید.
D2-9:	هاله هم مهربان است هم و لخرج. وقتی سپیده نمایشگاه گذاشته بود، همان روز اول [از سپیده] [تقویم جیبی/دستبند نقره] خرید.
D2-10:	کامپیوتر شاهین با این که نسبتاً قدیمی بود هنوز خوب کار می‌کرد، ولی دیروز وسط یک کار فوری هنگ کرد و دیگر روشن نشد. شاهین هم مجبور شد [از دوستش که در پایتخت مغازه داشت] [آیفن نو/لپتاپ نو] بخرد.
D2-10:	کامپیوتر شاهین با این که نسبتاً قدیمی بود هنوز خوب کار می‌کرد، ولی دیروز وسط یک کار فوری هنگ کرد و دیگر روشن نشد. شاهین هم مجبور شد [از دوستش] [آیفن نو/لپتاپ نو] بخرد.
D2-10:	شاهین از دوستش خواست هر جور می‌شود کامپیوترش را تعمیر کند اما وقتی دید کامپیوترش تعمیر شدنی نیست مجبور شد [از دوستش که در پایتخت مغازه داشت] [آیفن نو/لپتاپ نو] بخرد.
D2-10:	شاهین از دوستش خواست هر جور می‌شود کامپیوترش را تعمیر کند اما وقتی دید کامپیوترش تعمیر شدنی نیست مجبور شد [از دوستش] [آیفن نو/لپتاپ نو] بخرد.
D2-11:	آقای پارسا مدتها بود برای سفرهای مکرر به مناطق کوهستانی به دنبال یک ماشین شاسی‌بلند بود. او قبل از بالا رفتن قیمت‌ها موفق شد [از برادرزنی که نمایشگاه ماشین داشت] [پاترول دو در/وانت پیکاپ] بخرد.
D2-11:	آقای پارسا مدتها بود برای سفرهای مکرر به مناطق کوهستانی به دنبال یک ماشین شاسی‌بلند بود. او قبل از بالا رفتن قیمت‌ها موفق شد [از برادرزنی] [پاترول دو در/وانت پیکاپ] بخرد.
D2-11:	برادرزن آقای پارسا نمایندگی فروش ماشین داشت. به محض پایین آمدن قیمت‌ها آقای پارسا موفق شد [از برادرزنی که آدم بانصافی بود] [پاترول دو در/وانت پیکاپ] بخرد.
D2-11:	برادرزن آقای پارسا نمایندگی فروش ماشین داشت. به محض پایین آمدن قیمت‌ها آقای پارسا موفق شد [از برادرزنی] [پاترول دو در/وانت پیکاپ] بخرد.
D2-12:	چند روز بود که حتی یک نفر هم داخل مغازه نشده بود. وقتی جمعه فروش فوق العاده گذاشتند جوان‌ترین فروشنده توانست [به اولین مشتری که قصد خرید نداشت] [بخچال-فریزر ساید بای ساید/کاناپه‌ی تختخوابشو] بفروشد.
D2-12:	چند روز بود که حتی یک نفر هم داخل مغازه نشده بود. وقتی جمعه فروش فوق العاده گذاشتند جوان‌ترین فروشنده توانست [به اولین مشتری] [بخچال-فریزر ساید بای ساید/کاناپه‌ی تختخوابشو] بفروشد.

D2-12:	چند روز بود که هیچ فروشی نداشتند. وقتی پنجشنبه اولین مشتری وارد مغازه شد، مجربترین فروشنده به سراغ او رفت و توانست [به آن مشتری که قصد خرید هم نداشت] [پنچال سایید بای سایید/میل تختخوابشو] بفروشد.
IDL	
D2-12:	چند روز بود که هیچ فروشی نداشتند. وقتی پنجشنبه اولین مشتری وارد مغازه شد، مجربترین فروشنده به سراغ او رفت و توانست [به آن مشتری] [پنچال سایید بای سایید/میل تختخوابشو] بفروشد.
IDC	
D2-13:	صبا دختر زرنگی است. وقتی پول لازم داشت و وسائل قدیمی اش را برای فروش گذاشت و همان روز اول [به زن صاحبخانه که آدم ولخرجی بود] [پنچال گل مرغی/کارد و چنگال نقره] فروخت.
INL	
D2-13:	صبا دختر زرنگی است. وقتی پول لازم داشت و وسائل قدیمی اش را برای فروش گذاشت و همان روز اول [به زن صاحبخانه] [پنچال گل مرغی/کارد و چنگال نقره] فروخت.
INC	
D2-13:	صبا وقتی پول لازم داشت همه همسایه ها را دعوت کرد و وسائل قدیمی اش را برای فروش گذاشت. او که دختر زرنگی است بلافاصله [به یکی از همسایه ها که آدم ولخرجی بود] [پنچال گل مرغی/کارد و چنگال نقره] فروخت.
IDL	
D2-13:	صبا وقتی پول لازم داشت همه همسایه ها را دعوت کرد و وسائل قدیمی اش را برای فروش گذاشت. او که دختر زرنگی است بلافاصله [به یکی از همسایه ها] [پنچال گل مرغی/کارد و چنگال نقره] فروخت.
IDC	
D2-14:	بهادر شاید دیپلم نداشته باشد ولی کارش را خوب بلد است. چند وقت پیش در اوج کساد بازار [به یک مهندس معمار جوان و تازمکار] [ساختمان نیمه کاره/آپارتمان کلنگی] فروخت.
INL	
D2-14:	بهادر شاید دیپلم نداشته باشد ولی کارش را خوب بلد است. چند وقت پیش در اوج کساد بازار [به یک مهندس معمار] [ساختمان نیمه کاره/آپارتمان کلنگی] فروخت.
INC	
D2-14:	چند وقت پیش در اوج کساد بازار گذر یک مهندس معمار به بنگاه افتاده بود. بهادر هم که کارش را خیلی خوب بلد است [به آن مهندس معمار جوان و تازمکار] [ساختمان نیمه کاره/آپارتمان کلنگی] فروخت.
IDL	
D2-14:	چند وقت پیش در اوج کساد بازار گذر یک مهندس معمار به بنگاه افتاده بود. بهادر هم که کارش را خیلی خوب بلد است [به آن مهندس معمار] [ساختمان نیمه کاره/آپارتمان کلنگی] فروخت.
IDC	
D2-15:	نمایشگاه بسیار شلوغ و پرازدهام بود. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه ها که بی تابی می کردند] [بستنی قیفی/آبنبات چوبی] داد.
INL	
D2-15:	نمایشگاه بسیار شلوغ و پرازدهام بود. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه ها] [بستنی قیفی/آبنبات چوبی] داد.
INC	
D2-15:	نمایشگاه بسیار شلوغ بود. بچه ها با پدر مادر هایشان سروکله می زدند. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه ها که بی تابی می کردند] [بستنی قیفی/آبنبات چوبی] داد.
IDL	
D2-15:	نمایشگاه بسیار شلوغ بود. بچه ها با پدر مادر هایشان سروکله می زدند. کارمند نمایشگاه که از سروصدا کلافه شده بود [به بچه ها] [بستنی قیفی/آبنبات چوبی] داد.
IDC	
D2-16:	همگی برای عیددینی رفته بودند به خانه ی خانم و آقای افتخاری. خانم افتخاری موقع رفتن [به همه ی بچه ها که منتظر عیدی بودند] [تخم مرغ شانسی/تخم مرغ رنگی] داد.
INL	
D2-16:	همگی برای عیددینی رفته بودند به خانه ی خانم و آقای افتخاری. خانم افتخاری موقع رفتن [به همه ی بچه ها] [تخم مرغ شانسی/تخم مرغ رنگی] داد.
INC	
D2-16:	همه ی بچه ها برای عیددینی رفته بودند به خانه ی پدر بزرگ و مادر بزرگ. مادر بزرگ موقع رفتن [به همه ی بچه ها که منتظر عیدی بودند] [تخم مرغ شانسی/تخم مرغ رنگی] داد.
IDL	
D2-16:	همه ی بچه ها برای عیددینی رفته بودند به خانه ی پدر بزرگ و مادر بزرگ. مادر بزرگ موقع رفتن [به همه ی بچه ها] [تخم مرغ شانسی/تخم مرغ رنگی] داد.
IDC	
D2-17:	کولر از کار افتاده بود و هوا همین طور گرم تر می شد. آبدارچی وقتی گرما به اوج رسید [به مشتری ها که از گرما کلافه بودند] [شربت بهلیمو/شربت آلبالو] داد.
INL	
D2-17:	کولر از کار افتاده بود و هوا همین طور گرم تر می شد. آبدارچی وقتی گرما به اوج رسید [به مشتری ها] [شربت بهلیمو/شربت آلبالو] داد.
INC	
D2-17:	کولر از کار افتاده بود و مشتری ها ابراز نارضایتی می کردند. آبدارچی وقتی اعتراضها به اوج رسید [به مشتری ها که از گرما کلافه بودند] [شربت بهلیمو/شربت آلبالو] داد.
IDL	
D2-17:	کولر از کار افتاده بود و مشتری ها ابراز نارضایتی می کردند. آبدارچی وقتی اعتراضها به اوج رسید [به مشتری ها] [شربت بهلیمو/شربت آلبالو] داد.
IDC	
D2-18:	ماشین کنار خیابان بود. بعد از مدتی راننده از ماشین پیاده شد و [از مردی جوانی که منتظر ایستاده بود] [تفنگ شکاری/چاقوی ضامن دار] گرفت.
INL	
D2-18:	ماشین کنار خیابان بود. بعد از مدتی راننده از ماشین پیاده شد و [از مرد جوانی] [تفنگ شکاری/چاقوی ضامن دار] گرفت.
INC	
D2-18:	مردی کنار خیابان ایستاده بود. راننده بعد از مدتی از ماشین پیاده شد و [از آن مرد جوان که بسیار درشت هیكل بود] [تفنگ شکاری/چاقوی ضامن دار] گرفت.
IDL	
D2-18:	مردی کنار خیابان ایستاده بود. راننده بعد از مدتی از ماشین پیاده شد و [از آن مرد جوان] [تفنگ شکاری/چاقوی ضامن دار] گرفت.
IDC	

D2-19:	ورود به سالن با کفش ممنوع بود و دانش آموزان برای وارد شدن به سالن [از مسوول سالن که مرد جوانی بود] [روکش پلاستیکی/دمپایی روفرشی] گرفتند.	INL
D2-19:	ورود به سالن با کفش ممنوع بود و دانش آموزان برای وارد شدن به سالن [از مسوول سالن] [روکش پلاستیکی/دمپایی روفرشی] گرفتند.	INC
D2-19:	مسوول سالن گفت ورود به سالن با کفش ممنوع است. دانش آموزان هم برای وارد شدن به سالن [از مسوول سالن که مرد جوانی بود] [روکش پلاستیکی/دمپایی روفرشی] گرفتند.	IDL
D2-19:	مسوول سالن گفت ورود به سالن با کفش ممنوع است. دانش آموزان هم برای وارد شدن به سالن [از مسوول سالن] [روکش پلاستیکی/دمپایی روفرشی] گرفتند.	IDC
D2-20:	ایستگاه خلوت بود. دختر و پسری که کنار ایستگاه ایستاده بودند [از دستفروشی که از آنجا رد می‌شد] [بلیت اتوبوس/کاغذ باطله] گرفتند.	INL
D2-20:	ایستگاه خلوت بود. دختر و پسری که کنار ایستگاه ایستاده بودند [از یک دستفروش] [بلیت اتوبوس/کاغذ باطله] گرفتند.	INC
D2-20:	دستفروشی کنار ایستگاه مشغول کار بود. دختر و پسری که منتظر اتوبوس بودند [از آن دستفروش که ریش بلندی داشت] [بلیت اتوبوس/کاغذ باطله] گرفتند.	IDL
D2-20:	دستفروشی کنار ایستگاه مشغول کار بود. دختر و پسری که منتظر اتوبوس بودند [از آن دستفروش] [بلیت اتوبوس/کاغذ باطله] گرفتند.	IDC

## Questionnaire "phrases à trous" 3 : Exp. D3

	Item	Ind
D3-1:	دو روز بعد از آغاز کار، سخنگوی شورای شهر [به شهردار تهران] [یک طرح جامع/یک پروژه جدید] [پیشنهاد کرد]	Yek
D3-1:	دو روز بعد از آغاز کار، سخنگوی شورای شهر [به شهردار تهران] [طرح جامعی/پروژه جدیدی] [پیشنهاد کرد]	i
D3-1:	دو روز بعد از آغاز کار، سخنگوی شورای شهر [به شهردار تهران] [طرح جامع را/پروژه جدید را] [پیشنهاد کرد]	Ra
D3-2:	دراولین روز سفر خود، رئیس‌جمهور [به نخست‌وزیر کشور همسایه] [یک برنامه توسعه مشترک/یک پروژه همکاری دوجانبه] [پیشنهاد کرد]	Yek
D3-2:	دراولین روز سفر خود، رئیس‌جمهور [به نخست‌وزیر کشور همسایه] [برنامه توسعه مشترکی/پروژه همکاری دوجانبه‌ای] [پیشنهاد کرد]	i
D3-2:	دراولین روز سفر خود، رئیس‌جمهور [به نخست‌وزیر کشور همسایه] [برنامه توسعه مشترک را/پروژه همکاری دوجانبه را] [پیشنهاد کرد]	Ra
D3-3:	بعد از صرف شام، سرآشپز [برای مهمان‌ها] [بسته‌ای/کیکی] [آورد]	Yek
D3-3:	بعد از صرف شام، سرآشپز [برای مهمان‌ها] [بسته را/کیک را] [آورد]	i
D3-3:	بعد از صرف شام، سرآشپز [برای مهمان‌ها] [یک بسته/یک کیک] [آورد]	Ra
D3-4:	وقتی کف زدن حضار تمام شد، مسوول جلسه [به سخنران] [شاخه گلی/دسته گلی] [داد]	Yek
D3-4:	وقتی کف زدن حضار تمام شد، مسوول جلسه [به سخنران] [شاخه گل را/دسته گل را] [داد]	i
D3-4:	وقتی کف زدن حضار تمام شد، مسوول جلسه [به سخنران] [یک شاخه گل/یک دسته گل] [داد]	Ra
D3-5:	بعد از چندساعت مشورت، داوران تصمیم گرفتند [به کارگردان جوان] [مدال افتخار را/جایزه ویژه را] [بدهند]	Yek
D3-5:	بعد از چندساعت مشورت، داوران تصمیم گرفتند [به کارگردان جوان] [یک مدال افتخار/یک جایزه ویژه] [بدهند]	i
D3-5:	بعد از چندساعت مشورت، داوران تصمیم گرفتند [به کارگردان جوان] [مدال افتخاری/جایزه ویژه‌ای] [بدهند]	Ra
D3-6:	یک روز قبل از تعطیلات، کارمندان [برای مدیر شرکت] [یک کارت تبریک/یک لوح تقدیر] [فرستادند]	Yek
D3-6:	یک روز قبل از تعطیلات، کارمندان [برای مدیر شرکت] [کارت تبریکی/لوح تقدیری] [فرستادند]	i
D3-6:	یک روز قبل از تعطیلات، کارمندان [برای مدیر شرکت] [کارت تبریک را/لوح تقدیر را] [فرستادند]	Ra

D3-7:	چند روز بعد از شروع ترم دوم، مدیر مدرسه
Yek	[برای مادر پدراها] [یک دعوتنامه/یک پرسشنامه] [فرستاد]
D3-7:	چند روز بعد از شروع ترم دوم، مدیر مدرسه
i	[برای مادر پدراها] [دعوتنامه‌ای/پرسشنامه‌ای] [فرستاد]
D3-7:	چند روز بعد از شروع ترم دوم، مدیر مدرسه
Ra	[برای مادر پدراها] [دعوتنامه را/پرسشنامه را] [فرستاد]
D3-8:	در دقایق آخر جلسه سوم کلاس، استاد ریاضی
Yek	[بین بچه‌ها] [یک کتاب/یک جزوه] [پخش کرد]
D3-8:	در دقایق آخر جلسه سوم کلاس، استاد ریاضی
i	[بین بچه‌ها] [کتابی/جزوه‌ای] [پخش کرد]
D3-8:	در دقایق آخر جلسه سوم کلاس، استاد ریاضی
Ra	[بین بچه‌ها] [کتاب را/جزوه را] [پخش کرد]
D3-9:	بعد از زنگ تفریح آخر، دانش‌آموزان
Yek	[به معلم] [یک نقاشی/یک کار دستی] [هدیه دادند]
D3-9:	بعد از زنگ تفریح آخر، دانش‌آموزان
i	[به معلم] [نقاشی‌ای/کار دستی‌ای] [هدیه دادند]
D3-9:	بعد از زنگ تفریح آخر، دانش‌آموزان
Ra	[به معلم] [نقاشی را/کار دستی را] [هدیه دادند]
D3-10:	برای تولد دهمسالگی، پدربزرگ
Yek	[به دوقولوها] [یک تیرکمان/یک تفنگچه] [هدیه داد]
D3-10:	برای تولد دهمسالگی، پدربزرگ
i	[به دوقولوها] [تیرکمانی/تفنگچه‌ای] [هدیه داد]
D3-10:	برای تولد دهمسالگی، پدربزرگ
Ra	[به دوقولوها] [تیرکمان را/تفنگچه را] [هدیه داد]
D3-11:	وقتی به پای کوه رسیدند، سرپرست
Yek	[به سرگروه] [یک نقشه/یک دوربین] [داد]
D3-11:	وقتی به پای کوه رسیدند، سرپرست
i	[به سرگروه] [نقشه‌ای/دوربینی] [داد]
D3-11:	وقتی به پای کوه رسیدند، سرپرست
Ra	[به سرگروه] [نقشه را/دوربین را] [داد]
D3-12:	وقتی کارهای پذیرش تمام شد، متصدی آزمایشگاه
Yek	[برای همراه بیمار] [یک ظرف آزمایش/یک لباس مخصوص] [آورد]
D3-12:	وقتی کارهای پذیرش تمام شد، متصدی آزمایشگاه
i	[برای همراه بیمار] [ظرف آزمایشی/لباس مخصوصی] [آورد]
D3-12:	وقتی کارهای پذیرش تمام شد، متصدی آزمایشگاه
Ra	[برای همراه بیمار] [ظرف آزمایش را/لباس مخصوص را] [آورد]
D3-13:	وقتی ورودی سالن خلوت شد، مربی
Yek	[به متصدی] [یک گواهی‌نامه/یک شکایت‌نامه] [داد]
D3-13:	وقتی ورودی سالن خلوت شد، مربی
i	[به متصدی] [گواهی‌نامه‌ای/شکایت‌نامه‌ای] [داد]
D3-13:	وقتی ورودی سالن خلوت شد، مربی
Ra	[به متصدی] [گواهی‌نامه را/شکایت‌نامه را] [داد]
D3-14:	در روز آخر مدرسه، بچه‌ها
Yek	[به بابای مدرسه] [یک دفتر خاطرات/یک آلبوم عکس] [دادند]
D3-14:	در روز آخر مدرسه، بچه‌ها
i	[به بابای مدرسه] [دفتر خاطراتی/آلبوم عکسی] [دادند]
D3-14:	در روز آخر مدرسه، بچه‌ها
Ra	[به بابای مدرسه] [دفتر خاطرات را/آلبوم عکس را] [دادند]
D3-15:	قبل از حمله پلیس، فعالان محیط زیست
Yek	[بین حاضران] [یک آگهی/یک اعلامیه] [پخش کردند]
D3-15:	قبل از حمله پلیس، فعالان محیط زیست
i	[بین حاضران] [آگهی‌ای/اعلامیه‌ای] [پخش کردند]
D3-15:	قبل از حمله پلیس، فعالان محیط زیست
Ra	[بین حاضران] [آگهی را/اعلامیه را] [پخش کردند]

## Questionnaire "phrases à trous" 3 : Exp. D4

Item	Ind
مدیر هتل چندین بار سفارش کرد حتماً [سر میز] [آش/گل] [بگذارند]	D4-1: Nu
مدیر هتل چندین بار سفارش کرد حتماً [سر میز] [آش رشته تیریزی/گل ارکیدۀ سفید] [بگذارند]	D4-1: Nu-M
صاحب رستوران گفت شاگردش رفته است [از انبار] [آرد/دوغ] [بیاورد]	D4-2: Nu
صاحب رستوران گفت شاگردش رفته است [از انبار] [آرد گندم سیوس‌دار/دوغ گازدار آبلی] [بیاورد]	D4-2: Nu-M
وزیر کشاورزی اعلام کرد این وزارتخانه به علت کمبود تولید داخلی [از هند] [برنج/خامه] [وارد خواهد کرد]	D4-3: Nu
وزیر کشاورزی اعلام کرد این وزارتخانه به علت کمبود تولید داخلی [از هند] [برنج باسماتی اعلاء/خامه رژیمی کمچرب] [وارد خواهد کرد]	D4-3: Nu-M
سختگوی دولت تاکید کرد دولت برای مبارزه با آلودگی [در تهران] [ماسک/بنزین] [توزیع خواهد کرد]	D4-4: Nu
سختگوی دولت تاکید کرد دولت برای مبارزه با آلودگی [در تهران] [ماسک تنفسی کش‌دار/بنزین مرغوب وارداتی] [توزیع خواهد کرد]	D4-4: Nu-M
ترانه دیروز گفت وحید قول داده است [از شمال] [پنیر/گلیم] [بیاورد]	D4-5: Nu
ترانه دیروز گفت وحید قول داده است [از شمال] [پنیر محلی گوسفندی/گلیم ابریشم ماشینی] [بیاورد]	D4-5: Nu-M
وزیر بهداشت و درمان اعلام کرد که این وزارتخانه برای مبارزه با سوءتغذیه [در مدارس] [شیر/نان] [توزیع خواهد کرد]	D4-6: Nu
وزیر بهداشت و درمان اعلام کرد که این وزارتخانه برای مبارزه با سوءتغذیه [در مدارس] [شیر پاکتی ویتامینه/نان شیرمال شکلاتی] [توزیع خواهد کرد]	D4-6: Nu-M
آروز گفت دلش می‌خواهد [توی ساندویچ] [تخم‌مرغ/خیارشور] [بگذارند]	D4-7: Nu
آروز گفت دلش می‌خواهد [توی ساندویچ] [تخم‌مرغ پخته پوست‌کنده/خیارشور حلقه‌حلقه شده] [بگذارند]	D4-7: Nu-M
سازمان هلال احمر اعلام کرد قبل از رسیدن فصل سرما [به اردوگاه] [بیسکویت/کیسه‌خواب] [می‌فرستند]	D4-8: Nu
سازمان هلال احمر اعلام کرد قبل از رسیدن فصل سرما [به اردوگاه] [بیسکویت ساقه‌طلایی مغزدار/کیسه‌خواب برزنتی ضدآب] [می‌فرستند]	D4-8: Nu-M
مسؤول برگزاری کنفرانس از دستیارش خواهش کرد [توی پوشه] [ماژیک/مداد] [بگذارند]	D4-9: Nu
مسؤول برگزاری کنفرانس از دستیارش خواهش کرد [توی پوشه] [ماژیک فلورسنت پاکشو/مداد قرمز تراشیده] [بگذارند]	D4-9: Nu-M
گویا آقای قنبری از پدرمادرها خواسته است [به خوابگاه] [زیرپوش/کنسرو] [بفرستند]	D4-10: Nu
گویا آقای قنبری از پدرمادرها خواسته است [به خوابگاه] [زیرپوش آستین‌بلند پشمی/کنسرو خورش قرم‌سبزی] [بفرستند]	D4-10: Nu-M

## Questionnaire "phrases à trous" 3 : Exp. T1

Item	Ind
مادربزرگ با نگرانی گفت شنیده است [خواهر پرستو] [شوهر مریم را] [ناراحت کرده است/خوشحال کرده است]	T1-1: A
مادربزرگ با نگرانی گفت شنیده است [خواهر پرستو] [شوهر مریم را که تازه از لباس عزا درآمده] [ناراحت کرده است/خوشحال کرده است]	T1-1: B
مادربزرگ با نگرانی گفت شنیده است [خواهر پرستو] [شوهر مریم را که تازه از لباس عزا درآمده را] [ناراحت کرده است/خوشحال کرده است]	T1-1: B'
مادربزرگ با نگرانی گفت شنیده است [مهمانی پرستو] [شوهر مریم را] [ناراحت کرده است/خوشحال کرده است]	T1-1: C
مادربزرگ با نگرانی گفت شنیده است [مهمانی پرستو] [شوهر مریم را که تازه از لباس عزا درآمده] [ناراحت کرده است/خوشحال کرده است]	T1-1: D
مینا می‌گوید آزاده عصبانی است چون [علیرضا] [بچه را] [بیدار کرده است/خسته کرده است]	T1-2: A
مینا می‌گوید آزاده عصبانی است چون [علیرضا] [بچه را که یک و نیم درجه تب دارد] [بیدار کرده است/خسته کرده است]	T1-2: B
مینا می‌گوید آزاده عصبانی است چون [علیرضا] [بچه که یک و نیم درجه تب دارد را] [خسته کرده است/بیدار کرده است]	T1-2: B'
مینا می‌گوید آزاده عصبانی است چون [سروصدا] [بچه را] [خسته کرده است/بیدار کرده است]	T1-2: C
مینا می‌گوید آزاده عصبانی است چون [سروصدا] [بچه را که یک و نیم درجه تب دارد] [خسته کرده است/بیدار کرده است]	T1-2: D
علی گفت پروین خانم خیلی معذب شد وقتی [پدربزرگ] [مهمان‌ها را] [از خواب پراند/از جا پراند]	T1-3: A
علی گفت پروین خانم خیلی معذب شد وقتی [پدربزرگ] [مهمان‌ها را که داشتند تلوزیون نگاه می‌کردند] [از خواب پراند/از جا پراند]	T1-3: B
علی گفت پروین خانم خیلی معذب شد وقتی [پدربزرگ] [مهمان‌ها که داشتند تلوزیون نگاه می‌کردند را] [از خواب پراند/از جا پراند]	T1-3: B'
علی گفت پروین خانم خیلی معذب شد وقتی [زنگ آیفن] [مهمان‌ها را] [از خواب پراند/از جا پراند]	T1-3: C
علی گفت پروین خانم خیلی معذب شد وقتی [زنگ آیفن] [مهمان‌ها را که داشتند تلوزیون نگاه می‌کردند] [از خواب پراند/از جا پراند]	T1-3: D
منوچهر گفت خسته است چون چند ساعت طول کشید تا [مادربزرگ] [بچه‌ها را] [بخواباند/بخنداند]	T1-4: A
منوچهر گفت خسته است چون چند ساعت طول کشید تا [مادربزرگ] [بچه‌ها را که بی‌وقفه گریه‌زاری می‌کردند] [بخواباند/بخنداند]	T1-4: B
منوچهر گفت خسته است چون چند ساعت طول کشید تا [مادربزرگ] [بچه‌ها را که بی‌وقفه گریه‌زاری می‌کردند] [بخواباند/بخنداند]	T1-4: B'
منوچهر گفت خسته است چون چند ساعت طول کشید تا [نوار قصه] [بچه‌ها را] [بخواباند/بخنداند]	T1-4: C
منوچهر گفت خسته است چون چند ساعت طول کشید تا [نوار قصه] [بچه‌ها را که بی‌وقفه گریه‌زاری می‌کردند] [بخواباند/بخنداند]	T1-4: D
احمد می‌گوید فقط مادربزرگ می‌داند چرا [دوست محسن] [خواهر نازنین را] [رنجانده است/گریانده است]	T1-5: A
احمد می‌گوید فقط مادربزرگ می‌داند چرا [دوست محسن] [خواهر نازنین را که تازه کنکور قبول شده] [رنجانده است/گریانده است]	T1-5: B
احمد می‌گوید فقط مادربزرگ می‌داند چرا [دوست محسن] [خواهر نازنین که تازه کنکور قبول شده را] [رنجانده است/گریانده است]	T1-5: B'
احمد می‌گوید فقط مادربزرگ می‌داند چرا [عکس محسن] [خواهر نازنین را] [گریانده است/رنجانده است]	T1-5: C

T1-5:	احمد می گوید فقط مادر بزرگ می داند چرا
D	[عکس محسن] [خواهر نازنین را که تازه کنکور قبول شده] [گریانده است/رنجانده است]
T1-6:	نگهبان می گوید هیچکس باور نمی کند
A	[نظافتچی] [مدیر عامل را] [غافلگیر کرده باشد/عصبانی کرده باشد]
T1-6:	نگهبان می گوید هیچکس باور نمی کند
B	[نظافتچی] [مدیر عامل را که روی میز مشغول رقصیدن بوده] [غافلگیر کرده باشد/عصبانی کرده باشد]
T1-6:	نگهبان می گوید هیچکس باور نمی کند
B'	[نظافتچی] [مدیر عامل که روی میز مشغول رقصیدن بوده را] [غافلگیر کرده باشد/عصبانی کرده باشد]
T1-6:	نگهبان می گوید هیچکس باور نمی کند
C	[جاروبرقی] [مدیر عامل را] [غافلگیر کرده باشد/عصبانی کرده باشد]
T1-6:	نگهبان می گوید هیچکس باور نمی کند
D	[جاروبرقی] [مدیر عامل را که روی میز مشغول رقصیدن بوده] [غافلگیر کرده باشد/عصبانی کرده باشد]
T1-7:	رضا گفت حتی خانم سلیمانی هم زیر خنده زد وقتی
A	[حسابدار] [آبدارچی را] [ترساند/رقصاند]
T1-7:	رضا گفت حتی خانم سلیمانی هم زیر خنده زد وقتی
B	[حسابدار] [آبدارچی را که مشغول جارو کردن بود] [ترساند/رقصاند]
T1-7:	رضا گفت حتی خانم سلیمانی هم زیر خنده زد وقتی
B'	[حسابدار] [آبدارچی که مشغول جارو کردن بود را] [ترساند/رقصاند]
T1-7:	رضا گفت حتی خانم سلیمانی هم زیر خنده زد وقتی
C	[زردوبرق] [آبدارچی را] [ترساند/رقصاند]
T1-7:	رضا گفت حتی خانم سلیمانی هم زیر خنده زد وقتی
D	[زردوبرق] [آبدارچی را که مشغول جارو کردن بود] [ترساند/رقصاند]
T1-8:	مادر همیشه می خندد وقتی مریم تعریف می کند چه طور آن روز
A	[پزشک درمانگاه] [مامور انتظامی را] [سرگرم کرد/سردرگم کرد]
T1-8:	مادر همیشه می خندد وقتی مریم تعریف می کند چه طور آن روز
B	[پزشک درمانگاه] [مامور انتظامی را که با دقت همه جا را می گشت] [سرگرم کرد/سردرگم کرد]
T1-8:	مادر همیشه می خندد وقتی مریم تعریف می کند چه طور آن روز
B'	[پزشک درمانگاه] [مامور انتظامی که با دقت همه جا را می گشت را] [سرگرم کرد/سردرگم کرد]
T1-8:	مادر همیشه می خندد وقتی مریم تعریف می کند چه طور آن روز
C	[کارت پایان خدمت] [مامور انتظامی را] [سرگرم کرد/سردرگم کرد]
T1-8:	مادر همیشه می خندد وقتی مریم تعریف می کند چه طور آن روز
D	[کارت پایان خدمت] [مامور انتظامی را که با دقت همه جا را می گشت] [سرگرم کرد/سردرگم کرد]
T1-9:	معلمها تعریف کردند که بچهها گریهکنان منتظر بودند تا
A	[مامور آتش نشانی] [شاگرد اول کلاس را] [نجات دهد/مداوا کند]
T1-9:	معلمها تعریف کردند که بچهها گریهکنان منتظر بودند تا
B	[مامور آتش نشانی] [شاگرد اول کلاس را که بی حال روی زمین افتاده بود] [نجات دهد/مداوا کند]
T1-9:	معلمها تعریف کردند که بچهها گریهکنان منتظر بودند تا
B'	[مامور آتش نشانی] [شاگرد اول کلاس که بی حال روی زمین افتاده بود را] [نجات دهد/مداوا کند]
T1-9:	معلمها تعریف کردند که بچهها گریهکنان منتظر بودند تا
C	[تنفس مصنوعی] [شاگرد اول کلاس را] [نجات دهد/مداوا کند]
T1-9:	معلمها تعریف کردند که بچهها گریهکنان منتظر بودند تا
D	[تنفس مصنوعی] [شاگرد اول کلاس را که بی حال روی زمین افتاده بود] [نجات دهد/مداوا کند]
T1-10:	سپیده می گوید قاضی امیدوار است
A	[وکیل متهم] [خانواده مقتول را] [تشویق کند/راضی کند]
T1-10:	سپیده می گوید قاضی امیدوار است
B	[وکیل متهم] [خانواده مقتول را که چندماه است تحت فشار هستند] [تشویق کند/راضی کند]
T1-10:	سپیده می گوید قاضی امیدوار است
B'	[وکیل متهم] [خانواده مقتول که چندماه است تحت فشار هستند را] [تشویق کند/راضی کند]
T1-10:	سپیده می گوید قاضی امیدوار است
C	[اتفاقات اخیر] [خانواده مقتول را] [تشویق کند/راضی کند]
T1-10:	سپیده می گوید قاضی امیدوار است
D	[اتفاقات اخیر] [خانواده مقتول را که چندماه است تحت فشار هستند] [تشویق کند/راضی کند]

T1-11:	بچه‌ها می‌گویند مسؤل اردو مطمئن است
A	[بابای مدرسه] [مدیر مدرسه را] [منصرف می‌کند/قانع می‌کند]
T1-11:	بچه‌ها می‌گویند مسؤل اردو مطمئن است
B	[بابای مدرسه] [مدیر مدرسه را که تصمیم گرفته است به اردو بیاید] [منصرف می‌کند/قانع می‌کند]
T1-11:	بچه‌ها می‌گویند مسؤل اردو مطمئن است
B'	[بابای مدرسه] [مدیر مدرسه که تصمیم گرفته است به اردو بیاید را] [منصرف می‌کند/قانع می‌کند]
T1-11:	بچه‌ها می‌گویند مسؤل اردو مطمئن است
C	[ترافیک مسیر] [مدیر مدرسه را] [منصرف می‌کند/قانع می‌کند]
T1-11:	بچه‌ها می‌گویند مسؤل اردو مطمئن است
D	[ترافیک مسیر] [مدیر مدرسه را که تصمیم گرفته است به اردو بیاید] [منصرف می‌کند/قانع می‌کند]
T1-12:	گزارشگر حاضر در سالن می‌گوید کسی انتظار نداشت
A	[مجری برنامه] [وزیر ارشاد را] [به گریه بیاندازد/به خنده بیاندازد]
T1-12:	گزارشگر حاضر در سالن می‌گوید کسی انتظار نداشت
B	[مجری برنامه] [وزیر ارشاد را که برای توزیع جوایز آمده بود] [به گریه بیاندازد/به خنده بیاندازد]
T1-12:	گزارشگر حاضر در سالن می‌گوید کسی انتظار نداشت
B'	[مجری برنامه] [وزیر ارشاد که برای توزیع جوایز آمده بود را] [به گریه بیاندازد/به خنده بیاندازد]
T1-12:	گزارشگر حاضر در سالن می‌گوید کسی انتظار نداشت
C	[فیلم منتخب] [وزیر ارشاد را] [به گریه بیاندازد/به خنده بیاندازد]
T1-12:	گزارشگر حاضر در سالن می‌گوید کسی انتظار نداشت
D	[فیلم منتخب] [وزیر ارشاد را که برای توزیع جوایز آمده بود] [به گریه بیاندازد/به خنده بیاندازد]
T1-13:	رئیس دانشکده می‌گفت دانشجویان باور نمی‌کردند
A	[استاد راهنما] [استاد مشاور را] [ناامید کند/متقاعد کند]
T1-13:	رئیس دانشکده می‌گفت دانشجویان باور نمی‌کردند
B	[استاد راهنما] [استاد مشاور را که یکی از سختگیرترین استادان است] [ناامید کند/متقاعد کند]
T1-13:	رئیس دانشکده می‌گفت دانشجویان باور نمی‌کردند
B'	[استاد راهنما] [استاد مشاور که یکی از سختگیرترین استادان است را] [ناامید کند/متقاعد کند]
T1-13:	رئیس دانشکده می‌گفت دانشجویان باور نمی‌کردند
C	[صورت جلسه] [استاد مشاور را] [ناامید کند/متقاعد کند]
T1-13:	رئیس دانشکده می‌گفت دانشجویان باور نمی‌کردند
D	[صورت جلسه] [استاد مشاور را که یکی از سختگیرترین استادان است] [ناامید کند/متقاعد کند]
T1-14:	پلیس می‌گوید به نظر همسایه‌ها ممکن نیست
A	[پستی] [دربان را] [مریض کرده باشد/مسموم کرده باشد]
T1-14:	پلیس می‌گوید به نظر همسایه‌ها ممکن نیست
B	[پستی] [دربان را که مرد قوی و ورزشکاری است] [مریض کرده باشد/مسموم کرده باشد]
T1-14:	پلیس می‌گوید به نظر همسایه‌ها ممکن نیست
B'	[پستی] [دربان که مرد قوی و ورزشکاری است را] [مریض کرده باشد/مسموم کرده باشد]
T1-14:	پلیس می‌گوید به نظر همسایه‌ها ممکن نیست
C	[پیغیاف] [دربان را] [مریض کرده باشد/مسموم کرده باشد]
T1-14:	پلیس می‌گوید به نظر همسایه‌ها ممکن نیست
D	[پیغیاف] [دربان را که مرد قوی و ورزشکاری است] [مریض کرده باشد/مسموم کرده باشد]
T1-15:	مهرنوش گفت خواهرش فکر می‌کند
A	[سرایدار ساختمان] [آقای آذرمنش را] [آزار می‌دهد/اذیت می‌کند]
T1-15:	مهرنوش گفت خواهرش فکر می‌کند
B	[سرایدار ساختمان] [آقای آذرمنش را که همیشه مشغول مطالعه است] [آزار می‌دهد/اذیت می‌کند]
T1-15:	مهرنوش گفت خواهرش فکر می‌کند
B'	[سرایدار ساختمان] [آقای آذرمنش را که همیشه مشغول مطالعه است] [آزار می‌دهد/اذیت می‌کند]
T1-15:	مهرنوش گفت خواهرش فکر می‌کند
C	[صدای موتورخانه] [آقای آذرمنش را] [آزار می‌دهد/اذیت می‌کند]
T1-15:	مهرنوش گفت خواهرش فکر می‌کند
D	[صدای موتورخانه] [آقای آذرمنش را که همیشه مشغول مطالعه است] [آزار می‌دهد/اذیت می‌کند]
T1-16:	آقای علیزاده گفت مدیر کارخانه حدس می‌زد
A	[نام‌رسان] [رئیس‌دفتر را] [بی‌هوش کرده باشد/خفه کرده باشد]
T1-16:	آقای علیزاده گفت مدیر کارخانه حدس می‌زد
B	[نام‌رسان] [رئیس‌دفتر را که بعد از ظهرها پشت میز می‌خوابید] [بی‌هوش کرده باشد/خفه کرده باشد]

T1-16:	آقای علیزاده گفت مدیر کارخانه حدس می‌زد
B'	[نام‌رسان] [رئیس‌دفتر که بعد از ظهرها پشت میز می‌خوابید را] [بی‌هوش کرده باشد/خفه کرده باشد]
T1-16:	آقای علیزاده گفت مدیر کارخانه حدس می‌زد
C	[گازکربنیک] [رئیس‌دفتر را] [بی‌هوش کرده باشد/خفه کرده باشد]
T1-16:	آقای علیزاده گفت مدیر کارخانه حدس می‌زد
D	[گازکربنیک] [رئیس‌دفتر را که بعد از ظهرها پشت میز می‌خوابید] [بی‌هوش کرده باشد/خفه کرده باشد]
T1-17:	خبرنگار می‌گوید گروه‌های بدون معطلی تاکید کرد
A	[زندانی فراری] [سرباز وظیفه را] [زخمی کرده است/مجروح کرده است]
T1-17:	خبرنگار می‌گوید گروه‌های بدون معطلی تاکید کرد
B	[زندانی فراری] [سرباز وظیفه را که در حال نگهداری بوده] [زخمی کرده است/مجروح کرده است]
T1-17:	خبرنگار می‌گوید گروه‌های بدون معطلی تاکید کرد
B'	[زندانی فراری] [سرباز وظیفه را که در حال نگهداری بوده را] [زخمی کرده است/مجروح کرده است]
T1-17:	خبرنگار می‌گوید گروه‌های بدون معطلی تاکید کرد
C	[ترکش خمپاره] [سرباز وظیفه را] [زخمی کرده است/مجروح کرده است]
T1-17:	خبرنگار می‌گوید گروه‌های بدون معطلی تاکید کرد
D	[ترکش خمپاره] [سرباز وظیفه را که در حال نگهداری بوده] [زخمی کرده است/مجروح کرده است]
T1-18:	شیرین فکر می‌کند برخلاف آنچه همه می‌گویند
A	[خواهر مهشید] [برادر مازیار را] [کمک کرد/معالجه کرد]
T1-18:	شیرین فکر می‌کند برخلاف آنچه همه می‌گویند
B	[خواهر مهشید] [برادر مازیار را که از نوجوانی افسردگی شدید داشت] [کمک کرد/معالجه کرد]
T1-18:	شیرین فکر می‌کند برخلاف آنچه همه می‌گویند
B'	[خواهر مهشید] [برادر مازیار را که از نوجوانی افسردگی شدید داشت را] [کمک کرد/معالجه کرد]
T1-18:	شیرین فکر می‌کند برخلاف آنچه همه می‌گویند
C	[معجون مهشید] [برادر مازیار را] [معالجه کرد/کمک کرد]
T1-18:	شیرین فکر می‌کند برخلاف آنچه همه می‌گویند
D	[معجون مهشید] [برادر مازیار را که از نوجوانی افسردگی شدید داشت] [معالجه کرد/کمک کرد]
T1-19:	سینا گفت پزشک اورژانس امیدوار بود هرچه زودتر
A	[پرستار] [مریض را] [آرام کند/ساکت کند]
T1-19:	سینا گفت پزشک اورژانس امیدوار بود هرچه زودتر
B	[پرستار] [مریض را که از درد فریاد می‌کشید] [آرام کند/ساکت کند]
T1-19:	سینا گفت پزشک اورژانس امیدوار بود هرچه زودتر
B'	[پرستار] [مریض که از درد فریاد می‌کشید را] [آرام کند/ساکت کند]
T1-19:	سینا گفت پزشک اورژانس امیدوار بود هرچه زودتر
C	[مسکن] [مریض را] [آرام کند/ساکت کند]
T1-19:	سینا گفت پزشک اورژانس امیدوار بود هرچه زودتر
D	[مسکن] [مریض را که از درد فریاد می‌کشید] [آرام کند/ساکت کند]
T1-20:	سمیرا می‌گوید متصدی سالن گفت مطمئن باشید
A	[عروسک‌گردان] [دانش‌آموزان را] [هیجان‌زده می‌کند/سرگرم می‌کند]
T1-20:	سمیرا می‌گوید متصدی سالن گفت مطمئن باشید
B	[عروسک‌گردان] [دانش‌آموزان را که گرسنه و خسته بودند] [هیجان‌زده می‌کند/سرگرم می‌کند]
T1-20:	سمیرا می‌گوید متصدی سالن گفت مطمئن باشید
B'	[عروسک‌گردان] [دانش‌آموزان که گرسنه و خسته بودند را] [هیجان‌زده می‌کند/سرگرم می‌کند]
T1-20:	سمیرا می‌گوید متصدی سالن گفت مطمئن باشید
C	[خیمه‌شب بازی] [دانش‌آموزان را] [هیجان‌زده می‌کند/سرگرم می‌کند]
T1-20:	سمیرا می‌گوید متصدی سالن گفت مطمئن باشید
D	[خیمه‌شب بازی] [دانش‌آموزان را که گرسنه و خسته بودند] [هیجان‌زده می‌کند/سرگرم می‌کند]

## Questionnaire Jugement d'acceptabilité : Exp. Coordination

Item	Ind
مریم تو انباری مادر بزرگ این آلبوم عکسو پیدا کرده	A1-1:C
مریم تو انباری مادر بزرگ چندتا کاست ویدئو و این آلبوم عکسو پیدا کرده	A1-1:Coor-ord1
مریم تو انباری مادر بزرگ این آلبوم عکسو و چندتا کاست ویدئو پیدا کرده	A1-1:Coor-ord2
مریم تو انباری مادر بزرگ چندتا کاست ویدئو پیدا کرده و این آلبوم عکسو	A1-1:Post-ord1
مریم تو انباری مادر بزرگ این آلبوم عکسو پیدا کرده و چندتا کاست ویدئو	A1-1:Post-ord2
به نظر من فقط این کفش کتونیا رو باخودت بیار	A1-2:C
به نظر من فقط به دست لباس راحتی و این کفش کتونیا رو باخودت بیار	A1-2:Coor-ord1
به نظر من فقط این کفش کتونیا رو و به دست لباس راحتی باخودت بیار	A1-2:Coor-ord2
به نظر من فقط به دست لباس راحتی باخودت بیار و این کفش کتونیا رو	A1-2:Post-ord1
به نظر من فقط این کفش کتونیا رو باخودت بیار و به دست لباس راحتی	A1-2:Post-ord2
برای ثبت نام کافی است فرم تکمیل شده را برای ما پست کنید	A1-3:C
برای ثبت نام کافی است یک قطعه عکس ۶x۴ و فرم تکمیل شده را برای ما پست کنید	A1-3:Coor-ord1
برای ثبت نام کافی است فرم تکمیل شده را و یک قطعه عکس ۶x۴ برای ما پست کنید	A1-3:Coor-ord2
برای ثبت نام کافی است یک قطعه عکس ۶x۴ برای ما پست کنید و فرم تکمیل شده را	A1-3:Post-ord1
برای ثبت نام کافی است فرم تکمیل شده را برای ما پست کنید و یک قطعه عکس ۶x۴	A1-3:Post-ord2
آزاده مجبور شده برای قسط خونه حلقه ازدواجشو بفروشه	A1-4:C
آزاده مجبور شده برای قسط خونه به تعداد سگه و حلقه ازدواجشو بفروشه	A1-4:Coor-ord1
آزاده مجبور شده برای قسط خونه حلقه ازدواجشو و به تعداد سگه بفروشه	A1-4:Coor-ord2
آزاده مجبور شده برای قسط خونه به تعداد سگه بفروشه و حلقه ازدواجشو	A1-4:Post-ord1
آزاده مجبور شده برای قسط خونه حلقه ازدواجشو بفروشه و به تعداد سگه	A1-4:Post-ord2
اگه پول کم داری فقط قرص قلب بابا رو بگیر	A1-5:C
اگه پول کم داری فقط به بسته استامینافن و قرص قلب بابا رو بگیر	A1-5:Coor-ord1
اگه پول کم داری فقط قرص قلب بابا رو و به بسته استامینافن بگیر	A1-5:Coor-ord2
اگه پول کم داری فقط به بسته استامینافن بگیر و قرص قلب بابا رو	A1-5:Post-ord1
اگه پول کم داری فقط قرص قلب بابا رو بگیر و به بسته استامینافن	A1-5:Post-ord2
برای امشب این کیک شکلاتی رو گرفتم	A1-6:C
برای امشب به کیلو بستنی وانیلی و این کیک شکلاتی رو گرفتم	A1-6:Coor-ord1
برای امشب این کیک شکلاتی رو و به کیلو بستنی وانیلی گرفتم	A1-6:Coor-ord2
برای امشب به کیلو بستنی وانیلی گرفتم و این کیک شکلاتی رو	A1-6:Post-ord1
برای امشب این کیک شکلاتی رو گرفتم و به کیلو بستنی وانیلی	A1-6:Post-ord2
اگه می‌خوای مینو رو با خودت ببری جعبه مدارنگیاشو بردار	A1-7:C
اگه می‌خوای مینو رو با خودت ببری چندتا ورق کاغذ و جعبه مدارنگیاشو بردار	A1-7:Coor-ord1
اگه می‌خوای مینو رو با خودت ببری جعبه مدارنگیاشو و چندتا ورق کاغذ بردار	A1-7:Coor-ord2
اگه می‌خوای مینو رو با خودت ببری چندتا ورق کاغذ بردار و جعبه مدارنگیاشو	A1-7:Post-ord1
اگه می‌خوای مینو رو با خودت ببری جعبه مدارنگیاشو بردار و چندتا ورق کاغذ	A1-7:Post-ord2
نادر از شیراز فقط این گلیم رو آورده	A1-8:C
نادر از شیراز فقط به قالی درب و داغون و این گلیم رو آورده	A1-8:Coor-ord1
نادر از شیراز فقط این گلیم رو و به قالی درب و داغون آورده	A1-8:Coor-ord2
نادر از شیراز فقط به قالی درب و داغون آورده و این گلیم رو	A1-8:Post-ord1
نادر از شیراز فقط این گلیم رو آورده و به قالی درب و داغون	A1-8:Post-ord2

از حراج این پیرهن مردونه رو خریدم	A1-9:C
از حراج دوتا شلوارجین و این پیرهن مردونه رو خریدم	A1-9:Coor-ord1
از حراج این پیرهن مردونه رو و دوتا شلوارجین خریدم	A1-9:Coor-ord2
از حراج دوتا شلوارجین خریدم و این پیرهن مردونه رو	A1-9:Post-ord1
از حراج این پیرهن مردونه رو خریدم و دوتا شلوارجین	A1-9:Post-ord2
دیروز با پریسا رفتیم دست‌دوم فروشی، من این پالتو رو خریدم ولی پریسا چیزی نخرید	A1-10:C
دیروز با پریسا رفتیم دست‌دوم فروشی، من به کت مخمل نسبتاً نو و این پالتو رو خریدم ولی پریسا چیزی نخرید	A1-10:Coor-ord1
دیروز با پریسا رفتیم دست‌دوم فروشی، من این پالتو رو و به کت مخمل نسبتاً نو خریدم ولی پریسا چیزی نخرید	A1-10:Coor-ord2
دیروز با پریسا رفتیم دست‌دوم فروشی، من به کت مخمل نسبتاً نو خریدم و این پالتو رو ولی پریسا چیزی نخرید	A1-10:Post-ord1
دیروز با پریسا رفتیم دست‌دوم فروشی، من این پالتو رو خریدم و به کت مخمل نسبتاً نو ولی پریسا چیزی نخرید	A1-10:Post-ord2
امروز توی خیابون این تراول چکا رو پیدا کردم	A1-11:C
امروز توی خیابون چند دسته اسکناس هزاری و این تراول چکا رو پیدا کردم	A1-11:Coor-ord1
امروز توی خیابون این تراول چکا رو و چند دسته اسکناس هزاری پیدا کردم	A1-11:Coor-ord2
امروز توی خیابون چند دسته اسکناس هزاری پیدا کردم و این تراول چکا رو	A1-11:Post-ord1
امروز توی خیابون این تراول چکا رو پیدا کردم و چند دسته اسکناس هزاری	A1-11:Post-ord2
گفت چیزی لازم نیست بیاریم فقط کلید خونه رو بردار زود بریم	A1-12:C
گفت چیزی لازم نیست بیاریم فقط به مقدار پول نقد و کلید خونه رو بردار زود بریم	A1-12:Coor-ord1
گفت چیزی لازم نیست بیاریم فقط کلید خونه رو و به مقدار پول نقد بردار زود بریم	A1-12:Coor-ord2
گفت چیزی لازم نیست بیاریم فقط به مقدار پول نقد بردار و کلید خونه رو زود بریم	A1-12:Post-ord1
گفت چیزی لازم نیست بیاریم فقط کلید خونه رو بردار و به مقدار پول نقد زود بریم	A1-12:Post-ord2
من هنوز باید حوله تنیمو بذارم تو چمدون	A1-13:C
من هنوز باید به دست ملافه و حوله تنیمو بذارم تو چمدون	A1-13:Coor-ord1
من هنوز باید حوله تنیمو و به دست ملافه بذارم تو چمدون	A1-13:Coor-ord2
من هنوز باید به دست ملافه بذارم تو چمدون و حوله تنیمو	A1-13:Post-ord1
من هنوز باید حوله تنیمو بذارم تو چمدون و به دست ملافه	A1-13:Post-ord2
برای بچه‌ها این جاکلیدی‌ها را سوغاتی آورده‌ام	A1-14:C
برای بچه‌ها چند بسته شکلات و این جاکلیدی‌ها را سوغاتی آورده‌ام	A1-14:Coor-ord1
برای بچه‌ها این جاکلیدی‌ها را و چند بسته شکلات سوغاتی آورده‌ام	A1-14:Coor-ord2
برای بچه‌ها چند بسته شکلات سوغاتی آورده‌ام و این جاکلیدی‌ها را	A1-14:Post-ord1
برای بچه‌ها این جاکلیدی‌ها را سوغاتی آورده‌ام و چند بسته شکلات	A1-14:Post-ord2
به دستور پزشک مادرم شب‌ها فقط غذای رژیمی مخصوصش را می‌خورد	A1-15:C
به دستور پزشک مادرم شب‌ها فقط کمی ماست و غذای رژیمی مخصوصش را می‌خورد	A1-15:Coor-ord1
به دستور پزشک مادرم شب‌ها فقط غذای رژیمی مخصوصش را و کمی ماست می‌خورد	A1-15:Coor-ord2
به دستور پزشک مادرم شب‌ها فقط کمی ماست می‌خورد و غذای رژیمی مخصوصش را	A1-15:Post-ord1
به دستور پزشک مادرم شب‌ها فقط غذای رژیمی مخصوصش را می‌خورد و کمی ماست	A1-15:Post-ord2
افسانه برای عروسی این کفش پاشنه بلند رو به نگار قرض داده	A1-16:C
افسانه برای عروسی به دست لباس شب و این کفش پاشنه بلند رو به نگار قرض داده	A1-16:Coor-ord1
افسانه برای عروسی این کفش پاشنه بلند رو و به دست لباس شب به نگار قرض داده	A1-16:Coor-ord2
افسانه برای عروسی به دست لباس شب به نگار قرض داده و این کفش پاشنه بلند رو	A1-16:Post-ord1
افسانه برای عروسی این کفش پاشنه بلند رو به نگار قرض داده و به دست لباس شب	A1-16:Post-ord2
امیرحسین بعد از ظهر نشست جلوی تلوزیون همه تخمه‌ها رو خورد	A1-17:C
امیرحسین بعد از ظهر نشست جلوی تلوزیون به کاسه گوجه‌سبز و همه تخمه‌ها رو خورد	A1-17:Coor-ord1
امیرحسین بعد از ظهر نشست جلوی تلوزیون همه تخمه‌ها رو و به کاسه گوجه‌سبز خورد	A1-17:Coor-ord2

امیرحسین بعد از ظهر نشست جلوی تلوزیون به کاسه گوجه‌سبز خورد و همه تخمه‌ها رو	A1-17:Post-ord1
امیرحسین بعد از ظهر نشست جلوی تلوزیون همه تخمه‌ها رو خورد و به کاسه گوجه‌سبز	A1-17:Post-ord2
فراموش کردم برای سپهر چیزی بخرم ولی برای مژگان این تابلو را خریدم	A1-18:C
فراموش کردم برای سپهر چیزی بخرم ولی برای مژگان چندتا کتاب و این تابلو را خریدم	A1-18:Coor-ord1
فراموش کردم برای سپهر چیزی بخرم ولی برای مژگان این تابلو را و چندتا کتاب خریدم	A1-18:Coor-ord2
فراموش کردم برای سپهر چیزی بخرم ولی برای مژگان چندتا کتاب خریدم و این تابلو را	A1-18:Post-ord1
فراموش کردم برای سپهر چیزی بخرم ولی برای مژگان این تابلو را خریدم و چندتا کتاب	A1-18:Post-ord2
می‌شه لطفاً قبل از رفتن سیگار فندک منو بیاری	A1-19:C
می‌شه لطفاً قبل از رفتن به زیرسیگاری و سیگار فندک منو بیاری	A1-19:Coor-ord1
می‌شه لطفاً قبل از رفتن سیگار فندک منو و به زیرسیگاری بیاری	A1-19:Coor-ord2
می‌شه لطفاً قبل از رفتن به زیرسیگاری بیاری و سیگار فندک منو	A1-19:Post-ord1
می‌شه لطفاً قبل از رفتن سیگار فندک منو بیاری و به زیرسیگاری	A1-19:Post-ord2
قرار شد ما فقط گاز پیکنیک‌مون رو بیاریم	A1-20:C
قرار شد ما فقط ده دوازدهتا سیخ کباب و گاز پیکنیک‌مون رو بیاریم	A1-20:Coor-ord1
قرار شد ما فقط گاز پیکنیک‌مون رو و ده دوازدهتا سیخ کباب بیاریم	A1-20:Coor-ord2
قرار شد ما فقط ده دوازدهتا سیخ کباب بیاریم و گاز پیکنیک‌مون رو	A1-20:Post-ord1
قرار شد ما فقط گاز پیکنیک‌مون رو بیاریم و ده دوازدهتا سیخ کباب	A1-20:Post-ord2

## Questionnaire Jugement d'acceptabilité : Exp. Vides parasites

Item	Ind
تو رو نمی‌دونم ولی من تا حالا نشده شلوار بدون اینکه پُرو کنم بخرم	A2-1:Gap
تو رو نمی‌دونم ولی من تا حالا نشده بدون اینکه پُرو کنم شلوار بخرم	A2-1:Gap-C
قارچ آگه بچه‌ها دوست دارن بریز تو سالاد	A2-2:Gap
آگه بچه‌ها قارچ دوست دارن بریز تو سالاد	A2-2:Gap-C
مامانم جوراب پشمی با اینکه اصلاً دوست نداره تو سرمای کانادا مجبور بود پاش کنه	A2-3:Gap
مامانم با اینکه جوراب پشمی اصلاً دوست نداره تو سرمای کانادا مجبور بود پاش کنه	A2-3:Gap-C
خب پالتو آگه نمی‌پوشه براش نخیریم	A2-4:Gap
خب آگه پالتو نمی‌پوشه براش نخیریم	A2-4:Gap-C
شله‌زرد با اینکه همسایه‌مون امسال پخت برای ما نیارده	A2-5:Gap
با اینکه همسایه‌مون امسال شله‌زرد پخت برای ما نیارده	A2-5:Gap-C
آلبالو از وقتی همکارم برام از باغشون میاره دیگه نمی‌خرم	A2-6:Gap
از وقتی همکارم برام از باغشون آلبالو میاره دیگه نمی‌خرم	A2-6:Gap-C
دختر سارا فارسی با اینکه می‌فهمه هنوز بلد نیست حرف بزنه	A2-7:Gap
دختر سارا با اینکه فارسی می‌فهمه هنوز بلد نیست حرف بزنه	A2-7:Gap-C
من بستی سنتی با اینکه خیلی دوست دارم به خاطر دیابتم اصلاً نباید بخورم	A2-8:Gap
من با اینکه بستی سنتی خیلی دوست دارم به خاطر دیابتم اصلاً نباید بخورم	A2-8:Gap-C
فیلم ترسناک با اینکه اصلاً دوست ندارم دستجمعی بدم نمیدانم نگاه کنم	A2-9:Gap
با اینکه فیلم ترسناک اصلاً دوست ندارم دستجمعی بدم نمیدانم نگاه کنم	A2-9:Gap-C
شطرنج با اینکه خیلی خوب بلد بودم از بس بازی نکردم یادم رفته	A2-10:Gap
با اینکه شطرنج خیلی خوب بلد بودم از بس بازی نکردم یادم رفته	A2-10:Gap-C

## Questionnaire Jugement d'acceptabilité : distracteurs

## phrases bien formées (BF) et mal formées (MF)

Item	Ind
لطفاً برای حامد به توپ فوتبال چهل تیکه بخر	F-1:BF
سعید جان این گلدون کریستالو از جلوی پنجره بردار	F-2:BF
از خانم آذر خواهش کنید روی میز مدیرعامل یک دسته گل تازه بگذارد	F-3:BF
به آقای فاتح یادآوری کنید یک کارت تیریک برای خانم مهندس بفرستد	F-4:BF
به حمیدرضا سفارش کن نون سنگک خشخاشی برامون بخره	F-5:BF
از علی بپرس به سرایدار پول داد یا نه	F-6:BF
آوا گفت کتابها رو به شیوا داده	F-7:BF
نیلوفر حدس می‌زنه موبایلشو تو اتوبوس جا گذاشته باشه	F-8:BF
می‌خواستم همکار جدیدم را به سپیده معرفی کنم	F-9:BF
آقای مدیر این قضیه را از یکی از آشنایانشان شنیده‌اند	F-10:BF
فکر کنم ماندانا نمی‌دونه قاشق و چنگالا کجان	F-11:BF
خواهشمند است از خوردن و آشامیدن در این سالن پرهیز نمایید	F-12:BF
ظرف‌های یک بار مصرف و کیسه‌های پلاستیکی بالای جان محیط زیست هستند	F-13:BF
خوردن یک یا دو وعده میوه و سبزیجات تازه در روز برای سلامتی لازم است	F-14:BF
در و پنجره‌ها رو خوب ببندین ممکنه امشب طوفان شه	F-15:BF
همه با اینکه خوابیده سروصدا نکن	F-1:MF
خورش ریواس ترش برای اینکه نشود می‌توان هنگام پخت یک قاشق شکر به آن اضافه کنی	F-2:MF
به علت نقص فنی قبل از اینکه بیش از نیمی از دستگاه‌ها ترخیص شوند به کارخانه عودت می‌دهیم	F-3:MF
مهسا رو با وجود اینکه نیست قراره بگو بیا	F-4:MF
قبل از هر چیز باید گفتن از شما دیدن خرسند گردیدم	F-5:MF
پارکینگ اگر خالی شدند می‌توانید وارد شوید	F-6:MF
خب به من بگو آگه به پول داشتی احتیاج حتماً	F-7:MF
محسن دارد لازم با اینکه زیاد نگرفت	F-8:MF
گمان می‌کنم تعارف کافی به اندازه نکردیم که نیامد شاید	F-9:MF
اگر مورد نظر دستگاه خاموش می‌باشد تماس بگیرید مجدداً	F-10:MF
یک مداد و کاغذها بیاورید تا این نقشه برایتان بکشم	F-11:MF
امروز صبح منوچهر و یک کبک و خروس را دیدم آواز می‌خواند	F-12:MF
یادش رفته تلوزیون و این رادیوها خاموش کنه	F-13:MF
باید این میز و صندلی‌ها بیرون بچینیم	F-14:MF
تمامی پروازهای امشب به علت شدید طوفان و باد کنسل شده‌اند	F-15:MF

## Questionnaire "phrases à trous" 1 : Exp. D1

Ind	Item
D1-1: IDC	Madame Soltani n'était pas riche quand son neveu se maria. Lorsqu'elle hérita des propriétés de son mari, elle octroya [un collier/un appartement]à son neveu ]
D1-1: IDL	Madame Soltani n'était pas riche quand son neveu se maria. Lorsqu'elle hérita des propriétés de son mari, elle octroya [un collier/un appartement]à son neveu qui avait eu un enfant récemment]
D1-1: INC	Madame Soltani était une femme généreuse. Lorsqu'elle hérita des propriétés de son mari, elle octroya [un collier de vraies perles de gros calibre /un appartement deux pièces de 150m2 neuf][à une de ses sœurs]
D1-1: INL	Madame Soltani était une femme généreuse. Lorsqu'elle hérita des propriétés de son mari, elle octroya [un collier/un appartement]à une de ses sœurs qui avait eu un petit-enfant récemment]
D1-2: IDC	La seule personne qui s'intéressait à la bibliothèque de Monsieur Tehrani était Amirali, le fils du voisin. Alors, parfois Monsieur Tehrani octroyait [quelques chemises d'homme à carreaux/quelques livres précieux et coûteux][à Amirali]
D1-2: IDL	La seule personne qui s'intéressait à la bibliothèque de Monsieur Tehrani était Amirali, le fils du voisin. Alors, parfois Monsieur Tehrani octroyait [quelques chemises/quelques livres][à Amirali qui était un bon élève]
D1-2: INC	Les enfants de Monsieur Tehrani ne manifestaient aucun intérêt pour sa bibliothèque. Alors, parfois Monsieur Tehrani octroyait [quelques chemises d'homme à carreaux/quelques livres précieux et coûteux] [à son neveu ]
D1-2: INL	Les enfants de Monsieur Tehrani ne manifestaient aucun intérêt pour sa bibliothèque. Alors, parfois Monsieur Tehrani octroyait [quelques chemises/quelques livres] [à son neveu qui était un bon élève]
D1-3: IDC	Le nouveau maire, qui a été élu par les jeunes, a annoncé quelques mois plus tard qu'il octroiera [une (pièce de) Peugeot 405/un bout de terrain dans le nord de la ville] [aux jeunes gens pieux]
D1-3: IDL	Le nouveau maire, qui a été élu par les jeunes, a annoncé quelques mois plus tard qu'il octroiera [une (pièce de) Peugeot/un bout de terrain] [aux jeunes gens qui veulent se marier]
D1-3: INC	Le nouveau maire, qui avait été élu avec le slogan de résolution de la crise du logement, a annoncé quelques mois après sa prise de fonction qu'il octroiera [une (pièce de) Peugeot 405/un bout de terrain dans le nord de la ville] [à chaque jeune couple]
D1-3: INL	Le nouveau maire, qui avait été élu avec le slogan de résolution de la crise du logement, a annoncé quelques mois après sa prise de fonction qu'il octroiera [une (pièce de) Peugeot/un bout de terrain] [aux jeunes gens qui veulent se marier]
D1-4: IDC	Après l'examen du malade, le médecin prescrivit un sérum, mais quand il quitta la chambre, l'infirmier fit manger [quelques grandes bouchées de tartine au fromage/quelques cuillères à soupe d'huile de ricin] [au malade]
D1-4: IDL	Après l'examen du malade, le médecin prescrivit un sérum, mais quand il quitta la chambre, l'infirmier fit manger [un morceau de pain/quelques cuillères d'huile] [au pauvre malade qui se tordait de douleur]
D1-4: INC	L'infirmier ne fut pas du même avis que le médecin. Quand le médecin quitta la chambre, elle fit manger [quelques grandes bouchées de tartine au fromage/quelques cuillères à soupe d'huile de ricin] [au malade]
D1-4: INL	L'infirmier ne fut pas du même avis que le médecin. Quand le médecin quitta la chambre, elle fit manger [un morceau de pain/quelques cuillères d'huile] [au pauvre malade qui se tordait de douleur]
D1-5: IDC	La mère de Kiyarach était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [un grand verre de smoothie vitaminé/un grand verre de jus de carotte frais] [à Kiyarach]

D1-5: IDL	La mère de Kiyarach était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [un verre de sirop à l'herbe de Sainte Sophie/un verre de jus de carotte] [à Kiyarach qui étudiait toute la journée]
D1-5: INC	Zohreh était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [un grand verre de smoothie vitaminé/un grand verre de jus de carotte frais] [à son fils]
D1-5: INL	Zohreh était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [un verre de sirop à l'herbe de Sainte Sophie/un verre de jus de carotte] [à son fils qui étudiait toute la journée]
D1-6: IDC	Grand-mère n'avait pas d'appétit, elle était maigre et faible. Hier soir, les filles ont réussi avec beaucoup de difficultés à (qu'elles) fassent manger [un peu de frites avec du ketchup/un peu de viande grillée et des légumes] [à Grand-mère]
D1-6: IDL	Grand-mère n'avait pas d'appétit, elle était maigre et faible. Hier soir, les filles ont réussi avec beaucoup de difficultés à (qu'elles) fassent manger [une assiette de riz aux lentilles/un bol de soupe de poulet] [à Grand-mère qui était pâle et fatiguée]
D1-6: INC	Les filles avaient essayé tous les repas et avaient tout tenté jusqu'à hier soir quand enfin elles ont réussi à (qu'elles) fassent manger [un peu de frites avec du ketchup/un peu de viande grillée et des légumes] [à Grand-mère]
D1-6: INL	Les filles avaient essayé tous les repas et avaient tout tenté jusqu'à hier soir quand enfin elles ont réussi à (qu'elles) fassent manger [une assiette de riz aux lentilles/un bol de soupe de poulet] [à Grand-mère qui était faible et maigre]
D1-7: IDC	Depuis des années, Parastou, la fille de Madame et Monsieur Nikfar, est partie en Allemagne. Chaque année avant le nouvel an, ils envoient [un paquet de dix kilos de fruits secs et de gâteaux/un paquet de dix kilos de nourriture et de livres] [pour Parastou]
D1-7: IDL	Depuis des années, Parastou, la fille de Madame et Monsieur Nikfar, est partie en Allemagne. Chaque année avant le nouvel an, ils envoient [un paquet de vêtements/un paquet de nourriture] [pour Parastou qui étudie à Stuttgart]
D1-7: INC	À ce moment de l'année, Madame et Monsieur Nikfar font leurs courses. Chaque année avant le nouvel an, ils envoient [un paquet de dix kilos de fruits secs et de gâteaux/un paquet de dix kilos de nourriture et de livres] [pour leur fille]
D1-7: INL	À ce moment de l'année, Madame et Monsieur Nikfar font leurs courses. Chaque année avant le nouvel an, ils envoient [un paquet de vêtements/un paquet de nourriture] [pour leur fille qui étudie à Berlin]
D1-8: IDC	Le collègue de Maryam n'avait pas obtenu son visa et Maryam qui avait décidé d'y aller toute seule, au cours du voyage envoya [plusieurs peintures de lieux incontournables de la ville/plusieurs cartes postales de lieux incontournables de la ville] [pour son collègue]
D1-8: IDL	Le collègue de Maryam n'avait pas obtenu son visa et Maryam qui avait décidé d'y aller toute seule, à la première occasion envoya [une carte de vœux / une carte postale] [pour sa collègue qui était une femme susceptible]
D1-8: INC	C'était la première fois que Maryam avait eu un visa européen et voyageait en Italie. À la première occasion, elle envoya [plusieurs peintures de lieux incontournables de la ville/plusieurs cartes postales de lieux incontournables de la ville] [pour Arezou]
D1-8: INL	C'était la première fois que Maryam avait eu un visa européen et voyageait en Italie. À la première occasion, elle envoya [une carte de vœux / une carte postale] [pour Arezou, sa jeune sœur]
D1-9: IDC	Haleh est à la fois gentille et dépensière. À l'exposition de Sepideh, le premier jour, elle acheta [un portefeuille en cuir et un sac à dos/une bague en agate et une chaîne] [de Sépideh]
D1-9: IDL	Haleh est à la fois gentille et dépensière. À l'exposition de Sepideh, le premier jour, elle acheta [un portefeuille en cuir] [de Sepideh qui ne pensait pas vendre]
D1-9: INC	Haleh est une fille dépensière et dépense tout son salaire pendant une semaine. Hier encore, elle acheta [un portefeuille en cuir et un sac à dos/une paire de boucles d'oreilles et un collier de perles] [de l'un de ses amis]
D1-9: INL	Haleh est une fille dépensière et dépense tout son salaire pendant une semaine. Hier encore, elle acheta [un sac à dos/une bague en agate] [de l'un de ses amis qui avait fait une vente promotionnelle]

D1-10: IDC	Chahin avait demandé à son ami de faire tout pour réparer son ordinateur, mais quand il s'est aperçu que son ordinateur est irrécupérable, il a été obligé de (qu'il) achète [un iphone bon marché de deuxième main/un (ordinateur) portable bon marché de deuxième main] [de son ami]
D1-10: IDL	Chahin avait demandé à son ami de faire tout pour réparer son ordinateur, mais quand il s'est aperçu que son ordinateur est irrécupérable, il a été obligé de (qu'il) achète [un iphone neuf/un (ordinateur) portable neuf] [de son ami qui avait un magasin dans la capitale]
D1-10: INC	L'ordinateur de Chahin fonctionnait bien même s'il était vieux, mais hier au milieu d'un travail urgent, il a beugué, et Chahin a été obligé de (qu'il) achète [un iphone plutôt bon marché de deuxième main/un (ordinateur) portable plutôt bon marché de deuxième main] [de l'un de ses amis]
D1-10: INL	L'ordinateur de Chahin fonctionnait bien même s'il était vieux, mais hier au milieu d'un travail urgent, il a beugué, et Chahin a été obligé de (qu'il) achète [un iphone bon marché/un (ordinateur) portable bon marché] [de son ami qui avait un magasin dans la capitale]
D1-11: IDC	Le beau-frère de Monsieur Parsa fut concessionnaire de vente de véhicules. Dès la baisse des prix, Monsieur Parsa réussit à (qu'il) achète [une Mini Minor gris zéro kilomètre/un Nissan Patrol quatre portes zéro kilomètre] [de son beau-frère]
D1-11: IDL	Le beau-frère de Monsieur Parsa fut concessionnaire de vente de véhicules. Dès la baisse des prix, Monsieur Parsa réussit à (qu'il) achète [une Mini Minor/un Nissan Patrol] [de son beau-frère qui était un homme naïf]
D1-11: INC	Depuis quelques temps, Monsieur Parsa chercha une 4x4 pour ses voyages en régions montagneuses. Avant la hausse des prix, il réussit à (qu'il) achète [une Mini Minor gris zéro kilomètre/un Patrol quatre portes noir zéro kilomètre] [de l'un de ses collègues]
D1-11: INL	Depuis quelques temps, Monsieur Parsa chercha une 4x4 pour ses voyages en régions montagneuses. Avant la hausse des prix, il réussit à (qu'il) achète [une Mini Minor/un Nissan Patrol] [de l'un de ses collègues qui avait besoin d'argent]
D1-12: IDC	Ils n'avaient pas de vente depuis des jours. Jeudi, quand un client entra dans le magasin, le vendeur le plus expérimenté a réussi à (qu'il) vende [un lave-vaisselle compact Samsung/un canapé-lit entièrement en cuir] [à ce client]
D1-12: IDL	Ils n'avaient pas de vente depuis des jours. Jeudi, quand un client entra dans le magasin, le vendeur le plus expérimenté a réussi à (qu'il) vende [un aspirateur/un canapé en cuir] [à ce client qui n'avait même pas l'intention de faire des achats]
D1-12: INC	Depuis des jours, personne n'entrait dans le magasin. Quand ils ont fait une vente exceptionnelle le vendredi, le plus jeune vendeur a réussi à (qu'il) vende [un réfrigérateur côte à côte haut de gamme Samsung/un canapé-lit luxe entièrement en cuir] [au premier client]
D1-12: INL	Depuis des jours, personne n'entrait dans le magasin. Quand ils ont fait une vente exceptionnelle le vendredi, le plus jeune vendeur a réussi à (qu'il) vende [un réfrigérateur côte à côte /un canapé en cuir] [au premier client qui n'avait même pas l'intention de faire des achats]
D1-13: IDC	Quand Saba eut besoin d'argent, elle invita tous les voisins et mit en vente ses vieilles affaires. Comme elle était fûtée, durant quelques heures, elle vendit [un set de couteaux et de fourchettes en argent anciens/un service de table et quelques casseroles] [aux voisins]
D1-13: IDL	Quand Saba eut besoin d'argent, elle invita tous les voisins et mit en vente ses vieilles affaires. Comme elle était fûtée, durant quelques heures, elle vendit [un set de tasses et de soucoupes/un service de table] [à la voisine du dessous qui était une femme dépensière]
D1-13: INC	Saba est une fille fûtée. Quand elle avait besoin d'argent, elle a mis en vente ses vieilles affaires et dès le premier jour, elle vendit [un set de couteaux et de fourchettes en argent anciens /un service de table et quelques casseroles] [aux voisins]
D1-13: INL	Saba est une fille fûtée. Quand elle avait besoin d'argent, elle a mis en vente ses vieilles affaires et dès le premier jour, elle vendit [un set de tasses et de soucoupes/un service de table] [à la femme du propriétaire qui était une femme dépensière]
D1-14: IDC	Il y a quelques temps pendant que le marché immobilier stagnait, un architecte passa à l'agence. Bahador qui est très expérimenté dans son travail vendit [un bâtiment commercial de douze étages/un vieil immeuble de quatre étages] [à cet ingénieur architecte]
D1-14: IDL	Il y a quelques temps pendant que le marché immobilier stagnait, un architecte passa à l'agence. Bahador qui est très expérimenté dans son travail vendit [un bâtiment commercial/une vieille maison] [à cet ingénieur architecte jeune et débutant]
D1-14: INC	Bahador n'a peut-être pas de diplôme, mais il connaît bien son travail. Il y a quelques temps, à un moment où le marché immobilier stagnait, il vendit [un bâtiment commercial de douze étages/un vieil immeuble de quatre étages] [à un ingénieur architecte]

D1-14: INL	Bahador n'a peut-être pas de diplôme, mais il connaît bien son travail. Il y a quelques temps, à un moment où le marché immobilier stagnait, il vendit [un bâtiment commercial/une vieille maison ] [à un ingénieur architecte jeune et débutant]
D1-15: IDC	L'exposition était bondée. Les enfants se chamaillaient avec leurs parents. L'agent de l'exposition qui était excédé par le bruit des enfants donna [un fusil à vent et quelques cartouches/une feuille de papier et quelques crayons de couleur] [aux enfants]
D1-15: IDL	L'exposition était bondée. Les enfants se chamaillaient avec leurs parents. L'agent de l'exposition qui était excédé par le bruit des enfants donna [quelques chewing-gums /quelques bonbons] [aux enfants qui s'impatientaient]
D1-15: INC	L'exposition était vraiment bondée. L'agent de l'exposition qui était excédé par les bruits donna [un fusil à vent et quelques cartouches/une feuille de papier et quelques crayons de couleur] [aux enfants]
D1-15: INL	L'exposition était vraiment bondée. L'agent de l'exposition qui était excédé par les bruits donna [quelques chewing-gums /quelques bonbons] [aux enfants qui s'impatientaient]
D1-16: IDC	Tous les petits-enfants étaient allés chez Grand-père et Grand-mère pour la visite du nouvel an. Au moment de partir, Grand-mère donna [une poignée de pêches et d'abricots séchés/une poignée de sucettes et de chocolats] [à tous les petits-enfants]
D1-16: IDL	Tous les petits-enfants étaient allés chez Grand-père et Grand-mère pour la visite du nouvel an. Au moment de partir, Grand-mère donna [une poignée de pois chiches grillés/une poignée de bonbons] [à tous les petits-enfants qui attendaient leurs cadeaux de nouvel an]
D1-16: INC	Tout le monde était allé chez Madame et Monsieur Eftekhari pour la visite du nouvel an. Au moment de partir, Madame Eftekhari donna [une poignée de pêches et d'abricots séchés/une poignée de sucettes et de chocolats] [à tous les enfants]
D1-16: INL	Tout le monde était allé chez Madame et Monsieur Eftekhari pour la visite du nouvel an. Au moment de partir, Madame Eftekhari donna [une poignée de pois chiches grillés/une poignée de bonbons] [à tous les enfants qui étaient venus]
D1-17: IDC	La climatisation ne fonctionnait plus et les clients se plaignaient. Quand les contestations se sont aggravées, l'agent d'accueil donna [une tranche de pastèque fraîche sans pépins/un verre de sirop de sekandjebine glacé] [aux clients]
D1-17: IDL	La climatisation ne fonctionnait plus et les clients se plaignaient. Quand les contestations se sont aggravées, l'agent d'accueil donna [une tasse de thé/un verre de sirop] [aux clients qui étaient excédés de chaleur]
D1-17: INC	La climatisation ne fonctionnait plus et la chaleur montait. Quand la chaleur a atteint un pic, l'agent d'accueil donna [une tranche de pastèque fraîche sans pépins/un verre de sirop de sekandjebine glacé] [aux clients]
D1-17: INL	La climatisation ne fonctionnait plus et la chaleur montait. Quand la chaleur a atteint un pic, l'agent d'accueil donna [une tasse de thé/un verre de sirop] [aux clients qui étaient excédés de chaleur]
D1-18: IDC	Le conducteur s'est arrêté au carrefour et depuis la fenêtre (de la voiture), un homme prit [une grande enveloppe scellée/une liasse de billets de cinq mille toumans] [du conducteur]
D1-18: IDL	Le conducteur s'est arrêté au carrefour et depuis la fenêtre (de la voiture), un homme prit [une enveloppe fermée/une liasse de billets de mille toumans] [du conducteur qui était très en colère]
D1-18: INC	La voiture s'est arrêtée au carrefour et depuis la fenêtre (de la voiture), le conducteur prit [une grande enveloppe scellée/une liasse de billets de cinq mille toumans] [d'un jeune homme]
D1-18: INL	La voiture s'est arrêtée au carrefour et depuis la fenêtre (de la voiture), le conducteur prit [une enveloppe fermée/une liasse de billets de mille toumans] [d'un jeune homme qui est descendu de la voiture]
D1-19: IDC	Le portier a dit que l'accès à la salle est interdit en chaussures. Alors pour entrer dans la salle, les élèves prirent [une paire de chaussures jetables/une paire de chaussures pantoufles] [du gardien]
D1-19: IDL	Le portier a dit que l'accès à la salle est interdit en chaussures. Alors pour entrer dans la salle, les élèves prirent [une (paire de) chaussures neuves/une paire de chaussettes] [du gardien qui fut un jeune homme]

D1-19: INC	L'accès à la salle était interdit en chaussures et pour entrer dans la salle, les élèves prirent [une paire de chaussures jetables/une paire de chaussures pantoufles] [du gardien]
D1-19: INL	L'accès à la salle était interdit en chaussures et pour entrer dans la salle, les élèves prirent [une (paire de) chaussures neuves/une paire de chaussettes] [de l'homme qui fut responsable de la salle]
D1-20: IDC	Un homme entra dans la station et s'assit sur le banc. La fille et le garçon qui se tenaient debout à côté de la station prirent [un sac plastique noir/un bout de papier déchiré et un stylo] [de cet homme]
D1-20: IDL	Un homme entra dans la station et s'assit sur le banc. La fille et le garçon qui se tenaient debout à côté de la station prirent [un sac/un stylo] [de cet homme qui avait une barbe longue]
D1-20: INC	La fille et le garçon, qui se tenaient debout à côté de la station, semblaient inquiets. Avant l'arrivée du bus, ils prirent [un sac plastique noir/un bout de papier déchiré et un stylo] [d'un marchand ambulancier]
D1-20: INL	La fille et le garçon, qui se tenaient debout à côté de la station, semblaient inquiets. Avant l'arrivée du bus, ils prirent [un sac plastique/un bout de papier] [d'un marchand ambulancier qui passait par là]

### Questionnaire "phrases à trous" 2 : Exp. D2

Ind	Item
D2-1: IDC	Madame Soltani n'était pas riche quand son neveu se maria. Lorsqu'elle hérita des propriétés de son mari, elle octroya [collier(s) de perles/appartement(s) neuf(s)] [à son neveu]
D2-1: IDL	Madame Soltani n'était pas riche quand son neveu se maria. Lorsqu'elle hérita des propriétés de son mari, elle octroya [collier(s) de perles/appartement(s) neuf(s)] [à son neveu qui fut longtemps locataire]
D2-1: INC	Madame Soltani était une femme généreuse. Lorsqu'elle hérita des propriétés de son mari, elle octroya [collier(s) de perles/appartement(s) neuf(s)] [à son neveu]
D2-1: INL	Madame Soltani était une femme généreuse. Lorsqu'elle hérita des propriétés de son mari, elle octroya [collier(s) de perles/appartement(s) neuf(s)] [à son neveu qui fut longtemps locataire]
D2-2: IDC	Le neveu de Monsieur Tehrani était la seule personne qui lui rendait visite. Alors parfois Monsieur Tehrani octroyait [chemise(s) d'homme/livre(s) de photographie] [à son neveu]
D2-2: IDL	Le neveu de Monsieur Tehrani était la seule personne qui lui rendait visite. Monsieur Tehrani qui ne comptait plus sur ses enfants octroyait [chemise(s) d'homme/livre(s) de photographie] [à son neveu qui était étudiant en architecture]
D2-2: INC	Tous les enfants de Monsieur Tehrani avaient quitté l'Iran. Monsieur Tehrani qui détestait la solitude octroyait [chemise(s) d'homme/livre(s) de photographie] [à son neveu]
D2-2: INL	Tous les enfants de Monsieur Tehrani avaient quitté l'Iran. Monsieur Tehrani qui détestait la solitude octroyait [chemise(s) d'homme/livre(s) de photographie] [à son neveu qui lui rendait visite régulièrement]
D2-3: IDC	Le nouveau maire, qui a été élu par les jeunes, a annoncé quelques mois plus tard qu'il octroiera [dot(s) complète(s)/voyage(s) à la Mecque] [aux jeunes]
D2-3: IDL	Le nouveau maire, qui a été élu par les jeunes, a annoncé quelques mois plus tard qu'il octroiera [dot(s) complète(s)/voyage(s) à la Mecque] [aux jeunes qui veulent se marier]
D2-3: INC	Le nouveau maire, qui avait été élu avec le slogan de résolution de la crise du logement, a annoncé quelques mois après sa prise de fonction qu'il octroiera [dot(s) complète(s)/voyage(s) à la Mecque] [aux jeunes]

D2-3: INL	Le nouveau maire, qui avait été élu avec le slogan de résolution de la crise du logement, a annoncé quelques mois après sa prise de fonction qu'il octroiera [dot(s) complète(s)/voyage(s) à la Mecque] [aux jeunes qui veulent se marier]
D2-4: IDC	Après l'examen du malade, le médecin prescrivit un sérum, mais quand il quitta la chambre, l'infirmier fit manger [(du) thé tiède/(du) lait tiède] [au malade]
D2-4: IDL	Après l'examen du malade, le médecin prescrivit un sérum, mais quand il quitta la chambre, l'infirmier fit manger [(du) thé tiède/(du) lait tiède] [au pauvre malade qui se tordait de douleur]
D2-4: INC	L'infirmier ne fut pas de même avis que le médecin. Quand le médecin quitta la chambre, elle fit manger [(du) thé tiède/(du) lait tiède] [au malade]
D2-4: INL	L'infirmier ne fut pas de même avis que le médecin. Quand le médecin quitta la chambre, elle fit manger [(du) thé tiède/(du) lait tiède] [au pauvre malade qui se tordait de douleur]
D2-5: IDC	La mère de Kiyarach était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [pêche(s) séchée(s)/(de l') huile de ricin] [à Kiyarach]
D2-5: IDL	La mère de Kiyarach était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [pêche(s) séchée(s)/(de l') huile de ricin] [à Kiyarach qui étudiait toute la journée]
D2-5: INC	Zohreh était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [pêche(s) séchée(s)/(de l') huile de ricin] [à Kiyarach]
D2-5: INL	Zohreh était une femme maniaque et sévère. Pendant la période du concours, chaque matin, elle faisait manger [pêche(s) séchée(s)/(de l') huile de ricin] [à Kiyarach qui étudiait toute la journée]
D2-6: IDC	Grand-mère n'avait pas d'appétit, elle était maigre et faible. Hier soir, les filles ont réussi avec beaucoup de difficultés à (qu'elles) fassent manger [(du) lait aux amandes/(du) ragout de mouton] [à Grand-mère]
D2-6: IDL	Grand-mère n'avait pas d'appétit, elle était maigre et faible. Hier soir, les filles ont réussi avec beaucoup de difficultés à (qu'elles) fassent manger [(du) lait aux amandes/(du) ragout de mouton] [à Grand-mère qui était faible et maigre]
D2-6: INC	Les filles avaient essayé tous les repas et avaient tout tenté jusqu'à hier soir quand enfin elles ont réussi à (qu'elles) fassent manger [(du) lait aux amandes/(du) ragout de mouton] [à Grand-mère]
D2-6: INL	Les filles avaient essayé tous les repas et avaient tout tenté jusqu'à hier soir quand enfin elles ont réussi à (qu'elles) fassent manger [(du) lait aux amandes/(du) ragout de mouton] [à Grand-mère qui était faible et maigre]
D2-7: IDC	Depuis des années, Parastou, la fille de Madame et Monsieur Nikfar, est partie en Allemagne. Chaque année avant le nouvel an, ils envoient [habit(s) d'hiver/gâteau(x) fait(s) maison] [pour leur fille]
D2-7: IDL	Depuis des années, Parastou, la fille de Madame et Monsieur Nikfar, est partie en Allemagne. Chaque année avant le nouvel an, ils envoient [habit(s) d'hiver/gâteau(x) fait(s) maison] [pour leur fille qui étudie à Berlin]
D2-7: INC	À ce moment de l'année, Madame et Monsieur Nikfar font leurs courses. Chaque année avant le nouvel an, ils envoient [habit(s) d'hiver/gâteau(x) fait(s) maison] [pour leur fille]
D2-7: INL	À ce moment de l'année, Madame et Monsieur Nikfar font leurs courses. Chaque année avant le nouvel an, ils envoient [habit(s) d'hiver/gâteau(x) fait(s) maison] [pour leur fille qui étudie à Berlin]
D2-8: IDC	Le collègue de Maryam n'avait pas obtenu son visa et Maryam qui avait décidé d'y aller toute seule, au cours du voyage envoya [carte(s) de vœux 3D/carte(s) postale(s) touristique(s)] [pour son collègue]
D2-8: IDL	Le collègue de Maryam n'avait pas obtenu son visa et Maryam qui avait décidé d'y aller toute seule, à la première occasion envoya [carte(s) de vœux 3D/carte(s) postale(s) touristique(s)] [pour son collègue qui était une femme susceptible]

D2-8: INC	C'était la première fois que Maryam avait eu un visa européen et voyageait en Italie. À la première occasion, elle envoya [carte(s) de vœux 3D/carte(s) postale(s) touristique(s)] [pour son collègue]
D2-8: INL	C'était la première fois que Maryam avait eu un visa européen et voyageait en Italie. À la première occasion, elle envoya [carte(s) de vœux 3D/carte(s) postale(s) touristique(s)] [pour son collègue qui était une femme susceptible]
D2-9: IDC	Haleh est à la fois gentille et dépendière. À l'exposition de Sepideh, le premier jour, elle acheta [calendrier(s) de poche/bracelet(s) en argent] [de Sepideh]
D2-9: IDL	Haleh est à la fois gentille et dépendière. À l'exposition de Sepideh, le premier jour, elle acheta [calendrier(s) de poche/bracelet(s) en argent] [de Sepideh qui ne pensait pas vendre]
D2-9: INC	Haleh est une fille dépendière et dépense tout son salaire pendant une semaine. Hier encore, elle acheta [calendrier(s) de poche/bracelet(s) en argent] [de Sepideh]
D2-9: INL	Haleh est une fille dépendière et dépense tout son salaire pendant une semaine. Hier encore, elle acheta [calendrier(s) de poche/bracelet(s) en argent] [de Sepideh qui exposait]
D2-10: IDC	Chahin avait demandé à son ami de réparer son ordinateur, mais quand il s'est aperçu que son ordinateur est irrémédiable, il a été obligé de (qu'il) achète [iphone(s) neuf(s)/(ordinateur(s)) portable(s) neuf(s)] [de son ami]
D2-10: IDL	Chahin avait demandé à son ami de réparer son ordinateur, mais quand il s'est aperçu que son ordinateur est irrémédiable, il a été obligé de (qu'il) achète [iphone(s) neuf(s)/(ordinateur(s)) portable(s) neuf(s)] [de son ami qui avait un magasin dans la capitale]
D2-10: INC	L'ordinateur de Chahin fonctionnait bien même s'il était vieux, mais hier au milieu d'un travail urgent, il a beugué, et Chahin a été obligé de (qu'il) achète [iphone(s) neuf(s)/(ordinateur(s)) portable(s) neuf(s)] [de son ami]
D2-10: INL	L'ordinateur de Chahin fonctionnait bien même s'il était vieux, mais hier au milieu d'un travail urgent, il a beugué, et Chahin a été obligé de (qu'il) achète [iphone(s) neuf(s)/(ordinateur(s)) portable(s) neuf(s)] [de son ami qui avait un magasin dans la capitale]
D2-11: IDC	Le beau-frère de Monsieur Parsa fut concessionnaire de vente de véhicules. Dès la baisse des prix, Monsieur Parsa réussit à (qu'il) achète [Patrol(s) deux portes/camionnette(s) Pickup] [de son beau-frère]
D2-11: IDL	Le beau-frère de Monsieur Parsa fut concessionnaire de vente de véhicules. Dès la baisse des prix, Monsieur Parsa réussit à (qu'il) achète [Patrol(s) deux portes/camionnette(s) Pickup] [de son beau-frère qui était (une personne) équitable]
D2-11: INC	Depuis quelques temps, Monsieur Parsa chercha une 4x4 pour ses voyages aux régions montagneuses. Avant la hausse des prix, il réussit à (qu'il) achète [Patrol(s) deux portes/camionnette(s) Pickup] [de son beau-frère]
D2-11: INL	Depuis quelques temps, Monsieur Parsa chercha une 4x4 pour ses voyages aux régions montagneuses. Avant la hausse des prix, il réussit à (qu'il) achète [Patrol(s) deux portes/camionnette(s) Pickup] [de son beau-frère qui fut concessionnaire de ventes de véhicules]
D2-12: IDC	Ils n'avaient aucune vente depuis quelques jours. Jeudi, dès qu'un client entra dans le magasin, le vendeur le plus expérimenté a réussi à (qu'il) vende [réfrigérateur(s) côte à côte /lit(s) convertible(s)] [à ce client]
D2-12: IDL	Ils n'avaient aucune vente depuis quelques jours. Jeudi, dès qu'un client entra dans le magasin, le vendeur le plus expérimenté a réussi à (qu'il) vende [réfrigérateur(s) côte à côte /lit(s) convertible(s)] [à ce client qui n'avait même pas l'intention de faire des achats]
D2-12: INC	Depuis des jours, personne n'était entré dans le magasin. Ils ont fait une vente exceptionnelle le vendredi et le vendeur le plus jeune a réussi à (qu'il) vende [réfrigérateur(s) côte à côte /canapé(s) lit(s)] [au premier client]
D2-12: INL	Depuis des jours, personne n'était entré dans le magasin. Ils ont fait une vente exceptionnelle le vendredi et le vendeur le plus jeune a réussi à (qu'il) vende [réfrigérateur(s) côte à côte /canapé(s) lit(s)] [au premier client qui n'avait pas l'intention de faire des achats]
D2-13: IDC	Quand Saba eut besoin d'argent, elle invita tous les voisins et mit en vente ses vieilles affaires. Comme elle était fûtée, durant quelques heures elle vendit [assiette(s) émaillée(s) à motifs/service(s) d'argenterie] [à l'un des voisins]

D2-13: IDL	Quand Saba eut besoin d'argent, elle invita tous les voisins et mit en vente ses vieilles affaires. Comme elle était fûtée, durant quelques heures elle vendit [assiette(s) émaillée(s) à motifs/service(s) d'argenterie] [à l'un des voisins qui était dépendaire]
D2-13: INC	Saba est une fille fûtée. Quand elle avait besoin d'argent, elle a mis en vente ses vieilles affaires et dès le premier jour, elle vendit [assiette(s) émaillée(s) à motifs/service(s) d'argenterie] [à la femme du propriétaire]
D2-13: INL	Saba est une fille fûtée. Quand elle avait besoin d'argent, elle a mis en vente ses vieilles affaires et dès le premier jour, elle vendit [assiette(s) émaillée(s) à motifs/service(s) d'argenterie] [à la femme du propriétaire qui était dépendaire]
D2-14: IDC	Il y a quelques temps pendant que le marché immobilier stagnait, un architecte passa à l'agence. Bahador qui est très expérimenté dans son travail vendit [bâtiment(s) inachevé(s)/vieil (vieux) immeuble(s)] [à cet ingénieur architecte]
D2-14: IDL	Il y a quelques temps pendant que le marché immobilier stagnait, un architecte passa à l'agence. Bahador qui est très expérimenté dans son travail vendit [bâtiment(s) inachevé(s)/vieil (vieux) immeuble(s)] [à cet ingénieur architecte jeune et débutant]
D2-14: INC	Bahador n'a peut-être pas de diplôme, mais il connaît bien son travail. Il y a quelques temps, à un moment où le marché immobilier stagnait, il vendit [bâtiment(s) inachevé(s)/vieil (vieux) immeuble(s)] [à un ingénieur architecte]
D2-14: INL	Bahador n'a peut-être pas de diplôme, mais il connaît bien son travail. Il y a quelques temps, à un moment où le marché immobilier stagnait, il vendit [bâtiment(s) inachevé(s)/vieil (vieux) immeuble(s)] [à un ingénieur architecte jeune et débutant]
D2-15: IDC	L'exposition était bondée. Les enfants se chamaillaient avec leurs parents. L'agent de l'exposition qui était excédé par le bruit des enfants donna [glace(s) en cornet/(bonbon en) sucette(s)] [aux enfants]
D2-15: IDL	L'exposition était vraiment bondée. L'agent de l'exposition qui était excédé par les bruits donna [glace(s) en cornet/(bonbon en) sucette(s)] [aux enfants qui s'impatientsaient]
D2-15: INC	L'exposition était vraiment bondée. L'agent de l'exposition qui était excédé par les bruits donna [glace(s) en cornet/(bonbon en) sucette(s)] [aux enfants]
D2-15: INL	L'exposition était vraiment bondée. L'agent de l'exposition qui était excédé par les bruits donna [glace(s) en cornet/(bonbon en) sucette(s)] [aux enfants qui s'impatientsaient]
D2-16: IDC	Tous les enfants étaient allés chez Grand-père et Grand-mère pour la visite du nouvel an. Au moment de partir, Grand-mère donna [œuf(s) surprise/œuf(s) coloré(s)] [à tous les enfants]
D2-16: IDL	Tous les enfants étaient allés chez Grand-père et Grand-mère pour la visite du nouvel an. Au moment de partir, Grand-mère donna [œuf(s) surprise/œuf(s) coloré(s)] [à tous les enfants qui attendaient le cadeau de nouvel an]
D2-16: INC	Tout le monde était allé chez Madame et Monsieur Eftekhari pour la visite du nouvel an. Au moment de partir, Madame Eftekhari donna [œuf(s) surprise/œuf(s) coloré(s)] [à tous les enfants]
D2-16: INL	Tout le monde était allé chez Madame et Monsieur Eftekhari pour la visite du nouvel an. Au moment de partir, Madame Eftekhari donna [œuf(s) surprise/œuf(s) coloré(s)] [à tous les enfants qui attendaient le cadeau de nouvel an]
D2-17: IDC	La climatisation ne fonctionnait plus et les clients se plaignaient. Quand les contestations se sont aggravées, l'agent d'accueil donna [(du) sirop de verveine/(du) sirop de griotte] [aux clients]
D2-17: IDL	La climatisation ne fonctionnait plus et les clients se plaignaient. Quand les contestations se sont aggravées, l'agent d'accueil donna [(du) sirop de verveine/(du) sirop de griotte] [aux clients qui étaient excédés de chaleur]
D2-17: INC	La climatisation ne fonctionnait plus et la chaleur montait. Quand la chaleur a atteint un pic, l'agent d'accueil donna [(du) sirop de verveine/(du) sirop de griotte] [aux clients]
D2-17: INL	La climatisation ne fonctionnait plus et la chaleur montait. Quand la chaleur a atteint un pic, l'agent d'accueil donna [(du) sirop de verveine/(du) sirop de griotte] [aux clients qui étaient excédés de chaleur]

D2-18: IDC	Un homme était debout au bord de la rue. Après un moment, le conducteur est descendu de la voiture et prit [fusil(s) de chasse/couteau(x) verrouillable(s)] [de ce jeune homme]
D2-18: IDL	Un homme était debout au bord de la rue. Après un moment, le conducteur est descendu de la voiture et prit [fusil(s) de chasse/couteau(x) verrouillable(s)] [de ce jeune homme qui était très costaud]
D2-18: INC	La voiture était au bord de la rue. Après un moment, le conducteur est descendu de la voiture et prit [fusil(s) de chasse/couteau(x) verrouillable(s)] [d'un jeune homme]
D2-18: INL	La voiture était au bord de la rue. Après un moment, le conducteur est descendu de la voiture et prit [fusil(s) de chasse/couteau(x) verrouillable(s)] [d'un jeune homme qui attendait debout]
D2-19: IDC	Le responsable de la salle a dit que l'accès à la salle en chaussures est interdit. Alors pour entrer dans la salle, les élèves prirent [couverture(s) en plastique/chausson(s) pantoufle(s)] [du responsable de la salle]
D2-19: IDL	Le responsable de la salle a dit que l'accès à la salle en chaussures est interdit. Alors pour entrer dans la salle, les élèves prirent [couverture(s) en plastique/chausson(s) pantoufle(s)] [du responsable de la salle qui était un jeune homme]
D2-19: INC	L'accès à la salle était interdit en chaussures. Alors pour entrer dans la salle, les élèves prirent [couverture(s) en plastique/chausson(s) pantoufle(s)] [du responsable de la salle]
D2-19: INL	L'accès à la salle était interdit en chaussures. Alors pour entrer dans la salle, les élèves prirent [couverture(s) en plastique/chausson(s) pantoufle(s)] [du responsable de la salle qui était un jeune homme]
D2-20: IDC	Un marchand ambulant travaillait à côté de la station. La fille et le garçon qui attendaient le bus prirent [ticket(s) de bus/papier(s) brouillon(s)] [de ce marchand ambulant]
D2-20: IDL	Un marchand ambulant travaillait à côté de la station. La fille et le garçon qui attendaient le bus prirent [ticket(s) de bus/papier(s) brouillon(s)] [de ce marchand ambulant qui avait une barbe longue]
D2-20: INC	La station était déserte. La fille et le garçon qui étaient debout à côté de la station prirent [ticket(s) de bus/papier(s) brouillon(s)] [d'un marchand ambulant]
D2-20: INL	La station était déserte. La fille et le garçon qui étaient debout à côté de la station prirent [ticket(s) de bus/papier(s) brouillon(s)] [du marchand ambulant qui passait par là]
<b>Questionnaire "phrases à trous" 3 : Exp. D3</b>	
<b>Ind</b>	<b>Item</b>
D3-1: i	Deux jours après le démarrage des travaux, le porte-parole du Conseil Municipal [proposa] [un plan complet/un nouveau projet] [au maire de Téhéran]
D3-1: Ra	Deux jours après le démarrage des travaux, le porte-parole du Conseil Municipal [proposa] [le plan complet/le nouveau projet] [au maire de Téhéran]
D3-1: Yek	Deux jours après le démarrage des travaux, le porte-parole du Conseil Municipal [proposa] [un plan complet/un nouveau projet] [au maire de Téhéran]
D3-2: i	Pendant le premier jour de son voyage, le président [proposa] [un projet commun de développement /un projet de coopération bilatérale] [au premier ministre du pays voisin]

D3-2: Ra	Pendant le premier jour de son voyage, le président [proposa] [le projet commun de développement /le projet de coopération bilatérale] [au premier ministre du pays voisin]
D3-2: Yek	Pendant le premier jour de son voyage, le président [proposa] [un projet commun de développement /un projet de coopération bilatérale] [au premier ministre du pays voisin]
D3-3: i	Après le dîner, le chef cuisinier [apporta] [le paquet/le gâteau] [pour les invités]
D3-3: Ra	Après le dîner, le chef cuisinier [apporta] [un paquet/un gâteau] [pour les invités ]
D3-3: Yek	Après le dîner, le chef cuisinier [apporta] [un paquet/un gâteau] [pour les invités ]
D3-4: i	Quand l'applaudissement du public a cessé, le responsable de la réunion [donna] [la tige de fleur/le bouquet de fleurs] [au conférencier]
D3-4: Ra	Quand l'applaudissement du public a cessé, le responsable de la réunion [donna] [une tige de fleur/un bouquet de fleurs] [au conférencier]
D3-4: Yek	Quand l'applaudissement du public a cessé, le responsable de la réunion [donna] [une tige de fleur/un bouquet de fleurs] [au conférencier]
D3-5: i	Après quelques heures de délibération, les membres de jury ont décidé de (qu'ils [donnent] [une médaille d'honneur/une récompense spéciale] [au jeune réalisateur]
D3-5: Ra	Après quelques heures de délibération, les membres de jury ont décidé de (qu'ils [donnent] [une médaille d'honneur/une récompense spéciale] [au jeune réalisateur]
D3-5: Yek	Après quelques heures de délibération, les membres de jury ont décidé de (qu'ils [donnent] [la médaille d'honneur/la récompense spéciale] [au jeune réalisateur]
D3-6: i	Un jour avant les vacances, les employés [envoyèrent] [une carte de vœux /un prix d'honneur] [pour le directeur de l'entreprise]
D3-6: Ra	Un jour avant les vacances, les employés [envoyèrent] [la carte de vœux /le prix d'honneur] [pour le directeur de l'entreprise]
D3-6: Yek	un jour avant les vacances, les employés [envoyèrent] [une carte de vœux /un prix d'honneur] [pour le directeur de l'entreprise]
D3-7: i	Quelques jours après le début du deuxième semestre, le proviseur de l'école [envoya] [une invitation/un questionnaire] [pour les parents]
D3-7: Ra	Quelques jours après le début du deuxième semestre, le proviseur de l'école [envoya] [l'invitation/le questionnaire] [pour les parents]
D3-7: Yek	Quelques jours après le début du deuxième semestre, le proviseur de l'école [envoya] [une invitation/un questionnaire] [pour les parents]

D3-8: i	Aux dernières minutes de la troisième séance de cours, le professeur de mathématiques [distribua] [un fascicule /un livre] [parmi les enfants]
D3-8: Ra	Aux dernières minutes de la troisième séance de cours, le professeur de mathématiques [distribua] [le fascicule /le livre] [parmi les enfants]
D3-8: Yek	Aux dernières minutes de la troisième séance de cours, le professeur de mathématiques [distribua] [un fascicule /un livre] [parmi les enfants]
D3-9: i	Après la dernière récréation, les élèves [offrirent] [un collage/une peinture] [à l'instituteur]
D3-9: Ra	Après la dernière récréation, les élèves [offrirent] [le collage/la peinture] [à l'instituteur]
D3-9: Yek	Après la dernière récréation, les élèves [offrirent] [un collage/une peinture] [à l'instituteur]
D3-10: i	Pour l'anniversaire de (leur) dix ans, Grand-père [offrit] [un pistolet/un arc] [aux jumeaux]
D3-10: Ra	Pour l'anniversaire de (leur) dix ans, Grand-père [offrit] [le pistolet/l'arc] [aux jumeaux]
D3-10: Yek	Pour l'anniversaire de (leur) dix ans, Grand-père [offrit] [un pistolet/un arc] [aux jumeaux]
D3-11: i	Quand ils arrivèrent au pied de la montagne, le responsable [donna] [un appareil photo (des jumelles)/une carte] [au chef d'équipe]
D3-11: Ra	Quand ils arrivèrent au pied de la montagne, le responsable [donna] [l'appareil photo (les jumelles)/la carte] [au chef d'équipe]
D3-11: Yek	Quand ils arrivèrent au pied de la montagne, le responsable [donna] [un appareil photo (des jumelles)/une carte] [au chef d'équipe]
D3-12: i	Après avoir terminé les démarches administratives à l'accueil, l'agent de laboratoire [apporta] [un habit spécial/un récipient d'examen] [pour l'accompagnateur du malade]
D3-12: Ra	Après avoir terminé les démarches administratives à l'accueil, l'agent de laboratoire [apporta] [l'habit spécial/le récipient d'examen] [pour l'accompagnateur du malade]
D3-12: Yek	Après avoir terminé les démarches administratives à l'accueil, l'agent de laboratoire [apporta] [un habit spécial/un récipient d'examen] [pour l'accompagnateur du malade]
D3-13: i	Une fois que l'entrée s'est dépeuplée, l'entraîneur [donna] [une plainte /une attestation] [au responsable]
D3-13: Ra	Une fois que l'entrée s'est dépeuplée, l'entraîneur [donna] [la plainte/l'attestation] [au responsable]

D3-13: Yek	Une fois que l'entrée s'est dépeuplée, l'entraîneur [donna] [une plainte /une attestation] [au responsable]
D3-14: i	Au dernier jour de l'école, les enfants [donnèrent] [un album de photos/un cahier de souvenirs] [au gardien de l'école]
D3-14: Ra	Au dernier jour de l'école, les enfants [donnèrent] [l'album de photos/le cahier de souvenirs] [au gardien de l'école]
D3-14: Yek	Au dernier jour de l'école, les enfants [donnèrent] [un album de photos/un cahier de souvenirs] [au gardien de l'école]
D3-15: i	Avant l'attaque de la police, les activistes écologistes [distribuèrent] [un tract/un prospectus] [parmi le public]
D3-15: Ra	Avant l'attaque de la police, les activistes écologistes [distribuèrent] [le tract/le prospectus] [parmi le public]
D3-15: Yek	Avant l'attaque de la police, les activistes écologistes [distribuèrent] [un tract/un prospectus] [parmi le public]

### Questionnaire "phrases à trous" 3 : Exp. D4

Ind	Item
D4-1: Nu	Le directeur de l'hôtel insista plusieurs fois qu'obligatoirement, ils [mettent] [(de la) soupe/fleur(s)] [sur la table]
D4-1: NuMod	Le directeur de l'hôtel insista plusieurs fois qu'obligatoirement, ils [apporte] [(de la) soupe aux nouilles tabrizi/fleur(s) d'orchidée blanche(s)] [sur la table]
D4-2: Nu	Le propriétaire du restaurant a dit que son apprenti est parti pour (qu'il) [apporte] [(de la) farine/(de la) Dugh (boisson au lait fermenté)] [(de la réserve)]
D4-2: NuMod	Le propriétaire du restaurant a dit que son apprenti est parti pour (qu'il) [importera] [(de la) farine de blé complet/(de la) Dugh (boisson au lait fermenté) gazeuse d'Abali] [de la réserve]
D4-3: Nu	Le ministre de l'Agriculture déclara qu'en raison du manque de production nationale, ce ministère [importera] [(du) riz/(de la) crème] [de l'Inde]
D4-3: NuMod	Le ministre de l'Agriculture déclara qu'en raison du manque de production nationale, ce ministère [distribuera] [(du) riz Basmati de haute qualité/(de la) crème allégée ] [de l'Inde]
D4-4: Nu	Le porte-parole du Gouvernement déclara que pour combattre la pollution, (le gouvernement) [distribuera] [masque(s)/(de l')essence] [à Téhéran]
D4-4: NuMod	Le porte-parole du Gouvernement déclara que pour combattre la pollution, (le gouvernement) [apporte] [masque(s) respiratoire(s) à élastiques/(de l')essence importée de haute qualité] [à Téhéran]

D4-5: Nu	Hier, Taraneh a dit que Vahid a promis de (qu'il) [apporte] [(du) fromage/kilim(s)] [du Nord]
D4-5: NuMod	Hier, Taraneh a dit que Vahid a promis de (qu'il) [distribuera] [(du) fromage local au lait de brebis /Kilim(s) industriel(s) en soie] [du Nord]
D4-6: Nu	Le ministre de la Santé déclara que pour combattre la malnutrition, (ce ministre) [distribuera] [(du) lait/(du) pain] [dans les écoles]
D4-6: NuMod	Le ministre de la Santé déclara que pour combattre la malnutrition, (ce ministre) [distribuera] [(du) lait vitaminé en brique/(du) pain au lait chocolaté] [dans les écoles]
D4-7: Nu	Arezou a dit qu'elle a envie de (qu'elle) [mette] [œuf(s)/cornichon(s)] [dans le sandwich]
D4-7: NuMod	Arezou a dit qu'elle a envie de (qu'elle) [mette] [l'œuf cuit et épluché /le cornichon coupé en rondelles ] [dans le sandwich]
D4-8: Nu	Le Croissant Rouge annonça qu'avant l'arrivée de la saison froide, il [envoie] [biscuit(s)/sac(s) de couchage] [au camp]
D4-8: NuMod	Le Croissant Rouge annonça qu'avant l'arrivée de la saison froide, il [envoie] [biscuit(s) Saghehtalay fourré/sac(s) de couchage imperméable] [au camp]
D4-9: Nu	Le responsable de l'organisation de conférence demanda à son assistant de (qu'il) [mette] [feutre(s)/crayon(s)] [dans le dossier]
D4-9: NuMod	Le responsable de l'organisation de conférence demanda à son assistant de (qu'il) [mette] [feutre(s) fluorescent(s) effaçable(s) /crayon(s) rouge(s) taillé(s)] [dans le dossier]
D4-10: Nu	Il paraît que Monsieur Ghanbari a demandé aux parents (qu'ils) [envoient] [maillot(s) (de corps)/conserve(s)] [au dortoir]
D4-10: NuMod	Il paraît que Monsieur Ghanbari a demandé aux parents (qu'ils) [envoient] [maillot(s) (de corps) à manches longues en laine/conserve(s) de ragout de Ghormehsabzi] [au dortoir]
<b>Questionnaire "phrases à trous" 3 : Exp. T1</b>	
<b>Ind</b>	<b>Item</b>
T1-1: A	Grand-mère a dit avec inquiétude qu'elle a entendu que [le mari de Maryam] [a enchanté/a attristé] [la sœur de Parastou]
T1-1: B	Grand-mère a dit avec inquiétude qu'elle a entendu que [le mari de Maryam qui vient de sortir du deuil] [a enchanté/a attristé] [la sœur de Parastou]
T1-1: B'	Grand-mère a dit avec inquiétude qu'elle a entendu que [le mari de Maryam qui vient de sortir du deuil] [a enchanté/a attristé] [la sœur de Parastou]

T1-1: C	Grand-mère a dit avec inquiétude qu'elle a entendu que [le mari de Maryam] [a enchanté/a attristé] [la fête de Parastou]
T1-1: D	Grand-mère a dit avec inquiétude qu'elle a entendu que [le mari de Maryam qui vient de sortir du deuil] [a enchanté/a attristé] [la fête de Parastou]
T1-2: A	Mina dit qu'Azadeh est fâchée parce que [les enfants] [a fatigué/a réveillé] [Alireza]
T1-2: B	Mina dit qu'Azadeh est fâchée parce que [l'enfant qui fait 38,5° degré de fièvre] [a fatigué/a réveillé] [Alireza]
T1-2: B'	Mina dit qu'Azadeh est fâchée parce que [l'enfant qui fait 38,5° degré de fièvre] [a fatigué/a réveillé] [Alireza]
T1-2: C	Mina dit qu'Azadeh est fâchée parce que [l'enfant] [a réveillé/a fatigué] [le bruit]
T1-2: D	Mina dit qu'Azadeh est fâchée parce que [l'enfant qui fait 38,5° degré de fièvre] [a réveillé/a fatigué] [le bruit]
T1-3: A	Ali a dit que Madame Parvina a été très gênée quand [les invité(e)s] [fit sursauter/fit réveiller] [le Grand-père]
T1-3: B	Ali a dit que Madame Parvina a été très gênée quand [les invités qui étaient en train de regarder la télévision] [fit sursauter/fit réveiller] [le Grand-père]
T1-3: C	Ali a dit que Madame Parvina a été très gênée quand [les invités] [fit sursauter/fit réveiller] [la sonnette]
T1-3: D	Ali a dit que Madame Parvina a été très gênée quand [les invités qui étaient en train de regarder la télévision] [fit sursauter/fit réveiller] [la sonnette]
T1-4: A	Manouchehr a dit qu'il est fatigué parce que ça a duré quelques heures jusqu'à ce que [les enfants] [fasse rire/fasse coucher] [la Grand-mère]
T1-4: B	Manouchehr a dit qu'il est fatigué parce que ça a duré quelques heures jusqu'à ce que [les enfants qui pleurnichaient sans cesse] [fasse rire/fasse coucher] [la Grand-mère]
T1-4: C	Manouchehr a dit qu'il est fatigué parce que ça a duré quelques heures jusqu'à ce que [les enfants] [fasse rire/fasse coucher] [la cassette (de contes)]
T1-4: D	Manouchehr a dit qu'il est fatigué parce que ça a duré quelques heures jusqu'à ce que [les enfants qui pleurnichaient sans cesse] [fasse rire/fasse coucher] [la cassette (de contes)]
T1-5: A	Ahmad dit que seulement Grand-mère sait pourquoi [la sœur de Nazanin] [a fait pleurer/a fait souffrir (a contrarié)] [l'ami de Mohsen]
T1-5: B	Ahmad dit que seulement Grand-mère sait pourquoi [la sœur de Nazanin qui a été admise récemment au concours] [a fait pleurer/a fait souffrir (a contrarié)] [l'ami de Mohsen]

T1-5: C	Ahmad dit que seulement Grand-mère sait pourquoi [la sœur de Nazanin] [a fait vexer/a fait souffrir (a contrarié)] [la photo de Mohsen]
T1-5: D	Ahmad dit que seulement Grand-mère sait pourquoi [la sœur de Nazanin qui a été admise récemment au concours] [a fait vexer/a fait souffrir (a contrarié)] [la photo de Mohsen]
T1-6: A	Le gardien dit que personne ne va croire que [le directeur général] [aurait fâché/aurait surpris] [l'agent de nettoyage]
T1-6: B	Le gardien dit que personne ne va croire que [le directeur général qui était en train de danser sur la table] [aurait fâché/aurait surpris] [l'agent de nettoyage]
T1-6: C	Le gardien dit que personne ne va croire que [le directeur général] [aurait fâché/aurait surpris] [l'aspirateur]
T1-6: D	Le gardien dit que personne ne va croire que [le directeur général qui était en train de danser sur la table] [aurait fâché/aurait surpris] [l'aspirateur]
T1-7: A	Reza a dit que même Madame Soleimani a rigolé quand [l'agent d'accueil] [fit danser/fit craindre] [le comptable]
T1-7: B	Reza a dit que même Madame Soleimani a rigolé quand [l'agent d'accueil qui était en train de balayer] [fit danser/fit craindre] [le comptable]
T1-7: C	Reza a dit que même Madame Soleimani a rigolé quand [l'agent d'accueil] [fit danser/fit craindre] [la foudre]
T1-7: D	Reza a dit que même Madame Soleimani a rigolé quand [l'agent d'accueil qui était en train de balayer] [fit danser/fit craindre] [la foudre]
T1-8: A	Mère rit toujours quand Maryam raconte comment ce jour là [l'officier de police] [déconcentra/divertit] [le médecin de la clinique]
T1-8: B	Mère rit toujours quand Maryam raconte comment ce jour là [l'officier de police qui fouillait minutieusement partout] [déconcentra/divertit] [le médecin de la clinique]
T1-8: C	Mère rit toujours quand Maryam raconte comment ce jour là [l'officier de police] [déconcentra/divertit] [le certificat de service militaire accompli]
T1-8: D	Mère rit toujours quand Maryam raconte comment ce jour là [l'officier de police qui fouillait minutieusement partout] [déconcentra/divertit] [le certificat de service militaire accompli]
T1-9: A	Les professeurs racontèrent que les enfants en pleurant attendaient que [le meilleur élève de la classe] [guérisse/sauve] [le pompier]
T1-9: B	Les professeurs racontèrent que les enfants en pleurant attendaient que [le meilleur élève de la classe qui se trouvait par terre dans un état languissant] [guérisse/sauve] [le pompier]
T1-9: C	Les professeurs racontèrent que les enfants en pleurant attendaient que [le meilleur élève de la classe] [guérisse/sauve] [la respiration artificielle]

T1-9: D	Les professeurs racontèrent que les enfants en pleurant attendaient que [le meilleur élève de la classe qui se trouvait par terre dans un état languissant] [guérisse/sauve] [la respiration artificielle]
T1-10: A	Sepideh dit que le juge espère que [la famille de la victime] [satisfasse/encourage] [l'avocat de l'accusé]
T1-10: B	Sepideh dit que le juge espère que [la famille de la victime qui est sous pression depuis quelques mois] [satisfasse/encourage] [l'avocat de l'accusé]
T1-10: C	Sepideh dit que le juge espère que [la famille de la victime] [satisfasse/encourage] [les derniers événements]
T1-10: D	Sepideh dit que le juge espère que [la famille de la victime qui depuis des mois est sous pression] [satisfasse/encourage] [les derniers événements]
T1-11: A	Les enfants disent que le responsable de camp est sûr et certain que [le gardien de l'école] [convainc/dissuade] [le proviseur de l'école]
T1-11: B	Les enfants disent que le responsable de camp est sûr et certain que [le gardien de l'école] [convainc/dissuade] [le proviseur de l'école qui a décidé de venir au camp]
T1-11: C	Les enfants disent que le responsable de camp est sûr et certain que [l'embouteillage de l'itinéraire] [convainc/dissuade] [le proviseur de l'école]
T1-11: D	Les enfants disent que le responsable de camp est sûr et certain que [l'embouteillage de l'itinéraire] [convainc/dissuade] [le proviseur de l'école qui a décidé de venir au camp]
T1-12: A	Le journaliste présent dans la salle dit que personne ne s'attendait à ce que [le présentateur de l'émission] [fasse rire/fasse pleurer] [le ministre de la Culture]
T1-12: B	Le journaliste présent dans la salle dit que personne ne s'attendait à ce que [le présentateur de l'émission] [fasse rire/fasse pleurer] [le ministre de la Culture qui était venu pour distribuer les récompenses]
T1-12: C	Le journaliste présent dans la salle dit que personne ne s'attendait à ce que [le film sélectionné] [fasse rire/fasse pleurer] [le ministre de la Culture]
T1-12: D	Le journaliste présent dans la salle dit que personne ne s'attendait à ce que [le film sélectionné] [fasse rire/fasse pleurer] [le ministre de la Culture qui était venu pour distribuer les récompenses]
T1-13: A	Le président de la faculté disait que les étudiants ne croyaient pas que [le directeur d'études] [convainc/désespère] [le référent externe]
T1-13: B	Le président de la faculté disait que les étudiants ne croyaient pas que [le directeur d'études] [convainc/désespère] [le référent externe qui est un des plus sévères enseignants]
T1-13: C	Le président de la faculté disait que les étudiants ne croyaient pas que [le compte rendu] [convainc/désespère] [le référent externe]
T1-13: D	Le président de la faculté disait que les étudiants ne croyaient pas que [le compte rendu] [convainc/désespère] [le référent externe qui est un des plus sévères enseignants]

T1-14: A	La police dit que d'après les voisins ce n'est pas possible que [le facteur] [aurait intoxiqué/aurait rendu malade] [le portier]
T1-14: B	La police dit que d'après les voisins ce n'est pas possible que [le facteur] [aurait intoxiqué/aurait rendu malade] [le portier qui est un homme fort et sportif]
T1-14: C	La police dit que d'après les voisins ce n'est pas possible que [le Pif Paf (produit anti-insectes)] [aurait intoxiqué/aurait rendu malade] [le portier]
T1-14: D	La police dit que d'après les voisins ce n'est pas possible que [le Pif Paf (produit anti-insectes)] [aurait intoxiqué/aurait rendu malade] [le portier qui est un homme fort et sportif]
T1-15: A	Mehrnouche a dit que sa sœur pense que [le concierge de l'immeuble] [ennuie/embête] [Monsieur Azarmanech]
T1-15: B	Mehrnouche a dit que sa sœur pense que [le concierge de l'immeuble] [ennuie/embête] [Monsieur Azarmanech qui est toujours en train de lire]
T1-15: C	Mehrnouche a dit que sa sœur pense que [le bruit de la salle des machines] [ennuie/embête] [Monsieur Azarmanech]
T1-15: D	Mehrnouche a dit que sa sœur pense que [le bruit de la salle des machines] [ennuie/embête] [Monsieur Azarmanech qui est toujours en train de lire]
T1-16: A	Monsieur Alizadeh a dit que le directeur de l'usine soupçonnait que [le facteur] [aurait étranglé/aurait anesthésié] [le directeur du bureau]
T1-16: B	Monsieur Alizadeh a dit que le directeur de l'usine soupçonnait que [le facteur] [aurait étranglé/aurait anesthésié] [le directeur du bureau qui dormait les après-midi sur son bureau]
T1-16: C	Monsieur Alizadeh a dit que le directeur de l'usine soupçonnait que [le dioxyde de carbone] [aurait étranglé/aurait anesthésié] [le directeur du bureau]
T1-16: D	Monsieur Alizadeh a dit que le directeur de l'usine soupçonnait que [le dioxyde de carbone] [aurait étranglé/aurait anesthésié] [le directeur du bureau qui dormait les après-midi sur son bureau]
T1-17: A	Le journaliste dit qu'aussitôt le sergent a insisté que [le prisonnier évadé] [a touché/a blessé] [le (soldat) conscrit]
T1-17: B	Le journaliste dit qu'aussitôt le sergent a insisté que [le prisonnier évadé] [a touché/a blessé] [le (soldat) conscrit qui assurait un service de garde]
T1-17: C	Le journaliste dit qu'aussitôt le sergent a insisté que [l'éclat d'obus] [a touché/a blessé] [le (soldat) conscrit]
T1-17: D	Le journaliste dit qu'aussitôt le sergent a insisté que [l'éclat d'obus] [a touché/a blessé] [le (soldat) conscrit qui assurait un service de garde]
T1-18: A	Chirine pense que contrairement à ce que tout le monde raconte [la sœur de Mahchid] [guérit/aida] [le frère de Maziyar]

T1-18: B	Chirine pense que contrairement à ce que tout le monde raconte [la sœur de Mahchid] [guérit/aida] [le frère de Maziyar qui fut très dépressif depuis l'adolescence]
T1-18: C	Chirine pense que contrairement à ce que tout le monde raconte [la préparation de Mahchid] [aida/guérit] [le frère de Maziyar]
T1-18: D	Chirine pense que contrairement à ce que tout le monde raconte [la préparation de Mahchid] [aida/guérit] [le frère de Maziyar qui fut très dépressif depuis l'adolescence]
T1-19: A	Sina dit que le médecin d'urgence espérait qu'au plus vite [l'infirmier] [taise/calme] [le malade]
T1-19: B	Sina dit que le médecin d'urgence espérait qu'au plus vite [l'infirmier] [taise/calme] [le malade qui criait de douleur]
T1-19: C	Sina dit que le médecin d'urgence espérait qu'au plus vite [le calmant] [taise/calme] [le malade]
T1-19: D	Sina a dit que le médecin d'urgence espérait qu'au plus vite [le calmant] [taise/calme] [le malade qui criait de douleur]
T1-20: A	Samira dit que le responsable de la salle a dit que soyez sûrs et certains que [le marionnettiste] [divertira/excitera] [les élèves]
T1-20: B	Samira dit que le responsable de la salle a dit que soyez sûrs et certains que [le marionnettiste] [divertira/excitera] [les élèves qui avaient faim et étaient fatigués]
T1-20: C	Samira dit que le responsable de la salle a dit que soyez sûrs et certains que [le théâtre des marionnettes] [divertira/excitera] [les élèves]
T1-20: D	Samira dit que le responsable de la salle a dit que soyez sûrs et certains [le théâtre des marionnettes] [divertira/excitera] [les élèves qui avaient faim et étaient fatigués]

## Questionnaire Jugement d'acceptabilité : Exp. Coordination

Ind	Item
A1-1:C	Maryam a trouvé (cet album de photos) dans le débarras de Grand-mère
A1-1:Coor1	Maryam a trouvé (quelques cassettes vidéo et cet album de photos) dans le débarras de Grand-mère
A1-1:Coor2	Maryam a trouvé (cet album de photos et quelques cassettes vidéo) dans le débarras de Grand-mère
A1-2:C	A mon avis, n'apporte que (ces baskets)
A1-2:Coor1	A mon avis, n'apporte que (un pyjama et ces baskets)
A1-2:Coor2	A mon avis, n'apporte que (ces baskets et un pyjama)
A1-3:C	Pour l'inscription, il suffit de nous envoyer (le formulaire rempli)
A1-3:Coor1	Pour l'inscription, il suffit de nous envoyer (une photo de 4x6 et le formulaire rempli)
A1-3:Coor2	Pour l'inscription, il suffit de nous envoyer (le formulaire rempli et une photo de 4x6)
A1-4:C	Azadeh a été obligée de vendre (son alliance) pour payer le crédit de la maison
A1-4:Coor1	Azadeh a été obligée de vendre (quelques pièces en or et son alliance) pour payer le crédit de la maison
A1-4:Coor2	Azadeh a été obligée de vendre (son alliance et quelques pièces en or) pour payer le crédit de la maison
A1-5:C	Si tu n'as pas beaucoup d'argent, achète seulement (le comprimé de papa)
A1-5:Coor1	Si tu n'as pas beaucoup d'argent, achète seulement (un paquet d'acétaminophène et le comprimé de papa)
A1-5:Coor2	Si tu n'as pas beaucoup d'argent, achète seulement (le comprimé de papa et un paquet d'acétaminophène)
A1-6:C	Pour ce soir, j'ai pris (ce gâteau au chocolat)
A1-6:Coor1	Pour ce soir, j'ai pris (un kilo de glace à la vanille et ce gâteau au chocolat)
A1-6:Coor2	Pour ce soir, j'ai pris (ce gâteau au chocolat et un kilo de glace à la vanille)
A1-7:C	Si tu veux amener Minou avec toi, apporte (le paquet de ses crayons de couleur)
A1-7:Coor1	Si tu veux amener Minou avec toi, apporte (quelques feuilles et le paquet de ses crayons de couleur)
A1-7:Coor2	Si tu veux amener Minou avec toi, apporte (le paquet de ses crayons de couleur et quelques feuilles)
A1-8:C	Nader n'a apporté que (ces kilims-là) de Chiraz
A1-8:Coor1	Nader n'a apporté que (un tapis abîmé et ces kilims-là) de Chiraz
A1-8:Coor2	Nader n'a apporté que (ces kilims-là et un tapis abîmé) de Chiraz
A1-9:C	J'ai acheté (cette chemise) dans les soldes
A1-9:Coor1	J'ai acheté (deux pantalons en jean et cette chemise) dans les soldes
A1-9:Coor2	J'ai acheté (cette chemise et deux pantalons en jean) dans les soldes
A1-10:C	Hier, on est allé avec Parissa à la friperie, j'ai acheté (ce manteau), mais Parissa n'a rien acheté
A1-10:Coor1	Hier, on est allé avec Parissa à la friperie, j'ai acheté (une veste en velours relativement neuve et ce manteau), mais Parissa n'a rien acheté

A1-10:Coor2	Hier, on est allé avec Parissa à la friperie, j'ai acheté (ce manteau et une veste en velours relativement neuve), mais Parissa n'a rien acheté
A1-11:C	Aujourd'hui, j'ai trouvé (ces chèques de voyage) dans la rue
A1-11:Coor1	Aujourd'hui, j'ai trouvé (quelques liasses de billets de mille toumans et ces chèques de voyage) dans la rue
A1-11:Coor2	Aujourd'hui, j'ai trouvé (ces chèques de voyage et quelques liasses de billets de mille toumans) dans la rue
A1-12:C	Il a dit que nous n'avons rien à apporter, prends juste (la clé) et partons vite
A1-12:Coor1	Il a dit nous n'avons rien à apporter, prends juste (un peu d'argent liquide et la clé) et partons vite
A1-12:Coor2	Il a dit nous n'avons rien à apporter, prends juste (la clé et un peu d'argent liquide) et partons vite
A1-13:C	Je dois encore mettre (ma serviette de bain) dans la valise
A1-13:Coor1	Je dois encore mettre (un jeu de draps et ma serviette de bain) dans la valise
A1-13:Coor2	Je dois encore mettre (ma serviette de bain et un jeu de draps) dans la valise
A1-14:C	J'ai apporté comme cadeau (ces porte-clés) pour les enfants
A1-14:Coor1	J'ai apporté comme cadeau (quelques tablettes de chocolat et ces porte-clés) pour les enfants
A1-14:Coor2	J'ai apporté comme cadeau (ces porte-clés et quelques tablettes de chocolat) pour les enfants
A1-15:C	Selon les recommandations du médecin, les soirs, ma mère ne mange que (son repas allégé)
A1-15:Coor1	Selon les recommandations du médecin, les soirs, ma mère ne mange qu' (un peu de yaourt et son repas allégé)
A1-15:Coor2	Selon les recommandations du médecin, les soirs, ma mère ne mange que (son repas allégé et un peu de yaourt)
A1-16:C	Afsaneh a prêté (ces chaussures à talons) à Negar pour le mariage
A1-16:Coor1	Afsaneh a prêté (une robe de soirée et ces chaussures à talons) à Negar pour le mariage
A1-16:Coor2	Afsaneh a prêté (ces chaussures à talons et une robe de soirée) à Negar pour le mariage
A1-17:C	L'après-midi, Amir Hossein a mangé (toutes les graines de tournesol) devant la télévision
A1-17:Coor1	L'après-midi, Amir Hossein a mangé (un bol de reine-claude et toutes les graines de tournesol) devant la télévision
A1-17:Coor2	L'après-midi, Amir Hossein a mangé (toutes les graines de tournesol et un bol de reine-claude) devant la télévision
A1-18:C	J'ai oublié d'acheter un cadeau pour Sepehr, mais j'ai acheté (ce tableau) pour Mojgan
A1-18:Coor1	J'ai oublié d'acheter un cadeau pour Sepehr, mais j'ai acheté (quelques livres et ce tableau) pour Mojgan
A1-18:Coor2	J'ai oublié d'acheter un cadeau pour Sepehr, mais j'ai acheté (ce tableau et quelques livres) pour Mojgan
A1-19:C	Peux-tu m'apporter (ma cigarette-briquet) avant de partir s'il te plait
A1-19:Coor1	Peux-tu m'apporter (un cendrier et ma cigarette-briquet) avant de partir s'il te plait
A1-19:Coor2	Peux-tu m'apporter (ma cigarette-briquet et un cendrier) avant de partir s'il te plait
A1-20:C	Il faut juste que nous apportions (notre cuisinière à gaz de pique-nique)
A1-20:Coor1	Il faut juste que nous apportions (une douzaine de broches et notre cuisinière à gaz de pique-nique)
A1-20:Coor2	Il faut juste que nous apportions (notre cuisinière à gaz de pique-nique et une douzaine de broches)

**Questionnaire Jugement d'acceptabilité : Exp. Vides parasites**

Ind	Item
A2-1:Gap	Toi, je sais pas, mais moi, (pantalon) il ne m'est jamais arrivé d'acheter (de pantalon) sans (l')essayer
A2-2:Gap	(Champignons), si les enfants aiment (les champignons), mets (en) dans la salade
A2-3:Gap	Ma maman, (chaussettes en laine), bien qu'elle n'aime pas du tout (les chaussettes en laine), dans le froid de Canada, elle était obligée d'(en) mettre
A2-4:Gap	Eh ben, (manteau) s'il ne met pas de (manteau), ne lui (en) achetons pas
A2-5:Gap	(Cholehzard), bien que cette année notre voisin ait fait (du Cholehzard), il ne nous (en) a pas apporté
A2-6:Gap	(Griotte), depuis que mon collègue m'apporte (des griottes) de leur jardin, je n'(en) achète plus
A2-7:Gap	La fille de Sara, bien que, (persan), elle comprenne (le persan), elle ne peut pas encore (le) parler
A2-8:Gap	Moi, (glace au safran), bien que j'aime beaucoup (la glace au safran), à cause de mon diabète, je ne dois surtout pas (en) manger
A2-9:Gap	(Film d'horreur), bien que je n'aime pas du tout (les films d'horreur), j'aime bien (en) regarder en compagnie
A2-10:Gap	(Echec), bien que je savais bien jouer (aux échecs), je n'ai tellement pas joué que j'ai oublié

**Questionnaire Jugement d'acceptabilité : distracteurs (phrases mal formées (MF))**

Ind	Item
F-1:BF	Je pense que Mandana ne sait pas où sont les couteaux et les fourchettes
F-2:BF	S'il te plaît, achète un ballon de foot en patchwork pour Hamed
F-3:BF	Prière de ne pas manger ni boire dans cette salle
F-4:BF	Cher Saeid, enlève ce vase de cristal de devant la fenêtre
F-5:BF	Les assiettes jetables et les sacs plastiques nuisent gravement à l'environnement
F-6:BF	Demandez à Madame Azar de mettre un bouquet de fleurs fraîches sur le bureau du directeur
F-7:BF	Manger un ou deux fruits et légumes frais par jour est nécessaire pour la santé
F-8:BF	Rappelez à Monsieur Fateh d'envoyer une carte de vœux à Madame l'ingénieure
F-9:BF	Fermez bien les portes et les fenêtres, l'on prévoit une tempête cette nuit
F-10:BF	Demande à Hamid Reza de nous acheter du pain Sangak au pavot
F-11:BF	Demande à Ali s'il a donné de l'argent au concierge ou pas
F-12:BF	Ava a dit qu'elle a donné les livres à Chiva
F-13:BF	Niloufar pense qu'elle a oublié son portable dans le bus
F-14:BF	Je voulais présenter mon nouveau collègue à Sepideh
F-15:BF	Monsieur le proviseur a entendu parler de cette histoire par ses connaissances



## D Résultats complets des modèles mixtes

Dans cette Annexe, nous présentons les résultats complets de l'ensemble de nos modèles de régression mixtes calculés pour l'analyse de nos données expérimentales, dont le résultat pour les effets fixes a été rapporté dans le chapitre 5, ainsi que dans la section 6.2 du chapitre 6. Il s'agit plus précisément du résumé du modèle calculé dans R obtenu par la fonction *summary*.

### D.1 Expériences sur l'ordre des compléments

#### D.1.1 Expérience D1

Dans cette première expérience, pour neutraliser l'effet de l'ordre d'apparition des options sur l'écran, la génération aléatoire de cet ordre a été utilisée au lieu du contrebalancement. Suite à cette méthode, la distribution de l'ordre n'était pas équilibrée dans l'ensemble des données recueillies. Pour s'assurer que les effets trouvés se maintiennent au-delà de ce biais – c'est-à-dire l'effet de l'amorçage sur le choix de l'ordre dans la phrase complétée par le participant –, nous avons calculé un second modèle incluant ce biais, annoté comme **Prim**, comme variable à effet fixe. Les résultats de ce modèle, que nous présentons ici, est substantiellement identique au modèle rapporté dans le chapitre 5.

De plus, comme nous l'avons précisé dans le texte, pour cette expérience nous avons eu un jeu de données initial que nous avons été amenée à écarter par la suite, car, à cause d'une erreur de codage, chaque participant avait rempli une sélection de 16 items sur les 20 prévus. Nous présenterons dans cette Annexe également les résultats du modèle calculé à partir de ce jeu de données.

##### D.1.1.1 Codage des contrastes

<code>contrasts(D1\$Length)</code>		<code>contrasts(D1\$Given)</code>	
	<code>[,1]</code>		<code>[,1]</code>
Long	1	Given	1
Short	-1	New	-1

N.B. :    Long :    OI>OD    Given :    OI-Donné  
           Short :    OI<OD    New :    OI-Nouveau

##### D.1.1.2 Modèle présenté au chapitre 5 (Tableau 5.3)

Formula: Order~Length\*Given+(Length|Subj)+(Length|Item)+(1|Verb)

AIC	BIC	logLik	deviance
730.9770	780.3916	-354.4885	708.9770

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	1.1572	1.0757	
	LengthLong	1.1448	1.0700	-0.58
Item	(Intercept)	0.5953	0.7716	
	LengthLong	1.0046	1.0023	-0.81
Verb	(Intercept)	0.1536	0.3920	

Number of obs: 660, groups: Subj, 33; Item, 20; Verb, 7

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	2.4055	0.3956	6.080	1.2e-09	***
LengthLong	-1.9723	0.4300	-4.587	4.5e-06	***
GivenNew	-0.9925	0.3244	-3.059	0.00222	**
LengthLong:GivenNew	0.8252	0.4071	2.027	0.04264	*

---

Correlation of Fixed Effects:

	(Intr)	LngthL	GivnNw
LengthLong	-0.712		
GivenNew	-0.520	0.478	
LngthLng:GN	0.414	-0.553	-0.797

### D.1.1.3 Modèle supplémentaire avec le facteur *Prim*

Formula: Order~Length\*Given+Prim+(Length|Subj)+(Length|Item)+(1|Verb)

AIC	BIC	logLik	deviance
684.3398	738.2467	-330.1699	660.3398

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	1.22907	1.1086	
	LengthLong	1.16003	1.0770	-0.58
Item	(Intercept)	0.83550	0.9141	
	LengthLong	0.71316	0.8445	-0.73
Verb	(Intercept)	0.09479	0.3079	

Number of obs: 660, groups: Subj, 33; Item, 20; Verb, 7

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.9184	0.4112	4.665	3.08e-06	***
LengthLong	-1.9037	0.4235	-4.496	6.94e-06	***
GivenNew	-1.0033	0.3428	-2.927	0.00343	**
LengthLong:GivenNew	0.8861	0.4294	2.063	0.03908	*
PrimYes	2.0065	0.3130	6.411	1.44e-10	***

#### D.1.1.4 Modèle supplémentaire avec le jeu de données écarté alternatif

Formula: Order~Length\*Given+Prim+(Length|Subj)+(Length|Item)+(1|Verb)

AIC	BIC	logLik	deviance
896.9208	954.0597	-436.4604	872.9208

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	1.171e+00	1.082349	
	Length1	2.736e-01	0.523113	0.00
Item	(Intercept)	4.429e-01	0.665475	
	Length1	5.964e-02	0.244205	-0.82
Verb	(Intercept)	6.749e-06	0.002598	

Number of obs: 864, groups: Subj, 54; Item, 20; Verb, 7

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.89798	0.23587	3.807	0.000141	***
Length1	-0.73497	0.13269	-5.539	3.05e-08	***
Given1	0.15058	0.09216	1.634	0.102281	
Length1:Given1	-0.30550	0.09346	-3.269	0.001080	**
PrimYes	1.25151	0.22998	5.442	5.27e-08	***

#### D.1.2 Expérience D2

##### D.1.2.1 Codage des contrastes

contrasts(D2\$Length)		contrasts(D2\$Given)	
	[,1]		[,1]
Long	1	Given	1
Short	-1	New	-1

N.B. : Long : OI>OD      Given : OI=OD  
 Short : OI=OD      New : OI-Nouveau

### D.1.2.2 Modèle (Tableau 5.5)

Formula: Order~Length\*Given+(Length|Subj)+(1|Item)+(1|Verb)

AIC	BIC	logLik	deviance
270.2498	308.4473	-126.1249	252.2498

Random effects:

Groups Name	Variance	Std.Dev.	Corr
Subj (Intercept)	4.043e+00	2.010666	
Length1	5.830e-01	0.763541	1.00
Item (Intercept)	1.783e-05	0.004223	
Verb (Intercept)	2.805e+00	1.674741	

Number of obs: 515, groups: Subj, 26; Item, 20; Verb, 7

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.29921	0.83076	5.175	2.28e-07 ***
Length1	1.47901	0.30801	4.802	1.57e-06 ***
Given1	-0.04005	0.20299	-0.197	0.844
Length1:Given1	-0.05943	0.20212	-0.294	0.769

---

Correlation of Fixed Effects:

	(Intr)	Lnngth1	Given1
Length1		0.487	
Given1	-0.019	-0.039	
Lnngth1:Gvn1	-0.017	-0.047	0.481

### D.1.3 Expérience D3

Comme nous l'avons précisé dans l'introduction de cette Annexe, à cause d'une erreur de codage, nous avons été amenée à écarter les réponses de nos 21 premiers participants, pour lesquels la répartition des items dans les conditions ne s'est pas faite de manière unifiée. Nous présentons ici le résultat du modèle calculé avec le jeu de données comprenant l'ensemble des 60 participants.

#### D.1.3.1 Codage des contrastes

contrasts(D3\$Cond)

	Yek	Ra
I	0	0
Yek	1	0
Ra	0	1

**D.1.3.2** Modèle présenté au chapitre 5 (Tableau 5.7)

Formula: Order~Cond+(Cond|Subj)+(Cond|Item)

	AIC	BIC	logLik	deviance
	674.9849	740.5334	-322.4925	644.9849

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.2510	0.5010	
	CondI	1.3332	1.1546	0.08
	CondRa	2.1605	1.4699	-0.73 0.63
Item	(Intercept)	0.2876	0.5363	
	CondI	0.3219	0.5674	0.67
	CondRa	0.4047	0.6362	-0.84 -0.16

Number of obs: 584, groups: Subj, 39; Item, 15

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.4467	0.2206	2.025	0.0429 *
CondI	0.4727	0.3336	1.417	0.1565
CondRa	1.6611	0.3945	4.210	2.55e-05 ***

---

Correlation of Fixed Effects:

	(Intr) CondI
CondI	-0.113
CondRa	-0.641 0.356

**D.1.3.3** Modèle avec le jeu de données complémentaire

Formula: Order~Cond+(Cond|Subj)+(Cond|Item)

	AIC	BIC	logLik	deviance
	999.5914	1071.5939	-484.7957	969.5914

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.4946	0.7033	
	CondI	1.4605	1.2085	-0.18
	CondRa	1.8809	1.3715	-0.39 0.54
Item	(Intercept)	0.5456	0.7386	
	CondI	0.5121	0.7156	0.04
	CondRa	0.6788	0.8239	-0.99 0.12

Number of obs: 898, groups: Subj, 60; Item, 15

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.5291	0.2459	2.151	0.0314	*
CondI	0.1384	0.3079	0.449	0.6531	
CondRa	1.7875	0.3618	4.941	7.77e-07	***

---

Correlation of Fixed Effects:

	(Intr)	CondI
CondI	-0.227	
CondRa	-0.699	0.319

## D.1.4 Expérience D4

### D.1.4.1 Codage des contrastes

contrastes(D4\$Cond)

	Nu-Mod
Nu	0
Nu-Mod	1

### D.1.4.2 Modèle (Tableau 5.9)

Formula: Order~Cond+(Cond|Subj)+(1|Item)

	AIC	BIC	logLik	deviance
	648.4511	674.7926	-318.2256	636.4511

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	1.740	1.319	
	CondNu-M	1.407	1.186	-0.61
Item	(Intercept)	1.778	1.334	

Number of obs: 596, groups: Subj, 60; Item, 10

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	1.4550	0.4835	3.009	0.00262	**
CondNu-M	-1.4759	0.2660	-5.548	2.88e-08	***

---

Correlation of Fixed Effects:

	(Intr)
CondNu-M	-0.336

### D.1.5 Combinaison des Exp. D3 et D4

#### D.1.5.1 Modèle 1 (Tableau 5.11)

contrasts(D3&4\$Cond)

	Nu-mod	Indef	Ra
Nu	0	0	0
Nu-mod	1	0	0
Indef	0	1	0
Ra	0	0	1

Formula: Order~Cond+(Cond|Item)+(1|Subj)+(1|Verb)

	AIC	BIC	logLik	deviance
	1676.5214	1761.4688	-822.2607	1644.5214

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.39251	0.6265	
Item	(Intercept)	0.88763	0.9421	
	CondNu-M	0.06034	0.2456	1.00
	CondIndef	0.37887	0.6155	-0.81 -0.80
	CondRa	1.11411	1.0555	-1.00 -1.00 0.75
Verb	(Intercept)	0.09989	0.3160	

Number of obs: 1494, groups: Subj, 60; Item, 25; Verb, 8

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.2128	0.3667	-3.307	0.000943 ***
CondNu-M	1.2228	0.2178	5.615	1.96e-08 ***
CondIndef	1.7113	0.3935	4.349	1.37e-05 ***
CondRa	3.1923	0.3924	8.136	4.07e-16 ***

---

Correlation of Fixed Effects:

	(Intr)	CondBm	CondInd
CondNu-M	0.009		
CondIndef	-0.814	-0.005	
CondRa	-0.813	-0.003	0.763

#### D.1.5.2 Modèle 2 (Tableau 5.12)

contrasts(D3&4\$Cond)

```

      Indef Ra
Nu      0  0
Indef   1  0
Ra      0  1

Formula: Order~Cond+(Cond|Item)+(1|Subj)+(1|Verb)
      AIC      BIC    logLik  deviance
1322.4675 1378.4216 -650.2337 1300.4675
Random effects:
Groups Name          Variance Std.Dev.  Corr
Subj  (Intercept) 0.4375   0.6614
Item  (Intercept) 1.1978   1.0945
      CondIndef   0.4483   0.6695  -0.88
      CondRa      1.4628   1.2095  -1.00  0.86
Verb  (Intercept) 0.0997   0.3158
Number of obs: 1196, groups: Subj, 60; Item, 25; Verb, 8
Fixed effects:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.2939      0.4113  -3.146  0.00166 **
CondIndef    1.7955      0.4360   4.118  3.82e-05 ***
CondRa       3.2810      0.4341   7.558  4.08e-14 ***
---
Correlation of Fixed Effects:
      (Intr) CondInd
CondIndef -0.845
CondRa    -0.846  0.804

```

## D.1.6 Expérience T1

### D.1.6.1 Codage des contrastes

```

contrasts(D2$Animacy)           contrasts(T1$Length)
      [,1]                       [,1]
Anim      1                      Long      1
Inanim   -1                      Short    -1

```

N.B. : Anim : Suj-Animé      Long : OD>Suj  
       Inanim : Suj-Inanimé      Short : OD=Suj

**D.1.6.2 Modèle (Tableau 5.15)**

Formula: Order~ Animacy\*Length+(Animacy|Subj)+(Animacy|Item)

AIC	BIC	logLik	deviance
442.0283	490.5613	-211.0142	422.0283

Random effects:

Groups Name	Variance	Std.Dev.	Corr
Subj (Intercept)	2.49415	1.5793	
Animacy1	0.17063	0.4131	-0.73
Item (Intercept)	0.52343	0.7235	
Animacy1	0.01371	0.1171	0.81

Number of obs: 947, groups: Subj, 60; Item, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.74497	0.32463	11.536	<2e-16 ***
Animacy1	0.40028	0.17186	2.329	0.0199 *
Length1	-0.10136	0.15492	-0.654	0.5130
Animacy1:Length1	-0.01201	0.15470	-0.078	0.9381

---

Correlation of Fixed Effects:

	(Intr)	Anmcy1	Lngth1
Animacy1		0.038	
Length1	-0.044		-0.043
Anmcy1:Lng1	-0.019	-0.055	0.324

**D.2 Expériences de jugements d'acceptabilité**

Notons que dans les modèles présentés ici, l'échelle de Likert, [1,7], est remplacée par une échelle de [-3,+3].

**D.2.1 Expérience Vide parasite****D.2.1.1 Codage des contrastes**

contrasts(Gap\$Cond)

	Gap
Gap-control	0
Gap	1

### D.2.1.2 Modèle

Formula: Rate~Cond+(Cond|Subj)+(Cond |Item)

AIC	BIC	logLik	deviance
2636.334	2677.794	-1309.167	2618.334

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.88070	0.9385	
	CondGap	0.47549	0.6896	0.04
Item	(Intercept)	0.06896	0.2626	
	CondGap	0.59523	0.7715	0.17
Residual		1.48160	1.2172	

Number of obs: 740, groups: Subj, 74; Item, 10

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.912	0.151	12.662
CondGap	-1.424	0.272	-5.237

Correlation of Fixed Effects:

	(Intr)
CondGap	-0.004

## D.2.2 Expérience Coordination

### D.2.2.1 Modèle 1

contrasts(ET\$V1)

	Coor	Post
Cont	0	0
Coor	1	0
Post	0	1

Formula: Rate~V1+(1|Subj)+(V1|Item)

AIC	BIC	logLik	deviance
5184.330	5242.628	-2581.165	5162.330

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.81064	0.9004	
Item	(Intercept)	0.51977	0.7210	
	V1Coor	0.13904	0.3729	-0.66

V1Post 0.08871 0.2978 -0.89 0.65  
 Residual 1.62432 1.2745

Number of obs: 1480, groups: Subj, 74; Item, 20

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.9831	0.2061	9.621
V1Coor	-0.5771	0.1234	-4.675
V1Post	-0.5892	0.1130	-5.213

Correlation of Fixed Effects:

	(Intr) V1Coor
V1Coor	-0.568
V1Post	-0.649 0.656

### D.2.2.2 Modèle 2

contrasts(ETx2\$V1)-> V1 : Construction

	[,1]
Coor	1
Post	-1

contrasts(ETx2\$V2)-> V2 : Ordre

	[,1]
[+Ra]-[-Ra]	1
[-Ra]-[+Ra]	-1

Formula: Rate~V1\*V2+(V1|Subj)+(V1|Item)

	AIC	BIC	logLik	deviance
	4199.353	4255.197	-2088.677	4177.353

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subj	(Intercept)	0.89963	0.9485	
	V11	0.03283	0.1812	-0.28
Item	(Intercept)	2.71805	1.6487	
	V11	0.05435	0.2331	0.79
	Residual	1.55615	1.2475	

Number of obs: 1184, groups: Subj, 74; Item, 20

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	1.399795	0.386498	3.622

---

V11	0.005885	0.066981	0.088
V21	-0.219068	0.036505	-6.001
V11:V21	-0.187648	0.036386	-5.157

## Correlation of Fixed Effects:

	(Intr)	V11	V21
V11	0.558		
V21	0.000	0.000	
V11:V21	0.000	0.000	0.004

# Liste des tableaux

1.1	Suffixes personnels verbaux (pour <i>xordan</i> ‘manger’) . . . . .	26
1.2	Pronoms personnels clitiques . . . . .	27
4.1	Corpus BK : Liste partielle des verbes potentiellement ditransitifs . .	135
4.2	DS1 : Lemmes verbaux et leurs fréquences . . . . .	141
4.3	DS1 : Distribution de l’ordre par les variables $[\pm R\bar{A}]$ et $[\pm Nu]$ . . . . .	142
4.4	DS1 : Matrice de confusion à partir de l’hypothèse MDO . . . . .	143
4.5	DS1 : Distribution de l’ordre par la variable <b>COD-TYPE</b> . . . . .	145
4.6	DS1 : Matrice de confusion à partir de la variable <b>CODTYPE</b> . . . . .	145
4.7	DS1 : Distribution de la variable <b>COLL</b> par la variable <b>COD-TYPE</b> . . .	146
4.8	DS1 : Distribution de la variable <b>COLL</b> par la variable <b>COD-TYPE</b> . . .	147
4.9	DS2 : Distribution de l’ordre par les variables $[\pm R\bar{A}]$ et $[\pm Nu]$ . . . . .	148
4.10	DS2 : Distribution de l’ordre par la variable <b>COD-TYPE</b> . . . . .	148
4.11	DS2 : Matrice de confusion à partir de la variable <b>CODTYPE</b> . . . . .	149
4.12	DS2 : Matrice de confusion à partir de la variable <b>R<math>\bar{A}</math></b> . . . . .	149
4.13	DS2 : Matrice de confusion à partir de la variable <b>Nu</b> . . . . .	149
4.14	DS2 : Distribution de l’ordre par les variables <b>COLL</b> et <b>COD-TYPE</b> . . .	150
4.15	Mod-DS2 : Résumé des résultats . . . . .	154
4.16	Verbe <i>didan</i> ‘voir’ : Distribution de l’ordre par la variable <b>COD-TYPE</b> .	156
5.1	Exp. D1 : Exemple d’un item dans ses quatre conditions . . . . .	174
5.2	Exp. D1 : Distribution de l’ordre pour chaque condition . . . . .	175
5.3	Exp. D1 : Résumé des résultats du modèle optimal pour les effets fixes	177
5.4	Exp. D2 : Distribution de l’ordre pour chaque condition . . . . .	179
5.5	Exp. D2 : Résumé des résultats du modèle optimal pour les effets fixes	181
5.6	Exp. D3 : Distribution de l’ordre pour chaque condition . . . . .	187
5.7	Exp. D3 : Résumé des résultats du modèle optimal pour les effets fixes	188
5.8	Exp. D4 : Distribution de l’ordre pour chaque condition . . . . .	191
5.9	Exp. D4 : Résumé des résultats du modèle optimal pour les effets fixes	193
5.10	Exp. D3-D4 : Distribution de l’ordre pour chaque condition . . . . .	194

5.11	Exp. D3-D4-M1 : Résumé des résultats du modèle optimal pour les effets fixes . . . . .	194
5.12	Exp. D3-D4-M1 : Résumé des résultats du modèle optimal pour les effets fixes . . . . .	195
5.13	Exp. D1-D4 : % de OD-OI-V . . . . .	196
5.14	Exp. T1 : Distribution de l'ordre pour chaque condition . . . . .	203
5.15	Exp. T1 : Résumé des résultats du modèle optimal pour les effets fixes	204

# Table des figures

4.1	DS1 : Histogramme des verbes par fréquence totale du lemme en BK	142
4.2	DS2 : Boîte à moustache pour la longueur relative par l'ordre - COD Marqués . . . . .	151
4.3	DS2 : Boîte à moustache pour la longueur du SP par l'ordre - COD Nus . . . . .	152
4.4	DS2 : Boîte à moustache pour la longueur relative par l'ordre - COD Intermédiaires . . . . .	152
5.1	Exemple d'écran pour la tâche de rappel de phrase (Yamashita et Chang, 2001) . . . . .	163
5.2	Exemple d'écran pour la tâche de complétion de phrases dans notre paradigme . . . . .	165
5.3	Exp. D1 : Copie d'écran pour un item donné . . . . .	173
5.4	Exp. D1 : Distribution de l'ordre pour le lemme verbal . . . . .	176
5.5	Exp. D1 : % de l'ordre OD-OI-V . . . . .	176
5.6	Exp. D2 : Distribution de l'ordre pour le lemme verbal . . . . .	180
5.7	Exp. D2 : % de l'ordre OI-OD-V . . . . .	180
5.8	Exp. D3 : Copie d'écran pour un item donné . . . . .	186
5.9	Exp. D3 : Distribution de l'ordre . . . . .	187
5.10	Exp. D3 : Distribution de l'ordre pour le lemme verbal . . . . .	188
5.11	Exp. D4 : Distribution de l'ordre . . . . .	192
5.12	Exp. D4 : Distribution de l'ordre pour le lemme verbal . . . . .	192
5.13	La probabilité de l'ordre OD-OI en fonction du degré de détermina- tion du COD . . . . .	196
5.14	Exp. T1 : Copie d'écran pour un item donné . . . . .	202
5.15	Exp. T1 : % de l'ordre SOV . . . . .	203
6.1	Exemple d'échelle de Likert utilisée pour l'élicitation des jugements .	245
6.2	Exp. Vide parasite : Moyennes et intervalles de confiance à 95% par item . . . . .	248

6.3	Exp. Coordination : Moyennes et intervalles de confiance à 95%	. . . 254
-----	--	-----------

# Table des matières

Introduction	1
<b>I Problématique et méthodologie</b>	<b>7</b>
<b>1 Ordre des mots en persan</b>	<b>9</b>
1.1 Quelques éléments de la syntaxe du persan . . . . .	10
1.1.1 La direction de la tête . . . . .	11
1.1.2 La détermination nominale . . . . .	13
1.1.3 Le marquage différentiel de l'objet . . . . .	17
1.1.4 Les prédicats complexes . . . . .	22
1.1.5 La réalisation des arguments . . . . .	26
1.2 L'ordre canonique des constituants dans la phrase . . . . .	30
1.2.1 Phrase déclarative . . . . .	31
1.2.2 Phrase non déclarative . . . . .	36
1.2.3 Proposition subordonnée . . . . .	38
1.3 La variation de l'ordre dans la phrase déclarative . . . . .	40
1.3.1 L'ordre linéaire et la structure informationnelle . . . . .	43
1.3.2 Le placement de l'accent nucléaire . . . . .	45
1.3.3 Les ordres non canoniques . . . . .	47
1.4 Les traitements syntaxiques existants . . . . .	51
1.4.1 <i>Scrambling</i> dans le domaine perversal . . . . .	52
1.4.2 Le positionnement postverbal des constituants . . . . .	57
1.5 Conclusion . . . . .	59
<b>2 Étudier la variation de l'ordre des mots : État de l'art</b>	<b>61</b>
2.1 Qu'est-ce que la variation de l'ordre des mots? . . . . .	62
2.1.1 Phénomènes relevant de la variation de l'ordre . . . . .	62
2.1.2 La variation de l'ordre comme objet d'étude en syntaxe . . . . .	64
2.2 Évolution de l'étude de la variation de l'ordre . . . . .	65

2.2.1	La grammaire générative . . . . .	67
2.2.2	La typologie linguistique . . . . .	76
2.2.3	Les approches cognitivo-fonctionnelles . . . . .	78
2.3	Les problématiques principales liées à la variation de l'ordre des mots	83
2.3.1	Les facteurs influant le choix de l'ordre linéaire . . . . .	83
2.3.2	Langues SVO <i>vs.</i> Langues SOV . . . . .	94
2.4	Conclusion . . . . .	102
<b>3</b>	<b>Démarche et questions de recherche</b>	<b>105</b>
3.1	Méthodologie et approche adoptées . . . . .	106
3.1.1	Le choix méthodologique . . . . .	106
3.1.2	Les méthodes quantitatives en linguistique . . . . .	107
3.1.3	Les outils statistiques utilisés dans cette étude . . . . .	110
3.2	Questions de recherche . . . . .	111
3.2.1	Le choix des phénomènes étudiés . . . . .	112
3.2.2	Questions relatives à la syntaxe du persan . . . . .	113
3.2.3	Questions relatives aux débats théoriques sur les préférences .	121
3.3	Conclusion . . . . .	124
<b>II</b>	<b>Études empiriques</b>	<b>127</b>
<b>4</b>	<b>Variation de l'ordre des mots dans le corpus Bijankhan</b>	<b>129</b>
4.1	Le corpus utilisé . . . . .	130
4.2	L'ordre relatif entre le COD et le COI dans les phrases ditransitives	133
4.2.1	Constitution de l'échantillon . . . . .	134
4.2.2	Table de données préliminaire . . . . .	141
4.2.3	Table de données finale . . . . .	147
4.2.4	Analyse multifactorielle . . . . .	152
4.3	L'ordre relatif entre le sujet et le COD dans les phrases transitives	154
4.4	Conclusions . . . . .	156
<b>5</b>	<b>Études expérimentales de la variation de l'ordre des mots</b>	<b>159</b>
5.1	Le choix de notre paradigme expérimental . . . . .	160
5.1.1	Paradigmes expérimentaux existants . . . . .	160
5.1.2	Paradigme adopté pour cette étude . . . . .	163
5.2	Mise en place des expériences et récolte des données . . . . .	165
5.2.1	Recrutement de participants . . . . .	166
5.2.2	Déroulement des questionnaires . . . . .	166

5.2.3	Exécution de la tâche . . . . .	167
5.2.4	Annotation des résultats . . . . .	168
5.3	Les phrases ditransitives . . . . .	168
5.3.1	Expérience D1 : Les COD non marqués indéfinis . . . . .	169
5.3.2	Expérience D2 : Les COD non marqués nus modifiés . . . . .	178
5.3.3	Expérience D3 : Les COD non nus . . . . .	182
5.3.4	Expérience D4 : Les COD nus . . . . .	188
5.3.5	Combinaison des Expériences D3 et D4 . . . . .	193
5.3.6	Bilan . . . . .	196
5.4	Les phrases transitives . . . . .	197
5.4.1	Expérience T1 . . . . .	197
5.5	Conclusions . . . . .	205
 <b>III Discussion générale et apports théoriques</b>		 <b>209</b>
<b>6</b>	<b>Renouveler l'analyse de la structure syntagmatique du persan</b>	<b>211</b>
6.1	Hypothèse de la position syntaxique double de l'objet (PSDO) . . . . .	212
6.1.1	Objets générés dans deux positions distinctes . . . . .	214
6.1.2	Objets générés dans la même position . . . . .	218
6.2	Examen des arguments en faveur de l'hypothèse de la PSDO . . . . .	220
6.2.1	Fusion sémantique avec le verbe . . . . .	222
6.2.2	Portée des quantifieurs . . . . .	225
6.2.3	Liage des anaphores . . . . .	233
6.2.4	Autorisation des vides parasites . . . . .	242
6.2.5	Coordination . . . . .	251
6.2.6	Conclusion . . . . .	255
6.3	Structure hiérarchique vs. structure plate . . . . .	256
<b>7</b>	<b>Considérations sur les préférences de l'ordre des mots</b>	<b>263</b>
7.1	Les facteurs influant sur l'ordre linéaire . . . . .	263
7.1.1	Définitude . . . . .	264
7.1.2	Statut informationnel . . . . .	264
7.1.3	Poids grammatical . . . . .	265
7.1.4	Hiérarchie de la personne . . . . .	266
7.1.5	Dépendance sémantique . . . . .	267
7.1.6	Biais lexical du verbe . . . . .	268
7.2	La préférence "long-avant-court" en persan et son impact théorique . . . . .	268

---

<b>Conclusion</b>	<b>273</b>
<b>Bibliographie</b>	<b>278</b>
<b>Annexes</b>	<b>307</b>
A  Algorithme utilisé pour lemmatiser le corpus Bijankhan . . . . .	309
B  Les lemmes verbaux de la table de données DS1 . . . . .	311
C  Les stimuli de nos expériences . . . . .	313
D  Résultats complets des modèles mixtes . . . . .	355
D.1  Expériences sur l'ordre des compléments . . . . .	355
D.2  Expériences de jugements d'acceptabilité . . . . .	363
<b>Liste des tableaux</b>	<b>368</b>
<b>Table des figures</b>	<b>370</b>

## **La variation de l'ordre des constituants dans le domaine préverbal en persan : approche empirique**

Cette thèse propose une étude quantitative de la variation de l'ordre des constituants en persan avec un intérêt particulier pour l'ordre relatif entre le COD et le COI étant donné son rôle crucial dans les analyses de la structure du SV. Afin de remédier à une lacune empirique dont souffre l'étude de la syntaxe du persan, notre premier objectif est d'évaluer, à partir de données empiriques robustes, l'hypothèse largement admise selon laquelle il existe un ordre relatif canonique dichotomique entre les compléments verbaux, dépendant du marquage différentiel de l'objet (MDO). Notre second objectif, relatif à la linguistique générale et à la typologie, est de contribuer aux débats controversés sur les préférences translinguistiques de l'ordre des mots en étudiant, dans une langue SOV à structure mixte, les effets des facteurs tels que le poids (ou la longueur relative). Les résultats de nos études de corpus et de nos expériences montrent l'inadéquation du critère MDO pour expliquer l'ordre relatif entre le COD et le COI. Cette conclusion nous conduit à réfuter également l'hypothèse de la position syntaxique double de l'objet au profit d'une structure plate pour le SV. De plus, nos données révèlent une préférence "long-avant-court" subordonnée aux facteurs contribuant à la saillance, tels que la définitude, l'animéité et le rôle grammatical. Nous arguons que cette préférence échappe, en partie ou totalement, aux modèles du traitement se fondant sur la distance entre la tête et ses dépendants, alors qu'elle est compatible avec l'hypothèse selon laquelle dans les langues SOV l'accessibilité conceptuelle des constituants longs favorise leur production plus en amont dans la phrase.

**Mots clés :** Syntaxe, Persan, Ordre des mots, Études de corpus, Études expérimentales, Études des préférences, Poids grammatical, Production de la parole, Marquage différentiel de l'objet, Structure syntagmatique.

## **Word order variations in the preverbal domain in Persian : an empirical approach**

This thesis proposes a quantitative study of word order variations in Persian, focusing on the relative order between the direct object (DO) and the indirect object (IO). The latter plays a crucial role in the theoretical analyses of the VP, which in the absence of quantitative studies lack solid empirical underpinning. My first goal is to contribute to the study of Persian syntax by providing reliable data in order to evaluate the prevailing hypothesis according to which there exists a dual canonical relative order between the two objects triggered by the Differential Object marking (DOM). My second goal is to contribute to the ongoing debates on word order preferences in general linguistics and typology by bringing in data on an SOV language with mixed head-direction. To this end, I study the effect of factors such as grammatical weight (or relative length), which are claimed to influence the linear order across languages. First, the results of our corpus and experimental studies show that the DOM account of the relative order between the DO and the IO is flawed. Based on this conclusion, I also reject the two object positions hypothesis and plead for a flat structure view of the VP. Second, our data reveal a long-before-short preference, which is shown to depend on the effect of salience-enhancing factors such as definiteness, animacy and the grammatical role. I argue that while this preference is, either totally or partially, incompatible with the predictions of processing-oriented dependency-based models, it can be accounted for by production models assuming that the greater conceptual accessibility of longer constituents favors their early position in SOV languages.

**Keywords :** Syntax, Persian, Word order, Corpus studies, Experimental studies, Ordering preferences, Length-based effects, Sentence production, Differential object marking, Phrase structure.

UNIVERSITÉ SORBONNE NOUVELLE - PARIS 3

École doctorale 268 – Langage et langues : description, théorisation, transmission

UMR 7528 "Mondes iranien et indien"

Maison de la recherche, 4 rue des Irlandais, 75005 Paris