

Identification de constructions grammaticales en corpus : une approche quantitative de l'augmentation de valence¹

Florent Perek – Albert-Ludwigs Universität Freiburg & S.T.L. - Université Lille 3

Résumé : Cet article présente les résultats d'une étude sur corpus portant sur la valence verbale dans le cadre théorique des grammaires de construction. Nous décrivons une des techniques statistiques développées dans une étude précédente pour identifier des constructions grammaticales en se basant sur le potentiel de ces constructions à augmenter la valence des verbes. Cette technique a été testée sur le corpus ICE-GB pour deux constructions de l'anglais décrites dans la littérature. Notre étude montre que l'usage des verbes ne permet pas d'identifier l'augmentation de valence avec fiabilité et révèle les préférences constructionnelles inattendues de certains verbes, qui remettent en question la validité de certains principes théoriques dans une approche du langage basée sur l'usage.

Abstract : This paper presents the results of a corpus study on verbal valence in a construction grammar framework. We describe one of the statistical techniques developed in a previous study aiming at identifying grammatical constructions thanks to their potential to add arguments to the valence of verbs. This technique was tested on the ICE-GB corpus for two constructions of English described in the literature. Our study shows that verb usage do not allow us to identify valence augmentation with enough reliability and reveals the unexpected constructional preferences of some verbs, which challenges the validity of some theoretical principles in a usage-based approach to language.

1 Introduction

Dans cet article, nous décrivons puis évaluons une méthode d'identification de certaines constructions verbales dans un corpus sur la base de leur capacité à étendre la valence du verbe. La technique présentée est tirée d'une étude plus large (Perek 2008) dont l'objectif était de développer et tester des approches quantitatives pour l'identification automatique de constructions argumentales (Goldberg 1995) en corpus. Notre proposition s'inscrit dans la continuité des travaux de Stefanowitsch & Gries (2003), à qui elle emprunte le concept de collocation et l'approche quantitative générale, tout en se voulant orientée vers l'exploration de données et l'identification de nouvelles structures plutôt que sur la description de l'usage de structures précédemment reconnues.

Dans la section suivante, nous présentons le phénomène d'augmentation de valence et montrons sa pertinence pour la théorie des constructions argumentales de Goldberg (1995). Nous décrivons ensuite une approche quantitative de la valence des verbes et testons son applicabilité à l'identification du phénomène d'augmentation de valence subie par certains verbes, en l'appliquant à deux constructions argumentales manuellement identifiées dans le

¹ Nous tenons à remercier les relecteurs de cet article, Gilles Col et Jacques François, pour leurs conseils et leurs remarques constructives. Nous assumons la pleine responsabilité d'éventuelles erreurs subsistantes.

corpus ICE-GB. Nous verrons que les résultats montrent que les caractéristiques de l'usage des verbes en corpus ne permet pas d'atteindre ce but, et nous tenterons en conclusion d'en tirer les conséquences méthodologiques et théoriques.

2 Augmentation de valence et structures argumentales

Dans cette section, nous présentons le phénomène d'augmentation de valence et montrons sa pertinence pour la théorie des constructions argumentales.

Il est bien connu que la plupart des verbes peuvent être utilisés dans différentes *structures argumentales* spécifiant la *valence syntaxique*, c'est-à-dire « l'ensemble des arguments exprimés linguistiquement dans une proposition donnée » (Payne 1997 : 169–170, notre traduction) et la réalisation morphosyntaxique de ces arguments. Par exemple, *break* peut apparaître dans une construction transitive (*He broke the glass*) ou intransitive (*The glass broke*) ; dans le premier cas, la structure argumentale consiste en un argument *agent* et un argument *patient* (la valence syntaxique) respectivement réalisés comme sujet et objet direct ; dans le second cas, un seul de ces arguments, le *patient*, est réalisé comme sujet. Les arguments de la valence syntaxique sont choisis parmi les arguments de la *valence sémantique* du verbe, c'est-à-dire « l'ensemble des arguments nécessairement présents dans la scène dénotée par le verbe » (*ibid.*).

Il peut arriver cependant qu'un verbe subisse une *augmentation de valence*, c'est-à-dire que sa valence syntaxique contient des arguments ne provenant pas de la valence sémantique du verbe. C'est le cas par exemple des verbes de création, pouvant apparaître de manière productive dans la syntaxe à double objet direct où ils se voient adjoindre un argument *destinataire* :

- (1) Irene baked John a cake.
'Irene a préparé un gâteau à/pour John'
- (2) Sally painted John a picture.
'Sally a peint un tableau à/pour John'

Les verbes *bake* et *paint* ne présupposent pas d'argument *destinataire* dans leur autres emplois (principalement transitifs), pourtant (1) et (2) expriment la notion d'un transfert intentionnel de l'objet créé au référent du premier objet direct. Le phénomène d'augmentation de valence montre l'adéquation d'une approche constructionnelle des structures argumentales (Goldberg 1995) considérant que les structures argumentales sont des constructions indépendantes capables d'apporter elles-mêmes du sens. Les *constructions argumentales* (*argument structure constructions*) sont définies par Goldberg (1995) comme une association symbolique entre (i) côté sens, une représentation schématisée d'un événement incluant une liste de rôles sémantiques, et (ii) côté forme, les spécifications morphosyntaxiques pour réaliser ces rôles. Par exemple, la construction ditransitive² (Goldberg 1995, chapitre 6),

² Dans cet article, nous appelons « ditransitive » la construction composée de deux objets directs, c-à-d. deux groupes nominaux, à ne pas confondre avec la construction dite « *to*-dative », composée d'un objet direct

représentée schématiquement en (3), associe une configuration syntaxique contenant deux objets directs à une scène de changement de possession, dans laquelle les actions d'un *agent* (réalisé par le sujet) ont pour effet qu'un *destinataire* (premier objet direct) devient possesseur d'un *thème* (second objet direct).

(3)Forme : Sujet_X V Objet1_Y Objet2_Z
 Sens : X_{Agent} CAUSE (Y_{Destinataire} RECOIT Z_{Thème})

Une construction peut être « greffée » sur un verbe (pour emprunter la terminologie de François & Sénéchal 2008) si et seulement si (i) les rôles sémantiques du verbe sont compatibles avec ceux de la construction et (ii) le sens du verbe est compatible avec le sens constructionnel (selon un choix limité de relations sémantiques prédéfinies). Dans l'approche constructionnelle, les cas d'augmentation de valence tels que (1) et (2) sont issues de la greffe de la construction ditransitive sur un verbe transitif dont l'argument sujet et l'argument objet fusionnent respectivement avec les rôles *agent* et *thème* de la construction. L'argument *destinataire* est ajouté par la construction, et l'interprétation sémantique résulte de l'intégration du sens du verbe dans le sens schématique dénoté par la construction.

La même construction peut bien sûr être greffée sur des verbes de transfert trivalents comme *give*, *hand* ou *send*, dans ce cas il y a (i) une correspondance parfaite entre les arguments du verbe et les arguments de la construction et (ii) redondance entre la contribution sémantique du verbe et celle de la construction. Autrement dit, le sens constructionnel, associé à l'argument supplémentaire apporté par la construction, n'est visible que dans des exemples tels que (1) et (2). De tels exemples ont donc une importance primordiale pour motiver l'existence d'une construction : ils sont un indice que l'interprétation nécessite plus que ce que peut fournir la sémantique lexicale du verbe, ce qu'une approche constructionnelle explique par la contribution sémantique de la syntaxe.

3 Analyse des collexèmes distinctifs et augmentation de valence

Le phénomène d'augmentation de valence implique qu'il existe dans la distribution de certaines constructions argumentales des verbes capables d'apparaître avec deux syntaxes différentes : (i) dans une syntaxe « simple » ne contenant que des arguments de la valence sémantique du verbe, et (ii) dans une syntaxe « étendue » contenant les mêmes arguments plus un argument supplémentaire apporté par la construction. Dans cette section, nous présentons une méthode quantitative basée sur un développement de l'analyse collostructionnelle (Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004), qui permet de déterminer, pour un verbe donné, quelle est la valence syntaxique (syntaxe simple ou syntaxe

nominal et d'un groupe prépositionnel introduit par *to*. Comme nous le signale à juste titre J. François (c.p.), une certaine ambiguïté terminologique peut exister du fait que le terme « ditransitif » est parfois utilisé en grammaire française pour faire référence à la variante prépositionnelle (le français n'ayant pas de construction ditransitive au sens strict). Nous préférons toutefois conserver ce terme pour sa concision et par souci de continuité avec la littérature anglo-saxonne, et avertissons le lecteur que nous l'utilisons dans son sens restreint de « construction comprenant deux objets directs, c-à-d. non introduits par une préposition ».

étendue) la plus typique dans l'usage réel de la langue, tel que nous le présente un corpus. Nous présentons d'abord les fondamentaux de l'analyse collostructionnelle (Stefanowitsch & Gries 2003) avant de décrire comment nous appliquons cette méthode à notre étude de la valence syntaxique.

L'analyse collostructionnelle de Stefanowitsch & Gries (2003) vise à réhabiliter les corpus dans l'étude de la grammaire en offrant une approche rigoureuse des données quantitatives. Les fréquences d'occurrence en corpus sont potentiellement biaisées par des contingences extra-linguistiques, mais cette remarque n'est valide qu'en ce qui concerne les fréquences brutes. L'idée générale est que la dichotomie présence/absence d'un motif linguistique doit être envisagée comme une quadruple distinction prenant en compte la notion de significativité statistique. Considérer la présence ou l'absence d'un phénomène n'est pas suffisant : on doit aussi considérer si ce phénomène est significativement présent ou significativement absent. De façon générale, si un élément E est absent ou rare dans une structure S, cette absence/rareté doit être envisagée en fonction de la fréquence d'occurrence de E et de S dans tout le corpus et sera jugée d'autant plus significative que E est courant dans d'autres structures que S, et que S est courant avec d'autres éléments que E. L'analyse collostructionnelle est basée sur ces principes. Stefanowitsch & Gries (2003) présentent une série de procédures pour calculer l'attraction statistique entre des phénomènes linguistiques apparaissant dans un corpus. Ces méthodes partagent l'objectif des analyses collocationnelles, typiques des deux dernières décennies de linguistique de corpus néo-Firthienne, en les adaptant au cadre théorique des grammaires de construction, d'où le néologisme *collostruction*.

L'analyse des collexèmes distinctifs (*distinctive collexeme analysis*, cf. Gries & Stefanowitsch 2004) quantifie la force d'attraction statistique entre un lexème donné et une construction grammaticale par opposition à une autre construction. Dans les collostructions, une construction est vue comme un ensemble de positions (« slots ») remplies par des éléments lexicaux. L'objectif est de déterminer s'il existe une tendance statistique significative du lexème L à apparaître dans le slot S d'une construction C1 plutôt que d'une autre construction C2. Cette technique est utile pour comparer deux constructions fonctionnellement proches et est particulièrement adaptée à l'étude des alternances syntaxiques. Elle donne une indication de la préférence de certains verbes pour une construction plutôt qu'une autre (les deux options étant également possibles) et peut donc révéler des différences distributionnelles entre deux constructions quasi-équivalentes traduisant des différences sémantiques subtiles. Par exemple, Gries & Stefanowitsch (2004 : 104) utilisent cette technique pour étudier l'alternance dative en anglais, en déterminant les verbes préférant la construction ditransitive à double objet à la construction dative en *to* (et inversement), confirmant et étendant ainsi les résultats d'études non basées sur corpus. Nous proposons d'appliquer l'analyse des collexèmes distinctifs à l'étude de l'augmentation de valence.

Nous suggérons, pour tous les verbes apparaissant dans une construction argumentale donnée, de calculer la préférence du verbe pour la syntaxe simple (SS) plutôt que pour la

syntaxe étendue (SE) en utilisant l'analyse des collexèmes distinctifs. Pour chaque verbe, nous dressons une table de contingence croisant les fréquences d'occurrence du verbe dans chaque construction (voir Figure 1).

	Syntaxe Simple (SS)	Syntaxe Étendue (SE)	
Verbe V	F(V et SS)	F(V et SE)	F(V)
¬ V	F(¬V et SS)	F(¬V et SE)	F(¬V)
	F(SS)	F(SE)	

Figure 1 : Table de contingence pour la mesure de l'augmentation de valence

L'hypothèse nulle est que V n'a aucune préférence particulière, c-à-d. que V devrait être équitablement distribué sur SS et SE, en tenant compte de la fréquence totale de chaque construction. L'hypothèse nulle correspond donc à une fréquence attendue $F_{att}(V \text{ et SS}) = F(V) * F(SS) / (F(SS) + F(SE))$. Si la fréquence F(SS) observée dans le corpus excède la fréquence attendue, cela peut être considéré comme une preuve statistique d'une préférence de V pour SS ; à l'inverse, si elle y est inférieure, cela indique une préférence pour SE. La taille totale de l'échantillon et la fréquence de chaque phénomène affecte la force de l'association : plus l'échantillon est grand, plus réduite est la probabilité que la différence observée soit un artefact de l'échantillon. Afin de quantifier cette attraction/répulsion, on utilise un coefficient de corrélation qui mesure la probabilité que l'hypothèse nulle soit vraie compte tenu de la taille de l'échantillon³. Une valeur supérieure à 1,30103 (valeur seuil) traduit une préférence significative pour la syntaxe simple à un seuil de 5%, ce qui signifie qu'il y a moins de 5% de chances que l'hypothèse nulle soit vraie, c'est-à-dire que la disproportion d'occurrences de ce verbe dans la syntaxe simple (vs. la syntaxe étendue) soit due au hasard. Inversement, une valeur inférieure à -1,30103 traduit une répulsion significative pour la syntaxe simple, c'est-à-dire une préférence significative pour la syntaxe étendue, à un seuil de 5%. Si le résultat du test est compris entre ces deux valeurs, cela signifie qu'il y a plus de 5% de chances que l'hypothèse nulle soit vraie, autrement dit on ne peut pas déterminer de manière concluante s'il y a attraction ou répulsion significative (même si les fréquences observées divergent des fréquences attendues par l'hypothèse nulle).

Pour résumer, un résultat supérieur à 1,30103 traduit une préférence pour la syntaxe simple, tandis qu'une valeur inférieure à -1,30103 traduit une préférence pour la syntaxe étendue. Dans les deux cas, cela implique que la distribution du verbe dans le corpus n'est pas

³ Stefanowitsch & Gries (2003) préconisent le test de Fisher pour sa fiabilité sur les basses fréquences et appliquent une transformation supplémentaire que nous utilisons dans cette étude. Le résultat du test de Fisher renvoie une puissance négative de 10 (du type 10^{-x}). Plus le résultat du test est proche de 0, plus l'attraction est significative. Pour rendre le résultat plus lisible, Gries et Stefanowitsch proposent d'appliquer le logarithme à base 10, c-à-d. $\log(10^{-x})$, soit tout simplement x , et présentent ensuite une valeur positive en cas d'attraction (c'est-à-dire si la fréquence observée est supérieure à la fréquence attendue dans l'hypothèse nulle) et négative en cas de répulsion (c'est-à-dire si la fréquence observée est inférieure à la fréquence attendue dans l'hypothèse nulle). Par exemple, dans le cas d'un test donnant comme résultat 10^{-3} , $\log(10^{-3}) = -3$, on présentera donc la valeur 3 en cas d'attraction et -3 en cas de répulsion.

due au hasard et qu'il existe donc une raison expliquant ces préférences statistiques, ce qui est assimilé dans les collostructions à un fait grammatical.

4 Une étude quantitative de l'augmentation de valence

Dans la suite de l'article, nous présentons les résultats d'une étude dans laquelle nous testons la méthode proposée dans la section précédente sur deux constructions décrites dans la littérature, afin d'évaluer si elle peut être utilisée en tant que critère quantitatif indiquant si un verbe donné subit une augmentation de valence. Si tel est le cas, cette technique sera capable de mettre en évidence qu'une structure syntaxique donnée est une construction argumentale potentielle, en identifiant, par une méthode quantitative, des verbes utilisés de manière inhabituelle dans cette structure syntaxique.

4.1 Valence et usage

Il convient de souligner que l'augmentation de valence par adjonction d'un argument constructionnel n'est pas le seul cas où deux valences syntaxiques d'un même verbe peuvent être décrites comme une syntaxe simple et une syntaxe étendue. Le phénomène inverse de *réduction de valence* permet l'omission de certains arguments du verbe lorsque ceux-ci sont soit inconnus ou non pertinents (« objets nuls indéfinis », cf. Fillmore 1986), soit récupérable dans le contexte d'énonciation (« objets nuls définis », *ibid.*). La valence syntaxique ne contient alors qu'un sous-ensemble de la valence sémantique du verbe. Par exemple, le verbe *teach* comporte dans sa valence sémantique trois arguments, *teacher* 'professeur', *teachee* 'apprenant' et *topic* 'sujet' respectivement fusionnés aux arguments *agent*, *destinataire* et *thème* de la construction ditransitive (p.e. *She taught him French* 'Elle lui a enseigné le français'). Mais ce verbe est aussi couramment utilisé sans argument *teachee*, notamment si celui-ci est implicite et/ou s'il n'est pas pertinent de le mentionner explicitement dans la situation d'énonciation, p.e. *She taught French for twenty years* 'Elle a enseigné le français pendant vingt ans'.

La distinction entre augmentation et réduction de valence est parfois ténue, ce qui découle de la difficulté à identifier avec précision la valence sémantique du verbe. La différence majeure est de nature sémantique et concerne la scène générale dans laquelle le verbe est utilisé. Dans un cas de réduction de valence, l'utilisation du verbe dans la syntaxe réduite correspond au même type de scène que lorsque le verbe est utilisé dans la syntaxe étendue contenant toute sa valence sémantique, même si l'un des arguments est mis en retrait dans le premier cas. L'utilisation de *teach* ci-dessus dénote toujours un événement de transmission de savoir, que le verbe soit utilisé avec ou sans argument *teachee*, et ce dernier est toujours présupposé même s'il n'est pas explicitement mentionné dans la syntaxe réduite. En revanche, dans les cas d'augmentation de valence, le verbe est intégré à une nouvelle scène accueillant à la fois l'argument constructionnel et la scène dénotée par le verbe dans sa syntaxe simple mais néanmoins distincte de cette dernière. La différence est que l'argument manquant est latent en syntaxe simple dans le cas de la réduction de valence, tandis qu'il est normalement inexistant

dans le cas de l'augmentation de valence : il ne peut exister que si nous faisons l'effort de replacer l'action dans un contexte étendu, mais n'est en aucun cas systématiquement requis dans toutes les utilisations du verbe. Ceci constitue un critère pour décider si un verbe subit une augmentation ou une réduction de valence.

D'un point de vue valentiel, les verbes apparaissant dans une construction argumentale susceptible d'augmentation de valence se répartissent donc en deux types :

Type 1 : verbes dont la valence sémantique coïncide avec la valence syntaxique de la construction argumentale (c-à-d. de la syntaxe étendue). Ces verbes sont en principe capables d'alterner avec la syntaxe simple en subissant une réduction de valence;

Type 2 : verbes dont la valence sémantique ne contient qu'une partie des arguments de la construction, celle-ci apportant alors elle-même un argument supplémentaire à la valence syntaxique. Ces verbes peuvent apparaître en syntaxe étendue en subissant une augmentation de valence.

Dans cette étude, nous voulons vérifier si ces différences valentielles sont corrélées à des différences statistiques systématiques. Nous supposons que les préférences syntaxiques des verbes sont fortement contraintes par leur sémantique et que même si certains arguments peuvent être régulièrement omis (réduction de valence), ces verbes devraient présenter un alignement de leur valence syntaxique sur leur valence sémantique plus souvent que la moyenne comparativement aux autres verbes. En d'autres termes, nous nous attendons à ce que les verbes de type 1 préfèrent la syntaxe simple et que les verbes de type 2 la syntaxe étendue, malgré la possibilité pour ces derniers d'apparaître dans la syntaxe simple par réduction de valence.

Soulignons que notre méthode ne s'intéresse pas uniquement aux fréquences brutes mais prend aussi en compte la fréquence de chaque construction. Par exemple, la construction transitive étant plus fréquente que la construction ditransitive, il est prévisible d'un point de vue purement statistique que les fréquences d'utilisation des verbes soient biaisées vers la construction la plus fréquente. C'est l'éventuelle divergence par rapport à cette distribution « normale » qui est intéressante et révélatrice car elle est le signe d'une affinité particulière entre verbe et construction. Si l'analyse collocationnelle nous montre que cette divergence n'est pas due au hasard, elle est alors probablement causée par une contrainte grammaticale : en l'occurrence nous suggérons que la valence sémantique du verbe peut exercer une telle contrainte.

4.2 Constructions étudiées

Pour tester l'applicabilité de notre technique à la mesure de l'augmentation de valence, nous avons choisi deux constructions argumentales abondamment détaillées dans la littérature et capables d'apporter au moins un argument ne faisant pas partie de la valence sémantique du verbe, la construction ditransitive et la construction à mouvement causé.

La construction ditransitive (cf. Goldberg 1995, chapitre 6), présentée précédemment, désigne un événement de changement (causé) de possession, au cours duquel un *agent/cause* (sujet) provoque qu'un *destinataire* (premier objet direct) devient le possesseur d'un *thème* (second objet direct). La greffe de cette construction à des verbes normalement transitifs (notamment des verbes de création) apporte à la proposition un argument objet indirect (ou premier objet direct) qui désigne généralement le destinataire (ou bénéficiaire) du référent de l'objet direct de ces verbes, comme nous l'avons déjà exposé dans la section 2. Quelques verbes, comme *earn* et *win*, présentent un schéma d'alternance différent, il s'agit pourtant bien d'instances de cette construction. L'utilisation transitive de ces verbes implique déjà un argument présentant les caractéristiques d'un destinataire, réalisé par le sujet (p.e. *He earned an award*) ; cet argument est également agent responsable du gain (ou du moins en partie). La construction ditransitive réalise cet argument par le premier objet direct, et ajoute un argument sujet agent qui est le plus souvent une cause abstraite et non un agent animé, car la dissociation de l'agent responsable du gain et de son bénéficiaire va à l'encontre de la sémantique du verbe, dans laquelle ces deux arguments sont en général confondus (p.e. dans *earn one's living*⁴ 'gagner sa vie'). Le seul véritable cas d'agent humain attesté dans le corpus est tiré d'un commentaire sportif (ex. 4) : les règles du football créent en effet un contexte particulier dans lequel les actions d'un joueur (*agent*) ont pour effet de faire bénéficier l'équipe adverse (*destinataire*) d'une action supplémentaire (*thème*).

(4) Back with **Gary Pallister who just let the ball run across the touch-line to earn Manchester United a throw-in** about ten yards inside their own half on the near touch-line [s2a-003 84:1:A]

La construction à mouvement causé (*caused-motion construction*, cf. Goldberg 1995, chapitre 7) désigne un événement de changement (causé) de lieu, au cours duquel un *agent* provoque le mouvement d'un *patient* selon une *trajectoire*, c'est-à-dire toute expression d'une destination (*into the house, on the ground*), d'une origine (*from home, off the table*), d'un chemin (*through the door, between the trees*), ou de toute combinaison de ces trois types d'éléments. Cette construction fusionne avec les verbes transitifs de deux manières : (i) elle peut apporter un argument oblique de type locatif pour encoder la trajectoire du changement de lieu de l'argument objet direct du verbe transitif (p.e. *Bill kicked the ball* 'Bill frappe la balle' vs. *Bill kicked the ball to Bo* 'Bill passe la balle à Bo'), ou (ii) si l'objet direct du verbe transitif peut être vu comme un argument *destination* (ou *destinataire*), elle peut réaliser cet argument comme oblique, et l'argument apporté par la construction est alors, d'un point de vue sémantique, le *patient/thème* (p.e. *Mary fed the pigs* 'Marie a nourri les cochons' vs. *Mary fed potato peelings to the pigs* 'Marie a donné des épluchures de pommes de terre à manger aux cochons').

⁴ Je remercie Gilles Col pour avoir porté cet exemple à mon attention.

Ces deux constructions peuvent se greffer de deux manières différentes sur les verbes transitifs, ce qui ne nous empêche pas de voir l'augmentation de valence comme une alternance entre une syntaxe simple et une syntaxe étendue dans les deux cas. En effet, quel que soit le schéma de réalisation concerné, l'argument apporté par la construction est, d'un point de vue syntaxique, un deuxième objet direct pour la construction ditransitive et un oblique locatif pour la construction à mouvement causé.

En utilisant la méthode décrite dans la section 3, pour tous les verbes apparaissant dans ces deux constructions, nous mesurons dans le corpus ICE-GB⁵ la préférence pour la syntaxe simple par opposition à la syntaxe étendue. Bien que ICE-GB ne soit pas un très grand corpus (1 million de mots), son principal avantage est qu'il dispose d'annotations fournissant la structure syntaxique complète de chaque phrase et les types de constructions utilisées, ce qui nous permet d'identifier avec une grande précision et un traitement manuel minime les occurrences des trois structures syntaxiques concernées, à savoir :

- La construction ditransitive, étiquetée *ditr* dans le corpus ; 1088 instances.
- La construction à mouvement causé, c'est-à-dire une structure transitive suivie d'un oblique locatif (Sujet-Verbe-ObjetDirect-Oblique_{Locatif}). Nous nous limitons dans cette étude aux obliques prépositionnels, qui constituent la majorité des cas. Pour éviter de recueillir des cas de compléments circonstanciels locatifs, nous sélectionnons (i) les propositions étiquetées *montr* (*monotransitive*) et contenant un syntagme prépositionnel portant l'annotation *phras* (introduisant un argument du verbe⁶), et (ii) les propositions étiquetées *ctr* (*complex transitive*) avec un attribut de l'objet de type prépositionnel⁷. Enfin nous ne gardons que les résultats contenant une préposition locative⁸. Après un tri manuel pour retirer les occurrences non-locatives, nous obtenons 1728 instances.
- La construction transitive (Sujet-Verbe-ObjetDirect). Nous sélectionnons simplement les propositions étiquetées *monotransitive* en retirant les cas de compléments obliques contenant une préposition étiquetée *phras* ; 34452 instances.

⁵ *International Corpus of English – Great Britain*; ICE-GB est un corpus général d'anglais britannique de 1 million de mots, couvrant à la fois langue parlée (60%) et écrite dans de nombreux registres.

⁶ Dans la documentation du corpus, cet attribut est défini de manière si vague que rien ne nous garantit *a priori* que tous les cas pertinents seront identifiés, surtout dans le cas de verbes n'apparaissant pas normalement avec un syntagme prépositionnel, c'est-à-dire précisément les cas qui nous intéressent dans cet article. Toutefois, comme nous le verrons dans la suite de cette section, les résultats sont satisfaisants à cet égard car ils contiennent bel et bien des cas d'augmentation de valence. Nous avons donc foi en cette stratégie d'extraction, qui nous a de plus épargné une longue phase d'analyse manuelle tout en garantissant une précision supposément maximale et un rappel plus que satisfaisant.

⁷ Beaucoup de syntagmes prépositionnels dénotant la destination d'un changement de lieu sont en effet annotés 'attribut de l'objet' dans le corpus.

⁸ Par ordre de fréquence décroissante dans le corpus : *to, on, into, from, in, through, over, out of, upon, towards, off, across, against, around, between, onto, under, round, down, on to, away from, up to, up, within, past, behind, inside, above, along, beyond, down to, out, in to, below, near, outside*.

Dans les trois cas, nous ne sélectionnons que les compléments d'objet direct de type nominal. Nous avons ensuite identifié manuellement la distribution verbale de chaque construction. Pour chaque verbe dans chaque distribution, nous dressons la table de contingence présentée en Figure 1 et calculons le degré d'alternance avec la méthode présentée précédemment.

4.3 Résultats

Le Tableau 1 montre les verbes attirés par la syntaxe simple pour les deux constructions. Nous pouvons remarquer que relativement peu de verbes atteignent le seuil significatif. Pour la construction ditransitive, *read*, *find* et *play* illustrent des cas d'augmentation de valence : la construction apporte clairement un rôle *destinataire* à un verbe à deux arguments. *get* est un verbe extrêmement polysémique, mais un examen approfondi des exemples révèle qu'il s'agit dans la plupart des cas du sens d'obtention, ce qui correspond aussi à une augmentation de valence (p.e. *let me get a drink* vs. *let me get you a drink*). La pauvreté sémantique de *make* et *do* et leur utilisation fréquente en tant que verbes support en font des exemples moins clairs. Pour la construction à mouvement causé, *lose* et *read* sont les seuls cas évidents. L'utilisation de *want* dans la construction se rapproche de l'une des extensions polysémiques de la construction décrites par Goldberg (1995 : 76), dans laquelle les conditions de satisfaction du verbe impliquent le changement de lieu (comme p.e. avec *promise*) : si la volonté du sujet est accomplie, alors seulement le patient subit le changement de lieu. Enfin, la présence de *provide* va à l'encontre de notre hypothèse : la sémantique de ce verbe suppose toujours un destinataire, mais il préfère pourtant la syntaxe simple. Un examen des occurrences de ce verbe dans la construction transitive révèle que l'argument destinataire absent reçoit souvent une interprétation d'objet nul indéfini, comme dans l'exemple (5) :

- (5) Because of her industrial lead she could provide the technology and services the rest of the world wanted. [w2b-018 76]

Ditransitive (Suj-V-Obj vs. Subj-V-Obj1-Obj2)				Mouvement causé (Suj-V-Obj vs. Subj-V-Obj-Loc)			
Verb	Fq. syntaxe simple	Fq. syntaxe étendue	Attraction	Verb	Fq. syntaxe simple	Fq. syntaxe étendue	Attraction
do	1919	11	14,43	want	303	1	5,27
make	790	2	8,26	provide	247	3	2,7
get	1704	23	5,8	lose	146	1	2,2
find	313	2	2,48	read	198	4	1,48
read	198	1	1,83				
play	208	2	1,37				

Tableau 1 : Verbes préférant la syntaxe simple

En examinant la liste des verbes n'ayant pas atteint le seuil d'attraction ou de répulsion (force d'attraction $> -1,30103$ et $< 1,3103$), nous y trouvons beaucoup de verbes transitifs que nous nous attendions à trouver dans le Tableau 1. Pour la construction ditransitive, il s'agit notamment de verbes de création (*build*, *write*, *draw*, *design*, *cook*) et d'obtention (*win*, *purchase*, *afford*) pour lesquels la notion de transfert (ou d'absence de transfert) vers un

troisième participant est introduite par la construction. Dans le cas de la construction à mouvement causé, on peut dégager entre autres deux grandes catégories : (i) des verbes d'application de force (*strike, twist, cut, brush, hammer, wipe*), pour lesquels le mouvement est une conséquence typique de l'événement désigné par le verbe mais n'est pas présupposé, et (ii) des verbes introduisant un élément modalisateur (*admit, allow, help, order*) impliquant que le changement de lieu est rendu possible voire facilité par l'agent (cf. exemples 6 et 7 ci-dessous).

(6)[...] treating such people as ill and **admitting** them to hospital for extended periods of time, is counter-productive and harmful in the long run [...] [w1a-007 93:1]

(7)Consistency and clarity were the essential features of the president's Gulf policy when he **ordered** 250,000 men into Saudi Arabia two months ago. [w2e-010 20:1]

Tous ces verbes sont plus fréquents (en terme de fréquence brute) dans la syntaxe simple que dans la syntaxe étendue ; en fait beaucoup d'entre eux n'apparaissent qu'une seule fois dans la syntaxe étendue alors qu'ils sont comparativement courants dans la syntaxe simple, ce qui nous apparaît intuitivement comme un signe que les exemples dans la syntaxe étendue relèvent du phénomène d'augmentation de valence. Toutefois l'analyse collostructionnelle ne délivre pas un résultat permettant de conclure que le verbe préfère la syntaxe simple, ce qui peut signifier (i) soit que la distribution de ce verbe ne diverge que faiblement ou pas du tout de la distribution des constructions (c'est le cas de *write* et *leave* dans la construction ditransitive) et il y a donc une forte chance qu'elle soit due au hasard, (ii) soit que la distribution diverge de la distribution normale en terme de ratio (c'est le cas de *permit* et *purchase* dans la construction ditransitive), mais le verbe n'étant pas suffisamment fréquent en général, le corpus ne délivre pas suffisamment d'occurrences pour tirer une conclusion fiable sur ses préférences statistiques. Notons qu'un problème similaire peut être constaté pour des verbes trivalents qui devraient être attirés par la syntaxe étendue de par leur sémantique mais qui n'atteignent pas le seuil significatif : *bring, deliver, pass, serve* et *supply* pour la construction ditransitive, et *award, deliver, distribute, forward, incorporate, offer, pass, pay, return* et *sell* pour la construction à mouvement causé.

Parmi les verbes préférant la syntaxe étendue, nous trouvons en majorité des verbes que nous nous attendions à trouver dans cette catégorie. Pour la construction ditransitive, il s'agit de prédicats triadiques requérant comme premier objet direct le destinataire d'un objet ou d'un message et dénotant un quelconque transfert littéral (*give, offer, send*) ou métaphorique (*tell, show, ask, teach*). Nous observons un résultat similaire pour la construction à mouvement causé, avec une plus large variété de prédicats triadiques (ce qui est probablement dû au fait que l'argument destination de cette construction est applicable à une plus large variété de verbes que l'argument destinataire de la construction transitive), généralement des verbes de transfert : *bring, hand, fax, give, lend, offer, place, post, put, take* et *throw*. Ces observations ne sont guère étonnantes dans la mesure où une construction argumentale a tendance à attirer des verbes se conformant à sa sémantique.

En revanche, nous trouvons aussi certains verbes dont l'utilisation ditransitive peut être décrite comme relevant de l'augmentation de valence puisque leur utilisation transitive ne présuppose pas de rôle *destinataire* (comme par exemple *buy*⁹ et *earn*) ; contrairement à notre hypothèse, ces verbes préfèrent cependant la construction ditransitive apportant ce rôle. *allow*, *deny*, *guarantee*, *promise*, *refuse* et *wish* sont des cas semblables : dans leur utilisation transitive, l'objet direct est le plus souvent un fait qui est autorisé, souhaité, garanti, promis ou refusé, indépendamment de tout *destinataire* (ou plutôt ici *bénéficiaire*) potentiel ; cet argument est apporté par la construction ditransitive (cf. exemples 8 à 10 ci-dessous).

(8) Well certainly Kevin Campbell uh got the better of Mabbutt on that occasion and that **allowed** Merson the chance to swing that cross in. [s2a-015 175:1:A]

(9) There is a declining group of individuals who signed a waiver clause **denying** them their legal rights of claiming unfair dismissal and of redundancy payments. [w1b-024 30:3]

(10) I **promised** you a funny letter today. [w1b-007 14:1]

Notons que dans la construction ditransitive, l'argument *thème* de ces verbes est un objet concret plus souvent que dans la construction transitive, ce qui produit clairement une interprétation de transfert futur ou hypothétique. Nous trouvons de nombreux autres exemples similaires avec la construction à mouvement causé, notamment des verbes d'application de force (*kick*, *knock*, *pull*, *punch*, *push*, *shove*, *slap*, *stir*), dont l'emploi transitif n'évoque pas nécessairement une trajectoire, qui est introduite par le concept de mouvement causé.

4.4 Discussion

Nous avons pu constater que notre approche se heurte à un écueil majeur dans le cas des deux constructions choisies : la quantité de données disponibles. Il semble en effet que la plupart des verbes concernés par l'augmentation de valence ne soient pas assez fréquents dans le corpus pour fournir une preuve statistique suffisante de leur préférence pour la construction transitive. Une solution serait d'utiliser un corpus plus grand, dans lesquelles ces verbes seraient en théorie proportionnellement plus fréquents. C'est en effet tout ce que signifie l'absence de significativité : les données disponibles étant insuffisantes pour attester une corrélation entre deux phénomènes, il faut augmenter la taille de l'échantillon (cf. Kilgarriff 2005). Cette question pourrait être approfondie par des travaux ultérieurs sur un corpus plus important en se focalisant sur les verbes qui n'ont pas atteint le seuil significatif. Une telle étude devra néanmoins aborder le problème de l'obtention de la fréquence des structures syntaxiques, ce qui peut s'avérer difficile sur un corpus plus grand qui ne comportera probablement pas d'annotations syntaxiques comme ICE-GB.

Si on laisse de côté ces résultats non significatifs et que nous nous tournons vers les préférences effectivement attestées par nos données, nous permettent-elles de constater des différences statistiques entre les verbes de type 1 et 2, tels que nous les avons définis dans la

⁹ Il est également possible pour *buy* d'attribuer le rôle *vendeur* au premier objet direct. Toutefois une vérification manuelle révèle qu'il s'agit vraisemblablement du rôle *destinataire* pour tous les exemples du corpus.

section 4.1 (c-à-d. respectivement les verbes subissant une augmentation de valence et ceux subissant une réduction de valence) ? Il semble en effet que les deux types de verbes tendent à se distinguer, mais pas de manière aussi systématique que nous aurions pu le penser. Quelques verbes de type 1 sont significativement attirés par la syntaxe simple et de nombreux verbes de type 2 sont attirés par la syntaxe étendue. Deux faits posent cependant problème :

1. La préférence de *provide* pour la syntaxe simple;
2. La préférence de verbes de type 1 (selon nos critères) pour la syntaxe étendue : par exemple *allow, buy, deny, earn, guarantee, promise, refuse* et *wish* pour la construction ditransitive, et *kick, knock, pull, punch, push, shove, slap, stir* pour la construction à mouvement causé.

Sans être nécessairement incongrus (par exemple le déplacement d'un objet est un résultat attendu lorsque l'on y exerce une force, cf. *push* et *kick*), ces résultats indiquent que notre hypothèse concernant l'alignement des préférences syntaxiques des verbes sur leur valence sémantique doit être rejetée.

5 Conclusion

En conclusion, malgré un mince succès, cette technique ne nous a pas permis d'identifier l'augmentation de valence dans le corpus avec précision, car contrairement à ce que nous avions supposé, il n'a pas de différence systématique d'usage entre les verbes concernés par l'augmentation de valence et les autres. Malgré une certaine tendance à l'alignement de la valence syntaxique sur la valence sémantique (par exemple transitivité pour les verbes de création et ditransitivité pour les verbes de transfert), de nombreuses idiosyncrasies subsistent, chaque verbe présentant son propre profil syntaxique, présentant parfois des préférences non prévues par leur valence sémantique. Notamment nous avons pu noter que si beaucoup de verbes sont indubitablement triadiques d'un point de vue conceptuel, ils le sont à des degrés divers dans l'usage de la langue, comme le montre l'absence d'une claire préférence de certains pour une valence syntaxique triadique. De la même façon, certaines catégories de verbes affichent une préférence marquée pour une syntaxe triadique, sans être clairement triadiques d'un point de vue conceptuel, au sens où un troisième argument n'est jamais présupposé dans les utilisations diadiques de ces verbes. Une conception de la valence sémantique comme l'ensemble des arguments présents dans tous les usages du verbe n'anticipe pas ces préférences syntaxiques.

Ces résultats prennent un sens particulier dans un modèle de la grammaire basé sur l'usage (*usage-based approach*; cf. Langacker 1987). Partant de la découverte que l'usage, la variation et le changement linguistiques sont intimement liés (cf. Bybee 2006), la linguistique de l'usage rejette l'autonomie de la grammaire (compétence) par rapport à l'usage (performance) et défend l'idée que la grammaire ne préexiste pas à l'usage mais en est le produit, dans le sens où chaque « événement d'usage » a un impact potentiel sur le système linguistique des locuteurs. Cette étude montre clairement que le critère statistique ne fonctionne pas toujours pour déterminer la valence sémantique, qui semble donc être une

notion indépendante de l'usage, ce qui met en question son statut dans un tel modèle. Cette conclusion rejoint celle de Thompson & Hopper (2001 : 50, notre traduction), qui rejettent les notions de valence et de structure argumentale et suggèrent de les remplacer par « une théorie empirique basée sur l'usage, considérant l'ensemble des contextes propositionnels dans lesquels les verbes peuvent être utilisés, sans tenter d'imaginer des scènes et des participants ». Nous plaidons en faveur d'une théorie de la réalisation des verbes véritablement basée sur l'usage et fondée sur l'analyse de données de corpus et nous espérons, avec cet article, avoir contribué, même modestement, à l'esquisse d'une telle entreprise.

Références bibliographiques

- BYBEE, Joan, 2006, « From usage to grammar: The mind's response to repetition », *Language*, 82.4, 711–733.
- FILLMORE, Charles, 1986, « Pragmatically Controlled Zero Anaphora », *BLS*, 12, 95–107.
- FRANÇOIS, Jacques & SENECHAL, Morgane, 2008, « Le sémantisme propre des cadres prédicatifs et la polysémie des verbes de production de parole », in J.-M. MERLE (éd.), *La prédication, Faits de Langue 31-32*, 325–335.
- GOLDBERG, Adele, 1995, *Constructions: a construction grammar approach to argument structure*, Chicago, University of Chicago Press.
- GRIES, Stefan & STEFANOWITSCH, Anatol, 2004, « Extending collocation analysis. A corpus-based perspective on 'alternations' », *International Journal of Corpus Linguistics*, 9.1, 97–129.
- LANGACKER, Ronald, 1987, *Foundations of Cognitive Grammar. Vol. I.*, Stanford, Stanford University Press.
- KILGARRIFF, Adam, 2005, « Language is never, ever, ever, random », *Corpus Linguistics and Linguistic Theory*, 1.2, 263–276.
- PAYNE, Thomas, 1997, *Describing Morphosyntax*, Cambridge, Cambridge University Press.
- PEREK, Florent, 2008, *Towards a constructional approach to automatic argument structure acquisition: the case of oblique phrases*, Mémoire de Master, Université Lille III.
- STEFANOWITSCH, Anatol & GRIES, Stefan, 2003, « Collocations: Investigating the interaction of words and constructions », *International Journal of Corpus Linguistics*, 8.2, 209–243.
- THOMPSON, Sandra & HOPPER, Paul, 2001, « Transitivity, clause structure, and argument structure: Evidence from conversation », in J. BYBEE and P. HOPPER (éds.), 2001, *Frequency and the emergence of linguistic structure*, Amsterdam, John Benjamins, 27–60.