

Notes aux validateurs des cadres de sous-catégorisation obtenus à partir de corpus

PAULA CHESLEY

*Linguistics Department, University at Buffalo, USA & ATILF-CNRS, France
pchesley@buffalo.edu*

Abstract

Ce document fournit des critères d'évaluation aux validateurs de notre travail sur l'acquisition de cadres de sous-catégorisation à partir de corpus.

1 Indications générales

Ce travail porte sur les cadres de sous-catégorisation dits “canoniques”, c'est-à-dire avant des transformations telles que la passivation. Il ne faut donc pas chercher à effectuer des telles transformations pour faire en sorte que le cadre supposé soit un cadre effectif.

1. On vous fournit les cadres supposés de notre travail ainsi que deux jugements à faire. Pour simplifier l'évaluation, on vous demande de répondre de façon catégorique : le cadre supposé est-il oui ou non un cadre effectif ? Si c'est un cadre, vous devez supprimer le “n”, sinon, vous devez supprimer le “o” après chaque cadre.
2. Voici la légende des cadres :
 - *se*: verbe pronominal;
 - *adj_att*: adjectif attributif;
 - *cod*: complément d'objet direct;
 - *icl*: un verb à montée ou à contrôlé;
 - *marqueur + fcl/icl*: une proposition subordonnée (fcl) ou infinie (icl) introduite par le marqueur *marqueur*;
 - *PP[p]*: un groupe prépositionnel introduit par la préposition *p*.
3. Les compléments d'objets indirects sont inclus dans le groupe prépositionnel *PP[à]*. Si vous en voyez dans une phrase exemplaire et que le cadre supposé comprend ce groupe prépositionnel (et que le reste du cadre vous semble bon), il s'agit d'un cadre effectif.
4. Vous aurez une phrase exemplaire qui provient de notre corpus pour vous aider à prendre une décision sur le statut d'un cadre supposé. Toutefois cela n'indique pas qu'il faille valider le cadre uniquement dans le contexte de cette phrase. Un manque de phrase exemplaire pour un cadre supposé indique qu'une phrase avec ce cadre n'a pas été trouvée dans notre corpus.
5. L'ordre linéaire des cadres par rapport aux exemples n'est pas à prendre en compte. Par exemple, si le cadre est “dire que + fcl à qqn” et que la phrase exemplaire se trouve être “dire à qqn que”, il s'agit bien du même cadre.

6. Si le cadre proposé vous semble incomplet, c'est-à-dire qu'il ne peut pas être un cadre sans la présence d'une autre position syntaxique ne figurant pas dans le cadre (e.g., un cadre avec un PP mais sans complément d'objet direct qui vous semble obligatoire étant donné le PP), ce cadre doit être considéré comme mauvais. De même, si le cadre proposé semble moins détaillé que le cadre effectif (par exemple, nous proposons un cadre sans sujet impersonnel alors que le cadre doit prendre un sujet impersonnel), cela ne constitue pas un bon cadre non plus.
7. Si le verbe proposé dans le cadre et le verbe dans la phrase exemplaire ne viennent pas du même lemme, il faut valider le cadre par rapport au verbe dans le cadre.
8. Nous vous conseillons de faire l'évaluation dans le plus petit nombre de séances possible, l'effectuer en une ou deux séances serait l'idéal. Une seule séance risque de vous fatiguer.
9. Merci de ne pas passer plus d'une minute sur un cadre proposé et de me dire combien de temps vous avez mis au total pour faire ce travail.

2 Critères de sous-catégorisation

Cette partie vise à donner aux validateurs des critères aussi stricts que possible de la sous-catégorisation, sachant qu'il n'y a pas de consensus sur ce point parmi les linguistes.

2.1 Critères sémantiques

Plusieurs études ont révélé que les critères syntaxiques ne fournissent pas de réponses définitives pour ce qui constitue un cadre de sous-catégorisation (Koenig et al. (2003), Schütze (1995), Boons et al. (1976)). C'est pourquoi nous choisissons de fonder nos critères de sous-catégorisation sur la sémantique.

Koenig et al. font remarquer aussi que “. . . most events occur at a certain location, at a certain time, and for the benefit of somebody. Constituents which encode event locations, times, or beneficiaries can co-occur with most event-denoting verbs” (p. 73). En fait, certains des modificateurs sémantiques (*adjuncts*) sont nécessaires aussi pour la réalisation d'un événement car ce dernier doit se passer à un lieu précis et à un moment précis. Le problème devient celui de trancher entre les arguments obligatoires et les modificateurs obligatoires. Pour mieux cerner cette distinction, Koenig et al. font l'observation suivante :

Two differences between (semantically obligatory) arguments and (semantically obligatory) adjuncts emerge from these observations: (i) adjunct participant roles are common to most verbs, (ii) argument participant roles are lexically required to bear additional properties aside from those which are characteristic of the role. . . argument roles are specific to both an individual event-type and a restrictive class of event-types to which to meaning of the verb belongs (ibid).

Étant donné ces observations, nous adoptons le critère suivant pour un *argument sémantique*, qui est le *Semantic Specificity Criterion* (SSC) de Koenig et al. (2003):

- (1) If r is an argument participant role of predicate P denoted by verb V , then r is specific to V and a restricted class of verbs.

Il ne reste plus qu'à faire le lien entre la sémantique et la syntaxe. Nous donnons donc la définition suivante d'un *argument syntaxique*:

- (2) Si r constitue, selon le SSC, un argument sémantique, du prédicat P dont la dénotation est V , alors la réalisation syntaxique s de r est un argument syntaxique de V .

En bref, pour cette évaluation, un argument syntaxique est la réalisation en syntaxe d'un argument sémantique, ou alors du manque d'arguments sémantiques. Ainsi établissons-nous une définition d'un *cadre de sous-catégorisation* :

- (3) Un cadre de sous-catégorisation du verbe V est un ensemble d'arguments syntaxiques de V qui constitue une proposition grammaticale.

Veillez utiliser l'ensemble de définitions données en (1, 2 et 3) comme des indications lors de valider les cadres qui vous sont proposés.

2.2 Critères syntaxiques

Si les définitions de la section précédente fournissent un premier pas vers des critères fiables de cadres, elles ne recouvrent pas tous les cas épineux que l'on voit dans la syntaxe. Parmi ces derniers nous relevons les verbes pronominaux, les groupes prépositionnels introduits avec une préposition particulière p et les arguments facultatifs. Dans les ressources lexicales il n'y a souvent pas un traitement homogène pour ces deux cas. Une nouvelle entrée d'un verbe pronominal peut être créée uniquement lorsqu'il y a un changement de sens assez important entre les formes pronominales et non-pronominales (telle semble être la méthode des tables du LADL), ou lorsque l'usage pronominal est assez courant dans la langue (c'est la méthode des dictionnaires classiques et bilingues). Nous vous demandons d'utiliser ce dernier critère quand il s'agit d'évaluer un cadre de sous-catégorisation qui comprend une réalisation pronominale.

Les groupes prépositionnels peuvent aussi poser des problèmes aux définitions (1, 2 et 3). Or, souvent un verbe sélectionne un groupe prépositionnel introduit par une préposition précise, et non pas toutes les prépositions qui peuvent réaliser un argument sémantique. Par exemple un argument sémantique d'objectif locatif (une sous-classe de la classe d'arguments locatifs) peut sélectionner parmi *à*, *vers*, *jusqu'à* et *chez*. Cependant le verbe *se diriger* ne semble pas sélectionner *à* (? "Le bateau se dirige au port."). C'est un phénomène que la sémantique ne semble pas pouvoir prendre en compte. Comme nos cadres sont acquis à partir de corpus, de telles occurrences risquent d'être rares. De plus, il existe des cadres de sous-catégorisation archaïques qui ont une réalisation avec un groupe prépositionnel non standard. Ce phénomène arrive surtout avec les prépositions *de* et *à*.

Quant aux arguments facultatifs, Boons et al. (1976) indique que "seulement 2% environ des verbes ont des compléments obligatoires" (p. 194). Cela veut dire que vous verrez probablement des cadres supposés avec des arguments de moins qu'un cadre canonique.

Dans les trois cas problématiques détaillés plus haut, ainsi que pour d'autres mais surtout pour les cadres archaïques, nous vous invitons à chercher le cadre dans les *constructions* du *Trésor de la Langue française informatisé* (TLFi). Ces dernières sont visualisables dans la version en ligne qui se trouve à l'URL suivant :

<http://atilf.atilf.fr/tlf.htm> . Il suffit de choisir *Construction* du menu déroulant rose sous la rubrique *Peindre les objets suivants*. Si vous n'êtes pas d'accord avec le jugement du TLFi, nous vous conseillons de décider vous-mêmes le statut d'un cadre proposé.

References

- [Boons et al.1976] J.-P. Boons, A. Guillet, and C. Leclère. 1976. *La Structure des phrases simples en français: constructions intransitives*. Droz, Genève.
- [Koenig et al.2003] J.-P. Koenig, G. Mauner, and B. Bienvenue. 2003. Arguments for Adjuncts. *Cognition*, 89:67–103.
- [Schütze1995] C. Schütze. 1995. PP attachment and argumenthood. pages 95–151. MIT Working Papers in Linguistics, 26.