



ISSN 1724-0700

ISSN en ligne 2260-8087

## *Corpus Pattern Analysis et Classes d'objets : différences théoriques et retombées pratiques de deux approches de description du lexique*

**Adriana Orlandi**

Université de Modène et de Reggio d'Émilie, Italie  
adriana.orlandi@unimore.it

<https://orcid.org/0000-0002-6638-265X>

**Marco Fasciolo**

Sorbonne Université, France  
marco.fasciolo@sorbonne-universite.fr

Reçu le 10-02-2021 / Évalué le 22-03-2021 / Accepté le 07-04-2021

### Résumé

Dans cette contribution, nous souhaitons comparer deux approches (théories, méthodes, techniques) lexicographiques conçues dans le cadre du TAL, à savoir la Corpus Pattern Analysis et la Théorie des Classes d'objets. En particulier, nous analyserons la différence théorique entre les notions de *pattern* et de *schéma prédicatif*, et sur le plan applicatif nous évaluerons la capacité des deux approches à discriminer les différentes acceptions d'un mot polysémique. Pour ce faire, nous prendrons en examen les *patterns* du verbe italien *seguire* extraits d'une base de données se basant sur la CPA, la T-Pas, et nous les comparerons avec les schémas prédicatifs du verbe *suivre* définis selon la Théorie des Classes d'objets.

**Mots-clés :** pattern, schéma prédicatif, Corpus Pattern Analysis, Théorie des classes d'objets, T-Pas

### *Corpus Pattern Analysis e Classes d'objets: differenze teoriche e ricadute pratiche di due approcci alla descrizione del lessico*

### Riassunto

In questo contributo ci proponiamo di confrontare due approcci (teorie, metodi, tecniche) lessicografici concepiti nell'ambito del TAL, ovvero la *Corpus Pattern Analysis* e la *Théorie des Classes d'Objets*. In particolare, analizzeremo la differenza teorica tra le nozioni di *pattern* e di *schéma prédicatif*, mentre, sul piano applicativo, valuteremo la capacità dei due approcci di discriminare le diverse accezioni di una parola polisemica. Per fare ciò, prenderemo in esame i *pattern* del verbo *seguire* estratti da un database realizzato secondo i principi della CPA, la risorsa T-Pas, e li confronteremo con gli schemi predicativi del verbo francese *suivre* definiti secondo la *Théorie des Classes d'Objets*.

**Parole chiave:** *pattern*, schema predicativo, *Corpus Pattern Analysis*, Teoria delle « classi di oggetti », T-Pas

## Corpus Pattern Analysis and *Classes d'objets*: theoretical differences and practical implications of two approaches to lexicon description

### Abstract

The aim of this paper is to compare two lexicographic approaches (theories, methods, techniques) conceived within the NLP framework, namely the *Corpus Pattern Analysis* and the *Théorie des Classes d'Objets*. In particular, we will analyse the theoretical difference between the notions of *pattern* and *schéma prédicatif*, and we will evaluate the capacity of the two approaches to discriminate the different meanings of a polysemic word. In order to do so, we will examine the patterns of the Italian verb *seguire* extracted from a database created according to the principles of CPA, the T-Pas resource, and compare them with the predicative patterns of the French verb *suivre* defined according to the *Théorie des Classes d'Objets*.

**Keywords:** pattern, *schéma prédicatif*, Corpus Pattern Analysis, *Théorie des classes d'objets*, T-Pas

### 1. Introduction : la *Théorie des Classes d'Objets* et la *Corpus Pattern Analysis*

La Théorie des Classes d'Objets (dorénavant *TCO*) est une approche lexicographique, française, conçue au sein du laboratoire *LLI*, ensuite *LDI*, par M. Mathieu-Colas et G. Gross, comme un développement de l'idée de lexique-grammaire de M. Gross. Une première synthèse de cette approche est offerte par Le Pesant et Mathieu-Colas (1998), mais l'idée est précédente : cf. par exemple, Gross (1992, 1994, 1995). Pour une présentation récente, nous mentionnerons Gross (2012), Fasciolo et Gross (2020). Dans cette contribution, nous nous proposons de comparer la *TCO* avec la *Corpus Pattern Analysis* (dorénavant *CPA*) à la base de la ressource T-Pas (Jezek *et al.*, 2014). Pour une présentation de la *CPA*, nous renvoyons à Hanks (2004, 2013), Chinkova et Hanks (2010).

L'idée centrale de la *TCO* est que la signification d'un mot doit être étudiée dans le cadre de la *phrase*. Or, il y a deux acceptions pertinentes de *phrase*. La première acception - mise en avant par la *TCO* - conçoit la phrase comme un pivot prédicatif (typiquement un verbe) saturé par ses arguments. La seconde acception - qui reste en arrière-plan dans la *TCO*, mais qui demeure cruciale - conçoit la phrase comme une architecture formelle GN □ GV au sens de l'analyse en constituants immédiats. Dans le cadre de la *TCO*, la meilleure façon pour étudier la signification d'un mot consiste à déployer une structure prédicative relevant de la première acception de phrase à l'intérieur de la structure formelle relevant de la seconde acception de phrase<sup>1</sup>. Ce faisant, on obtient des *schémas prédicatifs* comme les suivants :

<humain> *suivre* <voie de communication> *jusqu'à* <lieu>

<événement> *suivre* <événement>

Les étiquettes <...> identifient des classes d'arguments conceptuellement homogènes, occupant les positions distributionnelles majeures du noyau de la phrase. Un *emploi prédicatif*, dans le cadre de la TCO, est défini comme le paquet de propriétés impliquées par chaque schéma prédicatif. Pour un schéma prédicatif donné, par exemple, un prédicat entre dans certaines relations lexicales (et pas d'autres), affecte certains rôles aux arguments (et pas d'autres), implique une certaine *Aktionsart*, peut être lié à certaines traductions, etc. Chacune de ces propriétés est définie exclusivement par rapport à un schéma prédicatif donné. Encore une fois, pour les détails, nous renvoyons à Gross (2012), Fasciolo (2019b) et à Fasciolo et Gross (2020).

La TCO distingue explicitement deux niveaux de classes d'arguments : les *hyper-classes* et les *classes d'objets stricto sensu*. Les *hyper-classes* sont des catégories *ontologiques*, universelles, qui délimitent des champs de cohérence conceptuelle *a contrario* : à savoir, plus la marge d'incohérence est grande, plus l'hyper-classe est positivement définie et ontologiquement centrale. Voici quelques exemples. Un verbe comme *peser* a une marge d'incohérence très petite car il s'applique à toute <entité matérielle> par opposition aux entités non-matérielles : *J'ai pesé la table, mon chien et mon collègue* vs. *\*J'ai pesé ma joie*. Un verbe comme *tuer* a une marge d'incohérence majeure car il oppose les <vivants> aux non-vivants : *J'ai tué mon collègue et son chien* vs. *\*J'ai tué la table*. Un verbe comme *scandaliser*, finalement, a une marge d'incohérence encore plus grande car il distingue les <humains> des non-humains : *J'ai scandalisé mon collègue* vs. *\*J'ai scandalisé mon chien, \*J'ai scandalisé ma table*. Les *classes d'objets* au sens strict sont des classes *sémantiques* que le lexique circonscrit à l'intérieur des hyper-classes grâce à des faisceaux de *prédicats appropriés* : par exemple, *tailler* pour les <arbres>, *tondre* pour les <herbes>, *siroter* pour les <boissons>, etc. Ces classes d'objets sont extrêmement fines et sont les plus pertinentes pour la description lexicographique. Pour un approfondissement de la distinction entre hyper-classes et classes d'objets, on pourra se reporter à Prandi (1998), Gross (2008), Fasciolo (2018, 2019a, 2019b) et Fasciolo et Gross (2020).

Passons maintenant à la CPA :

*En ce qui concerne les verbes, chaque pattern dans la CPA se base sur la structure des rôles phrastiques de l'anglais tels qu'ils sont décrits dans les grammaires. (...) Chaque rôle phrastique est « peuplé » par un ensemble paradigmatique de collocations (...). Les collocations pertinentes d'un verbe dans un rôle phrastique*

*sont habituellement des noms partageant quelques traits fondamentaux de leur signifié, à savoir un type sémantique. Les types sémantiques sont stockés dans une ontologie superficielle [shallow ontology dans le texte, ndt] organisée de manière hiérarchique. (...) Par exemple, un des patterns du verbe translate, (...) est le suivant<sup>2</sup> :*

*PATTERN: [[Human]] translate ([[Document]]) (from [[Language 1]]) (into [[Language 2]]) (Chinkova, Hanks, 2010: 3)*

Le constat est vite fait : dans les deux cas, les notions de *valence* (ou structure argumentale) et de paradigme d'arguments jouent un rôle central. Les notions de *classes d'objets* et de *types sémantiques* d'une part, et les notions de *schéma prédicatif* et de *pattern* de l'autre semblent en effet très proches (cf. également Hanks, 2013 : 11 et Feltracco, 2016 : 70-71).

Il paraît donc légitime de se demander s'il existe des différences non marginales entre la *TCO* et la *CPA*. À notre avis, la réponse est positive. Ces différences, cependant, ne concernent pas tellement les résultats pratiques des deux approches, mais plutôt leurs présupposés théoriques et leurs implications méthodologiques. Dans la suite, nous contrasterons la *TCO* et la *CPA* sur deux points spécifiques :

- la différence entre les notions de *schéma prédicatif* et de *pattern* (§2)
- l'analyse d'une portion de la polysémie du verbe prédicatif *suivre/seguire* (§3).

## 2. *Pattern vs. Schéma prédicatif*

### 2.1. Qu'est qu'un *pattern* ?

La notion de *pattern* de la *CPA* hérite de celle de Hunston et Francis (2000) : une notion *corpus-driven*. Chinkova et Hanks (2010 : 1-2) insistent sur la déviation à partir de *patterns* normaux. Ces derniers sont induits par le corpus en se basant sur un présupposé partagé au sein des approches cognitivistes : à savoir, plus un phénomène est récurrent, plus il s'imprime dans l'esprit des parlants et plus il acquiert le statut d'objet linguistique saillant. Ce présupposé - explicité par exemple par Perek (2015 : 28-29) - est mobilisé par Hanks (2013 : 5) lorsqu'il distingue entre *saillance cognitive* et *saillance sociale* d'un phénomène. La première relève du caractère inoubliable d'une collocation, issu de son écart par rapport à une alternative normale. La seconde relève, justement, de la banalité des collocations normales : c'est-à-dire, tellement fréquentes qu'elles sont inaperçues par les individus. Un *pattern* (normal) au sens de la *CPA* est donc une collocation socialement saillante, enrichie par des types sémantiques (Hanks, 2013 : 92) : par exemple, *hasarder une hypothèse, une explication, une réponse*, etc<sup>3</sup>. Ce cadre

théorique a des implications non banales, nous semble-t-il, sur la conception de la valence.

Depuis que Tesnière (1959) a mis en avant cette notion, la séparation entre les arguments (ou actants) et les non-arguments a toujours été *la* question cruciale. La notion de valence est épistémologiquement solide si, et seulement si, pour une acception spécifique d'un verbe prédicatif, on peut établir *le* nombre de ses arguments. Dans la mesure où l'existence d'une limite tranchée entre arguments et non-arguments est remise en cause, parler de *valence d'un verbe* (par exemple) perd de pertinence.

Or, sur la base de ces prémisses, considérons un verbe prédicatif que les manuels classeraient comme tétra-valent : *vendre*, par exemple. Supposons, ensuite, que sa manifestation la plus fréquente en corpus - sa *saillance sociale* - soit avec seulement deux arguments. La question que nous posons est la suivante : à la lumière de la définition de *pattern* mentionnée ci-dessus, que faut-il en conclure à propos de la structure argumentale de *vendre* ?

La réponse est claire. Il n'y a pas quelque chose comme *le nombre* d'arguments de *vendre* : ce verbe n'est pas tétra-valent en soi, mais il est prototypiquement bivalent et il admet marginalement l'insertion d'arguments supplémentaires. Cela est ce que nous *montre* le corpus. Cette conclusion est parfaitement cohérente avec les prémisses et est tirée, par exemple, par Mereu (2017, 2020) et Mereu et Piunno (2019). Le problème, cependant, est que, par-là, la notion de valence se dissout car il n'y a plus des arguments et des non-arguments, mais des compléments accompagnant le verbe avec une fréquence majeure ou mineure. Or, selon nous, une telle conséquence est inévitable pour toute approche partageant la notion de *pattern* de la CPA. C'est précisément sur cet aspect que la TCO diverge de la CPA.

## 2.2. Qu'est-ce qu'un schéma prédicatif ?

Bien que la TCO soit centrée sur le pivot prédicatif - la source de la valence - la notion de *schéma prédicatif* hérite de la notion de *phrase* dans son acception distributionnelle et formelle (Harris, 1951). Cela se reflète sur la conception du *contexte*. Selon Gross (2010 : 191) :

*Le contexte n'est jamais une donnée immédiate quand il s'agit de textes, qui ne sont jamais des suites de phrases simples (...) : des prédicats accompagnés de la suite la plus longue de leurs arguments. Et pourtant la phrase simple [= le schéma prédicatif] est l'unité la plus importante des textes.*

La notion de *contexte* ci-dessus mentionnée n'est pas pragmatique : elle ne renvoie pas à une occurrence discursive ou textuelle, mais à la structure abstraite de la phrase - le *schéma prédicatif* - qui explicite *tous* les arguments du verbe. Cette structure - elle - n'est pas un fait du corpus, mais une donnée eidétique fonctionnant comme condition préalable pour interpréter les faits de corpus. En principe, rien n'empêche que les faits du corpus reproduisent parfaitement la structure du schéma prédicatif ; cependant, dans la pratique, cela n'arrive jamais. Typiquement, les faits de corpus se limitent à manifester des fragments d'un schéma prédicatif. C'est la raison pour laquelle Gross (2010 : *cit.*) affirme que les phrases simples sont les unités de base des textes, mais qu'elles ne sont pas, pour autant, *dans* les textes.

Nous retrouvons ici l'une des distinctions nobles de la linguistique : *langue* vs. *parole*. La notion de *schéma prédicatif* de la *TCO* relève du niveau de la *langue*, alors que la notion de *pattern* de la *CPA* relève de la *parole*. Prandi (2019) saisit clairement cette différence en concevant la phrase comme modèle de l'énoncé : en ses termes, la notion de *schéma prédicatif* relève du niveau de la *phrase-modèle*, alors que la notion de *pattern* relève du niveau l'*énoncé*. Ce bouleversement de cadre théorique a un impact sur la conception de la valence contraire à celui sous 2.1.

Considérons la question : *Quelle est la valence du verbe X ?* Dans le cadre de la *TCO*, pour répondre à cette question, il faut procéder en deux étapes. Premièrement, il faut explorer le corpus à la recherche d'attestations exhibant la suite de candidats arguments la plus longue : jusqu'ici, la *TCO* et la *CPA* procèdent en parallèle. Deuxièmement, une fois que cette suite a été identifiée, il faut évaluer le statut d'argument de chaque complément, *a priori* - indépendamment du corpus - avec une batterie de tests basés sur la cohérence conceptuelle. Pour une présentation de ces tests et une discussion de leur fonctionnement, nous renvoyons à Prandi (2004, 2007, 2013). Le point important, maintenant, est que la *TCO* et la *CPA* divergent par rapport à cette seconde étape.

Certes, dans la littérature issue de l'approche de la *CPA* (Jezek, 2018), on reconnaît que certains arguments peuvent ne pas être exprimés pour des raisons pragmatiques ou communicatives. Cependant, ce constat est souvent fait dans le cadre plus général d'une approche se proposant d'induire la valence *a posteriori* : à partir du corpus et de ses *patterns*. Or, c'est précisément cette démarche que la *TCO* n'adopte pas.

### 2.3. Le poids de la fréquence

Concevoir la structure argumentale à travers le prisme de la notion de *schéma prédicatif* - plutôt qu'à travers celui de la notion de *pattern* - signifie envisager

la valence comme *a priori* par rapport au corpus. Cette perspective implique que la fréquence d'un complément n'est pas pertinente pour déterminer son statut d'argument.

Supposons qu'en corpus le verbe *vendre* (*quelque chose à quelqu'un*) apparaisse fréquemment sans le destinataire, mais avec seulement le vendeur et le bien. Supposons qu'en corpus le verbe *parler* (*avec quelqu'un de quelque chose*) apparaisse fréquemment sans mentionner ni l'interlocuteur, ni le propos, et qu'il soit en revanche accompagné par des adverbiaux comme *à tort et à travers* ou *à voix basse*. Si nous raisonnons en termes de *pattern*, force en est d'en conclure que la valence prototypique de ces verbes est « SUJET<sub>VENDEUR</sub> + vendre + COD<sub>BIEN</sub> » et « SUJET<sub>LOCUTEUR</sub> + parler + ADVERBE<sub>MANIÈRE</sub> ». D'une part, le destinataire de la vente cesse d'être un argument prototypique de *vendre* ; de l'autre, la manière devient un argument prototypique de *parler*, à la place de l'interlocuteur et du propos. Si nous raisonnons en termes de *schéma prédicatif*, cette conclusion n'a aucun sens. D'un point de vue communicatif, dans *vendre un appartement* et *parler à voix basse*, l'expression du bien vendu et du volume de la voix peut être plus importante que l'expression du destinataire de la vente et de l'interlocuteur de l'acte de parole ; cependant, cela n'implique ni que le destinataire cesse d'être un argument de *vendre*, ni que le volume devienne un argument de *parler*. De même, au niveau de la fréquence, nous voyons les personnes plus souvent habillées que nues ; pourtant, nous n'en concluons ni que les habits sont des parties prototypiques du corps de la personne, ni que la distinction entre corps et habits est une question de *continuum*.

La notion de *schéma prédicatif* ne vise pas à minimiser les faits de fréquence qu'on vient d'évoquer ; plus simplement, elle nous invite à séparer deux critères logiquement indépendants : un critère de *pertinence communicative* d'un complément d'une part et un critère de *pertinence idéationnelle* de ce complément (son statut d'argument) de l'autre. Le premier légifère à propos de ce qu'il est pertinent d'omettre ou d'exprimer lorsqu'on énonce un procès pour communiquer un message ou pour l'insérer dans un texte : ce critère doit donc être étudié *a posteriori*, à partir du corpus. Le second légifère à propos des participants logiquement essentiels, ou pas, pour concevoir un procès *avant* son emploi pour communiquer un message ou son insertion dans un texte : ce critère doit donc être étudié *a priori*. Les modifications issues de la pertinence communicative tiennent pour acquise la structure argumentale établie par la pertinence idéationnelle.

### 3. Discrimination des acceptions : le cas de *seguire-suivre*

Au-delà des différences qui caractérisent les deux approches sur le plan théorique et méthodologique, la CPA et la TCO partagent une même visée applicative lexicographique : la réalisation de dictionnaires électroniques pour le traitement automatique du langage<sup>4</sup>. À ce niveau, la détermination des *patterns* et des *schémas prédicatifs* constitue le moyen privilégié pour discriminer les différentes acceptions d'un mot polysémique. Le travail sur les corpus a en effet montré l'impact de la polysémie sur la variété des structures argumentales possibles. Considérons par exemple les schémas prédicatifs du verbe *suivre* isolés suivant la TCO, et les *patterns* du verbe *seguire* dans la ressource T-Pas, isolés moyennant la CPA, et comparons-les<sup>5</sup>. Pour des raisons d'économie, nous ne considérerons que les acceptions de *seguire-suivre* en tant qu'« activité » (*s. qqn de près, s. un chemin, s. une règle, s. une explication, s. un patient, etc.*), laissant de côté les acceptions de *seguire-suivre* en tant que « relation » (*le mardi suit lundi, un chapitre suit un autre, etc.*). Aussi, nous nous bornerons à comparer les acceptions qui sont réellement partagées par les deux langues, écartant celles qui n'existent que dans l'une des deux (*ex. s. un produit, au sens de « continuer à le produire », ne se dit qu'en français*).

La première remarque que l'on peut formuler est que les schémas prédicatifs sont plus nombreux que les T-Pas (*patterns*). On compte en effet 8 T-Pas pour *seguire* et 16 schémas prédicatifs pour *suivre*, le double. Les T-Pas de *seguire* isolent les acceptions principales du verbe (la numérotation est la nôtre) :

- 1: [[Animate1]] segue [[Animate2]]
- 2: [[Animate]] segue [[Route]]
- 3: [[Human]] segue [[Concept | System]]
- 4: [[Human]] segue [[Concept]]
- 5: [[Human]] segue [[Activity]]
- 6: [[Human]] segue [[Event]]
- 7: [[Human1]] segue [[Human2]]
- 8: [[Human1 | Human Group1]] segue [[Human2 | Human Group2]]

Les schémas prédicatifs de *suivre*, en plus d'isoler les acceptions principales du verbe, définissent également quelques-unes de ce qu'un dictionnaire de langue classerait habituellement comme des sous-acceptions. Par exemple, la T-Pas 2 ci-dessus ([[Animate]] segue [[Route]]) identifie *grosso modo* la même acception que le schéma prédicatif suivant (nous ne reproduisons par commodité que le schéma suivi d'un synonyme, du domaine d'emploi et d'un exemple) :

*suivre*/N0:hum(métonymie intégrée sur voiture)/N1:loc<voies>/N2:jusqu'à loc  
 Sy : parcourir  
 D : véhicule  
 Ex : *On suit le boulevard jusqu'à la place*

Il s'agit de l'acception « parcourir un chemin ». On remarquera au passage que l'acception est tout de même un peu plus spécifique dans la *TCO* (l'agent est un être humain, idéalement un chauffeur) et que le schéma prédicatif est plus complet (il comprend également N2, la destination). Nous y reviendrons.

Ce qui nous intéresse ici de souligner est que la *TCO* fournit également deux autres emplois voisinant le sens de « parcourir un chemin », mais qui s'en différencient par un schéma prédicatif spécifique :

*suivre*/N0:hum(métonymie intégrée sur voiture)/N1:loc<point de repère/signal>/N2:jusqu'à loc  
 Sy : adopter orientation  
 D : véhicule  
 Ex : *On suit le nord / l'indication / la boussole jusqu'à la place*

*suivre*/N0:inc<moyens de transport/voies>/N1:loc<voies/côtes>  
 Sy : longer  
 D : véhicule  
 Ex : *Le train suit la route. La frontière suit le fleuve*

Les (sous-)acceptions mentionnées ci-dessus concernent respectivement l'action faite par un être humain d'adopter une orientation, et l'action faite par un moyen de transport ou un élément du paysage de longer une voie ou un autre élément du paysage. D'un point de vue conceptuel, toutes ces acceptions représentent autant de déclinaisons de l'acception principale « parcourir un chemin » ; elles ne sont pas présentes dans la base de données T-Pas.

La deuxième remarque que l'on peut avancer concerne le degré de précision des *patterns* vis-à-vis des schémas prédicatifs. En comparant les *patterns* de *suivre* avec les schémas prédicatifs de *suivre*, on constate que les schémas prédicatifs sont souvent plus précis et plus spécifiques que les T-Pas.

Tout d'abord, les schémas prédicatifs sont plus complets pour les raisons que nous avons illustrées *supra* (§ 2.2), c'est-à-dire qu'ils ne sont pas fondés sur la fréquence mais sur la valence en dehors du contexte d'utilisation. C'est le cas, comme nous l'avons vu, pour la T-Pas 2 ([[Animate]] segue [[Route]]), mais aussi pour les T-Pas 1 ([[Animate1]] segue [[Animate2]]) et 7 ([[Human1]] segue [[Human2]]). Prenons la T-Pas 7. La définition de ce *pattern* est confiée à l'implicature suivante :

[[Human1]] sta dietro con la mente o nel ragionamento a [[Human2]]

(littér., « va derrière avec l'esprit ou dans le raisonnement »)

Le schéma prédicatif correspondant dans la *TCO* est :

*suivre*/N0:hum/N1:hum/N2:dans ina<raisonnement>

On voit bien que l'argument N2 (dans ina<raisonnement>, où <raisonnement> est une classe d'objets) est absent dans la T-Pas 7 ; il n'est présent qu'au niveau de l'implicature.

Ensuite, les classes d'objets qu'utilise la *TCO* pour définir les emplois de *suivre* ont la caractéristique de s'appliquer à des classes plus restreintes de référents par rapport aux types sémantiques utilisés par la *CPO*. L'extension des arguments étant plus limitée, le sens est souvent plus précis. Tel est le cas pour les T-Pas 3, 5, et 8.

Prenons en guise d'exemple la T-Pas 5 :

T-Pas 5 : [[Human]] segue [[Activity]]

Implicature : [[Human]] assiste o partecipa allo svolgersi di [[Activity]]

(littér., « assiste ou participe au déroulement de »)

Le schéma prédicatif correspondant dans la *TCO* est plus spécifique, car ses arguments ont une extension plus limitée :

*suivre*/N0:hum/N1:ina<spectacle télévisé>

Sy : regarder

D : général

Ex : *On suit un match, une série*

*suivre*/N0:hum/N1:ina<cours>

Sy : participer à, assister à

D : enseignement

Ex : *On suit un séminaire, un master, un stage*

*suivre*/N0:hum/N1:ina<traitement>

Sy : se soumettre à, subir un

D : médecine

Ex : *On suit une cure, un régime*

La T-Pas 5 isole un signifié plus large par rapport aux trois schémas prédicatifs ci-dessus, car, comme le dit l'implicature, le *pattern* [[Human]] segue [[Activity]] s'applique tant aux situations où l'individu assiste au déroulement d'une activité (ex. *Paul suit un match à la télé*), qu'aux situations où l'individu participe activement au déroulement d'une activité (ex. *Paul suit un cours d'anglais*).

En revanche, les schémas prédicatifs isolés par la *TCO* distinguent plusieurs cas de figure, comme le montrent bien la classe d'objets N1, le synonyme et l'exemple. Il s'agit d'acceptations plus restreintes.

Pour finir, nous aimerions souligner le rôle clé du domaine d'emploi dans le cas de la *TCO*. Le domaine d'emploi est prévu aussi dans la description des T-Pas, mais au moment de l'extraction des *patterns* à partir de la base de données, les domaines n'étaient pas fournis, ce qui nous empêche de faire une comparaison avec T-Pas sur ce point. Dans la *TCO*, la présence du domaine d'emploi est fondamentale. Prenons la T-Pas 8 :

T-Pas 8 : [[Human1 | Human Group1]] segue [[Human2 | Human Group2]]  
 Implicature: [[Human1 | Human Group1]] mostra attenzione in favore di, assiste, aiuta [[Human2 | Human Group2]]  
 (littér., « est attentif vis-à-vis de, assiste, aide »)

Or, considérons la perspective de la *TCO* et imaginons d'appliquer des domaines tels que la médecine et l'enseignement aux types sémantiques [[Human1 | Human Group1]] et [[Human2 | Human Group2]], appelés *hyper-classes* dans la *TCO*. Les domaines permettent de restreindre l'extension des hyper-classes, donnant lieu aux schémas d'arguments suivants :

*suivre*/N0:hum<soignant>/N1:<patient>

Sy : soigner, surveiller

D : médecine

Ex : *Le médecin suit le patient*

*suivre*/N0:hum<professeur>/N1:<étudiant>

Sy : surveiller

D : enseignement

Ex : *Le professeur suit ses élèves*

Par rapport à la T-Pas 8, les schémas prédicatifs issus de la *TCO*, tout en sélectionnant la même acception du verbe *suivre*, sont beaucoup plus spécifiques. Remarquons que les étiquettes des classes d'objets se confondent avec des rôles : les humains dans le domaine médical, en effet, sont typiquement des soignants et des patients. Un autre exemple montrant l'importance des domaines d'emploi pour la *TCO* est le suivant (pour cet exemple, nous faisons référence au sens de *suivre* en tant que « relation » et pas en tant qu'« activité ») :

*suivre*/N0:<événement>/N1:<événement>

Sy : venir après

D : général

Ex : *Un beau coucher de soleil a suivi la pluie*

*suivre*/N0:<événement>/N1:<événement>

Sy : être concomitant à, aller en parallèle avec

D : économie

Ex : *L'offre n'a pas suivi la demande*

Les domaines diffèrent par rapport aux hyper-classes pertinentes. Dans le cadre du langage général, un événement est par exemple un phénomène atmosphérique, mais dans le cadre de l'économie un événement est une variation de valeur comme la hausse de la demande, des prix etc. La *TCO* porte un intérêt tout particulier vers la description des langages de spécialité (d'où l'importance des domaines d'emploi), des langages qui, au sens de G. Gross, sont vus comme autant d'îlots formant l'archipel du langage tout court.

## Conclusion

Notre comparaison entre la *TCO* et la *CPA* peut être résumée comme suit. Le *pattern* de la *CPA* est un outil qui est censé décrire le corpus *et*, en même temps, il est une donnée induite du corpus. Le *schéma prédicatif* de la *TCO*, lui aussi, est un outil qui est censé décrire le corpus, *mais* il n'est pas une donnée du corpus. Nous sommes confrontés, par-là, à une différence épistémologique majeure entre les deux approches, qui concerne leur rapport vis-à-vis du corpus.

Si le *pattern* est un outil pour décrire le corpus *et* une donnée du corpus, alors, dans le cadre de la *CPA*, le corpus contient en lui-même les conditions de possibilité de sa propre connaissance. Si le *schéma prédicatif* est un outil pour décrire le corpus, *mais* il n'est pas une donnée du corpus, alors, dans le cadre de la *TCO*, le corpus ne contient pas en lui-même les conditions de possibilité de sa propre connaissance. Ces dernières se situent au niveau de structures eidétiques - les *schémas prédictifs* - qui transcendent le corpus.

De ce point de vue, ce n'est pas surprenant que, dans le cadre de la *CPA*, on parle de la signification en termes d'événements ayant lieu en corpus (*i.e.* des énoncés) :

*Qu'est-ce que, alors, le signifié d'un mot ? Il est raisonnable de croire que, dans l'utilisation quotidienne de la langue, les signifiés sont des événements et non des entités (...). Est-ce que les signifiés existent en dehors des contextes*

*transactionnels dans lesquels ils sont utilisés ? Je dirais que non*<sup>6</sup>. (Hanks, 2013 : 73) [c'est nous qui soulignons].

Le schéma prédicatif de la *TCO*, en revanche, n'est pas un événement qui *se produit* en corpus, mais précisément une *entité* - un objet d'ordre eidétique stable et partagé - fonctionnant comme condition préalable pour donner du sens aux événements du corpus.

À bien y regarder, la *CPA* s'avère ainsi affectée par une curieuse contradiction. D'une part, elle se propose comme une théorie du lexique : donc des signifiés stablement partagés par les parlants. De l'autre, elle s'appuie sur l'acception de *signification* (*meaning*) mise en avant par Grice (cf. Hanks, 2013 : 88), qui relève du niveau conversationnel, pragmatique, et donc contingent. Cela, cependant, n'est pas totalement surprenant. Dans le cadre de la *CPA*, la seule façon pour penser le partage des signifiés est dans les termes des *patterns* d'énonciations les plus fréquents. En revanche, dans le cadre de la *TCO*, le partage des signifiés n'est pas ancré dans la régularité de *patterns énonciatifs*, mais dans le partage des *schémas prédicatifs*, qui sont des structures *a priori* présupposées par les différents *patterns* énonciatifs.

Sur le plan applicatif, la *CPA* pourrait tirer profit des travaux menés au sein du laboratoire LDI : la description des schémas prédicatifs étant plus fine dans la *TCO* que dans la *CPA*, cela augmenterait le pouvoir de désambiguïsation des *patterns*. Il serait également intéressant de comparer l'ontologie de la *CPA* avec les classes d'objets et les hyper-classes de la *TCO*. Puisque la description des signifiés verbaux s'appuie sur la description des éléments occupant la place d'argument, il est aisé de comprendre que le succès d'un outil électronique visant au traitement automatique du langage ne peut que dépendre d'une description fiable et soignée des classes d'objets ainsi que des types sémantiques.

Pour finir, il a été récemment démontré (Giacomini, Di Muccio-Failla, 2019) que les résultats obtenus par l'application de la méthode *CPA* pourraient être sensiblement améliorés en adoptant une approche « mixte » qui ne se base pas seulement sur l'examen des concordances, mais aussi sur l'extraction de collocations, ainsi que sur l'exploitation des dictionnaires déjà existants. Il est donc souhaitable que les approches et les méthodes existantes en matière d'analyse de corpus sortent des frontières nationales et entament un processus de connaissance réciproque.

## Bibliographie

Buvet, P.A., Grezka, A. 2009. « Les dictionnaires électroniques du modèle des classes d'objets ». *Langages*, 176/4, p. 63-79.

- Chinkova, S., Hanks, P. 2010. « Validation of Corpus Pattern Analysis - Assigning pattern numbers to random verb samples », Octobre.
- Fasciolo, M. 2018. « Les noms du fond, ou la fonction des noms dits 'sommitaux' ». *Congrès Mondial de Linguistique Française*, en ligne.
- Fasciolo, M. 2019a. « Faut-il distinguer le lexique de l'ontologie ? » *Langue Française*, 204/4, p. 21-35.
- Fasciolo, M. 2019b. « The syntax of the lexicon : classes of objets and uses of predicates ». *PhrasaLex*.
- Fasciolo, M., Gross, G. 2020. *La sintassi del lessico*. Torino: UTET.
- Feltracco, A. 2016. « T-PAS: Building a Corpus-based Reference Tool for Italian ». *RiCOGNIZIONI*, 3/6, p. 69-78.
- Giacomini, L., Di Muccio-Failla, P. 2019. « Investigating Semi-Automatic Procedures in Pattern-Based Lexicography ». In : *Electronic Lexicography in the 21<sup>st</sup> Century. Proceedings of the eLex 2019 conference*, p. 490-505.
- Gross, G. 1992. « Forme d'un dictionnaire électronique ». In : Clas, A. et Safar, H. (dir.) *L'environnement traductionnel*. Sillery : PUQ.
- Gross, G. 1994. « Classes d'objets et description des verbes ». *Langages*, 28/115, p. 15-30.
- Gross, G. 1995. « Une sémantique nouvelle pour la traduction automatique : les classes d'objets ». *La tribune des industries de la Langue et de l'information électronique*, 17-18-19, Paris.
- Gross, G. 2008. « Les classes d'objets ». *Lalies*, p. 111-165.
- Gross, G. 2010 « Sur la notion de contexte ». *Meta*, 1, 55, p. 187-198.
- Gross, G. 2012. *Manuel d'analyse linguistique*. Villeneuve d'Ascq : PUS.
- Hanks, P. 2004. « Corpus Pattern Analysis ». In : *Euralex Proceedings*. Vol. I, Lorient : Université de Bretagne-Sud, p. 87-98.
- Hanks, P. 2013. *Lexical Analysis*. Cambridge Mass : MIT Press.
- Hanks, P. et Pustejovsky, J. 2005. « A Pattern Dictionary for Natural Language Processing ». *Revue française de linguistique appliquée*, 10/2, p. 63-82.
- Harris, Z. 1951. *Methods in Structural Linguistics*. Chicago-Londres : UCP.
- Hunston, S., Francis, G. 2000. *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam - Philadelphie : John Benjamins.
- Jezek, E. 2018. « Partecipanti impliciti nella struttura argomentale dei verbi ». In : Dallabrida, S. et Cordin, P. (dir.) *La grammatica delle valenze. Spunti teorici, strumenti e applicazioni*. Firenze : Franco Casati Editore, p. 55-72.
- Jezek, E. Magnini, B., Feltracco, A., Bianchini, A. et Popescu O. 2014. T-Pas : « A resource of Typed Predicate Argument Structures for linguistic analysis and semantic processing ». In : Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik.
- Le Pesant, D., Mathieu-Colas, M. 1998. « Introduction aux classes d'objets ». *Langages*, 32/131, p. 6-33.
- Mereu, L. 2017. « La struttura argomentale in un approccio « usage-based » ». *Studi e saggi linguistici*, 55/2, p. 36-65.
- Mereu, L. 2020. *Semantica della frase*. Rome : Carocci.
- Mereu, L., Piuino, V. 2019. « The Argument Structure of Verbs of Hitting and Breaking in Italian ». *Lingue e linguaggio*, 18/1, p. 143-176.
- Perek, F. 2015. *Argument Structure in Usage-Based Construction Grammar*. Amsterdam-Philadelphie : John Benjamins.
- Prandi, M. 1998. « Contraintes conceptuelles sur la distribution : réflexions sur la notion de classe d'objets ». *Langages*, 32, 131, p. 34-44.
- Prandi, M. 2004. *The building blocks of meaning*. Amsterdam-Philadelphie : John Benjamins.

- Prandi, M. 2007. « Les fondements méthodologiques d'une grammaire descriptive de l'italien ». *Langages*, 3/167, p. 70-84.
- Prandi, M. 2013. *L'analisi del periodo*. Roma : Carocci.
- Prandi, M. 2019. Phrase et énoncé. De l'ordre symbolique à l'ordre indexical. In : Neveu, F. (dir.) *Proposition, phrase, énoncé*. Londres : ISTE éditions, p.131-154.
- Tesnière, L. 1959 [1966] *Éléments de syntaxe structurale* [2 éd.]. Paris : Klincksieck.

## Notes

1. Cela n'implique pas que la structure formelle de la phrase soit complètement isomorphe par rapport à la structure conceptuelle prédicats-arguments, mais que la première (qui relève d'une « syntaxe *au-delà* du lexique ») permet d'exhiber de la façon la plus claire la seconde (qui relève d'une « syntaxe *du* lexique »). Sur ce point, cf. Fasciolo et Gross (2020).
2. « Each verb pattern in CPA is based on the structure of English clause roles described in systemic grammar (...) Each clause role in a pattern is 'populated' by a paradigm set of collocations (...). The relevant collocations of a verb in a clause role are usually nouns that share some basic aspect of their meaning, which can be expressed as a **semantic type**. Semantic types are stored in a hierarchically structured **shallow ontology**. (...) For instance, one of the patterns of the verb *translate* (...) is as follows: (...) ».
3. Dans la terminologie de Hanks (2013 : 10-11), *hasarder une hypothèse* est un *pattern* « normal », alors que *hasarder une explication, une réponse* sont des « exploitations » de cette norme.
4. La CPA a inspiré trois projets lexicographiques concernant respectivement l'anglais, l'espagnol et l'italien : le PDEV (*Pattern Dictionary of English Verbs*), coordonné par Hanks (cf. Hanks et Pustejovsky, 2005), le projet *Verbario*, coordonné par le Tecling, Université pontificale catholique de Valparaíso (cf. <http://www.tecling.com/index.php?l=verbario>), le T-Pas, coordonné par Jezek (cf. Jezek *et al.*, 2014). La TCO est à la base d'un dictionnaire électronique du français (cf. Gross, 1992 et plus récemment Buvet et Grezka, 2009). Tous ces projets visent au traitement automatique du langage.
5. Il aurait été préférable de comparer les deux approches par des exemples relevant d'une seule langue, mais les applications de la CPA et de la TCO les plus consolidées portent sur des langues différentes : l'anglais, l'espagnol et l'italien pour la CPA, le français pour la TCO (cf. note 4). Les deux approches ne sont donc pas comparables à partir d'une seule et même langue. C'est pourquoi nous avons fait le choix de comparer le verbe *seguire*, dont les *patterns* étaient disponibles à la plateforme [tpas.fbk.eu](http://tpas.fbk.eu), avec le verbe *suivre*, dont les schémas prédicatifs nous ont été fournis par G. Gross, que nous tenons à remercier.
6. « What, then, is a word meaning ? It is reasonable to assume that in everyday use of language, meanings are events, not entities (...). Do meanings exist outside the transactional contexts in which they are used? I would argue that they do not ».