



HAL
open science

La reformulation comme un moyen de clarification des noms abstraits

Iris Eshkol-Taravella, Natalia Grabar

► **To cite this version:**

Iris Eshkol-Taravella, Natalia Grabar. La reformulation comme un moyen de clarification des noms abstraits. Georges Kleiber; Emilia Hilgert; Silvia Palma; Pierre Frath; René Daval. Les catégories abstraites et la référence, 6, ÉPURE - Éditions et Presses universitaires de Reims, 2018, Res per nomen. halshs-01968306

HAL Id: halshs-01968306

<https://shs.hal.science/halshs-01968306v1>

Submitted on 2 Jan 2019

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

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

La reformulation comme un moyen de clarification des noms abstraits

Iris Eshkol-Taravella

Université d'Orléans, LLL UMR7270

iris.eshkoltaravella@parisnanterre.fr

Natalia Grabar

CNRS, Université de Lille 3, STL UMR 8163

natalia.grabar@univ-lille3.fr

Introduction

La notion de nom *abstrait* est une notion complexe qui regroupe différentes catégories de noms, ce que montrent les différents travaux de recherches en linguistique.

Les premières classifications de noms abstraits sont proposées par Lyons (1980) et Kleiber (1994). Lyons (1977) distingue des entités matérielles et immatérielles selon leurs degrés d'abstraction : premier ordre (table, voiture, pomme), second ordre (événements et processus) et troisième ordre (propositions en dehors du temps et de l'espace, comme par exemple les noms de qualité ou de manière). D'autres études portent sur des catégories particulières au sein de cette classe de noms assez large. Van de Velde (1995) fait ainsi une analyse des noms de matières, de qualité, d'état, de couleurs et d'activité, Paykin (2002) se concentre sur les noms d'événement, d'état et de substances, alors que Tauveron (2011) propose d'établir un continuum au sein des noms d'action et de qualité. Les noms de qualité sont également étudiés par Beuseroy & Knittel (2007) et les noms d'activités par Haas & Huyghe (2010). Notons également les travaux sur l'étude des noms d'idéalités par Flaux & Stosic (2009) et des noms d'événement par Lee (2001), Prestini-Christophe (2006), Van de Velde (2006). Le plus souvent, ces travaux reposent sur la compétence linguistique des chercheurs et sur des tests linguistiques (transformation de structures syntaxiques, substitution d'un nom par un autre, etc.) et l'acceptabilité linguistique des exemples ainsi produits (Kleiber *et al.*, 2012).

Par ailleurs, dans le domaine du TAL (Traitement Automatique des Langues), plusieurs travaux sont également dédiés aux noms abstraits, mais visent leur détection automatique dans les textes. Il s'agit de la temporalité (Bittar, 2009 ; Parent *et al.*, 2008) et des noms d'événement (Amulphy *et al.*, 2010 ; Garnier-Oeliarisoa *et al.*, 2014). Cet intérêt pour les noms de temps et d'événement s'explique plus précisément par le besoin de traiter de gros volumes de données textuelles de manière systématique, et le développement de méthodes et d'applications de TAL pour y répondre. Ces noms se montrent intéressants pour l'extraction automatique d'informations ciblées.

La description et la classification de noms abstraits varient d'une recherche à l'autre. Une des raisons de cette hétérogénéité est la nature des recherches menées qui reste empirique et n'est pas basée sur les corpus et les exemples attestés.

Le travail proposé dans cet article fait partie du domaine de la linguistique de corpus : il est fondé sur l'étude des corpus pour vérifier des hypothèses ou déduire de nouvelles pistes de recherche linguistiques, tout en utilisant des outils informatiques. Deux corpus sont analysés : un corpus oral de discussions transcrites et un corpus dialogique de discussions sur le forum du Web. Les données des corpus étudiés guident la catégorisation et l'analyse effectuées.

En évitant d'utiliser le terme « abstrait », qui semble être problématique et vaste, notre étude s'appuie sur le postulat classique concernant les noms concrets, qui renvoient aux objets ou réalités que l'on peut toucher. Par conséquent, les noms communs restants s'appelleront les noms non concrets. Notre hypothèse de départ est que, les noms non concrets étant plus « abstraits » et plus théoriques, ils requièrent des explicitations complémentaires par biais de définition, de précision, d'exemplification, etc. Ce mécanisme apparaît dans le discours à travers le procédé de reformulation.

L'approche proposée exploite le processus d'annotation manuelle, qui n'est qu'une modélisation de phénomènes linguistiques sous forme d'étiquettes. En conséquence, l'information annotée est plus facilement repérable, calculable et analysable avec les outils.

La première partie de l'article est consacrée à la brève présentation de travaux sur le procédé de reformulation. Le corpus et la méthodologie de travail sont décrits dans la deuxième partie. Elle est suivie par l'analyse quantitative et qualitative des résultats obtenus.

Reformulation dans les travaux en linguistique

Beaucoup de travaux en linguistique se sont concentrés sur le procédé de reformulation. Ils sont basés sur l'étude des corpus écrits et des corpus oraux.

Les travaux sur les corpus écrits distinguent plusieurs notions proches de la reformulation : *glose*, *reprise*, *élaboration*. La notion de reformulation est tout d'abord liée à celle de *paraphrase* étudiée de différents points de vue par les chercheurs : en tenant compte de la situation d'énonciation (Culioli, 1984 ; Martin, 1976 ; Vezin, 1976 ; Fuchs, 1994), à travers les transformations linguistiques que subissent les segments paraphrasés à différents niveaux (morphologique, lexical, syntaxique, sémantique) (Melčuk, 1988 ; Vila *et al.*, 2011 ; Bhagat & Hovy, 2013) ou en fonction de la taille des entités impliquées (Flottum, 1995 ; Bouamor, 2012). Le terme de *glose* (Steuckardt & Niklas-Salminen, 2003 ; Steuckardt, 2005), issu de la tradition philologique, désigne un commentaire sur un mot et impose au premier segment d'être une unité lexicale. Un terme plus générique, celui de *reprise*, comprend « la pure et simple répétition d'un segment textuel » ainsi que les « différents degrés de ses reformulations » (Vion, 2006 : 11). Enfin, le projet ANNODIS (<http://redac.univ-tlse2.fr/corpus/annodis/>), consacré à la constitution du corpus de référence annoté en structures discursives, distingue une relation rhétorique appelée *élaboration* qui semble également se rapprocher du phénomène de reformulation sur certains points (Kleiber & Vassiliadou, 2009).

C'est au début des années quatre-vingts, dans le cadre des analyses des interactions verbales, que le terme *reformulation* apparaît dans les études de linguistes en Allemagne, France et Suisse (Gulich et Kotshi 1983, 1987 ; Roulet 1987 ; Rossari 1990, 1992, 1994). Corinne Rossari (1990-1994) distingue deux types de reformulation : les *reformulations paraphrastiques* introduites par les marqueurs comme *c'est-à-dire*, *autrement dit*, *je m'explique*, *en d'autres termes* etc., qui instaurent une relation d'équivalence sémantique entre les segments reformulés et les *reformulations non paraphrastiques*, introduites par les marqueurs comme *en somme*, *en tout cas*, *de toute façon*, *enfin* etc.) signalant un changement de perspective énonciative. Cette distinction a permis de préciser les relations entre les deux termes : *reformulation* et *paraphrase*. Tout acte de reformulation dans le discours n'introduit pas toujours une paraphrase. Ainsi, un travail important en linguistique est consacré à l'étude de différents marqueurs utilisés pour l'introduction de la reformulation, comme par exemple ces trois marqueurs : *c'est-à-dire* (Gulich & Kotschi, 1983 ; Beeching, 2007 ; Vassiliadou, 2008), *je veux dire* (Teston-Bonnard, 2008), *disons* (Hwang, 1993 ; Petit, 2009 ; Khachatourian, 2011 ; Franckel, 2016).

Parmi les travaux récents sur l'oral, on peut également citer le projet d'annotation multi-niveau de l'oral, Treebank Rhapsodie (<http://www.projet-rhapsodie.fr/>) qui voit dans une reformulation le procédé de corriger ou de préciser une précédente dénotation (Kahane & Pietrandrea, 2012).

La structure étudiée dans notre travail s'appuie sur la présence de marqueurs et doit être réalisée sous la forme suivante :

Segment 1 Marqueur *Segment 2*

où le *Segment 1* est reformulé par le *Segment 2*, alors que le marqueur de reformulation peut être instancié de trois manières : *c'est-à-dire*, *je veux dire*, *disons*, comme le montrent plus bas les exemples (1) et (2).

La reformulation observée dans de tels contextes, dépasse largement le phénomène de la paraphrase, qui présuppose une équivalence sémantique exacte entre les expressions reformulées, ou de la glose, qui est utilisée pour la définition d'une unité lexicale. Il s'agit des cas plus généraux, où le locuteur modifie le segment ou l'énoncé pour le définir, l'exemplifier, le préciser, l'expliquer, le redire, le dénommer ou encore le synthétiser. Par conséquent, le terme de *reformulation* est utilisé dans notre travail dans un sens large, pour notifier le procédé du locuteur qui modifie un segment, déjà produit dans son propre discours ou dans celui de son interlocuteur, par un autre segment. De manière générale, cette modification peut être effectuée avec ou sans l'emploi d'un marqueur, mais elle garde toujours un invariant sémantique permettant de reconnaître un lien sous-jacent entre les deux segments reformulés.

L'objectif de notre étude est de montrer que le locuteur ou le scripteur se sert du procédé de reformulation pour définir, expliquer, préciser, exemplifier, etc. une notion considérée comme « abstraite » donc vague.

Annotation

Données annotées en reformulation

Le travail présenté ici s'appuie sur l'analyse de deux corpus : le discours oral transcrit et le discours dialogique écrit que l'on peut rencontrer sur le Web dans les blogs et les forums.

Le premier corpus étudié est le corpus oral ESLO (Enquêtes Sociolinguistique à Orléans) (Eshkol-Taravella *et al.* 2012) composé de deux sous corpus : ESLO1, une enquête menée par les chercheurs britanniques à Orléans dans les années 70, et ESLO2, une enquête actuelle entreprise et menée par les chercheurs du Laboratoire Ligérien de Linguistique (LLL). Le corpus ESLO comprend différentes situations d'enregistrement : enquêtes face à face, interviews avec des personnalités, communications téléphoniques, etc. Pour pouvoir comparer les corpus ESLO1 et ESLO2, le module commun, les entretiens face à face, est sélectionné. Il s'agit d'une série de questions posées aux orléanais et portant sur leur identité, leur vie quotidienne, leur vision de la ville d'Orléans, etc. Le sous-corpus du travail comprend 260 entretiens transcrits d'ESLO1 totalisant 2 349 829 occurrences de mots et 308 entretiens d'ESLO2 totalisant 1 412 891 occurrences de mots. Les fichiers de transcription ne comportent pas de signes de ponctuation, ou de majuscule au début d'énoncé pour éviter l'anticipation de l'interprétation car en ponctuant, le transcripteur « suggère une analyse avant de l'avoir faite » (Blanche-Benveniste & Jeanjean 1987 : 142). A partir du sous-corpus de travail, 476 tours de parole (54 entretiens) d'ESLO1 et 394 tours de parole (30 entretiens) d'ESLO2 contenant un des trois marqueurs (*c'est-à-dire, je veux dire, disons*) sont extraits, comme dans les exemples (1) et (2) :

(1) *y a aussi l'attrait de l'achat qui se malheureusement aussi s'amenuise de plus en plus c'est-à-dire que j'achète mes animaux vivants* (eslo1-001)

(2) *vous avez une vie de quartier ici c'est-à-dire que vous alors faites les courses dans le quartier par exemple* (eslo2-004)

Le deuxième corpus provient du forum Doctissimo. Il s'agit de discussions portant sur les problèmes cardiaques et les douleurs du dos. Il comprend 17 443 fils de discussion, 101 728 messages et plus de 7 millions de mots. A partir de ce second corpus, 422 messages (8710 occurrences de mots), contenant les marqueurs étudiés, sont extraits, comme dans l'exemple (3) :

(3) *donc cela demande beaucoup d'investissement, c'est-à-dire il faut tout gérer, les aides à domicile, le changement des médecins, etc...* (forum)

Conventions d'annotation des reformulations

Les énoncés et messages contenant les trois marqueurs de reformulation étudiés (*c'est-à-dire, je veux dire, disons*) sont annotés manuellement selon les conventions prédéfinies portant sur les différentes relations linguistiques entre les deux segments reformulés. Le travail

d'annotation et de détection automatique des reformulations est décrit dans les travaux précédents (Eshkol-Taravella & Grabar, 2016). Ici ne sera présentée qu'une synthèse de ces travaux.

L'annotation proposée est multidimensionnelle. Elle porte d'une part sur les deux segments reformulés, mais aussi sur la relation établie par le marqueur de manière générale. Selon les spécifications, les informations annotées sont de différentes natures (syntaxique, lexicale, morphologique et pragmatique) : catégorie syntaxique (*N, A, V, Prep...*) ou type de constituant syntaxique (*NP, VP, AP, PP*) d'un segment ; modifications syntaxiques, comme la modification passif/actif ; relations lexicales entre les unités lexicales composant les deux segments (hyponymie, hyponymie, synonymie, antonymie, méronymie etc.) ; modifications lexicales effectuées dans le processus de reformulation (remplacement d'un segment ou d'une partie d'un segment par un autre, suppression d'un/des unités lexicales, ajout d'unités lexicales dans le deuxième segment) ; modifications morphologiques marquant le lien au niveau morphologique entre les unités lexicales de deux syntagmes (flexion, dérivation, composition) ; relations pragmatiques, c'est-à-dire la raison pour laquelle le locuteur remplace un segment dans son discours par un autre (définition, explication, exemplification, précision, dénomination, résultat, correction linguistique, correction référentielle, paraphrase). La typologie des relations pragmatiques a été inspirée par les travaux de Gülich & Kotschi (1987) et de Kanaan (2011) et par les observations faites sur les corpus étudiés. Ainsi, dans l'exemple (4), le groupe nominal *des interventions bénignes* est remplacé par un autre groupe nominal *des dents à extraire* dans le but d'exemplifier les propos précédents. Notons que Les erreurs d'orthographe et de frappe présentes dans les forums sont conservées dans les exemples.

(4) *En fait, avez-vous eu depuis votre RVA mécanique <NP1>des interventions "bénignes"</NP1> <MRP>c'est à dire</MRP> <NP2 rel_pragm= "exempl">des dents à extraire par exemple</NP2> et quels ont été des différents problèmes!'*
(forum)

Dans l'exemple (5), le scripteur précise ce qu'il entend par *tous ces médicaments*. Les deux segments reformulés entretiennent les rapports lexicaux de méronymie. La méronymie inclut les relations de partie vs tout mais aussi les relations par association, ce qui est le cas de cet exemple.

(5) *Mais ça va pas être facile de gérer <NP1>tous ces médicaments</NP1>, <MR>je veux dire</MR> <NP2 rel_lex="mero(médicaments/l'heure des prises, les interactions...les éventuels effets secondaires)" rel_pragm="prec">l'heure des prises, les interactions, plus les éventuels effets secondaires</NP2>, franchement je te plains* (forum)

Du corpus ainsi constitué on extrait les énoncés et les messages où l'un des segments reformulés appartient à une catégorie nominale *NP* (groupe nominal) ou *N* (nom employé sans déterminant). Pour pouvoir analyser quantitativement et qualitativement les noms non concrets reformulés, il est nécessaire de procéder à leur annotation.

Convention d'annotation des noms non concrets

La typologie pour marquer les noms non concrets est inspirée des travaux antérieurs sur les noms abstraits cités dans l'introduction de cet article et sur les observables des corpus étudiés :

- *action* : une action se réalise dans le temps mais elle reste ponctuelle. L'exemple (6) illustre cette catégorie avec deux actes médicaux, *la cryoablation* et *la radiofréquence*, reformulés par les groupes prépositionnels *par le froid* et *par la chaleur*, respectivement ;
- *activité* : une activité se réalise dans le temps également mais contrairement à l'*action* elle se fait dans la durée. Il peut s'agir, par exemple, d'un métier ou d'une occupation ludique ;

- *affect* : un affect concerne les aspects psychologiques et subjectifs de personnes, comme par exemple *stress*, *oppression*, *peur* ou *douleur*, comme dans l'exemple (7). Cette catégorie inclut les noms de sentiments également ;
- *comportement* : le comportement, ou l'attitude, correspond aux noms de manières de faire, comme, par exemple, dans *la façon à nous* ou *courage* ;
- *état* : un état exclut les affects et couvre des aspects moins subjectifs et parfois plus durables, comme les maladies ;
- *événement* : un événement se situe dans la durée et a lieu quelque part. Il peut être dynamique et continu ;
- *idée* : une idée est une représentation abstraite de la pensée, de l'imagination ou de la réflexion ;
- *langue* : la langue correspond aux langues naturelles telles que pratiquées par les humains ;
- *temps* : le temps regroupe les unités temporelles, comme les dates ou les périodes.

(6) *je cherche une liste d'établissement en france pratiquant <NP1 na="action">la cryoablation</NP1> pour une maladie de bouveret, <MR>c'est a dire</MR> <PP2 rel_pragm="prec">par le froid</PP2>, au lieu de <NP3 na="action">la radiofréquence</NP3> <MRP>c'est a dire</MRP> <PP4>par la chaleur</PP4>* (forum)

(7) *<N1 na="affect">oppression</N1> <MRP>c'est à dire</MRP> <NP2 rel_pragm="def">une sensation d'être serré dans un éteau</NP2>, sinon pour le syndrome de tietze, plus rare, la douleur est déclenchée par un simple pression sur le sternum* (forum)

Au total, neuf types de noms non concrets sont différenciés suite aux données analysées. Les noms concrets, également observés dans les corpus, sont les noms de personnes ou les groupes de personnes (*les Grecs, les Français*), les lieux, les objets, les organes et les œuvres matérielles (*livres, peintures, dessins...*), considérées uniquement du point de vue de leur matérialité.

Analyse des données

L'annotation effectuée a permis de procéder à l'analyse quantitative et qualitative de l'information balisée.

Analyse quantitative

De manière générale, ce sont les noms non concrets qui sont reformulés plus souvent (voir le Tableau 1). Cette observation se confirme encore plus dans le corpus de forum. Au total, les noms non concrets représentent 53 % dans ESLO et 84 % dans le forum. Ceci vérifie l'hypothèse de départ : les noms non concrets peuvent requérir plus d'explicitations que les noms concrets de la part des locuteurs et pour cette raison ils sont reformulés plus fréquemment.

	ESLO		FORUM	
	Fréquence	Pourcentage	Fréquence	Pourcentage
Noms concrets	21	47 %	16	16 %
Noms non concrets	24	53 %	86	84 %
Total	45	100 %	102	100 %

Tableau 1. Noms concrets et non concrets dans les deux corpus étudiés.

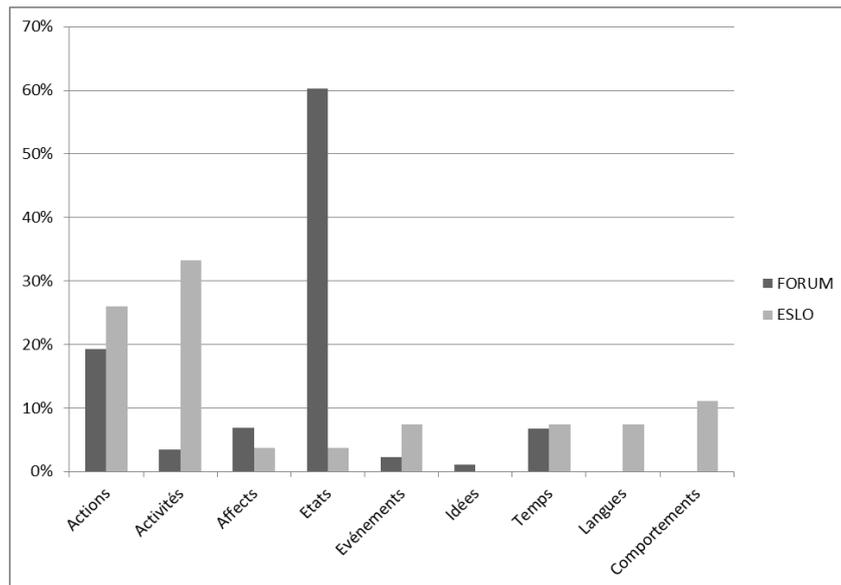


Figure 1. Distribution de noms non concrets selon les catégories et les corpus.

La Figure 1 montre la distribution de noms non concrets dans le corpus selon la typologie définie (actions, activités, affects, états, etc.). Cette distribution peut être corrélée avec le contenu des deux corpus étudiés :

- Forum est un corpus de discussions portant sur les maladies cardiaques, d'où l'apparition de beaucoup de noms d'état, ou de maladies ;
- ESLO est un corpus d'entretiens avec les Orléanais portant sur leur vie et sur le français parlé dans cette agglomération. Ceci peut expliquer la présence de nombreux noms d'activités, d'actions, d'événement, de comportements et de langue ;
- Une distribution très similaire de noms temporels dans les deux corpus est également observée.

Les résultats de différentes annotations effectuées peuvent être croisés, ce qui permet d'observer des corrélations suivantes avec les fonctions pragmatiques.

Les Figures 2 et 3 indiquent la distribution de relations pragmatiques (la raison pour laquelle le locuteur ou le scripteur effectue la reformulation) entre les deux segments reformulés où l'un des segments est un nom concret (*nc*) ou nom non concret (*nnc*). Selon la Figure 2, dans le corpus de forum, les scripteurs reformulent les noms non concrets pour apporter des définitions ou des précisions sur ceux-ci, alors que les noms concrets sont plutôt expliqués. Le scripteur reformule pour expliquer ce qu'il a dit précédemment. Dans le cas de l'explication, on peut remplacer souvent le marqueur de reformulation par *parce que*, comme dans les exemples (8) et (9) :

(8) *Il existe des techniques plus ou moins efficaces, mais j'ai entendu parler (sur le net) des bêtabloquants, or il paraît que c'est <NP1 nc="objet">des médicaments a vie et pour la vie</NP1>, <MRP>c'est-à-dire</MRP> <P2 rel_pragm= "explic">ils ne sont efficaces que lorsqu'ils sont pris tous les jours</P2>* (forum)

(9) *une phlébite se reconnaît par <NP1 nc="objet">un mollet dur chaud</NP1>, avec absence de "ballotement", <MRP>c'est à dire</MRP> <P2 rel_pragm= "explic">qu'il n'est pas souple au toucher</P2>.* (forum)

En outre, dans le corpus de forum, les noms concrets sont surtout les médicaments ou les maladies localisées dans une partie du corps. La reformulation porte donc sur les éléments les qualifiants considérées par le locuteur comme difficiles à comprendre. La reformulation sert, dans ces cas, à expliquer le type de médicament, à quoi il sert, comment le prendre et pourquoi ou bien le type de partie de corps, son emplacement, etc.

Dans les cas de précision, le scripteur en reformulant peut ajouter une information ou des détails dans le but d'éclaircir ce qui a été dit précédemment. Les noms non concrets semblent s'apprêter plus à ce processus dans le corpus de forum. Il s'agit souvent de noms d'état comme dans les exemples (10) et (11) :

- (10) *Par curiosité, j'ai repris ma tension ce soir, et j'étais à* <NP1 na="etat">15.3/9</NP1>, <MRP>c'est-à-dire</MRP> <AdvP2 rel_pragm="prec">plus, évidemment, qu'avec le médicament</AdvP2>. (forum)
- (11) *maintenant elle est très somnolente, seul* <NP1 na="etat">son bras gauche fonctionne</NP1> (<MRP>disons</MRP> plutôt <P2 rel_pragm="prec">la main et l'avant bras</P2>, l'épaule ne bouge pas) (forum)

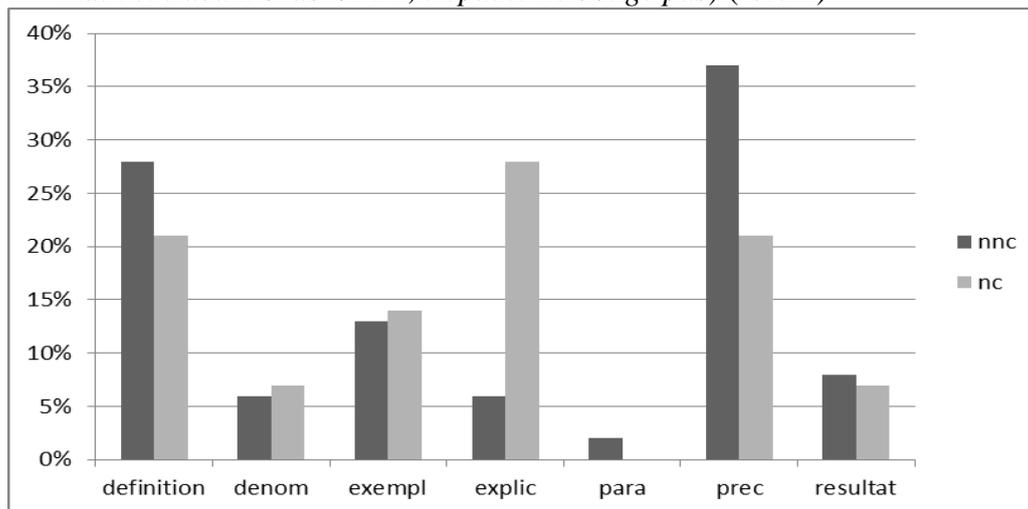


Figure 2. Distribution de relations pragmatiques des reformulations dans le corpus Forum.

Selon la Figure 3, la distribution de relations pragmatiques dans le corpus ESLO indique que :

- dans le corpus oral, comme dans le corpus de forum, le locuteur reformule les noms non concrets pour les définir. Ces noms peuvent aussi y être expliqués, paraphrasés ou bien présentés sous forme de résultat ;
- les noms concrets, comme dans le corpus de forum, sont reformulés pour être exemplifiés ou précisés.

De manière générale, la distribution de relations pragmatiques entre les noms non concrets est plus diversifiée dans le corpus ESLO. Cette observation peut être expliquée par la diversité des sujets abordés au cours de l'entretien, ce qui n'est pas le cas du forum consacré au domaine médical : même s'il existe une variété d'objets abordés, ils appartiennent à des types assez homogènes (maladies, traitements, médecins, qualité de vie...).

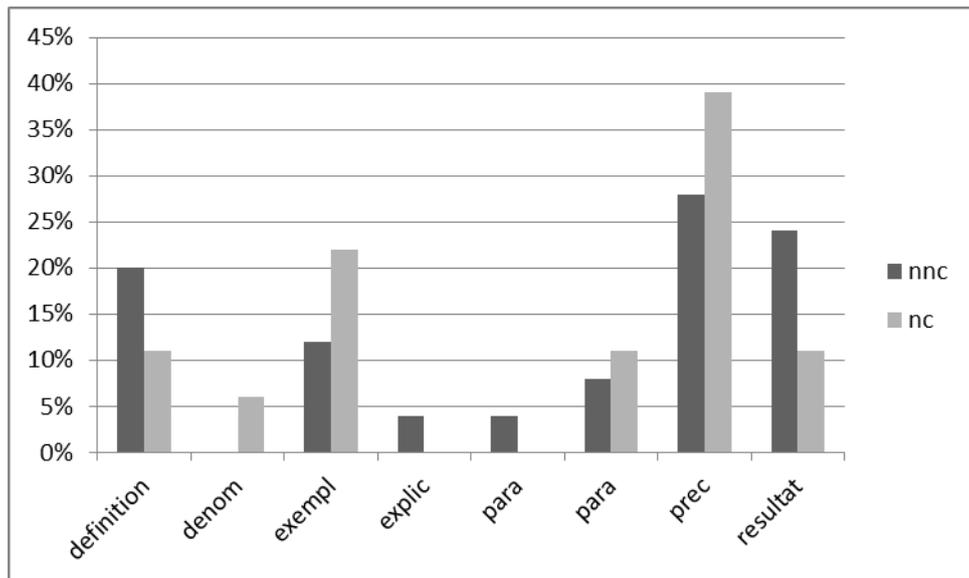


Figure 3. Distribution de relations pragmatiques des reformulations dans le corpus ESLO.

Analyse qualitative

L'analyse des faits observables du corpus n'est pas possible sans tenir compte du contexte dans lequel ce fait se produit. Ainsi, il a été observé quelques cas où un nom non concret devient concret grâce à la reformulation qui le contextualise. Ainsi, dans l'exemple (12) le nom non concret *psychologie* est redéfini par le nom de personne *les patients*.

(12) *toute la psychologie de de secteur en fait ce qu'on appelle la psychologie de secteur c'est-à-dire les les patients un peu hors les murs* (eslo2)

Il est intéressant de noter les cas où le nom non concret, placé après le marqueur de reformulation, sert à préciser ce qui a été dit précédemment. Il s'agit des formulations vagues considérées comme peu compréhensibles par le locuteur et c'est la raison pour laquelle il les reformule par une unité lexicale courte et plus précise. Dans la typologie proposée des relations pragmatiques, il s'agit d'un *résultat*. Dans ces cas, le marqueur de reformulation peut être remplacé par *en somme*, *en bref*, etc. Ainsi, dans l'exemple (13), le premier segment *tous ces bien passées à la maison* est précisé par le nom d'action *aucun faux mouvement*.

(13) *TU n ' as pas le même soin; et tous ces bien passées à la maison, par la je veux dire aucun faux mouvement, mais je n ' ai pas eu la qualité du kiné et cela a servi a rien* (forum)

Dans les autres cas observés, le locuteur ou scripteur utilisent la reformulation pour expliquer, préciser, exemplifier etc. un nom non concret qu'il considère imprécis ou flou. Dans l'exemple (14), le nom générique *la totale* est explicité par les actions *echo*, *doppler* et *test d'effort* qui apparaissent ici comme les hyponymes ou les exemples de *la totale* prescrit par le médecin.

(14) *Je suis allé voir un cardio et il m'a fait la totale c'est-à-dire echo, doppler et test d'effort.* (forum)

(15) *serai je capable de respecté les engagements que j'ai pris avec le kiné, c'est a dire de faire une activité de 3 fois par semaine* (forum)

Notons cependant que parfois ce n'est pas seulement le nom non concret qui est reformulé mais toute la prédication, où ce nom non concret sert d'appui. L'exemple (16) illustre ce type de situations. Ici, c'est le segment *respecter les engagements* qui est reformulé par le segment *faire une activité* avec la préposition *de*. Nous n'avons pas pris en compte ce type de reformulations, car elles dépassent la catégorie nominale.

(16) *serai je capable de respecté les engagements que j'ai pris avec le kiné, c'est a dire de faire une activité de 3 fois par semaine* (forum)

Conclusion

L'étude présentée dans cet article montre le fonctionnement de noms non concrets dans les reformulations introduites à l'aide de trois marqueurs *c'est-à-dire*, *je veux dire*, *disons* dans deux corpus (oral et écrit dialogique). L'hypothèse principale du travail est que les noms non concrets requièrent une reformulation car ils peuvent correspondre à des notions plus difficilement compréhensibles. La méthodologie est fondée sur l'annotation manuelle des reformulations et des noms reformulés. Une analyse qualitative et quantitative est proposée suite à l'annotation effectuée.

Neuf types de noms non concrets sont observés dans les données étudiées : actions, activités, affects, états, événements, idées, temps, langues et comportements. Cette typologie ne prétend pas être exhaustive mais, étant fondée sur des données langagières réelles, elle couvre une palette assez large de catégories de noms non concrets. Dans les deux corpus analysés, les noms non concrets sont reformulés plus fréquemment que les noms concrets. Ceci est vérifié plus largement dans le corpus dialogique du forum, où les discussions portent sur des questions de santé et concernent donc des notions spécifiques, comme les noms de maladies, de procédures ou de traitements. Les cas où la reformulation dépasse le nom non concret et porte sur toute la prédication sont également intéressants. Cependant, nous n'avons pas pris en compte ce type de reformulations.

Parmi les fonctions pragmatiques de la reformulation des noms non concrets, ce sont les définitions qui occupent une place importante quel que soit le corpus, ce qui confirme également l'hypothèse de départ. Il faut remarquer que les noms concrets peuvent aussi être reformulés. Dans ce cas, il s'agit surtout des explications et des exemplifications. Lorsque nous disposerons de données annotées plus étendues, il sera possible de faire des tests statistiques de significativité quant à l'association entre les corpus et les types de noms.

Les données analysées indiquent aussi que les locuteurs et les scripteurs peuvent effectuer la reformulation de noms non concrets de manière personnalisée et en fonction des situations d'énonciation et du contenu évoqué.

Références bibliographiques

- Arnulphy, B., Tannier, X., Vilnat A., 2010, « Les entités nommées événement et les verbes de cause-conséquence », *Actes TALN 2010*, Montréal, Canada.
- Bhagat, R., Hovy, E., 2013, "What is a paraphrase?", *Computational Linguistics*, 39(3), 463-472.
- Beauseroy, D., Knittel, M.-L., 2007, « Nombre et détermination : le cas des noms de qualité », *Rivista di linguistica*, 19, 231-262.
- Beeching, K., 2007, « La co-variation des marqueurs discursifs *bon*, *c'est-à-dire*, *enfin*, *hein*, *quand même*, *quoi* et *si vous voulez* : une question d'identité ? », *Langue française*, 154 (2), 78-93.
- Bittar, A., 2009, "Annotation of events and temporal expressions in French texts", *Proceedings of the Third Linguistic Annotation Workshop, ACL and AFNLP*, 48-51.
- Blanche-Benveniste, C., Jeanjean, C., 1987, *Le français parlé, transcription et édition*, Paris : Didier érudition.
- Bouamor, H., 2012, *Étude de la paraphrase sous-phrastique en traitement automatique des langues*, Thèse de doctorat, Université Paris Sud, Paris.
- Culioli, A., 1976, *Notes du séminaire de DEA*, Paris.
- Eshkol-Taravella, I., Grabar, N., 2016, « Reformulation à l'oral et dans le forum Web », *CMLF2016*, Tours, France.
- Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., Tellier, I., 2012, « Un grand corpus oral "disponible" : le corpus d'Orléans 1968-2012 », *TAL*, 52(3), 17-46.
- Flaux, N., Stosic, D., 2009, « Noms d'idéalités, prépositions et temporalité », in Arjoca-Ieremia, E., Avezard-Roger, C., Goes, J., Moline, E., Tihu, A. (éd.), *Temps, aspect et classes*

- de mots : études théoriques et didactiques. Actes du 7^{ème} colloque international de linguistique française et roumaine. Timisoara, 18-20 Mai 2009*, Arras : Artois Presses Université.
- Flottum, K., 1995, *Dire et redire. La reformulation introduite par « c'est-à-dire »*, Thèse de doctorat, Hogskolen i Stavanger, Stavanger.
- Franckel, J.-J., 2016, « Formes impératives de *dire* : *disons, dis, dites* et leurs variantes », in Rouanne, L., Anscombe J.-C. (éd.), *Histoires de dire*, Berne : Peter Lang, 131-154.
- Fuchs, C., 1994, *Paraphrase et énonciation*, Paris: Ophrys.
- Garnier-Oeliarisoa, D., Battistelli, D., Minel, J.-L., 2014, « Vers l'annotation automatique des noms prédicatifs d'évènements médiatiques : Perspectives actuelles », in Fryni Kakoyianni-Doa, *Penser le Lexique-Grammaire*, Paris : Honoré Champion, 413-426, [<https://halshs.archives-ouvertes.fr/halshs-00951182>].
- Gulich, E., Kotschi, T., 1983, « Les marqueurs de la reformulation paraphrastique », *Cahiers de linguistique française*, 5, 305-351.
- Gulich, E., Kotschi, T., 1987, « Les actes de reformulation dans la consultation La dame de Caluire », in Bange, P. (éd.), *L'analyse des interactions verbales. La dame de Caluire : une consultation*, Berne : Peter Lang, 15-81.
- Haas, P., Huyghe, R., 2010, « Les propriétés aspectuelles des noms d'activités », *Chronos*, 21, 103-118.
- Hwang, Y., 1993, « *Eh bien, alors, enfin* et *disons* en français parlé contemporain », *L'Information Grammaticale*, 57, 46-48.
- Kahane, S., Pietrandrea, P., 2012, « La typologie des entassements en français », *CMLF2012*, 1809-1828.
- Kanaan, L., 2011, *Reformulations, contacts de langues et compétence de communication : analyse linguistique et interactionnelle dans des discussions entre jeunes Libanais francophones*, Thèse de doctorat, Université d'Orléans, Orléans.
- Khatchatourian, E., 2011, « *Disons* et *pour ainsi dire*, deux marqueurs de la non prise en charge », in Dendale, P., Coltier, D. (éd.), *La prise en charge énonciative*, Louvain-la-Neuve : De Boeck Supérieur, 163-181.
- Kleiber, G., 1994, *Nominales*, Paris : Armand Colin.
- Kleiber, G., Benninger, C., Biermann Fischer, M., Gerhard-Krait, F., Lammert, M., Theissen, A. & Vassiliadou, H., 2012, « Typologie des noms : le critère *se trouver* + SP locatif », *Scolia*, 26, 105-130.
- Kleiber, G., Vassiliadou, H., 2009, « Sur la relation d'Elaboration : des approches intuitives aux approches formelles », *Journal of French Language Studies*, 19 : 2, 183-206.
- Lee, S.-H., 2001, *Les classes d'objets d'événements*, Thèse de doctorat, Université Paris 13.
- Lyons, J., 1980, *Sémantique linguistique*, Paris : Larousse.
- Martin, R., 1976, *Inférence, antonymie et paraphrase*, Paris : Klincksieck.
- Melčuk, I., 1988, « Paraphrase et lexique dans la théorie linguistique sens-texte », *Lexique*, 6, 13-54.
- Parent, G., Gagnon, M., Muller, P., 2008, « Annotation d'expressions temporelles et d'événements en français », *Actes de TALN2008*, [<http://talnarchives.atala.org/TALN/TALN-2008/taln-2008-long-004.pdf>], (« consulté le 5/03/2018 »).
- Paykin, K., 2002, « Événements, états et substances : un essai météorologique », *Chronos*, 10, 183-199.
- Petit, M., 2009, *Discrimination prosodique et représentation du lexique : application aux emplois des connecteurs discursifs*, Thèse de doctorat, Université d'Orléans, Orléans.
- Prestini-Christophe, M., 2006, « Une nouvelle grille de lecture : l'événement », *Pensée plurielle*, 13, 81-90.

- Rossari, C., 1990, « Projet pour une typologie des opérations de reformulation », *Cahiers de linguistique française*, 11, 345-359.
- Rossari, C., 1992, « De l'exploitation de quelques connecteurs reformulatifs dans la gestion des articulations discursives », *Pratiques*, 75, 111-124.
- Rossari, C., 1994, *Les opérations de reformulation : analyse du processus et des marques dans une perspective contrastive français-italien*, Berne : Peter Lang.
- Roulet, E., 1987, « Complétude interactive et connecteurs reformulatifs », *Cahiers de linguistique française*, 8, 111-140.
- Steuckardt, A., Niklas-Salminen, A., 2003, « Le mot et sa glose », *Langues et langages*, 9, Aix-en-Provence : Publications de l'Université de Provence.
- Steuckardt, A., 2005, « Les marqueurs formés sur *dire* », in Steuckardt, A., Niklas-Salminen A. (éd.), *Les Marqueurs de glose*, Aix-en-Provence : Publications de l'Université de Provence, 51-65.
- Tauveron, M., 2011, « Un classement des noms abstraits dans le lexique français : aux confins de l'action et des propriétés », *1st Meeting of Young Researchers in Romance Languages – La philologie romane aujourd'hui*, Madrid, 4 novembre 2011, [https://www.researchgate.net/publication/230725195_Tauveron_-_Aux_confins_de_l%27action_et_des_proprietes_-_Texte_integral], (« consulté le 5.03.2018 »).
- Teston-Bonnard, S., 2008, « En français parlé, *je veux dire* est-il toujours une marque de reformulation ? », in Le Bot, M-C., Schuwer, M., Richard, É. (éd.), *La reformulation*, Rennes : PUR, 51-68.
- Van de Velde, D., 1995, *Le spectre nominal*, Louvain-la-Neuve : Peeters Publishers.
- Van de Velde, D., 2006, *Grammaire des événements*, Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Vassiliadou, H., 2008, « Quand les voies de reformulation se croisent pour mieux se séparer : à savoir, autrement dit, c'est-à-dire, en d'autres termes », in Le Bot, M-C., Schuwer, M., Richard, É. (éds), *La reformulation*, Rennes : PUR, 35-50.
- Vezin, L., 1976, « Les paraphrases : étude sémantique, leur rôle dans l'apprentissage », *L'année psychologique*, 76 (1), 177-197.
- Vila, M., Antònia Mart, M., Rodríguez, H., 2011, « Paraphrase concept and typology. a linguistically based and computationally oriented approach », *Procesamiento del Lenguaje Natural*, 46, 83-90.