

Etude quantitative sur la présupposition*

Pascal Amsili
Laboratoire de Linguistique Formelle,
Université Paris Diderot

v.1, mai 2018

On rapporte ici une brève étude quantitative sur la présupposition, menée en annotant manuellement le plus grand nombre possible d’occurrences de déclencheurs de présupposition (effectifs) rencontrés dans un corpus littéraire. De cette étude très préliminaire, dont de nombreux aspects nécessitent des études plus poussées et des vérifications diverses¹, on présente dans ce document des comptages qui donnent une bonne indication générale sur la distribution des présuppositions, ainsi qu’une ébauche de lexique de la présupposition associé avec quelques informations de fréquence.

1 Méthode

Le texte est un extrait du roman « De la Terre à la Lune » (Jules Verne, 1865, domaine public, 28 chapitres). Le corpus n’a pas subi de pré-traitement autre qu’une tokenisation superficielle. Les informations quantitatives sur le corpus sont résumées dans la table 1.

Nombre de chapitres	20
Nombre de paragraphes	933
Nombre de phrases	2384
Nombre de tokens	48377
Nombre de types (mots-forme)	7566

TABLE 1: Description quantitative du corpus étudié

1.1 Processus d’annotation

L’annotation a été réalisée au cours d’un stage, au moyen de l’outil d’annotation Salto², qui permet une annotation phrase par phrase.

La personne en charge de l’annotation, linguiste de formation, a reçu quelques explications générales sur la notion de présupposition, et a travaillé avec une répartition

*Remerciement à Loïc Liegeois pour son aide technique, et surtout à Clotilde Chappé pour son patient et subtil travail d’annotation. Pour citer cet article: Amsili, P (2018); Étude quantitative sur la présupposition; manuscrit non publié, version 1, mai 2018; Université Paris Diderot, http://www.linguist.univ-paris-diderot.fr/~amsili/drafts/amsili18_psp_v1.pdf.

1. Il reste à mener des vérifications sur le caractère systématique des annotations, à mesurer leur robustesse par des mesures d’accord inter-annotateurs, etc.

2. <http://www.coli.uni-saarland.de/projects/salsa/>

des déclencheurs en 9 classes distinctes (voir section 1.2), chacune illustrée par quelques exemples, sans qu’un vocabulaire exhaustif ait été fourni.

On a demandé à la personne en charge de considérer la présupposition plutôt du point de vue phénoménologique : dans un premier temps, il s’agissait de repérer dans le texte les inférences qui semblaient relever du phénomène³. Dans un second temps, il était demandé de repérer le déclencheur responsable et de le ranger dans une des classes proposées ; enfin, en troisième lieu, on demandait aussi que les expressions déjà rencontrées comme déclencheurs mais qui, dans un contexte donné, ne semblaient pas déclencher de présupposition soient annotées aussi, comme « déclencheur non effectif ». Par exemple, alors que la plupart des descriptions définies sont clairement présuppositionnelles, il existe des emplois où il n’y a pas de présupposition d’existence, comme dans (1a) avec une expression idiomatique, ou dans (1b) où la description définie est employé de façon prédicative et non référentielle.

- (1) a. Léa veut prendre **le large**.
 b. Max va devenir **le tuteur de Fred**.

Les résultats présentés ici résultent d’une seule campagne d’annotation, et ne portent que sur les déclencheurs repérés comme « effectifs ». La campagne d’annotation a été suivie d’une phase de correction (erreurs évidentes d’étiquetage, empans mal délimités...) et surtout d’une phase de systématisation qui a consisté, pour chaque lexème ou construction annoté au moins une fois, à vérifier que toutes les occurrences pertinentes étaient repérées, qu’elles déclenchent effectivement une présupposition ou non. Par exemple, l’adverbe *encore* avait été initialement repéré 20 fois, comme déclencheur effectif, la phase de systématisation a conduit à ajouter 11 occurrences, sachant que l’adverbe apparaît au total 49 fois, certaines occurrences ont été considérées comme non effectives (en particulier dans des associations avec un comparatif : *plus grandes encore*). L’étude des occurrences annotées comme « non effectives », malgré l’intérêt qu’elle présente, est laissée à un travail ultérieur, le présent document se concentrant sur le comptage des occurrences réelles de la présupposition.

1.2 Catégories

Factifs	<i>savoir...</i>
Verbes aspectuels	<i>arrêter, continuer...</i>
Subordonnées temporelles	introduites par, p.ex., <i>avant, après, depuis...</i>
Adverbes de manière	<i>rapidement...</i>
Restrictions de sélection	p.ex. <i>veuve...</i>
Phrases clivées	<i>C’est une voisine qui m’a secouru...</i>
Quantificateurs	<i>J’ai écrit à tous les postiers de la ville...</i>
Descriptions définies	<i>Ma fille cadette...</i>
Noms propres	
Intonation	(plus précisément marques intonatives de focus, de contraste, etc)

TABLE 2: Liste (non exhaustive) de classes de déclencheurs de présupposition, selon Beaver and Geurts (2014).

Les classes retenues ont été élaborées à partir de la liste proposée dans (Beaver and Geurts, 2014), grossièrement traduite à la table 2, et qui correspond à un assez large

3. En particulier par leurs propriétés de projection — la « résistance à la négation », facile à tester dans de nombreux cas ; et leur caractère d’arrière-plan (ou *non-at-issue* si l’on préfère)).

consensus dans la littérature. Plus précisément, à partir de cette liste, nous avons proposé les catégories suivantes :

Prédicats factifs Il s’agit de la classe bien connue des verbes factifs, qu’ils soit cognitifs (*savoir*) ou émotifs (*regretter*), accompagnés des adjectifs ayant les mêmes valeurs dans les constructions impersonnelles (*il est (bien) connu que P*).

Prédicats aspectuels Le terme est un peu impropre car nous avons regroupé dans cette catégorie les *verbes* aspectuels comme *commencer*, *finir*, *interrompre*, *reprendre*, et les *adverbes* aspectuels comme *déjà*, *ne... plus*, *encore*, qui ne sont pas à proprement parler des prédicats.

Subordonnées temporelles On trouve dans cette classe toutes les subordonnées finies introduites par une conjonction de subordination temporelle (*quand*, *après*, *jusqu’à*, etc.) et les subordonnées infinitives introduites par une préposition temporelle (avec éventuellement le complémenteur *de*) : *après V_{inf}*, *avant de V_{inf}*...

Préconditions Ce terme a été choisi plutôt que « restriction de sélection » pour recouvrir à la fois des présuppositions lexicales comme celle qui est liée à *veuve* (qui présuppose au moins que le référent est adulte et de sexe féminin) et des présuppositions liées aux contenus lexicaux des mots de classe ouverte, en particulier les verbes, et qu’on présente souvent comme des préconditions : pour *réussir* il faut avoir essayé, pour *entrer* il faut être à l’extérieur, etc.

Les décomptes obtenus pour cette catégorie se sont révélés extrêmement peu fiables, sans doute en partie parce qu’il est difficile au cours de la tâche d’annotation de porter spécifiquement l’attention sur ces cas. Les données relevant de cette catégorie sont totalement exclues de la présente étude.

Clivées Cette classe regroupe les constructions de la forme *ce <être> XP que/qui Rel* (où le terme *XP* peut représenter un groupe nominal prépositionnel, mais aussi beaucoup d’autres types syntaxico-sémantiques). On ajoute les constructions appelées pseudo-clivées, dont le patron est *ce que/qui Rel*, *ce <être>*. Voir détails à la section B.

Descriptions définies Groupes nominaux définis, introduits en français par un article défini, un déterminant démonstratif, ou un adjectif possessif. La liste des formes considérées est donnée plus loin. Les cas où une description définie se combine avec *tout* (ou ses variantes) sont distingués dans les comptages.

Adverbes restrictifs Bien qu’il ne figurent pas dans la table 2, les adverbes de restriction font partie des déclencheurs de présupposition les plus étudiés, et de plus ils sont relativement fréquents, ce qui nous semble suffisant pour en faire une classe spécifique. La liste comprend, en plus de l’adverbe *seulement*, la construction en *ne... que*, et l’adjectif *seul* en fonction adverbiale.

Subordonnées causales Nous regroupons sous ce label pas nécessairement très heureux les subordonnées concessives introduites par *bien que*, ou par *si*, ainsi que les subordonnées en *puisque*. S’y ajoutent les subordonnées consécutives.

Adverbes additifs Bien qu’absents de la table 2, les adverbes comme *aussi* ou *de nouveau* sont très étudiés comme déclencheurs de présupposition dans la littérature, et il nous a semblé utile d’en faire une classe distinguée, même si *in fine* cette classe se révèle la moins nombreuse de notre collection.

Quelques remarques par rapport à la liste de (Beaver and Geurts, 2014) :

- **Quantificateurs** : nous n’avons pas pris en compte la catégorie en tant que telle, entre autre parce que le consensus sur le statut présuppositionnel des quantificateurs nous semble moins bien établi, et aussi parce qu’en français la forme la plus fréquente du quantificateur universel (*tous les N*) fait intervenir l’article défini lui-même potentiellement présuppositionnel, qu’on constate que l’adverbe *tout* (tous, toute, toutes) se combine avec toutes les descriptions définies de notre liste, et qu’il reste par conséquent à étudier la façon dont se construit en français la présupposition liée aux quantificateurs.
- **Adverbes de manière** : comme pour les quantificateurs, il nous semble que cette catégorie est un peu moins consensuelle, et potentiellement moins bien définie, c’est la raison pour laquelle nous ne l’avons pas du tout prise en compte.
- Nous avons choisi d’exclure les **noms propres** de l’annotation, essentiellement pour la raison qu’on peut les retrouver facilement, et même si de fait cela nous empêche à ce stade de tester si toutes les occurrences de noms propres sont présuppositionnelles.
- Enfin, l’**intonation**, pour des raisons évidentes puisque nous travaillons sur un corpus littéraire écrit, n’a pas pu être envisagée (sachant que le rôle de l’intonation comme déclencheur de présupposition n’est pas aussi fermement établi en français que c’est le cas pour l’anglais ou l’allemand).

2 Résultats

2.1 Données quantitatives

Le nombre total d’occurrences de déclencheurs effectifs repérés est d’environ 5 000, ce qui, rapporté au nombre total de tokens du texte, donne un rapport brut d’environ 1 présupposition pour 10 mots. Le détail des comptages par classe est donné à la table 3. La première colonne donne les décomptes absolus, la seconde la répartition des différents types parmi les présuppositions, la troisième donne le nombre d’occurrences de déclencheurs de présupposition de chaque classe rapporté au nombre de tokens du corpus.

Cd	Type	Nb Occ.	proportion	/# tokens
dd	Descriptions Définies	4616	92.69%	9.54%
pa	Prédicats Aspectuels	117	2.35%	0.24%
ar	Adverbes Restrictifs	64	1.29%	0.13%
sc	Subordonnées Causales	59	1.18%	0.12%
st	Subordonnées Temporelles	56	1.12%	0.12%
pf	Prédictat Factifs	31	0.62%	0.06%
cl	Clivées	24	0.48%	0.05%
aa	Adverbes Additifs	13	0.26%	0.03%
	total	4980		10.29%

TABLE 3: Comptage et répartition des présuppositions effectives relevées dans le roman « De la Terre à la Lune » (Jules Verne, 1865).

2.2 Liste de déclencheurs

Nous reprenons dans cette section les différentes classes de déclencheurs, en faisant une liste exhaustive des lexèmes trouvés avec leurs fréquences d’apparition. Nous indiquons

aussi la proportion d'occurrence de chaque déclencheur par rapport au nombre total d'occurrences de déclencheurs trouvés (le pourcentage n'est pas indiqué pour les hapax, la valeur étant d'environ 0.02 % ($\frac{1}{4980}$)). Nous discutons aussi la distribution des déclencheurs à l'intérieur de chaque classe, en gardant à l'esprit que les petits nombres dont nous disposons (sauf pour les descriptions définies) permettent seulement de dessiner des tendances qui devraient être confirmées dans de plus gros corpus.

Les descriptions définies sont de loin les plus nombreuses (table 4). On se contente d'une distinction entre celles qui sont introduites par un article défini au sens strict (*le, la, les*, ainsi que les variantes *l', aux, du, au, des*), qui sont largement majoritaires (environ 3/4 des occurrences), celles qui sont introduites par un adjectif possessif (15 formes), et celles qui sont introduites par un déterminant démonstratif (4 formes). On laisse à une étude ultérieure l'examen de la distribution plus fine selon le déterminant. Les descriptions définies des trois catégories peuvent se combiner avec avec *tout* (ou variante) (2). Ces cas représentent environ 70 occurrences mais ne sont pas décomptés à part dans le tableau.

(2) *toutes ces inventions, toutes nos habitudes, toute son attention, toutes les planètes*

TABLE 4: Comptage des occurrences de déclencheurs de la classe *Descriptions définies*

Lexème/construction	Nb occ.	distr.	proportion
article défini	3469	75 %	69.66 %
adjectif possessif	675	15 %	13.55 %
article démonstratif	472	10 %	9.48 %
<i>total descriptions définies</i>	4616	100 %	92.69 %

Les adverbes restrictifs (table 5) représentent environ 1,3 % des occurrences, avec une très forte sur-représentation de *ne... que* (environ deux tiers du total) qui n'est pourtant pas la forme la plus étudiée. Les formes *seulement* et *seul* (adverbe) représentent à elles deux environ un quart des occurrences.

TABLE 5: Comptage des occurrences de déclencheurs de la classe *Adverbes Restrictifs*

Lexème/construction	Nb occ.	distr.	proportion
<i>ne... que</i>	41	64 %	0.82 %
<i>seul</i>	11	17 %	0.22 %
<i>seulement</i>	9	14 %	0.18 %
<i>uniquement</i>	3	5 %	0.06 %
<i>total adverbes restrictifs</i>	64	100 %	1.29 %

Les présuppositions aspectuelles (table 6) sont les plus fréquentes après les descriptions définies, mais elles sont environ 40 fois moins nombreuses, et ne représentent plus que 2,35 % du nombre total d'occurrences. Comme nous l'avons dit plus haut, on peut distinguer les verbes, en nombre assez faible, parmi lesquels *commencer* et *terminer* dominent légèrement ; et les adverbes, où la distribution est bien plus déséquilibrée : si l'essentiel est formé par le système *encore-déjà-pas encore-ne... plus*, l'adverbe *encore* est 2 fois plus fréquent que chacun des autres.

TABLE 6: Comptage des occurrences de déclencheurs de la classe *Prédicats aspectuels*

Lexème/construction	Nb occ.	distr.	proportion
encore	31	26 %	0.62 %
ne... plus	13	11 %	0.26 %
terminer	12	10 %	0.24 %
déjà	12	10 %	0.24 %
commencer	12	10 %	0.24 %
ne... pas encore	9	8 %	0.18 %
demeurer	8	7 %	0.16 %
arrêter	8	7 %	0.16 %
toujours	4	3 %	0.08 %
de nouveau	3	3 %	0.06 %
cesser	2	2 %	0.04 %
prolonger	2	2 %	0.04 %
rester	1	1 %	
<i>total prédicats aspectuels</i>	117	100 %	2.35 %

Les subordonnées causales (table 7) forment une classe assez hétérogène, représentant en volume à peu près autant que la classe des subordonnées temporelles. L'hétérogénéité vient de ce que nous regroupons les subordonnées en *puisque*, relativement nombreuses, avec des constructions qui relèvent plutôt de la concession, comme les subordonnées en *bien que*, les conditionnelles à valeur concessive, les consécutives en *de sorte que* ou *de telle sorte que*, certaines subordonnées en *pour* ayant aussi une valeur concessive ; et que nous ajoutons à ces subordonnées des constructions concessives basées sur un adjectif, illustrées en (3). Ces constructions sont rarement évoquées dans la littérature sur la présupposition, mais leur sémantique concessive semble bien leur conférer une dimension présuppositionnelle.

- (3)
- a. Pourtant, **si** honorable, **si** extraordinaire **qu'**elle fût, cette majorité ne devait pas être l'unanimité.
 - b. D'ailleurs, **quelque** colossale **que** fût l'opération, elle ne dépassait point la limite des forces humaines.
 - c. L'assemblée, **quoique** très montée en faveur du héros français, resta un peu interdite devant cette audacieuse théorie.

TABLE 7: Comptage des occurrences de déclencheurs de la classe *Subordonnées Causales*

Lexème/construction	Nb occ.	distr.	proportion
puisque	23	39 %	0.46 %
si	18	31 %	0.36 %
quoique A	5	8 %	0.10 %
si A que	4	7 %	0.08 %
de sorte que	3	5 %	0.06 %
bien que	3	5 %	0.06 %
pour	1	2 %	
quelque A que	1	2 %	
quoi que	1	2 %	
<i>total subord. causales</i>	59	100 %	1.18 %

Les subordonnées temporelles (table 8) représentent environ 1,1 % des occurrences, avec une liste assez ouverte de subordonnants, parmi lesquels se détachent *quand* qui représente un tiers des occurrences, et *lorsque* (environ un quart). Il faut ajouter deux

occurrences de la construction *une fois Adj*, qui semble avoir aussi une valeur présuppositive qui n’a pas été documentée à notre connaissance. Il ne s’agit pas à proprement parler d’une subordonnée temporelle, mais plutôt d’une construction adverbiale temporelle, et si cette classe d’expressions n’est en général pas incluse dans les listes de déclencheurs de présupposition, on peut soupçonner que des recherches supplémentaires pourraient faire apparaître d’autres éléments de cette catégorie.

Les subordonnées introduites par *au moment où* et *du jour où* introduisent une ambiguïté d’annotation : elles comprennent une description définie (*à+le moment, de+le jour*), et de fait certaines occurrences ont été annotées comme des descriptions définies. Les comptages donnés ici sont sans doute sous-estimés pour cette raison.

TABLE 8: Comptage des occurrences de déclencheurs de la classe *Subordonnées Temporelles*

Lexème/construction	Nb occ.	distr.	proportion
quand	20	36 %	0.40 %
lorsque	11	20 %	0.22 %
tandis que	10	18 %	0.20 %
une fois A	3	5 %	0.06 %
après V	3	5 %	0.06 %
pendant	2	4 %	0.04 %
à mesure que	2	4 %	0.04 %
avant que	1	2 %	
au moment où	1	2 %	
du jour où	1	2 %	
depuis que	1	2 %	
dès que	1	2 %	
<i>total subord. temporelles</i>	56	100 %	1.12 %

Les constructions clivées (table 9) apparaissent en petit nombre, 24 occurrences au total, avec beaucoup de variation quant à la nature du XP extrait (et la présence d’ajouts variés). Nous consacrons la section B à une étude détaillée des différents cas et des effets présuppositionnels.

TABLE 9: Comptage des occurrences de déclencheurs de la classe *Constructions clivées*

Lexème/construction	Nb occ.	distr.	proportion
<i>ce</i> ⟨être⟩ <i>XP que/qui Rel</i>	22	92 %	0.44 %
<i>ce que/qui Rel, ce</i> ⟨être⟩	2	8 %	0.04 %
<i>total constr. clivées</i>	24	100 %	0.48 %

Les prédicats factifs sont présentés à la table 10. Les linguistes ont sans doute raison de considérer comme paragon de cette classe le verbe *savoir*, qui est le seul à avoir une fréquence significative (presque la moitié des occurrences, les effectifs étant assez faibles).

On observe que le nombre de déclencheurs différents est plutôt élevé (12 dans notre liste), et certains de cette liste ne sont pas souvent listés parmi les déclencheurs, et cela peut correspondre à des emplois particuliers. Par exemple, le verbe *avouer*, dans l’extrait (4), prend une valeur factive sans doute en partie à cause du tour particulier choisi par l’écrivain pour introduire l’information.

- (4) Je suis obligé d’avouer que, malgré mon ignorance, je connais fort exactement ce

petit détail astronomique (...)

TABLE 10: Comptage des occurrences de déclencheurs de la classe *Prédicats factifs*

Lexème/construction	Nb occ.	distr.	proportion
savoir	13	42 %	0.26 %
apprendre	4	13 %	0.08 %
comprendre	3	10 %	0.06 %
remarquer	2	6 %	0.04 %
étonner	2	6 %	0.04 %
calculer	1	3 %	
sentir	1	3 %	
reconnaître	1	3 %	
ignorer	1	3 %	
découvrir	1	3 %	
avoir raison	1	3 %	
avouer	1	3 %	
<i>total prédicats factifs</i>	31	100 %	0.62 %

Les adverbess additifs sont très abondamment étudiés dans la littérature, en tant que déclencheur de présupposition et en tant que particule sensible au focus, et on s’attendrait à en trouver un nombre élevé, mais il se trouve qu’en fait c’est la classe la moins représentée, ce que nous avons déjà eu l’occasion de constater au cours d’autres études (Amsili, 2016). La table 11 donne le nombre total d’occurrences, extrêmement faible.

TABLE 11: Comptage des occurrences de déclencheurs de la classe *Adverbes additifs*

Lexème/construction	Nb occ.	distr.	proportion
également	5	38 %	0.10 %
aussi	4	31 %	0.08 %
autre	2	15 %	0.04 %
à leur tour	1	8 %	
nouveau	1	8 %	
<i>total adverbes additifs</i>	13	100 %	0.26 %

3 Discussion

Le constat le plus facile à faire est celui de la quantité importante de descriptions définies : elles représentent près de 93% des cas de présupposition effective, et sont 40 fois plus nombreuses que les autres déclencheurs les plus fréquents (présuppositions aspectuelles), qui sont seulement 10 fois plus nombreuses que les moins représentées (adverbes additifs). Ces comptages confirment donc d’un côté le caractère massif de la présupposition en langue (on trouve environ 1 présupposition pour 10 mots, ce qui, dans ce texte où les phrases comptent en moyenne 20,3 mots, donne le chiffre de 2,1 présuppositions par phrase), mais montrent d’un autre côté que ce caractère massif est presque entièrement dû aux seules descriptions définies, dont le nombre peut par ailleurs être considéré comme sous-estimé, car nous n’avons pas compté les occurrences de noms propres, qui sont fréquemment considérés comme déclencheurs de présupposition aussi. Si l’on veut se restreindre aux présuppositions autres que les descriptions définies, la fréquence cumulée tombe à moins d’une présupposition pour 100 mots, cette moyenne cachant une certaine

disparité entre les plus fréquentes (environ 3 pour 1 000 mots, prédicats aspectuels) et les plus rares (3 pour 10 000 mots, adverbes additifs).

Ce déséquilibre a un impact évident pour la constitution de corpus annotés pour la présupposition (que ce soit dans un but de recherche quantitative ou pour alimenter des systèmes d'apprentissage automatique). Alors que pour les descriptions définies, la taille minimale nécessaire pour disposer d'un nombre statistiquement utile d'occurrences semble relativement facile à atteindre (et d'ailleurs des ressources orientées vers les descriptions définies ont été élaborées depuis déjà longtemps, par exemple par Vieira and Poesio (2000), ou pour le français par Gardent and Manuélian (2005)), les corpus de taille suffisante pour les autres déclencheurs doivent être nettement plus importants.

Si on ajoute l'idée que l'on cherche à couvrir empiriquement un ensemble suffisamment important d'occurrences pour étudier les cas de suspension ou de non projection des présuppositions, on mesure l'ampleur des données à collecter.

Ajoutons quelques remarques sur le lexique induit (voir section A). Si l'on se limite aux éléments du lexique qui en eux-mêmes sont des déclencheurs de présupposition, on trouve moins de 50 lexèmes, ce qui est assez peu dans un texte qui contient de l'ordre de 5 000 lemmes⁴. La répartition entre classes fermées (essentiellement adverbes ou locutions adverbiales et conjonctions de subordination) et classes ouvertes (verbes essentiellement) est légèrement en faveur du premier groupe (trois cinquièmes), mais il n'est pas sûr que ce soit représentatif.

À ce lexique on peut ajouter le répertoire des constructions présuppositionnelles, qui comprend, comme on s'y attendait, les descriptions définies et les clivées qui sont toujours mentionnées dans les inventaires, mais aussi tout l'éventail des constructions concessives que nous avons mises au jour, et ceci est sans doute un peu plus étonnant. D'une part, sur le plan des formes, on en déduit qu'il existe potentiellement un plus grand répertoire de présuppositions déclenchées non pas par un lexème particulier mais par une construction, et ce point nous semble mériter des études supplémentaires ; d'autre part, sur le plan du contenu, la quantité de formes qui ont émergé qui relèvent de la concession au sens large suggère que ce mécanisme de concession et son rapport avec la présupposition en général devraient sans doute être étudiés plus largement.

Ce travail d'annotation se veut une première étape vers un projet à une échelle bien plus grande, et parmi les produits dérivés de ce travail préliminaire figure un guide d'annotation qui sera prochainement publié. Par ailleurs, d'autres statistiques peuvent être produites à partir de ce corpus, que nous projetons de plus de rendre public assez rapidement.

Nous publions cette première étude car nous pensons que ces premières statistiques un peu grossières méritent d'être rendues publiques avant même l'achèvement de nos autres études sur ces données, ce qui nous permet aussi d'appeler à toutes suggestions utiles quant à la démarche générale d'annotation et à l'exploitation des données ainsi produites.

Références

- Amsili, Pascal (2016) 'De la présupposition obligatoire et autres considérations sémantiques.' Mémoire d'habilitation à diriger des recherches, Université Paris Diderot
- Beaver, David, and Bart Geurts (2014) 'Presupposition.' In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, winter 2014 ed. (Stanford)

4. Le nombre est approché, en l'absence d'une lemmatisation complète du corpus.

- Delin, Judy (1992) ‘Properties of *it*-cleft presupposition.’ *Journal of Semantics* 9(4), 289–306
- Gardent, Claire, and Hélène Manuélian (2005) ‘Création d’un corpus annoté pour le traitement des descriptions définies.’ *TAL (Traitement Automatique des Langues)*
- Prince, Ellen F. (1978) ‘A comparison of *wh*-clefts and *it*-clefts in discourse.’ *Language* 54(4), 883–906
- Vieira, Renata, and Massimo Poesio (2000) ‘An empirically-based system for processing definite descriptions.’ *Computational Linguistics* 26(4), 539–593

A Lexique induit

On donne ici la liste complète des déclencheurs (effectifs) rencontrés dans le corpus, accompagnés d’un exemple représentatif. La barre bleue est de longueur proportionnelle au nombre d’occurrences (un point carré correspond à une unité) pour faciliter le repérage des éléments importants.

à leur tour	. aa	De ces serviteurs que le Soleil maintient dans leur orbite elliptique par la grande loi de la gravitation, quelques-uns possèdent à leur tour des satellites.
à mesure que	- st	Aussi, le soir même, à mesure que les paroles s’échappaient des lèvres de l’orateur, elles couraient sur les fils télégraphiques, à travers les États de l’Union, avec une vitesse de deux cent quarante-huit mille quatre cent quarante-sept milles [Cent mille lieues. C’est la vitesse de l’électricité.]
apprendre	- pf	Mais, quelques jours plus tard, les dépêches apprenaient à l’Amérique que les souscriptions étrangères se couvraient avec un véritable empressement.
après V	- st	Après avoir rangé une suite d’anses riches en huîtres et en homards, le Tampico donna dans la baie d’Espiritu-Santo.
arrêter	— pa	Une nouvelle discussion assez vive allait s’engager, chacun préconisant son arme favorite, lorsque le président l’ arrêta net.
au moment où	. st	il faut que nous le suivions pendant son parcours jusqu’ au moment où il atteindra le but.
aussi	- aa	– Je le pense aussi , répondit Barbicane, c’est pourquoi je me propose de quadrupler cette longueur et de construire un canon de neuf cents pieds. »
autre	- aa	– Oui, répondit Barbicane, et pour une autre raison plus importante encore.
avant que	. st	– Mais la guerre, dit Barbicane, la guerre est impossible dans les circonstances actuelles, et, quoi que puisse espérer mon honorable interrupteur, de longues années s’écouleront encore avant que nos canons tonnent sur un champ de bataille.
avoir raison	. pf	« Mes braves amis, dit-il d’un accent inspiré, notre président a raison de donner à la question du projectile le pas sur toutes les autres !
avouer	. pf	Je suis obligé d’ avouer que, malgré mon ignorance, je connais fort exactement ce petit détail astronomique ; mais avant deux minutes vous serez aussi savants que moi.
bien que	- sc	Cependant, bien qu’ ils suivissent deux lignes divergentes, ces savants auraient fini par se rencontrer, en dépit de tous les axiomes de géométrie ; mais alors c’eût été sur le terrain du duel.
calculer	. pf	Béquilles, jambes de bois, bras articulés, mains à crochets, mâchoires en caoutchouc, crânes en argent, nez en platine, rien ne manquait à la collection, et le susdit Pitcairn calcula également que, dans le Gun-Club, il n’y avait pas tout à fait un bras pour quatre personnes, et seulement deux jambes pour six.
cesser	- pa	Cependant, ils durent se rassurer et cesser de craindre pour les générations futures, quand on leur apprit que, suivant les calculs de Laplace, un illustre mathématicien français, cette accélération de mouvement se renferme dans des limites fort restreintes, et qu’une diminution proportionnelle ne tardera pas à lui succéder.
commencer	— pa	« Mes chers camarades, dit ce chimiste distingué, je vais commencer par des chiffres irrécusables qui nous serviront de base.

comprendre	-	pf	On comprend aisément que ce n'était pas trop de douze cents fours pour liquéfier en même temps ces soixante mille tonnes de fonte.
de nouveau	-	pa	Mais, demanda Barbicane en insistant de nouveau , vous avez bien réfléchi?...
de sorte que	-	sc	L'appareil de chauffe et la cheminée se trouvaient aux deux extrémités du fourneau, de telle sorte que celui-ci était également chauffé dans toute son étendue.
découvrir	.	pf	Le Chaldéen Bérose découvrit que la durée de son mouvement de rotation était égale à celle de son mouvement de révolution, et il expliqua de la sorte le fait que la Lune présente toujours la même face.
déjà	—	pa	Cependant le nom de Michel Ardan circulait déjà dans la ville de Tampa.
demeurer	—	pa	Celui-ci ne le perdait pas de vue, et les regards de ces deux hommes demeuraient engagés comme deux épées frémissantes.
depuis que	.	st	Mais depuis que l'Amérique, la terre de la liberté, ne comptait plus que des hommes libres dans son sein, ceux-ci accouraient partout où les appelait une main-d'oeuvre largement rétribuée.
dès que	.	st	lui demanda Michel Ardan, dès qu' il furent seuls et du ton dont il eût parlé à un ami de vingt ans.
du jour où	.	st	Du jour où ce mot, vraiment trop vif, fut prononcé, la position devint intolérable.
également	-	aa	Il faut mentionner également un mortier formidable inventé par J.-T. Maston, membre distingué et secrétaire perpétuel du Gun-Club, dont le résultat fut bien autrement meurtrier, puisque, à son coup d'essai, il tua trois cent trente-sept personnes, – en éclatant, il est vrai!
encore	—	pa	Cependant, une question restait encore à décider :
étonner	-	pf	Voilà un mot que je m' étonne d'entendre dans cette enceinte!
ignorer	.	pf	Vous n' ignorez pas non plus que, d'une façon générale, la force de résistance des canons et la puissance expansive de la poudre sont illimitées.
lorsque	—	st	Lorsqu' il fit sa fameuse communication au Gun-Club, la colère du capitaine Nicholl fut portée à son paroxysme.
ne... pas encore	—	pa	Ressemblait-elle à la Terre au temps où l'atmosphère n' existait pas encore ?
ne... plus	—	pa	L'adversaire de Michel Ardan voulait encore parler, mais il ne pouvait plus se faire entendre.
ne... que	—	ar	– Tout cela n'est que trop vrai, répliqua J.-T. Maston avec une nouvelle véhémence.
nouveau	.	aa	Il fallut imprimer de nouvelles éditions.
pendant	.	st	« C'est désolant, dit un soir le brave Tom Hunter, pendant que ses jambes de bois se carbonisaient dans la cheminée du fumoir.
pour	.	sc	Il devenait donc manifeste que les Sélénites, pour vivre dans ces conditions, devaient avoir une organisation spéciale et différer singulièrement des habitants de la Terre.
prolonger	-	pa	La promenade triomphale du président se prolongea dans la soirée.
puisque	—	sc	Il est bien entendu que nous pouvons l'établir dans des conditions de résistance pour ainsi dire infinie, puisque il n'est pas destiné à être manoeuvré.
quand	—	st	Voilà ce que savaient parfaitement les membres du Comité quand le lendemain ils entrèrent en séance.
quelque A que	.	sc	D'ailleurs, quelque colossale que fût l'opération, elle ne dépassait point la limite des forces humaines.
quoi que	.	sc	aussi, le 6 octobre, le capitaine Nicholl, quoi qu' il en eût, s'exécuta vis-à-vis du président Barbicane, et celui-ci inscrivit sur ses livres, à la colonne des recettes, une somme de deux mille dollars.
quoique A	-	sc	L'assemblée, quoique très montée en faveur du héros français, resta un peu interdite devant cette audacieuse théorie.
reconnaître	.	pf	Mais bientôt on reconnut que c'était une mystification scientifique, et les Français furent les premiers à en rire.
remarquer	-	pf	Ainsi ils remarquèrent que, pendant la pleine Lune, le disque apparaissait dans certaines parties rayé de lignes blanches, et pendant les phases, rayé de lignes noires.
rester	.	pa	Cependant, une question restait encore à décider :
savoir	—	pf	Chacun savait que cette hauteur varie suivant la latitude du lieu où on l'observe.
sentir	.	pf	L'assemblée sentit que son président allait aborder le point délicat.

seul	—	ar	Mais les seules zones du globe pour lesquelles la Lune passe au zénith, c'est-à-dire vient se placer directement au-dessus de la tête de ses contemplateurs, sont nécessairement comprises entre les vingt-huitièmes parallèles et l'équateur.
seulement	—	ar	Elle est seulement hérissée de forts élevés contre les Indiens errants.
si	—	sc	Mais si les érudits savent parfaitement la fausse histoire de la poudre, peu de gens se rendent compte de sa puissance mécanique.
si A que	-	sc	mais si grandes que soient les difficultés, notre génie industriel en aura facilement raison.
tandis que	—	st	Aussi, tandis que diverses sociétés savantes promettaient d'envoyer des délégués à Tampa-Town, le bureau de Greenwich, réuni en séance, passa brutalement à l'ordre du jour sur la proposition Barbicane.
terminer	—	pa	Cette opération se termina le 8 juillet, et le coulage fut fixé au lendemain.
toujours	-	pa	ils rêvaient toujours de bombes gigantesques et d'obus incomparables.
une fois A	-	st	Les difficultés astronomiques, mécaniques, topographiques une fois résolues, vint la question d'argent.
uniquement	-	ar	quant à sa réussite, elle dépend uniquement de la puissance de l'engin employé.

B Constructions clivées

Les constructions clivées sont des déclencheurs de présupposition non lexicaux, puisqu'il s'agit de constructions. Cependant, il y a parmi les clivées une forte variabilité, et il nous a paru intéressant, à partir des occurrences trouvées dans le corpus, de proposer une typologie, basée sur la nature syntaxique de l'élément extrait, et indirectement sur la valeur présuppositionnelle. Rappelons que le modèle des constructions clivées est *ce <être> XP que/qui Rel*, la classification suivante se base donc sur la nature syntaxique de *XP*.

Les exemples discutés dans la littérature utilisent presque toujours un groupe nominal. On trouve de fait beaucoup de cas de ce type dans le corpus (5). Dans ce cas, la présupposition est la clôture existentielle de la proposition ouverte correspondant à la relative : par exemple *le génie des Américains prétendait conquérir quelque chose* pour (5a), ou *quelque chose peut l'éclipser* pour (5b). On remarquera que certaines occurrences sont ambiguës dans leur analyse : l'extrait (5i) peut s'analyser comme une clivée, et dans ce cas le pronom *ce* n'est pas référentiel, ou comme une phrase attributive, où le pronom *ce* est anaphorique. Dans le contexte de cet extrait c'est la lecture clivée qui a été retenue au cours de l'annotation.

- (5)
- a. **c'est** lui **que** le génie audacieux des Américains prétendait conquérir
 - b. **c'est** la Terre **qui** peut l'éclipser à son tour
 - c. **C'est** un grand problème **que** tu me poses là, mon digne président
 - d. **c'est** donc au total un puits large de soixante pieds **qu'**il faut creuser à une profondeur de neuf cents.
 - e. **Ce fut** ce bout de road-way **qui** fut prolongé jusqu'à Tampa-Town
 - f. **c'était** sa main **qui** précipitait, dans un abîme creusé par elle tout un Niagara, de métal en fusion.
 - g. **c'est** un autre **qui** portera la responsabilité de vos actes !
 - h. **c'est** précisément ma faiblesse **qui** fait ma force.
 - i. **c'était** une somme importante **que** risquait le capitaine dans son invincible entêtement.

Dans le cas suivant, le *XP* est un syntagme prépositionnel, en fonction de COI, et l'effet sémantique et présuppositionnel est le même qu'avec les exemples précédents.

- (6) a. **C'est** à un Français (...), M. Louis Ménard, **que** revient l'honneur de cette grande

découverte.

Le reste des occurrences concerne des situations où le *XP* est un complément circonstanciel : de manière (7), de lieu (8), de temps (9). Dans la plupart des cas, ces syntagmes seraient optionnels dans la version non clivée de la phrase, ou, pour le dire autrement, la relative ne comporte pas de lacune (*gap*), et la présupposition qui peut être construite est le plus souvent triviale : par exemple *il existe une manière dont on transforme la fonte en fer ductile* (7c), ou *il existe un endroit d'où notre projectile s'envolera* (8d), ou enfin le cas le plus trivial : *il existe un moment où Galilée expliqua les phénomènes...* (9b). Nous laissons à une discussion ultérieure la question de savoir s'il vaut mieux considérer que ces constructions sont présuppositionnelles même avec une présupposition trivialement satisfaite, ou plutôt que ces constructions ne sont pas présuppositionnelles. Voir (Prince, 1978; Delin, 1992).

- (7) a. **c'est** dans ce mouvement **que** l'astre des nuits présente ses différentes phases
b. **c'est** en le créant **que** l'homme s'est le plus rapproché du Créateur!
c. **C'est** en enlevant ce carbone et ce silicium par l'opération de l'affinage dans les fours à puddler **que** l'on transforme la fonte en fer ductile.
d. **c'est** au moyen d'une deuxième fusion **qu'**on l'épure
e. **Ce fut** ainsi **que** le *New York Herald* et la *Tribune* soutinrent le Texas
- (8) a. **Ce fut** là **que** le *Tampico* mouilla, le 22 octobre, à sept heures du soir
b. **Ce fut** sur ce rouet **que** reposèrent les premières assises de la maçonnerie
c. **c'est** en lui **qu'**elle se résume tout entière
d. **c'est** d'ici (...) **que** notre projectile s'envolera vers les espaces du monde solaire!
e. **c'est** donc dans cette plaine **que** s'élèveront nos magasins
- (9) a. **C'est** alors **que** les difficultés pratiques eussent paru insurmontables en tout autre pays que l'Amérique.
b. **Ce fut** alors **que** Galilée expliqua les phénomènes de lumière produits dans certaines phases par l'existence de montagnes (...)

Le corpus semble ne contenir que très peu de pseudo-clivées, dont le patron est *ce que/qui Rel, ce <être>*(10). La présupposition déclenchée est la proposition ouverte correspondant à la relative, sous une clôture existentielle. Par exemple, pour (10b), la présupposition est *que quelque chose rend notre globe peu confortable*.

- (10) a. tout **ce qu'**ils surent faire, **ce fut** de les nommer ainsi
b. Mais **ce qui** rend surtout notre globe peu confortable, **c'est** l'inclinaison de son axe sur son orbite.