# (J. Rebeyrolle)

# Titre

Revoir le titre

* Suggestion : « de la sous-spécification dans les titres d’articles scientifiques »

# Sommaire

Modifier le sommaire

* Ne pas descendre au niveau 4 : inutile d’avoir les sections A…
* Titre II.3.1 : on ne comprend pas
* Corriger : « III. Discussion sur nos résultats » en « Discussion sur nos résultats »  + pb de plan : pas de III.2 => à revoir

# Introduction

Remarque : vous abusez des anaphores encapsulantes cf. dans l’introduction « cet intérêt » : peut-être un effet indirect de votre sujet puisque ces anaphores présentent souvent un nom sous-spécifié à leur tête. En fait plus que d’en abuser vous ne les utilisez pas convenablement => rien ne permet dans ce qui précède de faire sémantique le lien avec le mot « intérêt ». Voir aussi page 10 « toutes ces questions » : on peut voir là une stratégie pour ne pas avoir à les formuler explicitement (vous demandez trop au lecteur, il faut l’accompagner).

§2 de l’intro : il aurait fallu réintroduire les notions de fonction informationnelle et de fonction d’attraction : vous n’en dites pas assez pour que le lecteur comprenne les arguments qui vous font privilégier une fonction plutôt que l’autre.

Votre introduction doit être modifiée : un résumé aussi détaillé du mémoire de M1 ne peut pas figurer en introduction. De plus vous ne partez pas de votre objet d’étude de M1 qui était le segment se trouvant à droite des deux points, autrement dit un sous-ensemble de titres.

Vous ne le dites qu’incidemment à la page 10 : « Tout d’abord, Nous pensons que la classe de nom ayant émergé dans notre premier travail peut gagner à être redéfinie par une autre approche, indépendante de la position des noms immédiatement après le double point, à la fois plus stricte et d’une couverture plus large. »

Pb de formulation en bas de la page 8 :

« Nous parlerons pour désigner cet emploi, en reprenant la terminologie de Legallois (2008), de noms sous-spécifiés (NSS). La question initiale est donc de savoir si nos noms seraient des NSS. »

* est-ce que ce n’est pas plutôt l’inverse : ces noms sont-ils des NSS ? pour y répondre on s’appuie sur Legallois.

Est-ce que vous avez vraiment observé ce que vous affirmez p.9 et qui pour moi est très contre-intuitif :

« On peut également voir que *problème* crée une liaison à son contenu spécificationnel dès le titre. Il pourra également créer des références anaphoriques en étant repris avec un démonstratif, *ce problème*, si ce n’est dans le titre du fait sa trop grande concision, dans le résumé ou le texte de la publication scientifique. »

Des formulations un peu trop tarabiscotées :

Nous faisons l’hypothèse, soutenue par notre intuition et notre connaissance du précédent corpus, que le premier nom que nous étudions immédiatement après le double point est la tête, aussi appelée noyau ou racine, du syntagme de premier niveau du segment après le double point et donc la tête du segment. Nous redéfinissons notre cible d’étude comme les têtes de segments et nous élargissons cette étude, en ne regardant plus seulement le segment après le double point, mais aussi le segment avant.

Nous élargissons également notre étude aux titres à un seul segment et aux titres à deux segments séparés par un autre signe de ponctuation que le double point. Notre étude portera donc sur toutes les têtes nominales des segments des titres à un ou deux segments. Dans l’exemple (3) ci-dessous, le titre est constitué de deux segments, délimités par un double point. Nous en mettons en gras la tête de chaque segment :

Il serait préférable d’être plus simple et plus direct.

* Proposition :

1. M1 : étude focalisée sur le segment se trouvant à droite du double point
2. M2 : extension de l’étude à tous les titres du corpus :
   1. Titres composés d’un seul segment (sans ponctuation)
   2. Titres composés de 2 segments (présence d’un signe de ponctuation quel qu’il soit)
3. objet considéré : la tête nominale de ces segments + 2 exemples

Problème du choix du vocabulaire :

Cf. par exemple p.11 : « On peut suspecter » non ! on fait l’hypothèse.

Qu’est-ce qui vous fait penser ce que vous écrivez à la suite :

« d’un faible contenu sémantique et que seule la prise en compte du contexte de la tête transdisciplinaire permet d’accéder à son sens complet. »

Les objectifs de l’étude sont clairs et bien formulés à la fin de l’intro.

# Chapitre 1

Le premier chapitre me semble convenir dans l’ensemble.

Quelques petites choses :

* Sous I.2 p. 17 : formulation très surprenante (revoir votre utilisation de *dessus* et modifier car très oral)

« Plus il est large, plus il donne une fondation solide pour la confirmation ou l’infirmation d’hypothèses dessus. »

* Question : est-ce qu’il y en a des titres ayant plus de 2 segments et donc plus de 2 têtes ? cf. « Nous nous limiterons aux titres avec au maximum une tête par segment. »
* Étonnant : tableau 2 Noms communs et noms propres : donc 2 têtes même dans les titres monosegmentaux ?
* Remarque page 20 : à rapprocher des chiasmes cf. article Tanguy & Rebeyrolle
* Restent beaucoup de coquilles (notamment accord)
* P. 27 : « Nous présentons dans le tableau (8) ci-dessous un extrait de notre résultat en prenant arbitrairement 10 têtes pour nos 25 domaines. » pourquoi « arbitrairement » ?
* P.29 « Plus le nombre de titres étudiés est faible, plus les résultats sont très sensibles. » que veut dire « sensibles » ?

# Chapitre 2

**Sélection des noms appartenant au lexique transdisciplinaire**

Un premier résultat qui est à peine discuté.

* p.31 :  Les 20 premières têtes des 94 classés par la médiane sont : *étude, analyse, cas, approche, exemple, enjeu, évolution, apport, rôle, modèle, réflexion, évaluation, outil, question, représentation, application, construction, introduction, histoire* et *développement*. La liste complète est fournie dans l’annexe *A3. Liste des têtes transdisciplinaires*. Aucun nom propre ne figure dans cette liste ce qui est logique. Il s’agit de noms communs abstraits ayant un faible contenu sémantique.
* Vous l’affirmez sans le démontrer or pour certains noms la question se pose par exemple pour *outil* que vous aviez discuté en M1

Comparaison à la liste des termes listés par Tutin : la méthode aurait plus de poids si vous partiez de la méthode de Tutin pour établir la vôtre (c’est sans doute comme cela que vous avez procédé)

* Page 32 : il faudrait commenter, vous vous contentez d’affirmer (sans pousser assez loin la réflexion) : c’est très superficiel.

Un second contrôle est de le confronter à la liste des noms du lexique transdisciplinaire des écrits scientifiques (LTES) établie par Tutin (2007, 2008). Sur les 94 têtes transdisciplinaires, 74 sont présentes dans le LTES soit 79 %.

* À commenter : c’est quand même beaucoup : il faut en dire quelque chose.

Les 20 têtes qui ne figurent pas dans le LTES sont : *enjeu, histoire, dynamique, regard, impact, retour, essai, politique, enseignement, note, formation, science, remarque, émergence, point, conception, méthodologie, discours, défi, jeu*. Il nous semble paradoxal que certains lemmes ne figurent pas dans le LTES, surtout ceux sémantiquement liés directement à la science comme méthodologie ou science.

* « paradoxal » : si vous voyez un paradoxe il faut dire lequel. Les termes de votre liste sont-ils tous des noms abstraits sous-spécifiés ?

Les autres peuvent avoir été considérés comme trop génériques : il en effet difficile de délimiter ce qui est propre à la science, le lexique transdisciplinaire des écrits scientifiques étant considéré comme un sous-ensemble d’un lexique abstrait général (Tutin, 2007).

**Comparaison segment 1 / segment 2**

Un deuxième résultat pas davantage discuté : mots du lexique transdisciplinaire présents dans les titres. Or il y aurait beaucoup de choses à creuser :

* Vous vous contentez de dire un mot seulement du second segment et de ses spécificités : il faudrait faire le lien avec le M1 (il s’agit là d’une des observations qui vous avait mis sur la piste des NSS)
* Dans mono, 6 N (sur 81) qui ne se trouvent pas dans la liste complète établie sur tout le corpus : *contrôle, fonction, notion, temps, transformation* et *valeur*.
* En dire quelque chose
* Dans segment 1, 5 N (sur 63) absent de la liste complète établie sur tout le corpus : *compte, contribution, culture, économie* et *identité*
* Que pouvez-vous en dire ?
* Dans segment 2, 19 N (sur 99) de plus : *condition, contexte, définition, démarche, donnée, illustration, leçon, limite, mode, mythe, paradoxe, parcours, piste, problématique, réalité, revue, source, synthèse* et *voie.*

Il faudrait aussi creuser ce que vous affirmez (sans l’étayer) page 33 « un très faible contenu sémantique »

Quelle est l’hypothèse qui justifie de comparer les têtes des 2 segments ? il faut la formuler pour que le lecteur sache où vous allez.

On ne comprend pas le 2e § de la page 33

Des formulations à simplifier :

Uniquement voit-on dans le sous-corpus des titres monosegmentaux poindre étude à 3 %.

Ce fait p.33 vous ne le prenez pas assez en considération :

un nom peut ainsi être employé de façon sous-spécifié ou non

**Section consacrée à la présentation des constructions spécificationnelles**

* Les exemples choisis ne sont pas toujours les plus percutants : prendre des exemples classiques, plus proches de la littérature à ce stade
* Compléter la phrase sous CS-V, CS-VI
* CS-VII = plutôt *le problème du recueil de données* que *le problème de recueillir les données*
* La distinction dont vous parlez de distinguer « N de N déverbal » ou N est un NSS des cas où N n’est pas un NSS se pose uniquement pour une extraction automatique
* Ce que vous dites au bas de la page 36 me paraît peu clair : de plus, l’exemple que vous donnez pour la CS-I est pour moi non acceptable (Le projet que l’État légifère contre le vapotage dans les lieux publics), de même que celui de CS-VII qui ne correspond pas à la construction (Le projet de loi contre le vapotage dans les lieux publics). On se demande quel est le but de ce § eu égard à votre objectif ?

C’est parce que vous ajoutez les constructions **CS-V, VI, VII** que vous faites tous ces détours => est-ce que cela vous est vraiment utile ? En tout cas, il faut le justifier (par exemple en disant : « Au vu des constructions que vous devez extraire vous passez par une analyse préalable de leurs propriétés syntaxiques… »).

Mais cela vous entraîne parfois un peu loin cf. ce que vous écrivez p.38 :

D’où la seconde hypothèse de considérer le contenu spécifiant comme un syntagme prépositionnel introduit par de incluant un syntagme verbal dont le verbe est à l’infinitif. Joseph Donato rappelle à ce propos que la « distinction entre syntagme et proposition n'était pas toujours très claire ni très systématique » **mais cette problématique vaste s’éloigne trop de notre sujet**. **Par symétrie, nous parlerons de proposition subordonnée infinitive, considérant le *de* comme un subordonnant équivalent à *que*.**

Cela vous conduit à des formulations quelque peu contestables et à des passages quelque peu obscurs :

Nous évitons le terme de *complétive* car, s’il signifie que la proposition peut occuper les fonctions d’un nom, il se rapporte directement au nom d’une fonction, celle de complément, alors qu’un nom, et d’autant plus les constructions spécificationnelles avec le verbe *être*, rapproche le contenu spécifiant de la fonction d’attribut ou complément attribut selon la terminologie de Delhay (2014). Nous privilégions donc deux appellations se référant à la catégorie morphosyntaxique d’un terme distinctif de chaque construction : la conjonction de coordination que dans un cas, l’infinitif dans l’autre. Comme l’indique Kalmbach (2019), *« on peut facilement mettre en parallèle les deux types de construction »* par une transformation : « le fait que le jury sélectionne les candidats » ➔ « le fait de sélectionner les candidats ». Néanmoins, l’auteur constate également une différence majeure : *« le sujet n’est pas exprimé dans l’infinitive. Par rapport à la construction conjonctive, la construction infinitive prend donc une valeur impersonnelle ou générale »* car en effet le sujet de *sélectionne*, *le jury,* est perdu dans la transformation.

Tout ce qui suit me semble très confus (suite page 38 et également page 39) : est-ce justifié ? Vous semblez dire que oui mais de manière tout de même un peu présomptueuse p. 41

Contrairement à ces travaux, dans un contexte averbal comme les titres, nous ne pouvons faire l’économie de pas considérer les CS sans verbe être conjuguée comme CS-VI et CS-VI. Nous cherchons donc les schémas suivant dans notre corpus :

NB : « ces travaux » n’est pas interprétable : le référent n’a pas été constitué en objet de discours (est-ce que d’ailleurs j’ai bien compris, vous parlez de Legallois et Roze qui écartent les constructions **CS-V, VI, VII**, c’est ça ?)

La justification pour l’étude des CS-VII est donnée en introduction : il faudrait y revenir ici.

**La question que je me pose est celle de savoir s’il était vraiment productif d’ajouter des constructions dont le repérage automatique pose des problèmes insolubles comme vous l’avez fait. Ce que nous vous avions conseillé était plutôt d’observer l’environnement des N identifiés afin de faire émerger d’éventuelles régularités.**

Soyez prudent :

Les deux ayant des étiquettes différentes dans Talismane, cela ne posera pas de problème.

Dites plutôt : « A priori, cela ne devrait pas poser de problème »

Comme le confirme la suite…

À propos des exemples 34 à 36 vous dites :

La paraphrase de ces exemples en CS-II, avec un infinitif, est immédiate : *Quelques problèmes d’analyser, le problème de regrouper, Problèmes de créer en multimédia. Analyse, regroupement, création sont des déverbaux dénotant une action ou une activité*.

* Ce n’est pas la construction II : NSS + être + de + inf soit *le problème est de regrouper des activités dans la modalisation ABC.* Ça marche pour 35 mais pas pour 34 et 36 où les paraphrases ne fonctionnent pas me semble-t-il ?

En revanche, ce que vous dites de (37) est juste.

Problème d'interprétation des enclos quadrangulaires de La Tène moyenne **découverts** en Flandre française : l'exemple de Borre

> le problème est d’interpréter les enclos quadrangulaires de La Tène moyenne **découverts** en Flandre française

La place que vous donnez à la CS-VII en l’absence d’autres constructions spécificationnelles présentes dans vos titres, vous entraîne sur un terrain très vaste qui nécessiterait une étude à part entière. Les relations qui se jouent dans le schéma N de N sont en effet extrêmement variées…

Autre piste que vous auriez pu explorer : étudier certains noms dans leur rapport au texte.

La comparaison interdomaine me semble insuffisante.

Vous vous aventurez sur des terrains linguistiques (relatives / complétives ou N de N) que vous ne pouvez pas maîtriser et vous laissez de côté les données… On voit finalement peu de titres de votre vaste corpus.

# fichier "2019-09-09-addendum..." (L. Tanguy)

Globalement il est très difficile de comprendre cette partie, il faudra faire un effort important pour la présentation orale. Il y a plusieurs problèmes méthodologiques (traitement des données, stats et analyse linguistique) que je détaille au fil du texte ci-dessous.

p2 : la partie sur les motifs émergeants est incompréhensible si vous ne détaillez pas plus et ne donnez aucun exemple. Expliquez ce que serait un motif, même s'il ne correspond pas exactement à ce que vous cherchez. Techniquement, ce sont des séquences connexes ou bien vous avez mis des "trous" dedans ? C'est vraiment difficile de lire ce passage.

p2 : la conclusion (dernière phrase de la section A) est un peu ambiguë : aucun motif n'existe avec les paramètres choisis. Il faudrait interpréter cela quand même. Par exemple, y a-t-il des cas-limites (qui passeraient avec un seuil plus bas) ? Etc.

p3 : la figure 1 n'est pas lisible, désolé. Par contre je vois que ce n'est pas un arbre comme indiqué, mais un treillis (je ne sais plus de quel type exactement, mais dans un arbre un nœud n'a qu'un seul parent, ce n'est pas le cas ici).

p4 : Pouvez-vous donner un exemple de ces deux séquences ? Et ça me paraît normal qu'une séquence aussi générique et fréquence que "Det N DE N" apparaisse comme fréquence, et qu'elle ne soit pas spécifique à votre classe de N. Là encore je n'ai pas compris : ce schéma autorise-t-il des insertions en plus de celles indiquées par les crochets ? J'imagine que oui étant donné que dans l'exemple 28 plus loin on voit un verbe, mais tout cela est très difficile à comprendre sans exemples en amont.

p6 : regroupez -*tion* et -*sion* en -*ion*

p6 : attention, je crois que vous dites un peu n'importe quoi au niveau morphologique... Je ne vois pas en quoi la liste des résidus que vous donnez se distingue par une modification du radical : *émergence, résistance, gouvernance* sont des dérivés normaux en -*ance* ; ordre n'est pas un déverbal de *ordonner* (c'est l'inverse) etc. Pour ne pas dire de bêtises, référez-vous à une ressource. Par exemple il y a Verbaction qui est exactement là pour ça :

<http://redac.univ-tlse2.fr/lexicons/verbaction.html>

Et surtout je pense que c'est le statut de nom d'action qui vous intéresse, quelque soit le procédé morphologique qui le relie au verbe apparenté. Ca tombe bien, c'est pour ça que Verbaction a été construit.

p8-9 : je ne comprends pas le tableau ni votre calcul de corrélation, désolé. Quelle mesure avez-vous utilisée pour obtenir les deux coefficients ? Pour rappel quand on parle de "corrélation" avec un tel coefficient, il s'agit soit de celui de Pearson, soit de Spearman (soit plus rarement de Kendall). Mais dans votre cas je ne vois pas quelles variables numériques vous avez croisées pour les calculer, puisque vous citez des variables qualitties... Votre tableau de la page 8 étant un tableau de contingence (enfin je crois), c'est un chi-deux qu'on attendrait, et après cela une mesure de la taille d'effet type phi ou V de Cramer qui donne effectivement un nombre entre 0 (pas de liaison entre les deux variables qualitatives) et 1 (liaison totale)

Et sinon si c'est un test de significativité il faut indiquer soit le degré de libertés (nombre d'individus - 2 pour les coefficients de corrélations évoqués ci-dessus) soit la valeur critique, soit la valeur-p.

p11 : qu'entendez-vous par "l'étendue est de 13" ? L'étendue d'une variable numérique est un intervalle (min-max)

p13 : "le second de a le même recteur que le premier, A, alors qu’il s’agit souvent de C" --> il s'agit souvent de B, non ?  Donnez svp un exemple de cas où c'est A qui est recteur du second "de.  Et donc au final vous avez appliqué le schéma linéaire normal en ignorant Talismane c'est bien cela ?

p13 : associer les têtes aux domaines ne demande pas d'apprentissage automatique mais encore une fois un simple calcul de spécificité type chi-deux.

Pour la conclusion, revoir la fin en fonction des calculs de liaison entre les variables qualitatives. **Légende :** AMELIORÉ BIEN COMMENTAIRE