

# Bilan des séances de regroupement

---

## I. Remarques sur notre travail et nouvelles pistes

Les séances de regroupement du 22 et 23 mai 2019 ont permis d'avoir un aperçu des trois stages et des trois projets de recherche effectués par les étudiants de deuxième année du Master LITL. Les stages permettent d'avoir un aperçu des problématiques industrielles et des possibilités d'applications en entreprise des connaissances et techniques apprises. Les projets de recherche permettent de s'intéresser à des problématiques scientifiques particulières du traitement automatique des langues. Nous avons bénéficié de notre côté des nombreuses remarques formulées par l'équipe pédagogique sur notre travail, nous les présentons dans cette première partie avec nos pistes de réflexions. Nous présentons dans la deuxième partie notre nouvel échéancier de travail.

La première remarque porte sur notre méthode globale de travail. Il a été pointé que nous nous attardions sur une question complexe, la nature sous-spécifiée de l'emploi des noms les plus fréquents et en position de racine dans les titres, au détriment d'étapes préliminaires plus simples : recensement de ces noms, étude des constructions les incluant et de la transdisciplinarité des noms et des constructions. Ces étapes, nous les avons prévues, mais il est vrai que nous ne les avons pas réalisées en entier. La question de la nature de l'emploi nous a beaucoup mobilisé car nous voulions déterminer à quel phénomène nous avons affaire. Nous avons décidé de revenir à une approche plus lexicale (Riegel, 1996, 2005 ; Legallois, 2008 ; Nakamura, 2017), en complétant ce travail d'identification des noms et l'étude de leur répartition entre les différentes disciplines scientifiques. Puis nous ferons de même avec les constructions, en élargissant à la syntaxe et à la sémantique notre approche, avant de statuer de façon nuancée sur la nature de l'emploi dans une partie plus avancée de notre travail, en précisant points de convergence et de divergence avec les travaux précédents.

En faisant cela, nous nous éloignons du point de vue orienté sémantique discursive (Flowerdew et Forest, 2015) qui peut poser problème : les titres pris isolément, sont des microdiscours où la capacité de référence quasi-pronominale de l'emploi sous-spécifié de noms n'est pas visible, l'espace étant trop court pour une reprise en anaphore ou cataphore comme dans l'exemple : « *Le problème des échanges commerciaux Chine-États-Unis. Un problème vital* ». Or, cette capacité de référence est une des trois fonctions clés de l'emploi sous-spécifié de noms selon Schmid (2000), avec la création de concepts temporaires et la catégorisation de ces concepts par le locuteur.

Une autre remarque qui a été formulée est que nous nous sommes trop inféodés à l'outil, en l'occurrence l'analyse en dépendances syntaxiques de Talismane, au lieu de mettre l'outil à notre service. A cela, nous répondons que les possibilités de l'outil, à la manière d'une lunette astronomique, ont rendu possible la perception de phénomènes dans notre corpus qui, en retour, demandent des hypothèses explicatives. Néanmoins, nous admettons qu'il est urgent de faire un pas de côté par rapport à l'outil pour deux raisons. La première est qu'il est faillible, comme dans l'exemple que nous avons montré lors de la séance de regroupement. La seconde est que s'il permet de percevoir de nouveaux phénomènes, l'élaboration d'hypothèses expliquant ces phénomènes doit venir d'une réflexion nourrie des travaux antérieurs, et non seulement de résultats techniques.

Une dernière piste est de chercher les constructions spécificationnelles qui établissent un lien entre le nom sous-spécifié et le contexte spécifiant, suivant le patron *Nom + [être] + de + infinitif* (Schmid, 2000 ; Legallois, 2008). La recherche de l'autre construction spécificationnelle la plus utilisée *Nom + [être] + que + proposition* n'avait pas donné de résultats sur notre corpus, nous avons écarté la première, à tort. Cela rejoint l'effort d'identification des noms et des structures déjà cité.

## II. Nouvel échéancier détaillé de juin à septembre 2019

### **Jusqu'au 12 juin : cycles de recherche itératif**

- Lectures complémentaires
- Développement complémentaire des outils
- Requêtes sur corpus
- Émission d'hypothèses

Ces quatre premières actions sont les étapes possibles d'une itération du processus de recherche. Chacune d'entre elles peut être le déclencheur d'un nouveau cycle. Par exemple, en observant notre corpus via des requêtes, on émet des hypothèses, on lit des travaux y ayant trait, développant éventuellement notre outil plus avant, pour enfin statuer sur l'hypothèse. A cette itération *corpus driven*, on peut opposer une itération *corpus based* où l'on part d'hypothèses, fondées sur des lectures, qui sont validées ou invalidées par des requêtes sur notre corpus.

Jusqu'au 12 juin, une direction doit être explorée plus en profondeur : les travaux adoptant un point de vue lexical sur les noms généraux sous-spécifiés (Riegel, 1996, 2005 ; Legallois, 2008 ; Nakamura, 2017) et non celui de la sémantique discursive (Flowerdew et Forest, 2015).

### **12 juin : réunion avec les directeurs pour la validation du plan du document final**

La validation entérine également la délimitation de notre problématique et des réponses apportées. À partir de cette date, les quatre actions précédentes sont reléguées en arrière-plan. Cette date marque la bascule entre la primauté de la recherche, l'exploration d'une problématique et de ses réponses ainsi que la lecture des travaux liés, et la primauté de la rédaction, la constitution du document présentant notre problématique et nos réponses en vue de les transmettre et d'être évalué. Si le cadre est arrêté, il sera encore nécessaire de finaliser les réponses apportées. De ce fait, la rédaction nécessitera parfois de vérifier le développement d'une hypothèse, une nouvelle requête, une nouvelle lecture ou un nouveau développement de nos outils, mais on veillera à rester dans le cadre arrêté par la réunion qui nous servira de fil conducteur.

### **Du 1<sup>er</sup> juin au 15 juillet (7 weekends) : rédaction du mémoire**

La rédaction a commencé avant la validation du plan, notamment par la rédaction de l'état de l'art au fil de notre travail de recherche. En plus du plan, certaines parties centrales ont déjà été rédigées pour expliciter et communiquer à nos directeurs notre problématique, notre démarche et notre liste d'hypothèses. Néanmoins, tout ce qui a été rédigé avant le 12 est susceptible de devoir être repris en profondeur.

### **15 juillet : rendu de la version initiale du document final**

*Du 15 juillet au 1<sup>er</sup> septembre, lecture par mes directeurs de la version initiale.*

### **Du 1<sup>er</sup> au 15 septembre (3 weekends) : corrections et améliorations suite aux retours**

Suite aux retours de nos directeurs, une nouvelle version sera produite pour les prendre en compte.

### **15 septembre : rendu de la version corrigée et améliorée du document final**

*Du 15 au 22 septembre, lecture par mes directeurs de la version corrigée et améliorée.*

### **Du 23 au 30 septembre : soutenance**

Éventuellement, une troisième version pourra être produite avant la soutenance si de nouveaux points à corriger ou à améliorer étaient soulevés par nos directeurs.