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 que nous présentons dans cette première partie avec nos pistes de réflexions. Nous présentons ensuite dans la deuxième partie notre nouvel échéancier.

Le premier retour 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 les avons en partie sautées et que la question de la nature de leur emploi nous a beaucoup mobilisé car nous voulions déterminer à quel phénomène nous avions affaire. Nous avons décidé de revenir à une approche plus lexicale (Riegel, 1996, 2005 ; Legallois, 2008 ; Nakamura, 2017), en complétant ce travail d’identification et d’étude de la répartition entre les différentes disciplines scientifiques. Puis nous ferons de même d’un point de vue lexico-syntaxique avec les constructions, avant de statuer de façon nuancée sur la nature de l’emploi dans une partie plus avancée de notre travail.

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

Un autre point qui a été formulé est que nous nous sommes trop inféodés à l’outil, en l’occurrence l’analyse en dépendances syntaxiques de Talismane, ou 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 ont provoqué la naissance d’une nécessité d’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 potentiellement 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’une hypothèse expliquant ces phénomènes doit venir d’une réflexion nourrie aux travaux antérieurs, et non seulement de résultats techniques.

Une dernière piste est de chercher les constructions spécificationnelles, celles 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’ayant pas donné de résultats dans notre corpus, nous avions écarté cette seconde, à tort. Cela rejoint l’effort d’identification 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 des outils complémentaires  
 - 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 primauté de la recherche, l’exploration d’une problématique et de ses réponses ainsi que des travaux liés, et primauté de la rédaction, la constitution d’un document présentant notre problématique et nos réponses d’être évaluer et de les transmettre.

La rédaction nécessitera parfois de vérifier un détail d’une hypothèse, une nouvelle requête, une nouvelle lecture ou un nouveau développement, mais on veillera à rester dans le cadre arrêté qui nous guidera.

**Du 1er juin au 15 juillet (7 weekends) : rédaction du mémoire**

La rédaction commence avant la validation du plan, notamment par l’état de l’art que nous avons commencé en amont, au fil de nos lectures. En plus du plan, certaines parties centrales ont déjà été rédigées pour expliciter notre problématique, notre démarche et dresser une 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 1er septembre, lecture par mes directeurs de la version initiale.*

**Du 1er au 15 septembre (3 weekends) : corrections 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.*

**Du 23 au 30 septembre : soutenance**

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