# Rapport de travail pour réunion du 7 novembre 2018

Ce document de travail présente un état des lieux de ma recherche en vue de la réunion du 7 novembre 2018.

Pour rappel, actions de la dernière réunion 26 septembre 2018 :

* Projeter les classes existantes de Schmid dans le corpus de titre pour…
* Détecter des constructions
* Étudier ces constructions par domaines

## 1. Les données

1) Il y avait des erreurs dans le balisage :

* <title id="xxx"><i>
* <i> tout seul sur une ligne
* </title>\n</title>
* </title> manquant

2) Il y avait des répétitions dans les données retournées par Talismane :

* pos = poslex
* -1 et -2 = -3 et -4 respectivement

3) Erreurs corrigés (les 2 premières) ou ignorées, de même pour les répétitions. Création d’un script pour les charger en mémoire.

## 2.Modèle en mémoire

Caractéristiques globales :

* 339 687 titres.
* 4 909 608 mots.

Word :

* form : str # forme
* lemma : str # lemme
* pos : str # part-of-speech
* info : str # informations morphosyntaxiques additionnelles
* gov : str # index du gouverneur ( ! attention, il faut faire -2 pour le n° et départ à 1)
* dep : str # type de relation de dépendance

Title :

* idt : str # id du titre
* year : str # année du document titré
* typ : str # type du document titré
* domains : str # domaines associés au document titré
* authors : str # auteurs du document titré
* text : str # titre sous forme d’une chaîne de caractère
* words : [Word] # suite des mots

## 3. Les shell nouns

#### Définitions

C’est un **emploi**: *shell noun = use-as-shell noun*.

À travers deux patrons grammaticaux suivants :

1. Déterminant + (pré-modificateur) + nom + that-clause ou wh-clause ou to-infinitive

Ex : The (deplorable) fact that I have no money.

1. Déterminant + (pré-modificateur) + nom + be + complementing that-clause ou wh-clause ou to-infinitive

Ex : The (big) problem was that I had no money.

Exemples les plus typiques et fréquents : case, chance, fact, idea, news, point, problem, position, reason, report, situation, thing

Termes utilisés dans la littérature pour décrire des idées se rapprochant du concept de Schmid : unspecific nouns, anaphoric nouns, general noun, low-content nouns.

Shell-content phrases  phrase dans laquelle est employé un shell noun.

Shell-content complexes  l’ensemble du shell noun / shell-content phrase ET ce vers quoi il pointe.

#### Transposition au français

Nous prenons comme hypothèse que l’on retrouve en français les traductions de ces mots comme des shell nouns :

1. Le fait déplorable que je n’ai pas d’argent
2. Le grand problème était que je n’avais pas d’argent

## 4. Les shell nouns dans notre corpus

Production d’un fichier ne contenant que les titres pour analyse avec AntConc.

Test de certains d’entre eux avec inspection visuelle des résultats. On cherche les deux patrons avec SN + que ou SN + est.

#### Fait

Problème qu’AntConc retourne les utilisations en tant que verbe.

Patron 1 SN + que : Pas d’utilisation en shell nouns détectée.

Patron 2 SN + est : En cherchant le second patron, *fait est* retourne un seul résultat qui n’est pas un shell noun.

* Quand une société devenue de fait est associée d’une société de

#### Idée

Patron 1 SN + que : 2 résultats seulement

* Démystification de l’idée que le réseau d’aide informelle
* Comprendre l’idée que se font les habitants de

Patron 2 SN + est : 0

#### Point

Patron 1 SN + que : 2 résultats qui ne sont absolument pas des shell nouns.

Patron 2 SN + est : 1 résultat qui n’est pas un shell noun.

#### Problème

Patron 1 SN + que : 1 résultat, non shell noun

* similitudes toponymiques galaïco-basques et le problème que posent certaines d'entre elles

Patron 2 SN + est : 2 résultats, non shell noun.

#### Situation

Patron 1 SN + que : 0 résultat.

Patron 2 SN + est : 0 résultat.

**Le problème est qu’il semble y avoir très peu d’emploi en tant que shell nouns dans ce corpus.**

## 5. Annexes

### A. Part-of-speech

Toutes les étiquettes possibles pour les catégories du discours.

* ADJ
* ADV
* ADVWH
* CC
* CLO
* CLR
* CLS
* CS
* DET
* DETWH
* ET
* I
* NC
* NPP
* P
* P+D
* P+PRO
* PONCT
* PRO
* PROREL
* PROWH
* V
* VIMP
* VINF
* VPP
* VPR
* VS

### B. Informations morphosyntaxiques additionnelles

* g=m|f  
  gender = male or female
* n=p|s  
  number = plural or singluar
* p=1|2|3|12|23|123  
  person = 1st, 2nd, 3rd (or a combination thereof if several can apply)
* poss=p|s  
  possessor number = plural or singular
* t=pst|past|imp|fut|cond  
  tense = present, past, imperfect, future or conditional. Verb mood is not included, since it is already in the postag.

### C. Dépendances

Toutes les étiquettes possibles pour les relations de dépendance.

* \_
* a\_obj
* aff
* arg
* ato
* ats
* aux\_caus
* aux\_pass
* aux\_tps
* comp
* coord
* de\_obj
* dep
* dep\_coord
* det
* mod
* mod\_rel
* obj
* p\_obj
* ponct
* prep
* root
* sub
* suj