REPRIS

Nous devons en premier lieu revenir sur notre travail, effectué durant la première année de master, sur les titres de publications scientifiques en français. Nous avions étudié trois schémas lexico-syntaxiques fréquents dans un corpus de titres de publications scientifiques.

Un schéma est défini par une séquence de tokens qui peut comporter des choix entre plusieurs tokens **A | B**, des tokens optionnels **[A]** et des répétitions de tokens **Ai-j**. Un token de schéma peut être une classe grammaticale (nom - N, adjectif qualificatif - ADJ, préposition - P, conjonction de coordination CC…), un sous-ensemble d’une classe (nom commun - NC), un lemme (*et*) ou un signe de ponctuation comme le double point ou le point-virgule.

À une définition de schéma correspond une à plusieurs réalisations. Par exemple, le schéma défini par **[DET] NC ADJ** a pour réalisationsla séquence d’un déterminant suivi d’un nom commun suivi d’un adjectif qualificatif ou la séquence d’un nom commun suivi d’un adjectif qualificatif.

Une séquence de tokens dans un titre, mots ou signes de ponctuation, correspond à un schéma lorsque la séquence du titre se conforme à une des réalisations possibles définies par le schéma. Ainsi le titre *Villes durables et changement climatique* correspond, entre autres, aux deux schémas **NC ADJ CC NC ADJ** et **NC ADJ et NC ADJ**. Le premier schéma n’utilise que des classes grammaticales comme tokens, alors que le second utilise comme tokens quatre classes grammaticales et un lemme, **et**. La séquence *union parfaite ou mariage impossible* correspond au premier schéma mais pas au second : *ou* peut être associé à **CC** mais pas à **et**. Pour les deux schémas, leur premier token, **NC**, est associé au mot *Villes* pour le premier exemple et à *union* pour le second.

Les trois schémas étudiés dans notre travail précédent ne correspondaient à chaque fois qu’à une partie du titre et non à son ensemble : un titre était sélectionné s’il contenait une séquence correspondant à un de nos schémas, mise en gras dans les exemples qui suivent. Nous donnons ci-dessous une description et une définition pour ces trois schémas.

* Description Un double point suivi d’un un syntagme nominal dont le nom est complémenté   
   par un syntagme prépositionnel  
  Définition **\_: NC P NC\_**Exemple (1) La société face au pouvoir dans le roman arabe moderne **: la voie** **religieuse comme alternative**
* Description Un double point suivi d’un syntagme prépositionnel dont le nom est   
   complémenté par un syntagme prépositionnel  
  Définition **\_: P NC P NC\_**Exemple (2) Couper les seins des femmes **: du supplice à la monstruosité**
* Description Un double point suivi d’un syntagme nominal constitué de deux noms   
   coordonnés  
  Définition **\_: NC CC NC\_**Exemple (3) La Fée Électricité **: espoirs et craintes** de la modernité

Les définitions sont données dans leurs formes minimales. Nous ne représentons pas la possibilité d’avoir des déterminants et des adjectifs qualificatifs pour les noms de chaque schéma dans des définitions étendues de ceux-ci, mais, par souci de simplification, nous écartons ces possibilités ici.