

Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux

Intervention au séminaire du Laboratoire CLILLAC-ARP du lundi 4 février 2019

Maxime Warnier

Université Toulouse Jean Jaurès et Université de Tours

Laboratoire : Cognition, Langues, Langage, Ergonomie

L'objectif du travail que nous présenterons est d'augmenter la clarté et la précision des spécifications techniques rédigées par les ingénieurs du CNES (Centre National d'Études Spatiales) préalablement à la réalisation de systèmes spatiaux. L'importance des spécifications et des exigences qui les composent pour la réussite des projets de grande envergure est en effet désormais très largement reconnue, de même que les principaux problèmes liés à l'utilisation de la langue naturelle (ambiguïtés, flou, incomplétude) sont bien connus. Dès lors, de nombreuses solutions, plus ou moins formalisées, ont été proposées et développées pour limiter les risques d'interprétation erronée – dont les conséquences potentielles peuvent se révéler extrêmement coûteuses – lors de la rédaction des exigences, allant des langages logiques aux guides de rédaction, en passant par des outils de vérification semi-automatique.

Nous pensons que pour qu'elle soit réellement adoptée par les ingénieurs du CNES (qui ne sont actuellement pas tenus de suivre de règles de rédaction), la solution que nous nous efforçons de mettre au point se doit d'être à la fois efficace (autrement dit, elle doit limiter sensiblement le risque langagier) et aisée à mettre en place (autrement dit, elle ne doit pas bouleverser trop profondément leurs habitudes de travail, ce qui la rendrait contre-productive). Une langue contrôlée (en anglais : Controlled Natural Language), c'est-à-dire un ensemble de règles linguistiques portant sur le vocabulaire, la syntaxe et la sémantique, nous paraît être une réponse idéale à ce double besoin – pour autant qu'elle reste suffisamment proche de la langue naturelle. Or, les langues contrôlées pour la rédaction technique déjà existantes que nous avons envisagées, bien qu'élaborées par des experts du domaine, ne nous semblent pas toujours pertinentes d'un point de vue linguistique : certaines règles sont trop contraignantes, certaines ne le sont pas assez, d'autres encore ne se justifient pas vraiment.

Nous voudrions donc définir une langue contrôlée pour la rédaction des exigences en français au CNES. L'originalité de notre démarche consiste à systématiquement vérifier nos hypothèses sur un corpus d'exigences (constitué à partir d'authentiques spécifications de projets spatiaux) à l'aide de techniques et d'outils de traitement automatique du langage existants, dans l'optique de proposer un ensemble cohérent de règles (nouvelles ou

inspirées de règles plus anciennes) qui puissent ainsi être vérifiées semi-automatiquement lors de l'étape de spécification et qui, surtout, soient conformes aux pratiques de rédaction des ingénieurs du CNES. Pour cela, nous nous appuyons notamment sur l'hypothèse de l'existence d'un genre textuel, que nous tentons de prouver par une analyse quantitative, ainsi que sur les notions de normalisation et normaison. Notre méthodologie combine les approches corpus-based et corpus-driven en tenant compte à la fois des règles imposées par deux autres langues contrôlées (dont l'adéquation avec des données réelles est discutée au travers d'une analyse plus qualitative) et des résultats offerts par des outils de text mining.