

Annotation d'article de recherche - Denis Maurel

Alice Millour

8 janvier 2016

1 Matériel

Nous avons récupéré 35 articles dont Denis Maurel figure parmi les auteurs datant d'entre 1990 et 2015. Ces articles ont été transformés au format txt avec la commande unix `pdftotxt` simplement pour le moment. Les articles 1997-002053ar.pdf (Un dictionnaire relationnel des noms propres liés à la géographie consulté par transducteurs) et 2003-art16_p183_192.pdf (Elaboration d'une cascade de transducteurs pour l'extraction des noms de personnes dans les textes) n'ont pas pu être transformés correctement. 33 articles sont utilisés.

2 Identification des indices

Objectif : cherchez des indices qui permettraient de résumer les résultats présentés dans les articles.

En première analyse, nous avons recherché un certain nombre de mots-clé (grâce à `grep`) afin d'identifier les plus pertinents dans les articles concernés. le nombre de papiers dans lequel chaque terme/groupe de terme est présent au moins une fois est calculé pour chaque mot-clé. (Les lettres grecques n'ont pas été cherchées car transformées en caractères numériques)

La liste de mots clés est la suivante : [conclusion ; [0-9][.,]\\+conclusion ; [0-9][.,]\\+résultat ; conclure ; résultat ; bilan ; résumé ; récapitulatif ; montrer ; démontrer ; comparer ; compar.* résultats ; comparant avec ; évaluer ; mesure ; f-mesure ; évaluation ; mesure d'évaluation ; accord ; accord observé ; accord inter-annotateur ; inter-annotateur ; exactitude ; précision ; rappel ; silence ; bruit ; score ; fiabilité ; biais ; faussé] Les formes conjuguées des verbes ont également été recherchées.

Les termes apparaissant dans au moins dix articles sont :

conclusion ; résultat ; résumé ; montre ; mesure ; rappel ; précision ; évaluation ; accord.

Le terme « **conclusion** » apparaît dans 27 articles, 19 fois sous la forme [0-9][.,]\\+conclusion correspondant à un titre de partie. Parfois « **Conclusion** » seul est titre de partie (cf. 2005-corela-1203-hs-2-une-ontologie-multilingue-des-noms-propres.pdf, 2011-corpus-2086-10-la-traduction-des-noms-propres-une-etude-en-corpus.pdf, Corpus2011.pdf)

Nous avons regardé en détail les articles sans section conclusion pour comprendre comment les résultats étaient amenés : On trouve en particulier « **Pour conclure** », « **Nous avons montré** », Partie « **Résultats et Perspectives** ».

Le terme « **résumé** » paraît également intéressant, car en début d'article, les conclusion sont parfois évoquées et certaines occurrences (cf. article JLC2009.pdf joint) renvoient directement vers un résultat.

Identification des « **faux positifs** » :

Une approche raisonnable semble être de commencer par identifier les blocs conclusifs donnant un récapitulatif des résultats obtenus afin de ne pas identifier par exemple des extraits du type :

« Parmi les avantages se trouve la possibilité de **comparer nos résultats** sur le français oral à des travaux multilingues sur l'écrit (voir 3.2). De plus, nous avons déjà connaissance des points délicats concernant **l'accord inter-annotateur**. Un **accord inter-annotateur** assez faible pour l'écrit semble être un inconvénient »

(Judith Muzerelle, Emmanuel Schang, Jean-Yves Antoine, Iris Eshkol, Denis Maurel, et al.. Annotation en relations anaphoriques d'un corpus de discours oral spontané en français. Congrès Mondial de Linguistique Française, CMLF'2012, Jul 2013, Lyon, France. 15 pp., 2013. <hal-00788164>)

(**accord inter-annotateur** n'est d'ailleurs présent que dans un seul document donc ne paraît pas vraiment pertinent ici)

3 Travail à réaliser sur Unitex

- Comme on l'a montré plus haut, la section conclusion paraît assez aisément identifiable (séquence chiffre, point, espace, « conclusion ») ou « Conclusion » en début de phrase, de paragraphe, ou expressions introduisant la conclusion.
- Parmi les indices qui paraissent intéressants, nous avons identifié les verbes « **montrer** », « **démontrer** », « **comparer** », « **évaluer** », l'expression « **compar.* résultats** » (3 occurrences). On pourrait envisager les verbes tels que « **présenter** », « **décrire** », conjugués au passé, ou bien « **permettre** », « **constater** » comme dans le cas de l'article JLC3009.pdf joint. Il semble possible de définir une liste de verbes reliés à la présentation de résultats et de créer des graphes des différentes conjugaisons possibles grâce à Unitex : On a par exemple :

montrer 7; montre 23; est montré 2; montrons 3; avons montré 4; a montré 4; a été montré 1

évaluer 4; évalue 4; est évalué 0; évaluons 3; avons évalué 2; a évalué 0; a été évalué 0

(Les chiffres ne désignent pas le nombre d'occurrences total mais le nombre de documents dans lesquels le terme est présent)

- On pourrait également faire un graphe qui sélectionne certaines figures en fonction de leur titre (exemple « Résultat de l'évaluation », ou les tableaux dont les entrées sont les mots-clé Rappel, Précision etc ...).