

Appel à soumission pour la revue Lexique n° 27

Titre : Ressources Lexicales Électroniques

Les ressources lexicales électroniques (RLE) sont au cœur des dispositifs de traitement automatique des langues : bases de données morphologiques (ex : mots construits, mots fléchis), syntaxiques (ex : parties du discours) ou sémantiques (ex : lexiques de « polarité », lexiques d'Entités Nommées) ont longtemps été constituées de façon essentiellement manuelle, au sein de cadres descriptifs et théoriques spécifiques (ex : Lexique-Grammaire, Lexique Génératif, Théorie Sens-Texte). Au cours des dix dernières années, deux tendances ont émergé dans le domaine du traitement automatique des langues. La première est l'augmentation de la taille, et donc de la couverture des ressources, soit par l'induction plus ou moins directe à partir de giga-corpus, soit par le recours à « l'annotation par les foules » (crowdsourcing). Les modèles probabilistes d'étiquetage en parties du discours induits de corpus de référence semblent ainsi avoir définitivement remplacé les lexiques d'étiquetage ou les règles d'analyse élaborés manuellement. De son côté, le crowdsourcing apparaît comme une stratégie efficace pour développer rapidement des ressources à large couverture. Enfin, les « plongements lexicaux » (word embeddings), qui reposent sur des profils distributionnels construits à partir de corpus de plusieurs centaines de milliards de mots, sont devenus la ressource à tout faire, alors que leur fondement théorique reste flou. La deuxième tendance qui a considérablement fait évoluer le domaine est liée à une interconnexion toujours croissante des données, grâce à des formats et protocoles standardisés. Il s'agit du mouvement des « données liées » (linked data), l'un des piliers du web sémantique. Dans ce cadre, des normes de structuration des RLE, comme ISO-LMF (Lexical Markup Framework), ou de façon plus large des standards de description de ressources comme RDF (Resource Description Framework) ont été adoptés par les acteurs du traitement automatique des langues, dans une logique toujours plus poussée de décentralisation et de dématérialisation de « services » de traitement de la langue, reposant sur des infrastructures techniques complexes. Toutefois, la qualité des ressources présentes au sein de ces infrastructures s'avère difficile à évaluer, notamment lorsqu'elles agrègent des ressources constituées selon des méthodes différentes. L'objectif du numéro 27 de la revue Lexique, à paraître en décembre 2020, est de recenser, évaluer, voire critiquer les Ressources Lexicales Électroniques aujourd'hui disponibles pour le traitement automatique des langues. Sont

attendues des présentations de bases de données lexicales, règles de détection, modèles induits de corpus, et tout autre dispositif visant à alimenter un système de traitement de la langue, quelle que soit la modalité (écrit, audio, vidéo) ou la langue traitée. Une attention particulière sera portée aux ressources intégrant différents niveaux de représentation (ex. : morphologie et sémantique, phonologie et morphologie, syntaxe et sémantique, etc.), ainsi que celles destinées au traitement de langues peu dotées, de domaines de spécialité, ou visant une population spécifique (ex. : apprenants langue seconde, variantes dialectales, diglossie). Sont également attendues des évaluations critiques de ressources, notamment en ce qui concerne la qualité des informations disponibles et par conséquent la précision des analyses qu'elles sous-tendent. Les soumissions détailleront la méthodologie de constitution, de structuration et de révision des ressources présentées. Une attention particulière sera apportée à la méthodologie d'évaluation (quantitative, qualitative) de ressources, ainsi qu'à des exemples concrets d'application. Dans tous les cas, les auteurs auront à cœur de souligner le gain descriptif ou théorique de la ressource présentée, au-delà des éventuels scores « objectifs » (score inter-annotateurs, rappel/précision, BLEU, etc.) ou des aspects techniques (ex. : norme ISO LMF, RDF). En conformité avec la philosophie FAIR (Findable, Accessible, Interoperable, Reusable) et les principes d'une science ouverte, les ressources présentées, ainsi que les corpus, jeux de données, et logiciels associés (ex. : scripts, interface de consultation, service web) devront être accessibles (dépôt logiciel). Dans la même logique, les ressources, données et logiciels afférents doivent disposer d'un identifiant pérenne, obtenu auprès de services tels que : ORTOLANG (<https://www.ortolang.fr/>), HUMA-NUM (<https://www.huma-num.fr/>) ou CLARIN (<https://www.clarin.eu/>). Dans le cas de ressources sous licence commerciale, les auteurs doivent proposer un accès éventuellement limité à une période d'essai (ex. : 30 jours).

Modalités de soumission des articles

Merci d'utiliser la feuille de style fournie, voir le site de la revue *Lexique* : <https://lexique.univ-lille.fr/soumission.html>

Déclaration d'intention de soumission : merci de faire parvenir à antonio.balvet@univ-lille.fr et contact-revue-lexique@univ-lille.fr une déclaration d'intention, comprenant un titre (provisoire), la liste des auteurs, une liste de mots-clés et un résumé de 2 500 signes maximum (voir dates plus bas).

Taille des articles : entre 35 000 et 50 000 signes (espaces compris).

Langues des articles : français ou anglais.

Envoi des articles aux deux adresses suivantes : antonio.balvet@univ-lille.fr et contact-revue-lexique@univ-lille.fr

Dates importantes

Déclaration d'intention : 31/1/2020

Date limite de réception des articles : 15/4/2020

Réponse des relecteurs (double relecture anonyme) : 30/6/2020

Date d'envoi final après éventuelles modifications : 15/9/2020