

Analyse d'un corpus médical en vue de l'élaboration d'un lexique médical aspectuo-temporel

Notre communication vise à caractériser et à spécifier l'expression temporelle et, plus spécifiquement, l'usage de l'aspect dans le langage médical. En effet, nombreuses sont les études ([1], [2], [3]) qui ont démontré l'intérêt d'explorer et de repérer les entités temporelles dans le domaine médical. Cette intégration de la référence temporelle permettrait notamment d'améliorer la sécurité du patient, la prise de décisions médicales, la reconnaissance du problème et la qualité des soins prodigués. Le traitement automatique du langage (TAL) constitue l'une des solutions pour traiter ces données médicales, souvent présentes sous la forme de larges amas de données non structurées, et pour effectuer ces actions de formalisation, de rechercher et d'extraction des informations aspectuo-temporelles en vue, par la suite, d'une analyse linguistique ([4]).

Si l'expression et les relations temporelles ont déjà été largement étudiées pour la langue médicale ([5], [6], [7]), l'aspect demeure une notion bien moins explorée ([8], [9], [10]) – malgré un regain d'intérêt ces dernières décennies en linguistique générale –, entre autres car son actualisation et sa perception varient d'une langue à l'autre ou que sa nature continue à faire débat. Or, nous pensons que la durée (d'un traitement, d'une maladie, d'une hospitalisation), la fréquence (de la prise de médicaments, de rechutes, de crises d'épilepsie), l'état (accompli vs inaccompli), la phase dans le procès (début ou fin d'un traitement, d'une maladie) – autant d'informations véhiculées par la notion d'aspect – contribueraient à l'enrichissement de l'information temporelle issue de documents médicaux en exploitant le point de vue interne porté sur le procès. Notre contribution s'interprète, en ce sens, comme une amorce dans ce vaste champ qu'il reste à investiguer pour produire un système capable d'exploiter les traits syntaxiques et sémantiques des expressions aspectuo-temporelles de façon optimale pour le domaine médical.

Nous définissons et décrivons l'aspect au regard de son actualisation dans le langage médical au moyen d'une analyse exploratoire approfondie d'un corpus d'environ 225.000 textes médicaux (protocoles opératoires, rapports radiologiques, lettres de sortie...), collectés dans six services d'un hôpital bruxellois (gastroentérologie, IRM, chirurgies bariatrique et abdominale, scanners et échographies) ([11]). Notre définition tente d'une part de fédérer, dans la mesure du possible, les différentes définitions disponibles dans la littérature dont la confrontation révèle la variété des notions qu'englobe la catégorie aspectuelle ([12], [13], [14], [15]). C'est également l'occasion d'élaborer une typologie pertinente et dont la granularité répond à nos besoins pour la langue médicale. D'autre part, cette définition tient compte des spécificités du langage médical. En effet, vu la sensibilité du domaine aux notions de temps et d'aspect, il nous semble intéressant de décrire d'un point de vue aspectuo-temporel l'usage de certaines formes et tournures phrastiques particulières, qui permettent d'assimiler la langue médicale à la notion de sous-langage ([16]) et qui sont symptomatiques du domaine (« *post transfusion* », « *pancréatite aiguë* »).

Cette étude approfondie nous permet de proposer un lexique reprenant un ensemble de marqueurs temporels ou aspectuels pertinents pour le domaine médical. Nous caractérisons chaque terme de ce lexique selon des critères syntaxiques et sémantiques afin de faciliter une éventuelle tâche de désambiguïsation, par exemple au moyen de patrons et graphes d'extraction.

Nous pourrions alors résoudre des difficultés telles que l'intégration des multiples représentations temporelles, les différentes natures de la temporalité (absolue ou relative), la diversité des indices et expressions linguistiques utilisés pour exprimer le temps et l'aspect.

Cette première étape est nécessaire pour répondre à une question plus générale qui serait « La prise en compte des informations aspectuelles permet-elle d'améliorer la représentation temporelle dans le langage médical et offre-t-elle une meilleure compréhension des événements médicaux ainsi que de leurs enchaînements et relations ? »

Références

[1] Tapi Nzali, M. D., Névéol, A. et Tannier X. (2015). Analyse d'expressions temporelles dans les dossiers électroniques patients. *TALN: Traitement Automatique des Langues Naturelles*, Jun 2015, Caen, France., 22ème Conférence sur le Traitement Automatique des Langues Naturelles.

[2] Bittar, A., Amsili, P. et Denis, P. (2011). French TimeBank : un corpus de référence sur la temporalité en français. Dans : Lafourcade, M. et Prince, V. (éds.). *TALN 2011 - Traitement Automatique des Langues Naturelles (Actes de la 18e conférence sur le Traitement Automatique des Langues Naturelles)*, Juin 2011, Montpellier, France. Laboratoire d'Informatique de Robotique et de Microélectronique, 1, 259-270.

[3] Tourille, J., Ferret, O., Névéol, A. et Tannier X. (2016). Extraction de relations temporelles dans des dossiers électroniques patient. *Traitement Automatique de la Langue Naturelle – TALN (Actes de la conférence conjointe JEP-TALN-RECITAL 2016)*, volume 2, 459-466.

[4] Enjalbert, P. et Victorri, B. (2005). Chapitre 2 : Les paliers de la sémantique. Dans : Enjalbert, P. (éd.). *Sémantique et traitement automatique du langage naturel*, Cachon : Lavoisier, 56-98.

[5] Reeves, R. M. et al. (2013). Detecting temporal expressions in medical narratives, *International Journal of Medical Informatics*, 82(2), 118-127.

[6] Sun, W., Rumshisky, A. et Uzuner, O. (2013) Temporal reasoning over clinical text: the state of the art, *Journal of the American Medical Informatics Association (JAMIA)*, 1–6.

[7] Zhou, L., et Hripcsak, G. (2007). Temporal reasoning with medical data — A review with emphasis on medical natural language processing, *Journal of Biomedical Informatics*, 40(2), 183-202.

[8] Oishi, A. et Matsumoto, Y. (1997). Automatic Extraction of Aspectual Information from a Monolingual Corpus. *ACL '98 Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, Juillet 1997, Madrid, Espagne.

[9] Dorr, B. J. et Gaasterland, T. (2007). Exploiting aspectual features and connecting words for summarization-inspired temporal-relation extraction, *Information Processing and Management*, 43, 1681-1704.

- [10] Pustejovsky, *et al.* (2003). TimeML: robust specification of event and temporal expressions in text, *New Directions in Question Answering*, AAAI Press, 28-34.
- [11] *iMediate : Interoperability of Medical Data through Information extraction and Term Encoding*. Projet sous la direction de C. Fairon, 2014-2016.
- [12] Coseriu E. (1980). Aspect verbal ou aspects verbaux. Dans : David, J. et Martin. R. (éds.). *La notion d'aspect*, Actes du colloque organisé par le Centre d'analyse syntaxique de l'université de Metz, Paris : Klincksieck, 13-25.
- [13] Bittar, A. (2010). *Building a TimeBank for French. A Reference Corpus Annotated According to the ISO-TimeML Standard*. Thèse de doctorat de linguistique théorique, descriptive et automatique sous la direction de Danlos, L., Amsili, P. & Denis, P. Paris : Université Paris Diderot.
- [14] Wilmet, M. (2010). *Grammaire Critique du Français*, 5e édition, Bruxelles : Duculot.
- [15] Niklas-Salminen, A. (2012). *Le verbe*. Paris : Armand Colin (Cursus. Lettres Grammaire).
- [16] Hirschman, L., et Sager, N. (1982). Automatic Information Formatting of a Medical Sublanguage. Dans : Kittredge, R. et Lehrberger, J. *Sublanguage. Studies of Language in Restricted Semantic Domains*, Berlin – New-York : Walter de Gruyter, 27-80.