Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240354

CLARA-MeD Tool – A System to Help Patients Understand Clinical Trial Announcements and Consent Forms in Spanish

Leonardo CAMPILLOS-LLANOS[a,1] Federico ORTEGA-RIBA,[a] Ana R. TERROBA,[b] Ana VALVERDE-MATEOS,[c] and Adrián CAPLLONCH-CARRIÓN[d]
[a] *ILLA, CSIC (Spanish National Research Council), Madrid, Spain*
[b] *Fundación Rioja Salud, Logroño, Spain*
[c] *UTM, Real Academia Nacional de Medicina de España (RANME), Madrid, Spain*
[d] *Centro de Salud Mental Retiro & Hospital Gregorio Marañón, Madrid, Spain*
ORCiD ID: Leonardo Campillos-Llanos https://orcid.org/0000-0003-3040-1756
Ana R. Terroba-Reinares https://orcid.org/0000-0003-1582-6481
Ana Valverde-Mateos https://orcid.org/0000-0003-1610-0770
A. Capllonch-Carrión https://orcid.org/0000-0001-9593-8621

Abstract. We present an NLP web-based tool to help users understand consent forms (CFs) and clinical trial announcements (CTAs) in Spanish. For complex word identification, we collected: 1) a lexicon of technical terms and simplified synonyms (14 465 entries); and 2) a glossary (70 547 terms) with explanations from sources such as UMLS, the NCI dictionary, Orphadata or the FDA. For development, we extracted entities from 60 CTAs, 60 CFs and 60 patient information documents (PIDs). To prepare definitions for new terms, we used ChatGPT and experts validated them (28.99% needed to be fixed). We tested the system on 15 new CTAs, 15 CFs, and 15 PIDs, and we achieved an average F1 score of 82.91% (strict match) and of 94.65% (relaxed). The tool is available at: http://claramed.csic.es/demo.

Keywords. Natural language processing, consumer health vocabulary, Spanish, clinical trials, consent forms

1. Introduction and Background

Improving health literacy has a positive impact on patients' outcomes across a wide range of conditions [1]. Enhancing the accessibility of health texts can avoid disinformation and promote patients' engagement in procedures and care [2]. This is highly relevant at present, since patients can access their Electronic Health Records (EHRs), but these may not be understandable to them [3]. EHRs, consent forms (CFs) and clinical trial announcements (CTAs) abound with jargon, acronyms/abbreviations and medical terms, which need to be explained by clinical professionals [4]. Simplification helps patients understand such types of texts [5], also when simplified by healthcare providers [6].

[1] Corresponding author: Leonardo Campillos-Llanos, ILLA, CSIC (Spanish National Research Council), c/Albasanz 26-28, 28037 Madrid, Spain; E-mail: leonardo.campillos@csic.es.

However, practitioners are not usually trained on the how to provide easy-to-understand information [7]. Readability formulas have not proved accurate for medical texts either [8]. Therefore, a solution is complementing human expertise with automatic methods [6].

Current endeavours resort to developing consumer health vocabularies or natural language processing (NLP) to perform automatic simplification [9-11]. NLP tools have been designed to improve patients' comprehension of EHRs and journal articles, or help content-creators produce simpler and more accessible texts [12-16]. Still, most tools are available for English, with few exceptions for other languages [7,13,17,18].

Herein we present an NLP web-based tool to help users understand CTAs and CFs in Spanish. After the user inputs a text, the system highlights difficult-to-read terms; when clicking on a word, a simpler synonym or explanation is shown. We next describe the development methods (§2), report our results (§3) and discuss our outcomes (§4).

2. Methods

We designed the system considering the features of available tools for English [12-16]. We collected a lexicon and glossary for Spanish and integrated them in an NLP pipeline.

2.1. Linguistic Resources

We first collected a patient lexicon (SimpMedLexSp) gathering technical terms (including acronyms/abbreviations) and their simplified synonym or paraphrase aimed at patients. The current version has 14 465 entries and we used these sources to create it:

- Eugloss: we used the Spanish terms from this glossary of technical and popular terms prepared by the Heymans Institute of Pharmacology.[2]
- Abbreviations/acronyms and full term forms in the MedLexSp lexicon [19].
- Colloquialisms included in the Dictionary of Medical Terms by RANME [20].
- Pairs of terms from a parallel (technical/laymen) corpus [21], which were extracted using paraphrase patterns adapted from [22].

Second, we compiled a medical glossary of 70 547 entries with not-simplified definitions from reliable sources, including the FDA.[3] Most term entries come from:

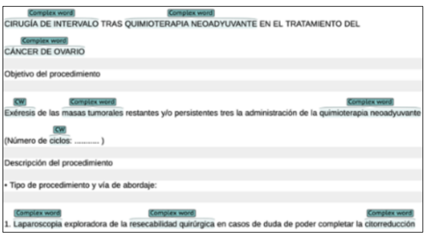
- Unified Medical Language System [23]: we extracted 43 822 term definitions.
- National Cancer Institute Dictionaries: we collected definitions from cancer-related and genetics-related terms[4] (respectively, 9181 and 251).
- Orphadata:[5] it contains definitions about rare diseases and is updated regularly (16 424 entries); we only used 2195 entries without too technical definitions.
- Dictionary of the Colombian Academy of Medicine (7025 entries).[6]

We updated the glossary with drug names from CTAs in the EudraCT clinical trials register[7] dating between 2010-2024. We then used a 60 CTAs from EudraCT, not used

[2] http://users.ugent.be/rvd-stich/eugloss/ES/lijsth.html [Accessed on 2021; the resource is not available now]
[3] www.fda.gov/patients/clinical-trials-what-patients-need-know/glosario-de-terminos [Accessed 5/3/2024]
[4] www.cancer.gov/publications/dictionaries/ [Accessed 5/3/2024]
[5] https://www.orphadata.com/ [Accessed 5/3/2024]
[6] http://www.idiomamedico.net/ [Accessed 5/3/2024]
[7] https://www.clinicaltrialsregister.eu/ [Accessed 5/3/2024]

for system development, and downloaded in January 2024 (41 080 tokens, avg. 673.76 per text). We also employed 60 CFs from organizations such as Fundación Rioja Salud and other medical societies (39 472 tokens, avg. 669.02 per text); and 60 patient information documents (PDIs) about grafts and chronic disorders (50 184 tokens, avg. 821.06).[8]

We input these 180 texts in the tool and output the recognized entities. Four co-authors of this work checked the annotations using BRAT [24] (Figure 1) to fix false positives, false negatives and scope errors (i.e., if the tool annotated multi-word entities in a wrong way: e.g., *severe pain* vs. *pain*). Two annotators revised each text, and the inter-annotator agreement (IAA) was computed using the F-measure. We thus selected the unrecognized entities from the revised annotations and included them in the glossary. We obtained definitions using OpenAI ChatGPT vs. 3.5 and this prompt: `Please, simplify or explain the following term: {}.` A medical doctor, a lexicographer and two documentalists validated the definitions. 1376 term entries were included in this way.



INTERVAL SURGERY AFTER NEOADYUVANT CHEMO-THERAPY IN OVARIAN CANCER TREATMENT

Purpose: Excision of remaining and/or persistent tumor masses after administration of neoadjuvant chemotherapy (Number of cycles:)

Type of procedure and approach: 1. Exploratory laparoscopy of surgical resectability in cases of doubt of being able to complete the cytoreduction

Figure 1. Sample of consent form revised with BRAT (left) and English translation (right).

2.2. NLP Pipeline

Firstly, SimpMedLexSp is used for complex word identification (CWI) and terms with a simple synonym for patients are given preference. Secondly, the medical glossary is used to detect the rest of terms and display term definitions. To avoid flagging too frequent words, we exclude terms above the 5000 frequency threshold using a word frequency list from the CORPES corpus.[9] A tailored exception list avoids flagging false positives (e.g., *tooth*). Figure 2 describes the workflow and Figure 3 shows a screenshot.

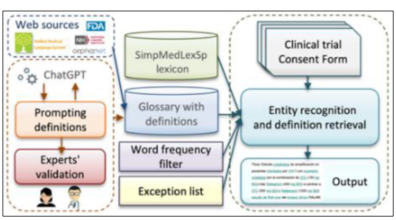


Figure 2. System components and workflow.

Figure 3. Screenshot of the current version.

2.3. Preliminary Evaluation

When fine results were obtained with the development sets (180 texts), 45 texts were used to measure the NLP system performance: 15 CTAs (6854 tokens, avg. 456.9 per

[8] www.ont.es/informacion-al-ciudadano-3/, www.saludcastillayleon.es/AulaPacientes [Accessed 2/4/2024]
[9] Available at: https://www.rae.es/corpes/ [Accessed 2/4/2024]

text), 15 CFs (9461 tokens, avg. 630.7 per text) and 15 PDIs (14 376 tokens, avg. 958.4). The evaluation metrics were precision, recall and F1-score (strict and relaxed settings).

3. Results

In the development set, we annotated 6294 entities (avg. 104.90 per text) in CTAs; 4053 entities (avg. 67.55 per text) in CFs; and 3330 entities in PIDs (avg. 55.5 per text). For all texts, the average IAA was of 84.42% (strict) and of 91.58% (relaxed), i.e., *almost perfect agreement*. After updating the lexicon, the pipeline achieved an average F1 score of 82.91% (strict) and of 94.65% (relaxed) on the held-out texts (Table 1).

Table 1. Evaluation results of the NLP pipeline on the test set (avg.: average; P: Precision; R: Recall).

	Reference entities (avg.)		Strict			Relaxed		
			P	**R**	**F1**	**P**	**R**	**F1**
CTAs (n=15)	1114 (74.27)		76.80	73.92	75.33	99.30	89.22	93.99
CFs (n=15)	851 (56.73)		92.05	91.19	91.62	98.70	95.52	97.08
PIDs (n=15)	1169 (77.93)		81.55	82.04	81.79	94.22	91.57	92.88
Total (n=45)	3134 (69.64)	**Avg.:**	83.47	82.38	82.91	97.41	92.10	94.65

In our analysis, 17.3% of errors were due to missing items in the lexicon, especially acronyms and drugs names in CTAs. Few errors were due to misspellings and tokenization (0.3%; e.g., **nacubatam, *intra abdominal*) or polysemy (0.2%; e.g., *appendix* must be flagged if it refers to a body part, not a text section). Coordinated entities (0.3%) were partly matched (e.g., *small intestine* in *large and small intestine*). The frequency filter caused 4.8% of false negatives; and the rest were mostly scope errors. A secondary outcome is the analysis of ChatGPT 3.5 for providing definitions of 1376 terms, out of which 399 needed to be amended (28.99%; that is, more than 2 out of 3 were suitable).

4. Discussion

In our data, CFs achieved higher scores, may be due to their repetitive structure and lower lexical variety. A limitation of our work is the need to regularly update the lexicons with new terms. We plan to explore transformer-based models [25] for the CWI task and detecting out-of-vocabulary items. Besides, we did not develop modules for lexical substitute generation or selection; ambiguous words or acronyms are not disambiguated (e.g., *mg* stands for 'myasthenia gravis' or 'milligram'). Moreover, our evaluation is limited: we need to test the tool with more texts (especially, EHRs), and assess the quality and correctness of the term definitions and synonyms. Evaluating the system with end-users would show its strengths and weaknesses and provide valuable feedback.

5. Conclusions

We introduced a new system to help patients understand trial protocols and consent forms in Spanish. In addition, it can assist healthcare professionals in improving the writing of these texts. Our evaluation needs to be improved with more text types (e.g. EHRs) and potential users. A demonstration system is available at: http://claramed.csic.es/demo.

Funding and Acknowledgements

CLARA-MeD project (PID2020-116001RA-C33) funded by AEI/MICIU/10.13039/50 1100011033/. We thank the Information Processing Unit (CSIC) for revising the lexicon.

References

[1] Walters R, Leslie SJ, Polson R, Cusack T, Gorely T. Establishing the efficacy of interventions to improve health literacy and health behaviours: a systematic review. BMC Public Health. 2020;20:1–17.
[2] Morrison AK, Glick A, Yin HS. Health literacy: implications for child health. Pediatr Rev. 2019;40(6):263–77, doi: 10.1542/pir.2018-0027
[3] Sørensen K, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health. 2015;25(6):1053–8, doi: 10.1093/eurpub/ckv043
[4] Donovan-Kicken E, Mackert M, Guinn TD, Tollison AC, Breckinridge B. Sources of patient uncertainty when reviewing medical disclosure and consent documentation. Patient Educ Couns. 2013;90(2):254–60
[5] Kim EJ, Kim SH. Simplification improves understanding of informed consent information in clinical trials regardless of health literacy level. Clinical Trials. 2015;12(3):232–236
[6] Choi YK, Kirchhoff K, Turner AM. Medical text simplification by medical trainees: a feasibility study. In: Proc. IEEE Int Conf Healthcare Inform (ICHI); 2016 Oct; Chicago IL: IEEE Press; 2016. p. 334–40
[7] Zilio L, et al. A lexical simplification tool for promoting health literacy. In: Proc. 1st READI2020 Workshop; 2020; Marseille, France: Association for Computational Linguistics; 2020. p. 70–6.
[8] Zeng-Treitler Q, et al. Text characteristics of clinical reports and their implications for the readability of personal health records. Stud Health Technol Inform. 2007;129.2:1117–21.
[9] Fang Y, et al. Participatory Design of a Clinical Trial Eligibility Criteria Simplification Method. Stud Health Technol Inform. 2021;281:984–88, doi: 10.1016/j.pec.2012.10.007
[10] Ondov B, Attal K, Demner-Fushman D. 2022. A survey of automated methods for biomedical text simplification. JAMIA. 2022;29(11):1976–88, doi: 10.1093/jamia/ocac149
[11] Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. JAMIA. 2008;15(4):473–83.
[12] Polepalli B, Houston T, Brandt C, Fang H, Yu H. Improving patients' electronic health record comprehension with NoteAid. Proc. MEDINFO; 2013 Aug; Copenhagen DK: IOS Press; 2013. p. 714-8
[13] Alfano M, Lenzitti B, Bosco GL, Muriana C, Piazza T, Vizzini G. Design, development and validation of a system for automatic help to medical text understanding. Int J Med Inform. 2020;138:104109
[14] Leroy G, Kauchak D, Haeger D, Spegman D. Evaluation of an online text simplification editor using manual and automated metrics for perceived and actual text difficulty. JAMIA open. 2022;5(2): ooac044.
[15] Kandula S., Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. In: Proceedings AMIA Annual Symp; 2010 Nov; Washington, DC, USA: AMIA; 2014. p. 366–70.
[16] August T, Wang L, Bragg J, Hearst MA, Head A, Lo K. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. ACM Trans Comput-Hum Interact. 2023;30(5):1–38, doi: 10.1145/3589955
[17] Koptient A, Londres M, Grabar N. Fine-Grained Simplification of Medical Documents. Stud Health Technol Inform. 2021;281:308–12, doi: 10.3233/SHTI210170
[18] Segura-Bedmar I, Núñez-Gómez L, Martínez-Fernández P, Quiroz M. Simplifying Drug Package Leaflets. In: Proceedings of the 7th Int. SMBM Conference; 2016; Potsdam, Germany; 2016. p. 20–8.
[19] Campillos-Llanos L. MedLexSp – A medical lexicon for Spanish natural language processing. J Biomed Sem. 2023;14(1):1–23, doi: 10.1186/s13326-022-00281-5
[20] RANME. Diccionario de Términos Médicos. 2011. Madrid: Panamericana.
[21] Campillos-Llanos L, Terroba-Reinares AR, Zakhir-Puig S, Valverde-Mateos A, Capllonch-Carrión A. Building a comparable corpus and a benchmark for Spanish medical text simplification. Proc Leng Nat. 2022;69(2):189–96, doi: 10.26342/2022-69-16
[22] Vydiswaran V, Mei Q, Hanauer D, Zheng K. Mining consumer health vocabulary from community-generated text. In: Proc. of AMIA Annual Symp; 2014 Nov; Washington, DC: AMIA; 2014. p. 1150–9.
[23] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic acids res. 2004;32(suppl_1):D267–D270, doi:10.1093/nar/gkh061
[24] Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. BRAT: a web-based tool for NLP-assisted text annotation. In: Proceedings 13th EACL; 2012 Apr; Avignon, France: ACL; 2012. p. 102–7.
[25] Vaswani A, et al. Attention is all you need. In: Proceedings of the 31st Int. Conf. on Neural Information Processing Systems; 2017 Dec 4–7; Long Beach, CA, USA: Curran Associates Inc. 2017. p. 6000–10.