Italian Medical Term Simplification: From Patient Information Leaflets to Simplified Language Resources

Maria Pia di Buono
University of Naples "L'Orientale"
Via Duomo, 219 - 80136 - Napoli (Italy)
mpdibuono@unior.it

Abstract

Terminological simplification in patient information leaflets (PILs) is implemented through a variety of linguistic strategies. Although these strategies help improve text comprehensibility, their overall impact remains limited. The complexity of PILs is still influenced by multiple factors, including frequent cross-referential terminology, the presence of subordinate clauses, lengthy sentences, and the use of domain-specific terms. This paper introduces I-MTS (Italian Medical Term Simplification), the first resource specifically developed for medical term simplification in Italian. I-MTS is designed to support research on lexical simplification and to facilitate the automatic adaptation of medical texts for non-expert audiences, thereby enhancing the readability and accessibility of health information in Italian.

Keywords: Medical Term Simplification, Italian Linguistic Resources, Health Literacy

1. Introduction

Promoting clear communication and facilitating access to medical information are central objectives in healthcare, as reflected in the Guideline on the Readability of the Labelling and Package Leaflet of Medicinal Products for Human Use issued by the European Medicines Agency (EMA).[1]

In line with this, also the International Plain Language Federation (IPLF) has developed guidance to enhance healthcare communication through the application of the ISO Plain Language Standard to health information.[2] Drawing on the widely adopted definition of plain language, the guide outlines a set of core principles designed to ensure that health information is relevant, findable, clear, and usable by patients and the public.

These efforts align with the broader aim of promoting health literacy (Simonds, 1974), intended as "the degree to which individuals and communities have the capacity to obtain, process, understand, evaluate, and apply information needed to make public health decisions that benefit the community" (Freedman et al., 2009). Thus, health literacy - encompassing both individual and collective dimensions - may play a central role in empowering patients to make informed decisions and actively participate in their own healthcare.

As an attempt to apply such indications to promote health literacy and improve healthcare communication, and with the aim of mitigating the use of medical jargon in patient information leaflets (PILs), simplifications are incorporated into package insert texts through a variety of linguistic strategies.

Examples[3] of these strategies include paraphrastic or explicative reformulations such as colestipol and colestiramina, described as *medicines used to reduce the intestinal absorption of fats*, as well as the use of clarifying parenthetical expressions such as *sexual desire* (libido).

Although these strategies contribute to improving text comprehensibility, their overall effect remains limited, as several factors continue to contribute to the complexity of PILs. Among these are the presence of subordinate clauses and lengthy sentences even in simplified texts, the inconsistency of term simplification throughout the text, the frequent use of cross-referential terminology - for example, *medicines belonging to the same group as* Akineton *ampoules* (anticholinergic drugs) - as well as the coexistence of domain-specific terminology.

Indeed, simplification in PILs tends to operate locally, targeting specific lexical items or short appositive phrases, while the overall lexical and syntactic complexity of the text often remains unchanged. Expressions such as quinolone antibiotics*, used to treat infections* illustrate how specialized terminology persists and how simplifications may rely on non-nominal constructions that can still challenge readers' comprehension, even within otherwise simplified contexts.

Thus, despite initiatives to promote health literacy and improve healthcare communication, PILs often remain overly complex, hindering patients' full understanding and underscoring the need for systematic approaches to identify and model simplification strategies at both the lexical and syntactic levels.

[1] EMA Guidelines - Revision 1, 12 January 2009

[2] IPLF - Clear Communication for Better Health

[3] In the examples, technical terms are underlined, and simplified expressions are shown in italics. When present, Italian examples are marked by quotation marks, and English translations are provided in parentheses.

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 182–189

To address these challenges, we introduce I-MTS (Italian Medical Term Simplification), the first resource specifically developed for the simplification of medical terminology in Italian. I-MTS has been designed to support research on lexical simplification and to facilitate the automatic adaptation of medical texts for non-expert audiences, thereby enhancing the readability and accessibility of health information in Italian.
The paper is structured as follows: Section 2 reviews related work in the field, Section 3 details the methodology employed at each stage, Section 4 presents an overview of the resulting linguistic resource, and Section 5 concludes with a discussion of the findings and their limitations.

2. Related Work

Several studies have addressed the problem of medical text simplification, exploring a range of topics and languages, from the development of simplified linguistic resources to the evaluation of simplification strategies and user comprehension. Among these, we highlight only a selection of the most relevant contributions.
The development of linguistic resources address both the identification of complex terminology and the creation of parallel data for training and evaluation to support medical text simplification. In this direction, some approaches aim to enhance interpretability by generating explicit explanations of specialized terms, for instance by leveraging semantic type information and predefined connectors to produce structured paraphrases (Kandula et al., 2010). Other work concentrates on the systematic identification of medical jargon in real-world texts, combining Natural Language Processing (NLP) models with the construction of expert-annotated datasets, such as MedJ, a corpus of over 18,000 sentences from electronic health record (EHR) notes designed to support research on jargon extraction and readability enhancement (Kwon et al., 2022). More recent efforts have moved towards the development of parallel corpora specifically tailored to simplification tasks, including resources composed of manually simplified educational materials, as Simplified Digestive Cancer (SimpleDC) (Rahman et al., 2024), to enable the investigation of advanced simplification approaches based on large language models, including fine-tuning, reinforcement learning, and prompt-based methods.
Alongside the development of linguistic resources, a growing body of work has also addressed the evaluation of medical text simplification. In this context, Devaraj et al. (2021) introduce a parallel corpus of technical and lay summaries in English, covering multiple clinical topics, and propose a novel evaluation metric based on likelihood scores derived from masked language models pretrained on scientific texts, thus providing an automatic approach to assessing simplification quality.
Complementary approaches have investigated the assessment and improvement of lexical complexity in medical texts, with a particular focus on terminology. For example, Abrahamsson et al. (2014) adapt a method for estimating word difficulty in the Swedish medical domain by combining frequency-based measures with more fine-grained features, such as substring frequency, in order to account for language-specific phenomena like compounding to identify and replace complex terms.
Similarly, other studies have explored the potential of existing consumer-oriented lexical resources, evaluating both their coverage and their suitability for simplifying specialized texts. In this line of work, Qenam et al. (2017) examine the comprehensibility and suitability of the Consumer Health Vocabulary (CHV) in terms of its ability to support the simplification of radiology reports, through analyses of lexical similarity and content alignment between technical terms and their preferred consumer equivalents.
Finally, Bala et al. (2020) investigate patient comprehension and perceptions of AI-generated simplified medical notes. Through comprehension questions and guided interviews, they assess perceived benefits and concerns. This research is closely related to the OpenNotes[4] initiative, which began as an effort to share medical notes with patients to enhance understanding and engagement. While OpenNotes has shown benefits in improving patient satisfaction and clinical outcomes, clinician concerns about potential misunderstandings and increased workload remain.

3. Methodology

Despite the limitations that PILs still present, they constitute a valuable source of term simplifications that can be leveraged to identify and extract simplified variants of technical terms for the development of a domain-specific linguistic resource.
In line with this perspective, rather than generating or manually validating simplifications, the present study adopts a data-driven methodology to extract real-world term-simplification (T-S) pairs as they naturally occur in the context of PILs.
For this purpose, we select authoritative data released under an open license to model intralinguistic equivalence between specialized terms and their simplified counterparts, and to develop a structured resource in which each T-S pair is enriched with additional metadata, such as variant forms and semantic classes.
The following subsections describe each step of the methodology in detail.

[4] https://www.opennotes.org/

3.1. Data Selection

AIFA[5], the Italian Medicines Agency, follows Open Data principles and distributes its resources under a CC-BY 4.0 license. Its datasets include pharmaceutical and "transparency lists"[6], which provide comprehensive electronic inventories of drugs with metadata such as active ingredient, packaging, and Marketing Authorisation (MA) codes. AIFA also maintains a public database[7] containing all summaries of product characteristics (SmPCs) and PILs approved by AIFA or EMA.

Using the AIFA resources, Giordano and Di Buono (2024) created the D-LeafIT corpus, formed by 1,819 Italian drug leaflets, of which 1,439 are for generic drugs and 380 for Class A drugs. The D-LeafIT corpus has served to evaluate the capability of large language models to provide reliable drug information to patients.

Subsequently, Manna et al. (2024) extracted a random subset of 600 entries to develop the Drug Information Mining in Italian (DIMMI) dataset as part of CALAMITA (Attanasio et al., 2024; Nissim et al., 2025), a shared task - sourced by a community effort - to test the LM abilities in Italian, encompassing several different type of tasks.

Besides, the commercial drug names and their correspondent PILs raw text, DIMMI contains a manually annotated gold standard (GS) designed for the shared-task evaluation. For each of the drug entry, the DIMMI GS provides a list of entities present in the related PIL, covering five classes: i) molecule; ii) dosage; iii) drug interaction; iv) usage; and v) side effects.

The DIMMI GS annotation was carried out by three annotators, following detailed guidelines that primarily addressed span detection rather than category assignment, the latter being relatively straightforward. The final inter-annotator agreement (IAA) reported in the paper indicates a good level of agreement across all categories.

In particular, we focus on three of the five categories annotated for each entry in DIMMI GS: drug interaction, side effects, and usage. This choice is motivated by two considerations. First, as reported by the authors, these categories achieved the highest IAA scores in the annotation, measured as average pairwise F1 for usage (0.68) and token-level agreement percentage for drug interaction (0.66) and side effect (0.78). Second, a randomized manual inspection of the data indicates that these three categories exhibit the highest frequency of terminological simplifications in PILs. The total number of DIMMI GS not-unique entities for the three selected categories are 49,016 (Table 1).

[5] AIFA - Homepage

[6] AIFA - Pharmaceutical and Transparency Lists

[7] AIFA - Banca Dati Farmaci

3.2. Intralinguistic Equivalence Modelling

Specifically designed to evaluate the ability of language models to answer targeted questions about drugs in Italian extracting information from PILs, the DIMMI GS annotation does not provide an explicit distinction between technical terms and their simplified variants.

Furthermore, as information in PILs are presented with an uneven level of details, different annotation choices regarding the levels of conceptual granularity across categories have been adopted in DIMMI (Manna et al., 2024).

In general, the minimal-span principle and the most-specific-concept principle are applied. For example, in "meningite criptococcica (*un'infezione micotica del cervello*)" (cryptococcal meningitis (*a fungal infection of the brain*)), only the most specific term, meningite criptococcica, is annotated.

In other cases, when only generic conceptual entities are presented, e.g., "*antibiotici*" (*antibiotics*), the information-preservation principle is applied.

As a result, the list of entities in the DIMMI GS may include both specialized terms and their simplified counterparts, annotated as two distinct entities. This implies that the intralinguistic equivalence between a term and its corresponding simplification, when present in a PIL, is not fully preserved.

To restore the equivalence between specialized terms and their simplified counterparts, we adopt a methodology similar to the one proposed by Speranza et al. (2021), which relies on the use of appositives to extract term simplifications in the Cultural Heritage domain.

Due to their semi-fixed syntactic patterns and high degree of semantic transparency, appositive constructions can be used as reliable linguistic indicators of simplified language (Speranza et al., 2021). Syntactic constructions of appositives, usually marked with the presence of brackets, commas, and hyphens, can be systematically described and exploited to identify the term and its corresponding simplified reference.

Although the apposition-based extraction method could, in principle, be applied to raw texts independently of a predefined list of entities, we deliberately anchor the process to the entities already annotated in the DIMMI GS. This choice allows us to (i) preserve and fully exploit the manual annotations available in DIMMI; (ii) reduce the potential noise in the apposition-based extraction of simplifications; (iii) ensure methodological consistency and comparability with the existing resource.

Apposition Extraction Before proceeding with apposition extraction, we examined the target entities in the DIMMI GS to identify entries that already contained parenthetical content. These cases re-

flect annotation variability, specifically instances in which entities were not annotated according to the minimal-span and most-specific-concept guidelines. Such variability can be attributed to the aforementioned challenges in span definition, which are also reflected in the IAA scores. We found that the annotated entities presenting a parenthetical content are about 19,90% of the total annotated entities, with a variability among categories (see Table 1). These entities undergo a cleaning procedure aimed at identifying and removing potential annotation inconsistencies. We evaluate them according to two criteria: (i) the entry must not begin with a punctuation mark; and (ii) its syntactic structure must correspond to that of a well-formed noun phrase in Italian. The former is implemented through regular expressions designed to detect leading punctuation characters and ten entries are removed. The latter is assessed via automatic part-of-speech (POS) tagging, by verifying that the internal structure of each span conforms to a well-formed Italian noun phrase pattern. Entries beginning with a preposition - an invalid structure for Italian noun phrases - are excluded (157 entities).

Violations of these criteria are interpreted as evidence of incomplete span annotation, suggesting that relevant elements of the entity are missing. Since such cases may compromise the reliability and consistency of subsequent analyses, these entities are excluded from further processing. The total number of retained entities is presented in Table 1.

For DIMMI GS entities without an annotated parenthetical content, concordance searches were conducted in the DIMMI PIL corpus using a context window that dynamically adapted to each occurrence. The inclusion of an entity and its surrounding content was guided by the presence or absence of parentheses, ensuring that the context captured was appropriate for the structure of each instance. In the absence of specific guidelines governing the simplification process, intralinguistic equivalence in PILs may be realized through appositive structures that vary in the order of their components and/or in the markers used to signal simplification.

Category	+ Par. [Ret.]	- Par.	Total
Drug Int.	979 [834]	9,690	10,669
Side Eff.	8,268 [7,149]	27,985	36,253
Usage	508 [414]	1,586	2,094
Total	**9,755 [8,397]**	**39,261**	**49,016**

Table 1: Number of DIMMI entities with (+ Par.) parenthetical content total and retained [Ret.] and without (- Par.) parenthetical content per each category

For instance, an equivalence between the term "acufene" (tinnitus) and the simplification "*ronzio alle orecchie*" (*ringing in the ears*) can be realized differently within different texts, as in the examples below.

"*ronzio alle orecchie* (acufene)"

"acufene (*ronzio alle orecchie*)"

To account for such structural variation in appositive constructions, we employ regular expressions to model the possible combinatorial patterns involving the targeted DIMMI entities.

Since not all PILs provide simplifications for every term, some of the entities in the selected class from the DIMMI GS may not belong to a T-S pair within a given text. To exclude these standalone entities from the analysis, entries that did not display markers of appositive constructions in their context were removed.

By verifying the presence of appositive structures, we were able to identify entities that qualify as candidates for inclusion in a T-S pair and to extract them, together with their appositive counterparts, from the unstructured texts in which they occur.

The total numbers of T-S pairs resulting from this step is presented in Table 2.

Pair Filtering The extracted appositive components do not always instantiate an intralinguistic equivalence by providing a simplified version of the target term. Instead, they may realize other types of semantic relations.

For example, in "traumi (ad esempio causati dallo sport)" [injuries (e.g., sports-related)], the parenthetical content has an illustrative function and does not introduce a semantic equivalence through a distinct, reusable concept. As such cases do not qualify as T-S pairs, candidate entities have been filtered on the basis of the semantic relations they encode.

Given that medical language is highly specialized and is characterized by recurrent structural patterns and domain-specific lexical usage, we define some selection criteria to filter out appositive elements that do not correspond to term simplifications from the list of candidate pairs.

Syntactic and lexical constraints are applied to identify the semantic relation existing between the term and the second component in the candidate pairs. In particular, we exclude appositive-construction pairs in which the parenthetical segment encodes relations other than intralinguistic equivalence. These cases include:

- Coordinated structures introduced by simple coordinating conjunctions: copulative ("e", and, "né", nor), expressing additive coordination; adversative ("ma", but), expressing contrast; disjunctive ("o", or), expressing alternation or exclusion; as well as complex conjunctions (e.g., "anche se", even if).

- Non-complement subordinate clauses (i.e., adverbial or peripheral clauses), which function as clause-level adjuncts and encode semantic relations external to the argument structure of the main clause. These include: (a) temporal ("quando / mentre / finché / dopo che / prima che"); (b) causal ("perché 7 poiché / siccome"); (c) purpose ("perché / affinché"); (d) concessive ("sebbene / benché"); (e) conditional ("se / qualora / purché"); (f) result/consecutive ("così. . . che / a tal punto che"); (g) comparative "(più/meno [. . .] che, tanto [. . .] quanto"); (h) limitative/exceptive ("a meno che / per quanto").
- Adverbial phrases introduced by exemplificatory markers such as "per esempio" or "ad esempio" (for example), which signal illustrative expansion rather than reformulation or equivalence.
- Prepositional phrases functioning as clausal- or event-level adjuncts. However, when such phrases contribute to intra-domain explicitation rather than circumstantial modification - by encoding conceptual specification, referential restriction, or therapeutic indication (e.g., anatomical specification or functional purpose) - they are retained. These phrases are typically signalled in Italian through the use of "per" and articulated forms of "di" (e.g., "del / della / dello").

We retain prepositional phrases when they instantiate domain-conventionalized relational patterns between ontological classes (e.g., drug interaction, usage). In such cases, the phrase encodes concept-level relational specification (e.g., therapeutic indication, anatomical localization) rather than event-level circumstantial modification.
The inclusion criterion is therefore based on the semantic class of the head noun and the ontological type of the complement, rather than on the preposition per se.
The criteria emerged from iterative corpus analysis and were subsequently operationalized as rule-based filters, combining data-driven generalizations with independently established morphosyntactic constraints of Italian. The number of pairs retained after the filtering phase is presented in Table 2.

Term-Simplification Alignment With the exception of retained prepositional phrases, which transparently encode simplification directionality, structural variation in appositive constructions prevents automatic identification of term *vs*. simplification within T-S pairs. Intralinguistic equivalence does not by itself determine role assignment. A dedicated reconciliation step is therefore introduced to align pair components and resolve directionality.
To differentiate entities representing technical terms from those representing simplifications, ensuring terminological validity and avoiding arbitrary term selection, T-S pair component were cross-checked against an external authoritative clinical reference (the Italian MSD Manual - Professional version[8]). The MSD Manual provides several resources including tables that report information about diseases and other topics. The MSD section headings and table titles were consulted for coverage verification[9]. These labels were subsequently normalized, lemmatized, and stripped of descriptive modifiers before comparison with our data. The resulting list constitutes our internal lexical benchmark suitable to identify the technical term within each T-S pair.
A total of 1005 headings were examined, of which 589 corresponded to identifiable disease-related terms after normalization.
To account for the other entries in our T-S pair list we use the Wikipedia list for the "Farmaci secondo la funzione" (medication classified for function) category[10]. The category tree, comprising 162 unique terms, was retrieved and subsequently applied to verify our entries. To identify further terms related to anatomy concepts, we relied on an external encyclopedic list available online, which is one of the few such resources and originates from a private hospital website[11].
The pair items that we were not able to check against an authorative source have been dismissed.
The number of pairs retained after the T-S alignment phase is presented in Table 2.

Data	Drug_Int.	Side_Eff.	Usage	Tot.
DIMMI	10,669	36,253	2,094	49,016
AE	3,122	9,851	691	13,609
PF	1,952	5,275	477	7,704
TSA	960	2,502	263	3,726

Table 2: Number of entities for the selected categories presented for each step of the intra-linguistic equivalence modelling: Apposition Extraction (AE); Pair Filtering (PF); Term-Simplification Alignment (TSA)

[8] Manuale MSD - Versione per i professionisti
[9] For the sake of methodological transparency, no expressive or structural elements of the source were reused, nor was any portion of the original database reproduced or redistributed. The resulting resource relies exclusively on independently curated data, enriched with original metadata and classification criteria.
[10] Wiki category list
[11] Gavazzeni website

3.3. Final Reconciliation

After identifying the technical term within each T-S pair using an external resources, all entries were processed through a final reconciliation procedure to systematically reconcile morphological, orthographic, and simplification variations while preserving functional distinctions.
First, we investigated instances of partial simplifications. By partial simplification, we refer to cases in which the simplification is only partially realized, with the noun phrase head expected in the simplified form being absent. As a result, partial simplifications are typically realized as prepositional phrases (e.g., *(allo stomaco)*) or adjective phrases rather than full noun phrases. As final step of the final reconciliation, prepositional- and adjective-phrase simplifications were completed where possible, allowing for a consistent representation of simplifications while retaining distinctions critical to downstream analyses.
Then, inflection-aware consolidation was performed: singular and plural forms were initially retained to capture maximal surface form variability and to account for differences in the paired simplification component. Lemma-level grouping was then applied to reduce purely inflectional redundancy without collapsing semantically or functionally distinct TS mappings.
Next, a surface-level normalization step addressed orthographic and capitalization variants, ensuring harmonization of superficial differences.
A final grouping and cleaning procedures are performed to remove duplicates, irrelevant entries, and formatting inconsistencies.
This multi-step procedure ensured that morphological, orthographic, and simplification variations were systematically reconciled, while maintaining the structural and functional integrity of each T–S pair. The number of pairs retained after the final reconciliation phase is presented in Table 3.

3.4. Semantic Classification

Finally, semantic classification organizes the terms into predefined categories, facilitating downstream analysis and evaluation. This systematic post-processing ensures the quality, accuracy, and usability of the extracted data for subsequent tasks. DIMMI annotation labels are directly derived from the titles of PIL sections. To support a more general and reusable annotation scheme for NLP and language modeling in Italian, we map these labels onto a smaller set of more general semantic classes, using the classification derived from the term control lists. Specifically, most of the entries classified as drug interaction in DIMMI GS are retrieved from the Wiki control list, as this section predominantly refers to interactions between drugs[12], and mapped to `medication`.

Similarly, entries in the usage class were retrieved from the disease control list and mapped to `disease`, as the corresponding section typically describes the medical conditions for which the drug is prescribed. Finally, we introduce the general class `body_part`, derived from the anatomy concept list. Unlike other mappings, this class is not tied to a single PIL section, as references to anatomical entities occur across different sections, most notably in side effects.
The introduction of this class reflects a shift from section-based labels to concept-based categories: entities denoting anatomical structures are grouped together regardless of their textual location, ensuring greater semantic consistency across the dataset.

4. I-MTS Overview

I-MTS (Italian Medical Term Simplification) is the first open-data resource specifically designed for the simplification of medical terminology in Italian[13]. It has been developed to support research on lexical simplification and to facilitate the automatic adaptation of medical texts for non-expert audiences, thereby improving the readability and accessibility of health information. The resource comprises 1356 T-S pairs in total[14] (Table 3).

Category	**# Entries**
Disease	956
Medication	312
Body_Part	88
Total	**1356**

Table 3: Number of I-MTS pairs for Disease, Medication, Body Part Classes

Each entry in I-MTS is structured to include different types of lexical and semantic information (Table 4). For each entry, the resource includes: (i) a classification into three classes (CLASS) (i.e., Disease, Medication, and Body Part); (ii) available term variants (VAR), such as orthographic variations (e.g., "sindrome di Stevens Johnson" and "sindrome di Stevens-Johnson") or linguistic variants (e.g., "infarto del miocardio" and "infarto miocardico"), and linguistic borrowings or Latin-based forms, e.g., "miastenia gravis"; (iii) hierarchical subtypes (TYPE

[12]While drugs may also interact with nutraceuticals, such cases are relatively rare in the DIMMI GS.

[13]https://github.com/banshee44/I-MTS

[14]It is worth stressing that the total number of final entries exceeds that of the control lists. This reflects the fact that our resource captures multiple contextualized or variant forms of a concept, whereas the control lists provide a more limited, canonical representation.

CLASS	TERM	TYPE OF	VAR	SIMPL.
Disease	Ulcera	N/A	N/A	Lesione
Disease	Ulcera esofagea	Ulcera	N/A	Lesione della parete interna dell'esofago
Medication	Inibitori della cox-2	Inibitori	N/A	Medicinali che riducono l'infiammazione
Body Part	Uretra	N/A	N/A	Tubo che trasporta l'urina dalla vescica

Table 4: Examples of I-MTS entries

OF) are specified. These are identified through the partial overlap of multi-word terms with single terms, when this involves the head of the compound, e.g., "ulcera esofagea" and "ulcera boccale" are classified as subtypes of "ulcera".
Finally, all corresponding simplifications for each term are provided, resulting from the consolidation of the different variants attested across PILs, where the same term may be simplified in different ways.

5. Conclusions and Future Work

The current version of I-MTS represents an initial step toward the development of a richer and more informative resource for the simplification of medical terminology in Italian. Given the breadth of the medical domain, further efforts will be required to extend coverage to specific subdomains. Nevertheless, even in its present form, the resource can contribute to improving patients' health literacy.
At the same time, the resource remains preliminary and data-driven, as it is based on the systematic collection of existing simplifications rather than on validated or experimentally tested strategies. No quantitative measures of readability or comprehension have been computed, nor have user studies or formal validations against a gold standard been conducted. In addition, the dataset does not include measures of annotation consistency or inter-annotator agreement, and it does not aim to exhaustively capture all relevant linguistic phenomena. As such, it should be regarded as an exploratory and illustrative resource.
Despite these limitations, the dataset provides a useful foundation for future research. Although the present study does not include a systematic analysis of simplification strategies, the collected data can support the identification of recurring patterns, as well as the training, evaluation, and fine-tuning of language models for domain-specific text simplification. By aggregating and organizing real-world simplifications, I-MTS offers a data-driven basis for the development of more consistent and potentially consensus-driven approaches to medical text simplification.
Future work will focus on expanding the resource through the inclusion of additional patient information leaflets, formalizing annotation layers aligned with standardized classification systems to improve interoperability, and providing a systematic description of the simplification strategies employed. These efforts aim to enhance the resource's utility for both NLP applications and the development of consistent, evidence-based approaches to medical text simplification.

6. Acknowledgements

This research was carried out within the 4MLP Project, funded by the University of Naples "L'Orientale" under the University Research Projects funding scheme (PRA 2025) - CUP C66125002860005.

7. Bibliographical References

Emil Abrahamsson, Timothy Forni, Maria Skeppstedt, and Maria Kvist. 2014. Medical text simplification using synonym replacement: Adapting assessment of word difficulty to a compounding language. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 57–65.

Giuseppe Attanasio, Pierpaolo Basile, Federico Borazio, Danilo Croce, Maria Francis, Jacopo Gili, Elio Musacchio, Malvina Nissim, Viviana Patti, Matteo Rinaldi, et al. 2024. Calamita: Challenge the abilities of language models in italian. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*, pages 1054–1063.

Sandeep Bala, Angela Keniston, and Marisha Burden. 2020. Patient perception of plain-language medical notes generated using artificial intelligence software: pilot mixed-methods study. *JMIR formative research*, 4(6):e16670.

Ashwin Devaraj, Byron C Wallace, Iain J Marshall, and Junyi Jessy Li. 2021. Paragraph-level simplification of medical texts. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2021, page 4972.

Darcy A Freedman, Kimberly D Bess, Holly A Tucker, David L Boyd, Arleen M Tuchman, and Kenneth A Wallston. 2009. Public health literacy

defined. *American journal of preventive medicine*, 36(5):446–451.

Luca Giordano and Maria Pia Di Buono. 2024. Large language models as drug information providers for patients. In *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health)@ LREC-COLING 2024*, pages 54–63.

Sasikiran Kandula, Dorothy Curtis, and Qing Zeng-Treitler. 2010. A semantic and syntactic text simplification tool for health content. In *AMIA annual symposium proceedings*, volume 2010, page 366.

Sunjae Kwon, Zonghai Yao, Harmon S Jordan, David A Levy, Brian Corner, and Hong Yu. 2022. Medjex: A medical jargon extraction model with wiki's hyperlink span and contextualized masked language model score. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2022, page 11733.

Raffaele Manna, Maria Pia Di Buono, and Luca Giordano. 2024. Dimmi-drug information mining in italian: A calamita challenge. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)*, pages 1144–1152.

Malvina Nissim, Danilo Croce, Viviana Patti, Pierpaolo Basile, Giuseppe Attanasio, Elio Musacchio, Matteo Rinaldi, Federico Borazio, Maria Francis, Jacopo Gili, et al. 2025. Challenging the abilities of large language models in italian: a community initiative. *arXiv preprint arXiv:2512.04759*.

Basel Qenam, Tae Youn Kim, Mark J Carroll, and Michael Hogarth. 2017. Text simplification using consumer health vocabulary to generate patient-centered radiology reporting: translation and evaluation. *Journal of medical Internet research*, 19(12):e417.

Md Mushfiqur Rahman, Mohammad Sabik Irbaz, Kai North, Michelle S Williams, Marcos Zampieri, and Kevin Lybarger. 2024. Health text simplification: An annotated corpus for digestive cancer education and novel strategies for reinforcement learning. *Journal of Biomedical Informatics*, 158:104727.

Scott K Simonds. 1974. Health education as social policy. *Health education monographs*, 2(1_suppl):1–10.

Giulia Speranza, Maria Pia Di Buono, and Johanna Monti. 2021. Terms and appositions: What unstructured texts tell us. In *International Conference on Automatic Processing of Natural-Language Electronic Texts with NooJ*, pages 219–230. Springer.