Temporal Structure in Clinical Narratives in Portuguese: Insights from Cross-Document Annotation

Ana Luísa Fernandes[*‡§], **Purificação Silvano**[*‡§] **Luís Filipe Cunha**[*†]
[*] Institute for Systems and Computer Engineering, Technology and Science (INESC TEC)
[‡] Faculty of Arts and Humanities, University of Porto
[†] Faculty of Sciences, University of Porto
[§] Centre for Linguistics of the University of Porto (CLUP)
Porto, Portugal
{ana.l.fernandes, purificacao.silvano, luis.f.cunha}@inesctec.pt

Abstract

Medical reports, by documenting disease progression and patient responses to treatment, form continuous narratives in which each new document adds a chapter to the patient's clinical story. Constructing coherent patient timelines requires identifying temporal relations across multiple medical reports that compose a patient's clinical journey. However, cross-document temporal annotation remains an underexplored area, largely due to the methodological and conceptual challenges it entails. This study addresses these challenges by investigating the identification and characterization of cross-document temporal relations in Portuguese medical records. For this purpose, cross-document annotation was performed on different types of reports (Group Consultation Reports, Discharge Reports, and General Reports) from patients diagnosed with Acute Myeloid Leukemia and followed at IPO-Porto, Portugal. Annotation was carried out using the Med2Story scheme, specifically designed to capture both temporal and medical information. Our results indicate that, although cross-document annotation of temporal information is more demanding in terms of both the annotation scheme and the annotation process, it enables the construction of coherent chronological representations of patients' clinical journeys. Furthermore, the analysis reveals key characteristics of these clinical narratives, including the predominance of nominal events and the prevalence of simultaneity as the most frequent temporal relation type.

Keywords: clinical narratives, temporal annotation, cross-document annotation, Electronic Health Records

1. Introduction

Information extraction from Electronic Health Records (EHRs) remains a particularly challenging task. These records contain extensive amounts of unstructured narrative text, which introduces major challenges for data organization, curation, management, and effective reuse in both clinical and research settings. Natural Language Processing (NLP) methods play an essential role in automating the retrieval, processing, and extraction of relevant biomedical information from these records (Irrera et al., 2024).

High-quality NLP methods require manually annotated clinical data for their development. However, the development of high-quality annotated corpora remains especially demanding in under-resourced languages such as Portuguese. This limitation is particularly evident in the study of diseases with poor prognosis, such as Acute Myeloid Leukemia (AML). AML is a malignancy of the bone marrow stem cells, typically associated with an unfavorable outcome (Shimony et al., 2023).

In such cases, understanding a patient's clinical trajectory depends on reconstructing the temporal sequence of events documented across multiple medical reports. In this context, cross-document temporal analysis, aimed at generating chronologically coherent patient timelines, can yield significant contributions to temporal semantics and, simultaneously, to clinical NLP applications. This process involves identifying and labeling events, capturing their morphosyntactic and semantic features, and establishing temporal relations between events that occur across different documents. It is, however, a demanding and labor-intensive task (Zhao et al., 2023), requiring fine-grained interpretation and consistent annotation across heterogeneous clinical narratives.

Although several studies have addressed temporal annotation in clinical narratives, these efforts focus primarily on the temporal annotation of single documents rather than inter-document relations (Sun et al., 2013b; Galescu and Blaylock, 2012). To the best of our knowledge, only Raghavan et al. (2014) have proposed a model for cross-narrative temporal ordering of medical events, although they did not present a systematic annotation scheme for this purpose. Some efforts have been made in other domains, such as Minard et al. (2015), whose Cross-Document Event Ordering Task aimed to build coherent timelines from English news articles.

As a matter of fact, the cross-document annotated datasets of clinical narratives are scarce as the annotation is a complex and cognitively demanding process. Clinical events are often dis-

Proceedings of the 8th Workshop on Clinical Natural Language Processing (Clinical NLP) @ LREC 2026, pages 300–312

tributed across multiple documents produced at different times and by different healthcare professionals, leading to inconsistencies in terminology, narrative structure and level of detail. Moreover, clinical text differs substantially from non-clinical text due to its domain-specific language and the frequent use of abbreviations, which significantly increases the difficulty of processing data (Moharasan and Ho, 2019). Another challenge arises from the high heterogeneity of EHRs, as their content and writing style vary considerably across hospitals (Zhu et al., 2023), and even across different services and departments within the same institution. Consequently, reconstructing a patient's clinical timeline requires identifying and aligning events across heterogeneous sources while preserving temporal coherence. The present study aims to demonstrate how the challenges of cross-document annotation can be addressed in order to capture the temporal organization of medical reports. Specifically, it seeks to: (1) Develop and apply methodological strategies to overcome the main difficulties associated with cross-document temporal annotation in clinical texts; (2) Identify and describe cross-document temporal properties in medical records written in Portuguese for nine patients diagnosed with Acute Myeloid Leukemia at the Portuguese Institute of Oncology of Porto (IPO-Porto).

The paper is organized as follows. Section 2 presents the related work, contextualizing previous studies and resources relevant to this research. Section 3 describes the corpus selected for annotation, the annotation scheme (3.1), and the annotation methodology adopted (3.2). Finally, Section 4 reports the qualitative and quantitative results, focusing on cross-document annotation, but also presenting results related to intra-document analysis. Finally, Section 5 summarizes the main findings of the paper and outlines directions for future work.

2. Related Work

Medical reports, by documenting the progression of disease and the patient's response to treatment, constitute continuous narratives in which each new document adds a chapter to the patient's clinical story. To make informed decisions, physicians frequently consult previous reports to reconstruct the patient's clinical trajectory. Although NLP in medicine has received increasing attention, subfields such as anonymization and the extraction of clinical concepts (e.g., pathologies or medications) have received far more focus than the analysis of temporal structures (Sun et al., 2013b). The i2b2 project (Sun et al., 2013a,b) was pioneering in making available the first fully anonymized clinical corpus annotated with temporal information. Subsequently, Bethard et al. (2015); Bethard and Parker (2016); Bethard et al. (2017) organized the Clinical TempEval challenges, which used the THYME corpus (Styler IV et al., 2014). In parallel, other studies adapted TimeML (Pustejovsky et al., 2003) to the clinical domain, including Galescu and Blaylock (2012) and Savova et al. (2009), who proposed versions tailored to the specificities of medical narratives. The CLEF project (Roberts et al., 2007) annotated 167 clinical records, focusing on intraphase relations and event–document creation time relations. More recently, the 2024 Shared Task on Chemotherapy Treatment Timeline Extraction introduced patient-level timelines as a novel challenge, involving tasks such as extracting relevant events, temporal expressions, and temporal relations from each document, and subsequently summarizing this information across documents (Yao et al., 2024).

In non-English-speaking contexts, important contributions include Campillos-Llanos et al. (2018), who developed the MERLOT corpus for French, and Campillos-Llanos et al. (2025), who introduced MEDSPANER, a medical and temporal named entity recognizer for Spanish, based on the TimeML framework and the HeidelTime system.

Despite the significant progress in intra-document temporal annotation, cross-document temporal annotation remains largely underexplored. Recent work has addressed event coreference resolution (Wu et al., 2020; Held et al., 2021; Eirew et al., 2021; Chen et al., 2025) and cross-document relation extraction (Yao et al., 2024; Jain et al., 2024). Gao et al. (2024a) proposed a cross-document coreference resolution model based on discourse information, integrating structural and semantic document representations through Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) and lexical chains. Within the SemEval-2015 Task 4, Minard et al. (2015) focused on cross-document event coreference resolution and cross-document temporal relation extraction to construct timelines from English news articles. Caselli and Vossen (2017) aimed to extract and classify events relevant to narratives, drawing from news documents distributed over time and clustered around a single seminal event or topic. More recently, Gao et al. (2024b) developed a large-scale, domain-agnostic dataset for cross-document event extraction. While recent studies have expanded the scope of cross-document information extraction, several persistent challenges continue to hinder progress in temporal annotation across documents. The task imposes a substantial cognitive burden on annotators, who must read and interpret multiple texts, retain fine-grained details, and determine whether different mentions refer to the same underlying event (Song et al., 2018). This process is particularly demanding when descriptions evolve over time or vary across sources, a common occurrence in clinical narratives, where similar procedures may be documented in distinct reports. Another obstacle arises from the absence of standardized guidelines for identifying and linking events across documents (Girardi et al., 2014), which undermines consistency and interoperability.

Building on the gaps identified in the literature, the present study aims to address the lack of research on cross-document temporal relations in clinical contexts. By doing so, this research contributes to advancing temporal annotation studies in the medical domain and to deepening the understanding of the temporal organization and narrative structure of Portuguese clinical records.

3. Dataset

The dataset to be annotated was extracted from a larger collection of 906 medical reports (MRs) written in European Portuguese, referring to 93 patients diagnosed with AML and followed at the IPO-Porto, Portugal. Access to these documents was authorized by the IPO-Porto Ethics Committee, framing the project within the institution's data management plan (Rb-Silva and Karimova, 2021).

The corpus included three types of reports: (1) Group Consultation Reports – MRs resulting from group consultations in which multiple specialists review the patient's clinical history as well as their clinical and analytical condition; (2) Discharge Reports – MRs documenting the patient's clinical evolution from hospital admission to medical discharge; (3) General Reports – MRs describing the patient's clinical trajectory up to the moment of writing, often requested to justify medical leave.

The examples presented in this paper correspond to adaptations of real sentences extracted from the MRs, which were previously pseudonymized to ensure that patient identification is impossible.

The creation of this dataset followed a set of predefined selection criteria to ensure consistency and representativeness. First, a maximum of 40 reports was set, deemed sufficient to ensure feasibility without compromising representativeness. Second, the inclusion of all three types of reports was ensured to guarantee typological diversity. Third, patients with more than two and fewer than five reports were selected, allowing for cross-document annotation while keeping homogeneity across patients. Finally, preference was given to documents with a narrative character, excluding those limited to laboratory results or complementary examinations.

Applying these criteria resulted in a subcorpus of 38 reports from 9 patients: 19 corresponding to Group Consultations, 16 to Discharges, and 3 to General Reports. The average document length was 157 words (standard deviation: 104). Within each patient's record, reports were organized and annotated in chronological order of document creation. A detailed distribution of report types per patient and their respective lengths is provided in the project's GitHub Repository[1].

3.1. Med2Story Annotation Scheme

The annotation scheme was designed to capture morphosyntactic, semantic, and domain-specific medical information, addressing the challenge of representing the temporal structure of a patient's clinical journey as described across multiple types of medical reports. The development of the annotation scheme was conducted in four main phases: (1) Foundational preparation, in which three existing annotation schemes were compared to identify their respective strengths and weaknesses, with the aim of replicating the strengths and avoiding the weaknesses; (2) Design and specification, which involved improving an existing temporal scheme and adding domain-specific medical tags; (3) Empirical validation, through a pilot annotation campaign involving multiple annotators; and (4) Consolidation and refinement, during which the scheme was further improved based on the validation results. The detailed methodology adopted in the design of the scheme is described in Fernandes et al. (2025b)

The annotation scheme is an extension of the *Text2Story* scheme (Silvano et al., 2021; Leal et al., 2022), developed on the basis of ISO 24617-1: 2012 . The scheme allows for the annotation of two types of structures Entity Structure and Link Structure, enabling the annotation of Events, Temporal Expressions, and Temporal Relations between Events, and between Events and Temporal Expressions.

Within the category of Events, a distinction is made between *General Events* and *Specialized Events*. General Events retain the original *Text2Story* attributes (namely *class*, *type*, *tense*, *aspect*, *polarity*, *vform*, *modality*, and *POS*) which are suitable for the linguistic description of clinical narratives. In contrast, Specialized Events incorporate domain-specific attributes tailored to the medical context, allowing for a more fine-grained and semantically enriched representation of clinical information. For example, in the sentence *"O hemograma mostrou leucocitose"* [*The blood count showed leukocytosis*], the events *"hemograma"* [*blood count*] and *"leucocitose"* [*leukocytosis*] would be annotated as Special-

[1] https://github.com/analuisacardosofernandes/Temporal-Structure-in-Clinical-Narratives

ized Events, whereas *"mostrou"* [*showed*] would be annotated as a General Event. Further details can be found in Fernandes et al. (2025a).

Figure 1 presents an overview of the annotation scheme. A detailed description of the scheme and

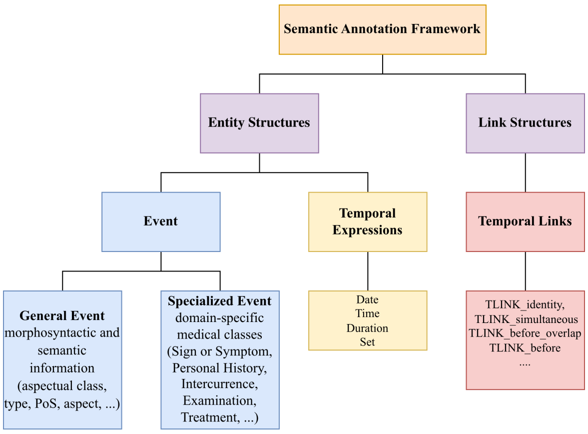

Figure 1: Annotation Scheme

its corresponding guidelines are available in the GitHub Repository.

3.2. Annotation Methodology

Annotation was performed by a single annotator with experience in semantic annotation and a background in Linguistics and Pharmaceutical Sciences. According to Roberts et al. (2009), the combination of linguistic and clinical knowledge in annotators tends to yield higher-quality annotations in this type of texts. The decision to employ a single annotator stemmed from the fact that morphosyntactic, semantic, and medical-domain annotation requires a highly specialized profile that is difficult to recruit. Furthermore, access to the original medical reports was restricted to a very limited number of researchers, preventing the inclusion of additional annotators.

To mitigate the limitations of having only one annotator, emphasis was placed on the clarification and validation of the guidelines, reviewed by annotators with experience in semantic annotation, and on the construction of a decision tree, validated by a hematology specialist. In cases of uncertainty regarding the classification of Specialized Events, the decision tree was consulted. Additionally, weekly meetings were held with a specialist in temporal semantics to discuss problematic cases, as well as regular meetings with the medical specialist to resolve domain-specific annotation questions. The validation process of the guidelines is described in detail in Fernandes et al. (2025b)

The annotation tool used was the INCEpTION platform (Klie et al., 2018). Within INCEpTION, nine documents (patient#1, patient#2, . . .) were created, each corresponding to the medical journey of an individual patient. Each document contains all the clinical reports associated with that patient's case. To improve clarity and facilitate the identification of report boundaries, each report within a document was separated by a blank line.

Annotation was carried out in two stages: first, events and temporal expressions, along with their attributes, were annotated across all reports, followed by the annotation of TLINKs. This approach follows Campillos-Llanos et al. (2018), who found that higher inter-annotator agreement (IAA) and overall annotation quality were achieved when the process was conducted in two steps: first annotating entities and their attributes, and then proceeding with relation annotation.

The reports of each patient were annotated in chronological order of their creation. Cross-document annotation followed a systematic procedure, as illustrated in the workflow presented in Figure 2. For instance, if the first report for a patient is a group consultation with a DOCTIME of *17.09.2023* and the next report is a Discharge Report with a DischargeTime of *28.01.2024*, this date must be linked to the DOCTIME of the first report by *TLINK_after*.

Additionally, when a Group Consultation or General Report was written during a hospitalization period, the DOCTIME of that report was linked to both the AdmissionTime and the DischargeTime recorded in the corresponding Discharge Report. For example, if the first report is a discharge summary with an AdmissionTime of *03.06.2022* and a DischargeTime of *06.08.2022*, and the second report is a group consultation with a DOCTIME of *05.06.2022*, this DOCTIME (*05.06.2022*) must be linked to the AdmissionTime (*03.06.2022*) by *TLINK_after* and to the DischargeTime (*06.08.2022*) by *TLINK_before*.

Whenever possible, a TLINK was also established between the first event of a given report and the last event of the preceding one. Events within a report were only linked to those of the previous report through *TLINK_identity*, except in cases where an event mentioned but not temporally located in the previous report could be temporally anchored by an event in the subsequent one.

All decisions were documented in the annotation guidelines, available in the GitHub Repository.

4. Results

With respect to the corpus[2] as a whole, a total of 1.859 markables were identified, of which 861

[2] Due to privacy and confidentiality constraints, the dataset cannot yet be shared publicly while ethical approvals and patient consent are being obtained.

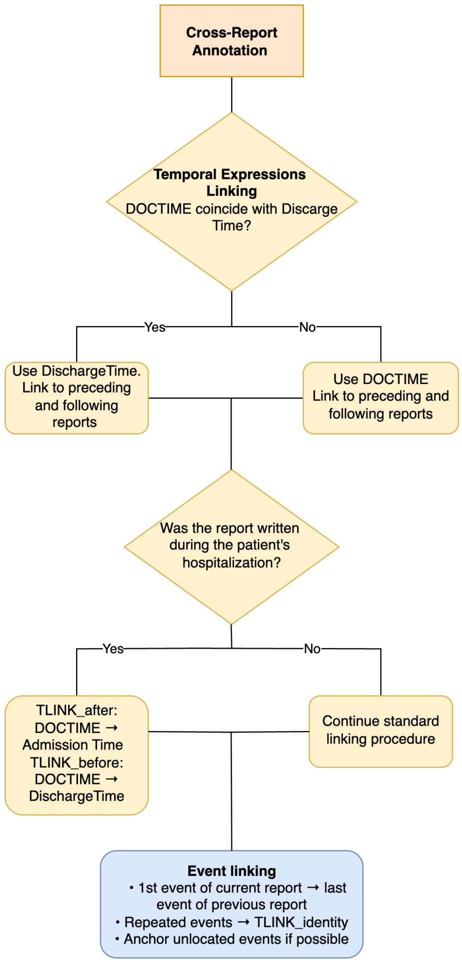

Figure 2: Cross-document annotation rules

(46%) correspond to Specialized Events, 728 (39%) to General Events, and 270 (15%) to Temporal Expressions. The main results are presented here, with detailed annotation results available in the GitHub Repository.

4.1. Events and Time Expressions

Among General Events, 50% (362) are nouns, 46% (335) verbs, and 4% (31) adjectives. Specialized Events reveal an even more marked predominance of nouns, accounting for 95.05% of all events (807 instances). Nominal expressions are used to refer to diagnoses and disease characterizations, symptoms, examinations and their results, procedures, and treatments. Adjectives constitute the second most frequent category in Specialized Events, representing 4.47% (38 instances). They serve mainly to qualify health states or clinical results, and also to express the modality through terms such as "provável" [*probable*]. Other categories occur only marginally: pronouns represent 1.40% (12 instances); verbs represent 0,23% (2 instances); and other categories, including adverbs (e.g., “o paciente está bem” [*the patient is well*]), account for 0.23%. The prevalence of eventive nouns appears to be characteristic of this textual genre, contrasting with journalistic narratives, where they occur less frequently (Silvano et al., 2024).

Table 1: Distribution of General Event Class

Class	Frequency	Percentage
Occurrence	166	22.80%
State	145	19.92%
I_Action	31	4.26%
Reporting	13	1.78%
None	373	51.24%

Regarding polarity in General Events, 96.98% of the cases (706 instances) display positive polarity, and only 3.02% (22 instances) negative polarity, mostly restricted to expressions such as "sem queixas" [*no complaints*]. This pattern aligns with the nature of medical reports, which predominantly describe clinical facts through affirmative statements, recording what is present rather than what is absent. Regarding Specialized Events, most events display positive polarity (836 instances), with only 25 exhibiting negation. Certain Specialized Event Classes show a comparatively higher rate of negation, notably Examination Result (e.g., “sem isolamentos bacteriológicos” [*no bacteriological isolations*]), followed by Personal History (e.g., “sem antecedentes pessoais de relevo” [*no relevant personal history*]), and Treatment Response (e.g., “sem recuperação hematológica” [*no hematological recovery*]).

General Events are mainly dynamic (Table 1, reflecting the narrative dimension of medical reports, where events advance the temporal progression of discourse (Caenepeel, 1989) (e.g., "iniciou protocolo de indução" [*he initiated induction protocol*]). States, though frequent, serve descriptive, characterizing the patient’s condition, providing personal data, or reporting examination results (e.g., "quadro recente caracterizado por astenia" [*recent condition characterised by astenia*]).

As shown in Table 1, 51.24% of the events were not assigned aspectual features (label *none*), since these were applied only to verbal events, and not to nominal ones, given that there is currently no established approach in the literature for the aspectual annotation of nominal events.

Reporting verbs (e.g., "dizer" [*to say*], "afirmar" [*to state*]) occur infrequently, suggesting that physicians seldom reproduce patients’ statements through indirect speech. Patient information is instead presented as factual data, devoid of enunciative responsibility. The same applies to symptoms, which, although subjective phenomena reported by the patient to the doctor, are expressed in an objective manner (e.g., "Instalou-se um quadro de sudorese nocturna, astenia e mialgias" [*a condition of night sweats, asthenia, and myalgia developed*]).

Table 2: Distribution of General Event Type

Type	Frequency	Percentage
Transition	165	22.66%
State	145	19.92%
Process	45	6.18%
None	373	51.24%

This absence of reported speech confirms observations made by Lysanets et al. (2017), who identify it as a distinctive feature of medical discourse. In the corpus, reporting events were limited to verbs such as "realçar" [*to emphasise*] and "referir" [*to note*], differing markedly from journalistic texts, which employ a range of enunciative strategies and direct speech. Silvano et al. (2024) report an average of 3.50 reporting events per news item.

No *Perception* events or *I_State* events (intensional states such as "acreditar" [*to believe*]) were found, reflecting the high degree of objectivity in medical descriptions. *I_Action* events occurred only in the construction "decide-se propor" [*it was decided to propose*]. Savova et al. (2009) also noted that the less common event classes in clinical narratives were *Perception* and *Reporting*.

Regarding General Event type (Table 2), *Transition* was the most frequent class, followed by *State*.

Concerning *VForm*, annotated only for non-finite verbs, 76 instances of the participle , 46 of the infinitive, and 20 of the gerund were identified. Finite forms were more common (586 occurrences). Only two instances of the future and one of the subjunctive were found. In terms of modality, "poder" [*may*] appeared six times and "dever" [*must*] once. The low frequency of modal forms reflects the communicative purpose of medical reports, which favor factual information over hypothetical or evaluative content.

Most events were expressed in the past tense, in line with the retrospective nature of clinical narratives. Lysanets et al. (2017) similarly note that the predominance of the past tense in medical case reports stems from their narrative purpose: to recount past events. Sixty-four events occurred in the present tense, typically describing the patient's current status, and only two in the future, both expressing epistemic modality rather than temporal reference (e.g., "será de realçar uma hiperbilirrubinemia" [*it should be noted that there is hyperbilirubinemia*]).

Regarding the Specialized Events, the most predominant was Treatment (Table 3).

The centrality of treatment reflects the very function of medical reports: all the information they contain ultimately serves to guide the clinician's therapeutic decisions. Group Consultation Reports systematically conclude with a treatment proposal; Discharge Reports necessarily include treatment-related references, since hospitalization itself is

Table 3: Distribution of Specialized Event Classes

Specialized Event Class	Frequency	Percentage
Treatment	167	19.40%
Examination Result	133	15.45%
Personal History	103	11.96%
Examination	99	11.50%
Sign or Symptom	83	9.64%
Characterization of the Disease	82	9.53%
Principal Diagnosis	66	7.66%
Intercurrence	56	6.50%
Treatment Response	51	5.92%
Medical Procedure	14	1.63%
Drug Administration Route	7	0.81%

linked to therapy (either for chemotherapy or for an intercurrent event arising from it); and General Reports incorporate treatment within the retrospective overview of the patient's medical course. In contrast, Drug Administration Route appears least frequently, with only seven occurrences limited to the specifications intravenous, intrathecal, and oral. This scarcity suggests that, although treatment constitutes the thematic axis around which the reports are organized, emphasis is placed on decision-making and outcomes rather than on the technical details of drug administration.

Finally, several abbreviations were identified in the corpus and subsequently clarified during the annotation process in an open-text field, in order to ensure terminological transparency (e.g., LMA ("Leucemia Mieloide Aguda" [*Acute Myeloid Leukemia*]). Most of these abbreviations correspond to disease names.

With regard to Temporal Expressions, the most frequent category was Date (Table 4). Expressions of this type mainly refer to dates of hospitalization, discharge, consultations, interventions, and to the beginning and end of treatments.

Table 4: Distribution of Temporal Expressions

Temporal Expressions	Frequency	Percentage
Date	76.29%	206
Duration	17.04%	46
Time	5.19%	14
Set	1.48%	4

The next most common category is Duration (e.g., "apiraxia sustentada durante 7 dias" [*apyrexia for 7 days*]), which essentially corresponds to the length of symptoms or treatments. This is followed by Time, and, finally, Set. The latter is generally associated with the frequency of medication administration or the periodic performance of examinations.

As for Temporal Function, most events were not assigned any annotation for this attribute (213 occurrences, 78.75%). Among the annotated cases, the most frequent label was DOCTIME (which refers to the date when the report was created) (6.65%, 18 occurrences), followed by Discharge-

Time and AdmissionTime (both 5.91%, 16 occurrences each), and, finally, AnchorTime (2.59%, 7 occurrences). This distribution suggests that only a small proportion of temporal expressions are directly associated with specific reference points in the clinical cycle (admission, discharge, or the document's reference time), while the majority are used descriptively, serving only to temporally situate events rather than perform an explicit temporal function.

4.2. TLINKs

TLINK_simultaneous was the most frequent (Table 5), reflecting the prevalence of events that occur concurrently or whose temporal boundaries cannot be meaningfully distinguished in context. This relation was also applied by default when no information enabled a more specific temporal localization. For example, in the sentence *"O hemograma mostrou leucocitose"* [*The blood count showed leukocytosis*], the events *"hemograma"* [*blood count*], *"mostrou"* [*showed*] and *"leucocitose"* [*leukocytosis*] would be linked with *TLINK_simultaneous* since the information regarding the beginning of "leucocitose is not available.

The high frequency of *TLINK_simultaneous* aligns with the high frequency of states, which naturally favor simultaneity links; however, dynamic events also frequently exhibit this relation, suggesting that medical reports, although they have a narrative character, often emphasize descriptive rather than sequential information.

Table 5: Distribution of TLINKs

TLINKs	Frequency	Percentage
TLINK_simultaneous	733	34%
TLINK_identity	345	16%
TLINK_after	337	15.63%
TLINK_isIncluded	286	13.27%
TLINK_includes	115	5.33%
TLINK_before	155	7.19%
TLINK_before_overlap	30	1.39%
TLINK_after_overlap	36	1.67%
TLINK_begins	40	1.85%
TLINK_begunBy	37	1.72%
TLINK_ends	12	0.56%
TLINK_endedBy	5	0.23%
TLINK_during	25	1.16%

TLINK_identity was the next most frequent, marking situations or temporal expressions that refer to the same event or the same time expression. Its recurrence is partly due to annotation rules based on the ISO standard, notably those governing light verb constructions (e.g., "fez tratamento" [*underwent treatment*]), where the verb and the nominal component are linked by *TLINK_identity* to represent a single event.

TLINK_after expresses chronological succession, and its relatively high frequency reflects the abundance of transition-type events, which inherently promote temporal sequencing.

Relations of inclusion also occurred frequently, capturing nested temporal relations, such as events occurring within a broader interval ("Fez quimioterapia de indução. Como intercorrências teve 2 síndromes febris" [*She underwent induction chemotherapy. As intercurrences, she experienced two febrile syndromes*]). Less frequent, though still relevant, were *TLINK_before* and overlap variations (*TLINK_before_overlap* and *TLINK_after_overlap*), typically associated with modifiers like "recentemente" [*recently*], which indicate partial temporal overlap. Finally, relations specifying explicit boundaries — *TLINK_begins*, *TLINK_begunBy*, *TLINK_ends*, *TLINK_endedBy*, and *TLINK_during* — most often occurred in medication contexts, such as the beginning and end of drug administration.

4.3. Cross-document Temporal Analysis Across Medical Reports

Table 6 of Appendix A presents the cross-document relations that were established. The predominant temporal relation between successive reports of the same patient is, as expected, *TLINK_identity*, since personal history, diagnosis, and treatment are repeated throughout the reports. The events most frequently associated with this relation belong to the Specialized Event and are predominantly nominal. The most common classes linked by this TLINK include Treatment, Principal Diagnosis, Characterization of the Disease, and Personal History. These results are consistent with the nature of clinical reports: Treatment is reiterated throughout the document sequence, being initially mentioned in the Group Consultation Report (when the patient is first admitted to the service), resumed during hospitalization for treatment, and referred to again in subsequent, more general reports. Similarly, the Principal Diagnosis, Acute Myeloid Leukemia, and the Characterization of the Disease tend to be recurrently linked across reports, as the patient's entire clinical history revolves around this main diagnosis. Personal History elements are also frequently repeated. The second most frequent relations are *TLINK_after* and *TLINK_before*, which occur predominantly between temporal expressions of the Date type, particularly between DOCTIMEs, between DischargeTime and DOCTIME, or among different DischargeTimes. These relations thus establish the chronological progression between reports. *TLINK_after* also appears linking General Events across different documents. Most of these events are nominal and therefore lack explicit aspectual specification. In cases where the General Event is verbal, *TLINK_identity* relations are mainly observed between two states, and *TLINK_before* between two occurrences of the transition type. Less frequent relations also emerge, such as *TLINK_is_included*. This relation was observed, for instance, when the last event in the first report was "chemotherapy" and the first event in the subsequent report was "discharge". In this case, the "discharge" event occurred during the "chemotherapy" period and is therefore considered to be included within it.

In what follows, an example of the cross-document annotation is provided, corresponding to the record of Patient 2. The record of Patient 2 comprised one *Discharge Report*, three *Group Consultation Reports*, and one *General Report*, amounting to a total of 201 annotated relations. Regarding cross-document annotation, the following *TLINKs* were established: (1) **Six *TLINK_after*** — two between *Group Consultation Report #1* and *Discharge Report #1*; one between *Group Consultation Report #2* and *Group Consultation Report #1*; one between *Group Consultation Report #3* and *Group Consultation Report #2*; and one between *General Report #1* and *Group Consultation Report #2*; (2) **Twenty-six *TLINK_identity*** — four between *Group Consultation Report #1* and *Discharge Report #1*; seven between *Group Consultation Report #2* and *Group Consultation Report #1*; six between *Group Consultation Report #3* and *Group Consultation Report #2*; and nine between *General Report #1* and *Group Consultation Report #3*; (3) **Four *TLINK_before*** — one between *Group Consultation Report #1* and *Discharge Report #1*; one between *Group Consultation Report #2* and *Group Consultation Report #1*; one between *Group Consultation Report #3* and *Group Consultation Report #2*; and one between *General Report #1* and *Group Consultation Report #3*.

4.3.1. Challenges in Cross-document Annotation

Cross-document annotation presents a range of challenges that stem from cognitive workload and the inherent complexity of certain relation types. In the present work, several key difficulties were encountered. First, the task imposes a high cognitive load on the annotator. To establish TLINKs, annotators must read, understand, and retain information distributed across multiple clinical reports. This process requires not only sustained attention and memory, but also a degree of domain knowledge to recognize whether differently phrased descriptions refer to the same underlying clinical event and to determine the appropriate temporal placement of those events. The need to mentally compare and integrate details across documents makes the annotation process particularly demanding.

Ambiguity in decision-making is another recurring challenge. Determining whether two expressions refer to the same real-world event is not always straightforward, especially when descriptions differ in specificity or narrative focus. This difficulty is compounded by the distribution of relevant information across documents: events are often mentioned only in fragments. For instance, one report may state that a patient has been "followed in this service for four years", while another specifies "followed in this service since April 2014", providing a clearer temporal anchor. As a result, cross-document annotation frequently involves linking partial and complementary pieces of information.

A further complication arises from evolving and sometimes conflicting clinical records. The same event may be described differently over time or by different healthcare professionals, reflecting updates, corrections, or documentation inconsistencies. In some cases, the same event was assigned different dates across reports — for example, a cesarean section listed as occurring in September 2014 in one report and in October 2014 in another. Such discrepancies challenge the establishment of *TLINK_identity*, as annotators must decide whether new or corrected details refer to the same event or to a distinct one. Finally, as annotation progresses, documents in INCEpTION become increasingly dense with marked entities and relations. This density can reduce the usability of annotation tools, making navigation more difficult and increasing the likelihood of overlooking inconsistencies.

5. Final Remarks

In this study, we aimed to identify and describe cross-document temporal properties in medical records. The findings show that, although cross-document temporal annotation is more demanding in terms of both the annotation scheme and the annotation process, it enables the construction of coherent chronological representations of patients' clinical journeys. The analysis also highlights key characteristics of these narratives, such as the predominance of nominal event expressions and the frequent use of simultaneity as the primary temporal relation type. Future work will include extending the annotation scheme to incorporate participant and discourse-relation layers, exploring automatic timeline generation from annotated data, and investigating automatic pre-annotation methods to support and accelerate the annotation process.

6. Limitations

This study presents some limitations that should be acknowledged. First, the annotation was carried out by a single annotator. Although measures were implemented to enhance the reliability and feasibility of the annotated data, as described in this article, the absence of inter-annotator agreement metrics may limit the ability to quantitatively assess annotation consistency and replicability.

Second, due to privacy and confidentiality constraints associated with clinical data, the annotated dataset cannot be publicly released. Although de-identification procedures were applied, the sensitive nature of the records prevents open access. This limitation reduces the reproducibility of the study and restricts external validation or reuse of the resource by the research community. In future work, this constraint will be addressed through the generation of synthetic clinical reports derived from the characterization of the original documents. Defining the structural and linguistic properties of this type of text, as undertaken in the present study, is therefore a necessary step toward enabling the creation of representative and shareable synthetic datasets.

Finally, the clinical reports analyzed originate from a single healthcare institution. As documentation practices, narrative styles, and temporal structuring in clinical reporting can vary across institutions and regions, the findings may not generalize to broader contexts. Future work should consider expanding the dataset to include documents from multiple sources in order to improve representativeness and enhance the general applicability of the annotation framework and results.

7. Acknowledgements

This work was supported by national funds through the Fundação para a Ciência e a Tecnologia (FCT), under PhD grant 2025.00440.BD. The authors also acknowledge support from the StorySense project (DOI: 10.54499/2022.09312.PTDC).

8. Bibliographical References

J. Vasco Barreto and Paulo Paiva. 2008. O registo clínico orientado por problemas = the problem oriented clinical record. *Medicina Interna: Revista da Sociedade Portuguesa de Medicina Interna*, 15(3):201–206.

Steven Bethard, Leon Derczynski, Guergana Savova, James Pustejovsky, and Marc Verhagen. 2015. SemEval-2015 task 6: Clinical TempEval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 806–814, Denver, Colorado. Association for Computational Linguistics.

Steven Bethard and Jonathan Parker. 2016. A semantically compositional annotation scheme for time normalization. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3779–3786, Portorož, Slovenia. European Language Resources Association (ELRA).

Steven Bethard, Guergana Savova, Martha Palmer, and James Pustejovsky. 2017. SemEval-2017 task 12: Clinical TempEval. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 565–572, Vancouver, Canada. Association for Computational Linguistics.

Peter Bramsen, Pranav Deshpande, Young-Kyoon Lee, and Regina Barzilay. 2006. Finding temporal order in discharge summaries. In *AMIA Annual Symposium Proceedings*, volume 2006, pages 81–85. American Medical Informatics Association.

Mimo Caenepeel. 1989. *Aspect, temporal ordering and perspective in narrative fiction*. Phd thesis in linguistics, University of Edinburgh, Edinburgh.

Leonardo Campillos-Llanos, Louise Deléger, Cyril Grouin, Thierry Hamon, Anne-Laure Ligozat, and Aurélie Névéol. 2018. A french clinical corpus with comprehensive semantic annotations: development of the medical entity and relation limsi annotated text corpus (merlot). *Language Resources and Evaluation*, 52(2):571–601. Online publication.

Leonardo Campillos-Llanos, Ana Valverde-Mateos, and Adrián Capllonch-Carrión. 2025. Hybrid natural language processing tool for semantic annotation of medical texts in spanish. *BMC Bioinformatics*, 26(1):7.

Tommaso Caselli and Piek Vossen. 2017. The event StoryLine corpus: A new benchmark for causal and temporal relation extraction. In *Proceedings of the Events and Stories in the News Workshop*, pages 77–86, Vancouver, Canada. Association for Computational Linguistics.

Xinyu Chen, Peifeng Li, and Qiaoming Zhu. 2025. Improving cross-document event coreference resolution by discourse coherence and structure. *Information Processing & Management*, 62(4):104085. [Online].

Alon Eirew, Arie Cattan, and Ido Dagan. 2021. WEC: Deriving a large-scale cross-document event coreference dataset from Wikipedia. In *Proceedings of the 2021 Conference of the North*

American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2498–2510, Online. Association for Computational Linguistics.

Ana Luísa Fernandes, Purificação Silvano, António Leal, Nuno Guimarães, Evelin Amorim, and Alípio Jorge. 2025a. Can ISO 24617-1 go clinical? extending a general-domain scheme to medical narratives. In *Proceedings of the 21st Joint ACL–ISO Workshop on Interoperable Semantic Annotation (ISA-21)*, pages 41–52, Düsseldorf, Germany. Association for Computational Linguistics.

Ana Luísa Fernandes, Purificação Silvano, António Leal, Nuno Guimarães, Rita Rb-Silva, Luís Filipe Cunha, and Alípio Jorge. 2025b. The incremental process of building an annotation scheme for clinical narratives in Portuguese: the contribution of human variation analysis. In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, pages 332–343, Vienna, Austria. Association for Computational Linguistics.

Lucian Galescu and Nathan Blaylock. 2012. A corpus of clinical narratives annotated with temporal information. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI 2012)*, pages 715–720, Miami, Florida. Association for Computing Machinery (ACM).

Qiang Gao, Bobo Li, Zixiang Meng, Yunlong Li, Jun Zhou, Fei Li, Chong Teng, and Donghong Ji. 2024a. Enhancing cross-document event coreference resolution by discourse structure and semantic information. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5907–5921, Torino, Italia. ELRA and ICCL.

Qiang Gao, Zixiang Meng, Bobo Li, Jun Zhou, Fei Li, Chong Teng, and Donghong Ji. 2024b. Harvesting events from multiple sources: Towards a cross-document event extraction paradigm. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 1913–1927, Bangkok, Thailand. Association for Computational Linguistics.

Christian Girardi, Manuela Speranza, Rachele Sprugnoli, and Sara Tonelli. 2014. CROMER: a tool for cross-document event and entity coreference. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3204–3208, Reykjavik, Iceland. European Language Resources Association (ELRA).

William Held, Dan Iter, and Dan Jurafsky. 2021. Focus on what matters: Applying discourse coherence theory to cross document coreference. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1406–1417, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ornella Irrera, Stefano Marchesin, and Gianmaria Silvello. 2024. Metatron: advancing biomedical annotation empowering relation annotation and collaboration. *BMC Bioinformatics*, 25:112.

Ann K. Irvine, Stephanie W. Haas, and Tessa Sullivan. 2008. Tn-ties: A system for extracting temporal information from emergency department triage notes. In *AMIA Annual Symposium Proceedings*, volume 2008, pages 328–332. American Medical Informatics Association.

ISO. 2012. *ISO 24617-1: 2012, Language resource management – Semantic annotation framework (SemAF) – Part 1: Time and events (SemAF-Time, ISO-TimeML)*. The International Organization for Standardization, Geneva. Editorial committee: James Pustejovsky (chair), Project leader and convenor: Kiyong Lee.

Monika Jain, Raghava Mutharaju, Kuldeep Singh, and Ramakanth Kavuluru. 2024. Knowledge-driven cross-document relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3787–3797, Bangkok, Thailand. Association for Computational Linguistics.

J.-C. Klie, M. Bugert, B. Boullosa, R. Eckart de Castilho, and I. Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics.

António Leal, Purificação Silvano, Evelin Amorim, Inês Cantante, Fátima Silva, Alípio Mário Jorge, and Ricardo Campos. 2022. The place of iso-space in text2story multilayer annotation scheme. In *Proceedings of the 18th Joint ACL–ISO Workshop on Interoperable Semantic Annotation within LREC 2022*, pages 61–70, Marseille, France. European Language Resources Association.

Yu-Kai Lin, Hsinchun Chen, and Randall A. Brown. 2013. Medtime: A temporal information extraction system for clinical narratives. *Journal of Biomedical Informatics*, 46(Suppl.):S20–S28.

Yuliia Lysanets, Halyna Morokhovets, and Olena Bieliaieva. 2017. Stylistic features of case reports as a genre of medical discourse. *Journal of Medical Case Reports*, 11(83). Online.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional description of text organization. *Text*, 8:243–291.

Anne-Lyse Minard, Manuela Speranza, Eneko Agirre, Itziar Aldabe, Marieke van Erp, Bernardo Magnini, German Rigau, and Rubén Urizar. 2015. SemEval-2015 task 4: TimeLine: Cross-document event ordering. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 778–786, Denver, Colorado. Association for Computational Linguistics.

Gandhimathi Moharasan and Tu Bao Ho. 2019. Extraction of temporal information from clinical narratives. *Journal of Healthcare Informatics Research*, 3(2):220–244.

James Pustejovsky, José M. Castaño, Robert Ingria, Roser Saurí, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering*, pages 28–34, Stanford, CA. AAAI Press.

Preethi Raghavan, Eric Fosler-Lussier, Noémie Elhadad, and Albert M. Lai. 2014. Cross-narrative temporal ordering of medical events. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 998–1008, Baltimore, MD. Association for Computational Linguistics.

Rita Rb-Silva and Yulia Karimova. 2021. amile: Application of text mining to clinical reports of patients with acute myeloid leukemia. Zenodo. Zenodo, January 28, 2021.

Angus Roberts, Richard Gaizauskas, Mark Hepple, George Demetriou, Yulan Guo, Andrea Setzer, and Bill Wheeldin. 2007. The clef corpus: semantic annotation of clinical text. In *AMIA Annual Symposium Proceedings*, pages 625–629.

Angus Roberts, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. 2009. Building a semantically annotated corpus of clinical texts. *Journal of Biomedical Informatics*, 42(5):950–966.

Guergana Savova, Steven Bethard, William Styler IV, James Martin, Martha Palmer, Julie Masanz, and Wayne Ward. 2009. Towards temporal relation discovery from the clinical narrative. In *AMIA Annual Symposium Proceedings*, pages 568–572.

Shira Shimony, Michael Stahl, and Richard M. Stone. 2023. Acute myeloid leukemia: 2023 update on diagnosis, risk-stratification, and management. *American Journal of Hematology*, 98(3):502–526. Epub 2023 Jan 13.

Purificação Silvano, António Leal, Fátima Silva, Inês Cantante, Fatima Oliveira, and Alípio Mario Jorge. 2021. Developing a multilayer semantic annotation scheme based on ISO standards for the visualization of a newswire corpus. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 1–13, Groningen, The Netherlands (online). Association for Computational Linguistics.

Purificação Silvano, Emanuel Amorim, António Leal, Inês Cantante, Adriana Jorge, Ricardo Campos, and Nanjiang Yu. 2024. Untangling a web of temporal relations in news articles. In *Proceedings of the 7th International Workshop on Narrative Extraction from Texts (Text2Story'24)*, pages 77–92, Glasgow. CEUR Workshop Proceedings (CEUR-WS.org).

Zhiyi Song, Ann Bies, Justin Mott, Xuansong Li, Stephanie Strassel, and Christopher Caruso. 2018. Cross-document, cross-language event coreference annotation using event hoppers. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

William F. Styler IV, Steven Bethard, Sean Finan, Martha Palmer, Sameer Pradhan, Piet C de Groen, Brad Erickson, Timothy Miller, Chen Lin, Guergana Savova, and James Pustejovsky. 2014. Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, 2:143–154.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013a. Annotating temporal information in clinical narratives. *Journal of Biomedical Informatics*, 46(Suppl 0):S5–S12. Epub 2013 Jul 19.

Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013b. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813. Epub 2013 Apr 5.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Lawrence L. Weed. 1968. Medical records that guide and teach. *The New England Journal of Medicine*, 278:593–600, 652–657. Online.

Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. CorefQA: Coreference resolution as query-based span prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6953–6963, Online. Association for Computational Linguistics.

Jiarui Yao, Harry Hochheiser, WonJin Yoon, Eli Goldner, and Guergana Savova. 2024. Overview of the 2024 shared task on chemotherapy treatment timeline extraction. In *Proceedings of the 6th Clinical Natural Language Processing Workshop*, pages 557–569, Mexico City, Mexico. Association for Computational Linguistics.

Jin Zhao, Nianwen Xue, and Bonan Min. 2023. Cross-document event coreference resolution: Instruct humans or instruct GPT? In *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages 561–574, Singapore. Association for Computational Linguistics.

Enwei Zhu, Qilin Sheng, Huanwan Yang, Yiyang Liu, Ting Cai, and Jinpeng Li. 2023. A unified framework of medical information annotation and extraction for chinese clinical text. *Artificial Intelligence in Medicine*, 142:1–12.

A. Cross-document relations

Patients / Reports	MR #2 → MR #1	MR #3 → MR #2	MR #4 → MR #3	MR #5 → MR #4	MR #5 → MR #3	MR #4 → MR #1	MR #4 → MR #2	MR #3 → MR #1
Patient 1	After–2 Identity–5							
Patient 2	After–2 Before–1 Identity–4	After–1 Before–1 Identity–7	After–1 Before–1 Identity–6	After–1 Before–1 Identity–4	Identity–5			After–1
Patient 3	After–2 Identity–6	After–1 Identity–1	After–2 Identity–6	After–2 Identity–1				
Patient 4	After–2 Identity–1	After–1 Before–1 Identity–5						
Patient 5	After–1 Before–1	After–2 Begun_By–1						Identity–2
Patient 6	After–1 Before–1 Identity–4	After–3 Identity–11 Simultaneous–1	After–2 Identity–17	After–2 Begins–1 Identity–17	Identity–1			Includes–1 Identity–1
Patient 7	After–2	After–1 Before–1 Identity–2	After–1 Is_Included–2 Identity–2	After–2 Identity–3		Identity–2	Identity–1	
Patient 8	After–1 Is_Included–1 Identity–10	After–2 Identity–12	After–2 Identity–8	After–2 Identity–7				Identity–1
Patient 9	After–2 Identity–7	After–2 Identity–6	After–2 Identity–1	After–2 Identity–3		Identity–2		Identity–18

Table 6: Cross-document relations across MRs