

# Temporal Annotation of Clinical Text

Danielle L. Mowery MS, Henk Harkema PhD, Wendy W. Chapman PhD

Department of Biomedical Informatics  
University of Pittsburgh, Pittsburgh, PA 15260, USA  
dlm31@pitt.edu, heh23@pitt.edu, wec6@pitt.edu

## Abstract

We developed a temporal annotation schema that provides a structured method to capture contextual and temporal features of clinical conditions found in clinical reports. In this poster we describe the elements of the annotation schema and provide results of an initial annotation study on a document set comprising six different types of clinical reports.

## 1 Introduction

Distinguishing between historical and recent conditions is important for most tasks involving retrieval of patients or extraction of information from textual clinical records. Various approaches can be used to determine whether a condition is historical or recent. Chapman et al. (2007) developed an algorithm called ConText that uses trigger terms like “history” to predict whether a condition is historical. Studies of ConText show that this approach is inadequate for determining whether a condition is historical, achieving recall of 67% and precision 74% on emergency department reports. Temporal modeling methods commonly reason about the temporality of an event with respect to absolute time and other temporally related events (Zhou et al., 2006; Chambers et al., 2007). Knowing the relative or absolute time the condition occurred can be useful in determining whether the condition is historical. However, we hypothesize that many clinical conditions in clinical reports are not modified by explicit temporal references.

To test this hypothesis and explore other types of information that may be useful in automatically distinguishing historical from recent clinical conditions in dictated clinical records, we developed a temporal annotation schema that accounts for explicit temporal expressions, temporal trigger terms,

and clinical reporting acts described in reports. Three annotators applied the schema to six types of reports. We measured inter-annotator agreement scores and obtained prevalence and distribution figures for the three annotation types.

## 2 Methods

### 2.1 Dataset

Our dataset is comprised of 24 clinical reports of six types dictated at the University of Pittsburgh Medical Center during 2007: discharge summaries, surgical pathology, radiology, echocardiograms, operative gastrointestinal, and emergency department reports. A physician pre-annotated the 518 clinical conditions in the reports and marked each one as recent or historical.

We developed our annotation schema using one of each report type (six reports). Annotators (authors HH, DM and WC) annotated the remaining 18 reports as described below.

### 2.2 Annotation Schema

For our temporal annotation study, each pre-annotated clinical condition was annotated with three types of information: temporal expression, trigger term, and clinical reporting act.

The set of **temporal expressions** (TEs) is taken from Zhou et al. (2006) and includes categories such as DATE AND TIME for explicit TEs and KEY EVENTS for TEs relative to significant clinical events. A given clinical condition is annotated with the category of the TE it is modified by. For example, in the sentence “*The stroke occurred on 1/5/2000*”, the condition “stroke” is annotated with category DATE AND TIME. There is also a category NO TEMPORAL EXPRESSION for annotating conditions that are not linked to a TE.

**Trigger terms** (TTs) are explicit signals (words and phrases) in text other than TEs that indicate

whether a condition is recent or historical (Chapman et al., 2007). If a condition co-occurs with a TT, it is annotated with TRIGGER: YES. For example, “pneumonia” in the sentence “*Films indicate pneumonia, which is new for this patient*” is annotated as TRIGGER: YES because “new” is a TT.

Error analyses of our previous studies indicate that the context in which a condition is mentioned in a report is potentially useful for prediction of a condition as recent or historical. Clinical reports consist of statements that group into segments according to the **clinical reporting act** (CRA) they describe, such as noting a past history and considering a diagnosis. CRAs are tightly correlated with report sections; however, sections are not consistent, and different CRAs can occur within a single section. We distinguish 16 CRAs. Each clinical condition is annotated with one CRA. For example, the condition “smoker” in the sentence “*She was a smoker*” is annotated SOCIAL HISTORY.

### 2.3 Analysis

To establish the level of inter-annotator agreement, we iteratively annotated groups of six reports (one of each type). After each iteration, we refined our annotation schema and guidelines. We analyzed annotations, overall and by report type, in the following way: 1) calculate inter-annotator kappa score, 2) measure prevalence of TT and TE categories, and 3) observe distribution of CRAs.

## 3 Results and Discussion

As shown in figure 1, average inter-annotator scores as measured by Cohen's kappa for TE, TT, and CRA (.68, .82 and .72 respectively) reached acceptable levels after three iterations and are expected to rise further with increased annotation experience and understanding of the guidelines.

Table 1 shows the prevalence of TEs and TTs across six report types, where prevalence is defined as the frequency of TE or TT found in a given report. Use of TEs across report types ranged from 0% to 52% whereas TTs were found less often at 0% to 34% by report genre. Table 2 plots the correlation between the CRA assigned to a clinical condition and the condition's classification as recent or historical. We found that there is a strong correlation for the most commonly occurring clinical reporting acts (PH, PR, and PO). We are therefore optimistic that CRAs can serve as an

informative feature for a statistical recent/historical classifier.

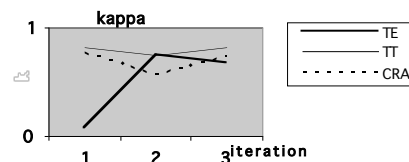


Figure 1. Average Cohen's kappa agreement for 3 iterations.

	DS	E	ED	GI	RAD	SP	O
TE	48(52)	0(0)	51(20)	2(10)	1(5)	8(36)	110(21)
TT	32(34)	0(0)	54(21)	1(5)	0(0)	6(27)	93(17)

Table 1. Prevalence, count (%), of TE and TT across report types, overall. DS: *discharge summary*, E: *echocardiogram*, ED: *emergency department*, GI: *operative gastrointestinal*, RAD: *radiology*, SP: *surgical pathology* and O: *overall*.

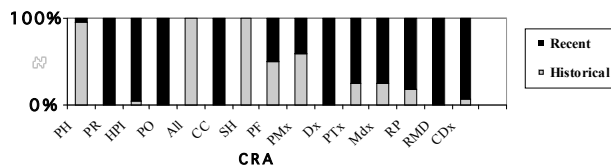


Table 2. Historical/recent distribution of CRAs. PH: *Past history*, PR: *Patient reporting*, HPI: *History of present illness*, PO: *Physician observing*, All: *Allergies*, CC: *Chief complaint*, SH: *Social history*, FH: *Family history*, PF: *Past Finding*, PMx: *Past medication*, Dx: *Diagnosis*, PTx: *Plan treatment*, Mdx: *Prescribing medication*, RP: *Referring problem*, RMD: *Refer to MD*, CDx: *Considering diagnosis*.

The finding that many conditions are associated with neither a TE nor a TT and study of ConText's limitations with such categories at the scope of the sentence suggests that additional features are necessary to discern a condition as recent or historical. Whereas temporality in discourse may follow a sequential chronology as narrative unfolds, references to past instances within clinical text are not easily resolved. We are optimistic that CRAs may help this issue and will focus our study to evaluate whether these three features are sufficient together.

## References

- L. Zhou, G. B. Melton, S. Parsons, G. Hripcsak. 2006. A temporal constraint structure for extracting temporal information from clinical narrative. *Journal of Biomedical Informatics*, 39(4):424-439
- N. Chambers, S. Wang, D. Jurafsky. 2007. Classifying Temporal Relations Between Events. In: *ACL-07*.
- W. Chapman, D. Chu, J. N. Dowling. 2007. ConText: An Algorithm for Identifying Contextual Features from Clinical Text. In: *ACL-07*.