How Availability of Explicit Temporal Cues Affects Manual Temporal Relation Annotation

Siim Orasmaa [a,1]

[a] *Institute of Computer Science, University of Tartu*
J. Liivi Str 2, 50409 Tartu, Estonia
siim.orasmaa@ut.ee

Abstract. We investigate how presence and absence of explicit temporal cues (temporal expressions and verb tenses past simple and present) affect the quality of manually provided temporal relation annotations in Estonian. Results support the hypothesis that there is higher inter-annotator agreement and less vagueness in determining temporal relations in the presence of explicit temporal cues.

Keywords. TimeML, temporal relations, inter-annotator agreement, Estonian

Introduction

Knowledge about temporal relations in text (e.g. about temporal ordering of events mentioned in text) supports many Natural Language Processing applications, including question answering, text summarization and machine translation. However, manual annotation of temporal relations has proven to be a very difficult task for humans, often characterized by relatively low inter-annotator agreement. As inter-annotator agreement between humans is considered as upper bound for performance of automatic annotation, it is an open question, how well this task can be automated [1].

The current work analyses manually annotated temporal relations (based on TimeML framework [2]) in Estonian, and investigates how the presence and absence of explicit temporal cues affect the inter-annotator agreement and vagueness in determining temporal relations. We consider the following types of explicit temporal cues:

- temporal expressions: expressions annotated with TimeML tag TIMEX3, such as *'29 July, 2013'* or *'on last Monday at 4 P.M.'*;
- tense information: Estonian present and past simple tenses;

In Estonian, verb tenses can be considered as having different degrees of explicit temporal information: past simple can be considered as most explicit (mostly used to

mark events occurring in the past), and present can be considered as least explicit (ambiguous between events culminating in the present, ongoing events, events taking place in the future, and generic statements with vague temporal boundaries).

We hypothesize that higher inter-annotator agreement and lower amount of "vague" relations used by annotators can be observed in the presence of explicit temporal cues (presence of temporal expressions and verbs in past simple). In order to test the hypothesis, we used morphological and dependency syntactic annotations available in the corpus to obtain subsets of temporal relation annotations characterized by the presence and absence of explicit temporal cues. We measured the inter-annotator agreement and the percentage of vague relations separately on each of the obtained subsets.

1. The Corpus

The corpus consists of 80 news articles (approx. 22,000 word tokens) from three Estonian newspapers: Maaleht, Postimees, and SL Õhtuleht. Three annotators participated in annotation of temporal relations and each article was annotated by 2 annotators [3]. Following the TempEval-2 guidelines [4], the annotation process was split into 4 subtasks:

event-timex. determine relations between events and temporal expressions in the same sentence;

event-dct. determine relations between events and document creation time;

main-events. determine relations between main events of two consecutive sentences;

events-in-sentence. determine relations between events in same sentence (intrasentential relations);

Like in TempEval-2, a simplified set of temporal relations was used: BEFORE, BEFORE-OR-OVERLAP, SIMULTANEOUS, IS INCLUDED, INCLUDES, OVERLAP-OR-AFTER, AFTER, VAGUE and IDENTITY.[2]

2. Experiments and Results

Although the corpus also contains nouns and adjectives annotated as event mentions, current work focuses on verb event mentions and temporal relations assigned to them. More specifically, we focus on verbs that are part of the syntactic predicate of a clause, because such verbs mostly have tense information[3], and they are likely governing temporal expressions (adverbials) of the clause.

Based on manual morphological and dependency syntactic annotations available in the corpus [3], we take out five subsets of event annotations, along with temporal relation annotations associated with these events[4]:

---

[2]In TempEval-2, a general relation OVERLAP was used instead of the elaborate relations SIMULTANEOUS, IS INCLUDED and INCLUDES. Elaborate relations were used in current work because annotators often found that the general relation OVERLAP was confusing and needed elaboration.

[3]A syntactic predicate can also consist of a single infinite verb without any tense information; however, these cases are rare in our corpus.

[4]A temporal relation was included in the subset if it connected two events in the subset or if it connected an event in the subset with a temporal expression (or with a document creation time).

0. All verb EVENTs;

1a. Event verbs in past simple tense;

1b. Event verbs in present tense;

2a. Event verbs governing an annotated temporal expression;

2b. Event verbs not governing any annotated temporal expression;

The quality of temporal relation annotation is estimated separately on each of the subsets using two measures: the proportion of VAGUE relations used by annotators, and the inter-annotator agreement on temporal relation annotation.

Table 1 shows proportions of VAGUE relations for different EVENT subsets, and Table 2 shows inter-annotator agreements (Cohen's Kappas) on annotating temporal relations on different EVENT subsets. Inter-annotator agreements were measured separately for each subtask (average kappa was calculated over all annotator pairs), and then total average kappa was calculated as a macro-average over all the averages of subtasks.

Table 1. Proportions of VAGUE temporal relations annotated on different EVENT subsets.

Event subset description	Total relation count	VAGUE relation count	Proportion of VAGUE relations
0. All verb EVENTs	9,756	1,773	18.2%
1a. EVENTs in past simple tense	3,054	109	3.57%
1b. EVENTs in present tense	4,246	1,215	28.6%
2a. EVENTs governing TIMEX	1,558	65	4.17%
2b. EVENTs not governing TIMEX	7,218	1,526	21.1%

Table 2. Inter-annotator agreements (Cohen's Kappas) on temporal relations on different EVENT subsets.

Event subset description	event-timex	event-dct	main-events	events-in-sentence	total avg κ
0. All verb EVENTs	0.287	0.419	0.417	0.327	0.362
1a. EVENTs in past simple tense	0.238	0.279	0.432	0.382	0.333
1b. EVENTs in present tense	0.333	0.227	0.311	0.215	0.272
2a. EVENTs governing TIMEX	0.283	0.489	0.655	0.473	0.475
2b. EVENTs not governing TIMEX	0.089	0.402	0.378	0.296	0.291

3. Discussion

The results indicate that the presence of explicit temporal cues (verbs in past simple and verbs governing a temporal expression) makes the temporal relation annotation task clearer (less than 5% of the relations were marked as "vague"), and also supports higher inter-annotator agreement (e.g. kappa 0.475 was measured for agreement on the relations associated with verbs governing a temporal expression). If these results are contrasted against the cases of limited/absent temporal cues (verbs in ambiguous present tense, and

verbs not governing any temporal expression), the difference clearly indicates that both inter-annotator agreement and vagueness encountered in assigning relations are dependent on the availability of explicit temporal cues.

However, the low inter-annotator agreement (and large amount of "vague" relations used) could have been the result of the limitations in annotation methodology. Guidelines on how to annotate temporal relations were not linguistically detailed, so in difficult cases, the annotators had no other guide than their own intuition. Also, a relatively large set of temporal relations was used (9 relations), which made further possible to use idiosyncratic annotation strategies. It remains an open question whether similar results can be obtained using fewer temporal relations (only relations *before, overlap, after*, like suggested in [5]), and with using more detailed linguistic instructions on how to annotate temporal relations.

It can also be argued that corpus composition (which genres of texts are annotated) plays an important role in temporal annotation [6]. Some texts, such as opinion articles, have less clear temporal structure, and thus are more difficult to annotate. This study did not consider how text genre affects the annotation, as it was limited to using a corpus with available manual syntactic annotations. It remains a future work to find out how text genre affects manual temporal relation assigning.

In conclusion, while it has been stated that the ultimate goal of the research is to design systems that can detect all temporal relations from text [1], results of this research indicate that the focus should be, perhaps, more on devising annotation strategies covering contexts with explicit temporal cues, where higher inter-annotator agreement amongst humans can be expected.

References

[1] M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, J. Moszkowicz, and J. Pustejovsky. *The TempEval challenge: identifying temporal relations in text*. Language Resources and Evaluation, pp. 161–179, 2009.

[2] *Language Resource Management - Semantic Annotation Framework (SemAF) -Part 1: Time and events*. ISO/TC 37/SC 4/WG 2. 2007.

[3] S. Orasmaa. *Towards an Integration of Syntactic and Temporal Annotations in Estonian*. In 9th Edition of the Language Resources and Evaluation Conference (LREC'14), Reykjavic, Iceland, May 26-31, 2014.

[4] TimeML Working Group. 2009. *TLINK Guidelines.* `http://www.timeml.org/tempeval2/ tempeval2-trial/guidelines/tlink-guidelines-081409.pdf`

[5] G. Marşic. *Syntactically Motivated Task Definition for Temporal Relation Identification*. In Special Issue of the TAL (Traitement Automatique des Langues) Journal on Processing of Temporal and Spatial Information in Language - Traitement automatique des informations temporelles et spatiales en langage naturel. ATALA (French Association for Natural Language Processing), vol. 53, no. 2, pp. 23-55, 2012.

[6] A. Bittar. *Building a TimeBank for French: a Reference Corpus Annotated According to the ISO-TimeML Standard*. PhD thesis. Université Paris Diderot, Paris, France, 2010.