A Examples of Table 1

The names of those categories in Table 1 are straightforward. Here we further provide examples for each of them in Example 8. Note that most of them are consistent with the definitions in the literature, with one exception for INTENTION. In TimeML (Pustejovsky et al., 2003a), there are two types of intentions, I-Action (e.g., *attempt, try* and *promise*) and I-State (e.g., *believe, intend* and *want*). But our definition of intention is the actual intent of these verbs. For example, in Example 8, ***e20*** and ***e21*** are INTENTION. This definition is more general so that verbs that are not I-Action or I-State can still create orthogonal axis of intention, e.g., the verb "allocated" in the sentence of ***e21***.

Example 8
[Orthogonal axis] INTENTION/OPINION I plan/want to (***e20:leave***) tomorrow. The mayor has allocated funds to (***e21:build***) a museum. I think he will (***e22:win***) the race.
[Parallel axis] HYPOTHESIS/GENERIC If I'm (***e23:elected***), I will cut income tax. If I'm elected, I will (***e24:cut***) income tax. Fruit (***e25:contains***) water. Lions (***e26:hunt***) zebras.
[Not on any axis] NEGATION The financial assistance from the Wolrd Bank is not (***e27:helping***). They don't (***e28:want***) to play with us. He failed to (***e29:find***) buyers.
[Other] STATIC/RECURRENT He (***e30:is***) brave. New York (***e31:is***) on the east coast. The shuttle will be (***e32:departing***) at 6:30am every day.

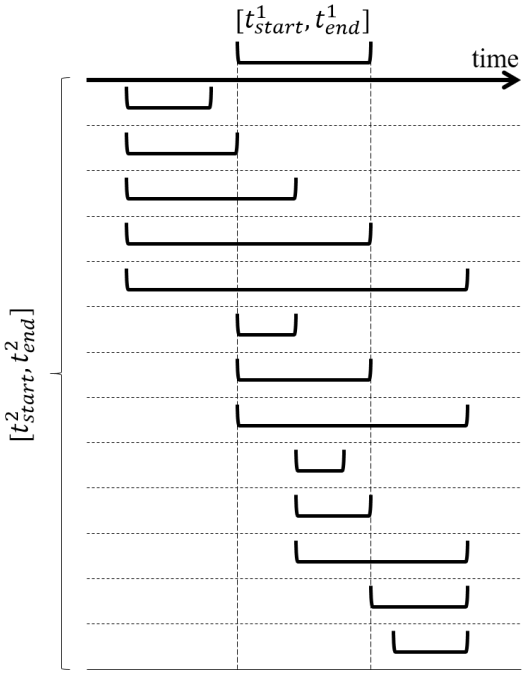

Figure 4: Thirteen possible relations between two events whose timescopes are $[t^1_{start}, t^1_{end}]$ and $[t^2_{start}, t^2_{end}]$ (from top to bottom): *after, immediately after, after and overlap, ends, included, started by, equal, starts, includes, ended by, before and overlap, immediately before* and *before*.

B *Anchorable* vs. *Actual*

As discussed in the paper, when we check if an event is *Anchorable* onto the main axis, it seems very similar to annotating whether an event is *Actual* in REALIS labeling. We have discussed the differences

in Sec. 2.3.3. To better understand them, we randomly selected 5 documents from RED (O'Gorman et al., 2016), where there are 314 events, 166 of which are verbs (we only handle verb events). Two experts annotated the anchorability of these 166 verb events independently without looking at the original REALIS annotation from RED, and they achieved a Cohen's Kappa of .88 in anchorability annotation, consistent with their Cohen's Kappa achieved on MATRES. To aggregate the result from two experts, we mark an event as *Anchorable* only when both experts labeled *Anchorable*. As for REALIS labeling in RED, we group GENERIC, HYPOTHETICAL, and HEDGED into a single label of *Non-Actual*.

		Anchorable	
		Yes	No
Actual	Yes	108	25
	No	0	33

Table 7: Comparison between anchrability and factuality on a subset of verb events randomly selected from RED.

The comparison between *Anchorable* and *Actual* is shown in Table 7. On this subset of 166 events, we did not see *Anchorable* events that are *Non-Actual* because such cases are indeed less frequent in practice; the only difference is that we annotated 25 events as *Non-Anchorable*, while RED annotated them as *Actual*. Among the 25 different cases, 11 are INTENTION, 4 are OPINION, 6 are STATIC, and 4 are NEGATION. Typical examples from each category are shown in Example 9. Note that if we calculate the McNemar's statistics based on Table 7, *Anchorable* and *Actual* are statistically different with $p \ll 0.001$.

Example 9: Typical cases that RED annotated *Actual* and we annotated *Non-Anchorable*.
Libya has since agreed to (***e33:pay***) compensation to the families of the Berlin disco victims as well as the families of the victims of the 1988 Pan Am 103 bombing over Lockerbie, Scotland, which killed 270 people, including 189 Americans. [**We think it is INTENTION**]
Gadhafi had long been ostracized by the West for (***e34:sponsoring***) terrorism, but in recent years sought to emerge from his pariah status by abandoning weapons of mass destruction and renouncing terrorism in 2003. [**We think it is OPINION**]
We need to resolve the deep-seated causes that have resulted in these problems, Premier Wen said in an interview with Hong Kong-(***e35:based***) Phoenix Television. [**We think it is STATIC**]
Fuel prices had been frozen for six years, but the government said it could no longer afford to (***e36:subsidize***) them. [**We think it is NEGATION**]

C Annotation Interface

The annotation interface was designed based on the web interface of CrowdFlower. In the anchorability annotation step (i.e., the first step), we show each crowdsourcer one event at a time, along with the full context of this event. Crowdsourcers only need to make a binary decision of Yes/No, as shown in Fig. 5.

The interface design for the relation annotation step (i.e., the second step) is tricky. As explained in Sec. 4.2, we need to ask two questions for each pair of events to figure out the actual TempRel: Q1=Is it possible that t^1_{start} is before t^2_{start}? Q2=Is it possible that t^2_{start} is before t^1_{start}? We notice in practice that asking Q1 and Q2 simultaneously (as shown in Fig. 6) gives annotators the wrong impression that there has to be one "yes" and one "no". Therefore, we decide to ask Q1 and Q2 separately. Specifically, we launch two separate tasks. One task only has Q1 (Task A), and the other only has Q2 (Task B), so that a same annotator is guaranteed not to see Q1 and Q2 simultaneously (as shown in Fig. 7).

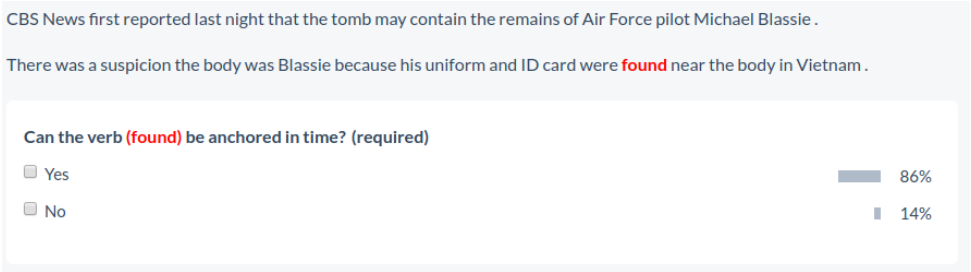

Figure 5: Annotation interface for the first step: temporal anchorability. The owner of the task can see the crowdsourcers' distribution of each answer (e.g., 86% and 14%), which is of course not available to crowdsourcers.

Figure 6: Tentative annotation interface for the second step: relation annotation. This design gives crowdsourcers the wrong impression to select one "yes" and one "no" for Q1 and Q2, leading to strong correlation between answers of Q1 and answers of Q2.

(a) Task A: Only ask Q1

(b) Task B: Only ask Q2

Figure 7: The final annotation interface, where Q1 and Q2 are posed in separate tasks so that a single annotator will not see both two questions simultaneously, forcing them to think the temporal relation carefully instead of simply putting the opposite answer to the other question.