Nominal pattern	KB Predicate	Domain	Range
"receiving"	hasWonPrize	wn_person	wn_award
"nomination"	hasWonPrize	wn_person	wn_award
"inauguration"	holdsPosition	wn_person	wn_politicalPost
"inauguration"	rdf:type	rdfs:resource	wn_inauguration
"presidency"	holdsPosition	wn_person	wn_politicalPost
"presidency"	isLeaderOf	wn_person	yagoLegalActorGeo
"death"	rdf:type	rdfs:resource	wn_death
"death"	diedIn	wn_person	wn_city

Table 2: Examples of lexico-syntactic patterns for temponym detection.

Therefore, when resolving a temponym, our methods also jointly resolve the following:

i the *entity mentions* in the context of the temponym are disambiguated;

ii the *time point or period* attached to the KB fact or event is propagated to the temponym as additional markup for the input text;

iii the *semantic type* of the temponym is determined (e.g., marriage, election, concert, etc.).

In the rest of this Section, we define several measures that capture the similarities between temponym candidates and KB facts, coherence of entities, temporal expressions and events. Then, we formulate the problem of resolving temponyms as three different Integer Linear Programs (ILP) with an objective to maximize the similarity and coherence. We also introduce several constraints to ensure that the selected mapping is meaningful.

5.1 Candidate Mappings Generation

One of the challenges of resolving temponyms is the large number of entities from the candidate temponyms obtained from Section 4.2. To address this problem we prune the candidates that have low potential of being resolved as temponyms. In this regard, we refer to the named entity dictionary from [18] to obtain a score for the candidates. Using this scoring we rank and select top-k candidates.

Consequently, we derive the candidate facts and events from Yago2. A fact from Yago2 is a candidate fact if it contains an entity from the above mentioned top-k selected candidates as a subject or an object. In addition, a Yago2 fact that is of type `event` is also a candidate mapping if the subject is one of the top-k entities. This gives us a set of candidate mappings for a temponym.

The final goal is to select the best matching mapping among the chosen candidates. For this purpose, given a temponym t and a fact f, we derive measures of relatedness for the mapping between t and f using diverse features such as textual, temporal, and semantic similarity as below:

$w\text{-}text_{tf}$**:** The jaccard string similarity between the tokens of t and f.

$w\text{-}sem_{tf}$**:** The semantic similarity score for the head noun of t and the predicate of f obtained from the pattern dictionary (as explained in Section 4.1)

$w\text{-}temp_{tf}$**:** The temporal similarity of f and the normalized dates in the context of t.

The temporal similarity $w\text{-}temp_{tf}$, between a temponym and a fact is estimated from the divergence between the distribution of the normalized dates in the context of the temponym, and the time scope of the fact. The time scope of a fact is converted to a yearly uniform distribution between the beginning and the end of the time scope. We implement these distributions in the form of histograms.

Definition 1 (**Temporal similarity**). *The temporal similarity between temponym t and fact f is defined as:*

$$w\text{-}temp_{tf} = 1 - JSD(H_t||H_f)$$

JSD (Jensen-Shannon Divergence) is the symmetric extension of the Kullback-Leibler Divergence, and H_t and H_f are the distributions of temporal information in the contexts of t and in the time scope of f. The Jensen-Shannon Divergence between H_t and H_f is defined as:

$$JSD(H_t||H_f) = \frac{1}{2}KL(H_t||M) + \frac{1}{2}KL(H_f||M)$$

Here, $M = \frac{1}{2}(H_t + H_f)$ and KL is the Kullback-Leibler divergence calculated as $KL(H_t||H_f) = \sum_i H_t(i) \log \frac{H_t(i)}{H_f(i)}$.

Using a linear combinations of the three measures described above we compute the relatedness score for fact-temponym mappings as below:

Definition 2 (**Fact-temponym relatedness**). *$w\text{-}rel_{tf}$ is a relatedness measure for temponym t and fact f:*

$$w\text{-}rel_{tf} = w\text{-}text_{tf} + w\text{-}sem_{tf} + w\text{-}temp_{tf}$$

In addition to fact-temponym relatedness, we also consider a probablistic prior coined *Mention-entity prior* denoted as $w\text{-}ned_{me}$ quantifying the similarity of an entity mention m and an existing canonical Wikipedia entity e. $w\text{-}ned_{me}$ is computed based on the frequency of a particular mention m appearing in the inlink anchor texts referring to specific entity e in Wikipedia as described in [18].

The features defined above are derived from a given single temponym and fact pair locally. We further make an important observation that a coherent text from a document contains entities, explicit TempEx's and temponyms that have high mutual relatedness in terms of their semantic and temporal properties. To exploit this, we introduce measures for semantic coherence between entities and temporal coherence between facts.

Definition 3 (**Entity-entity coherence**). *$w\text{-}coh_{ee'}$ is the precomputed Jaccard coefficient of two entities e and e':*

$$w\text{-}coh_{ee'} = \frac{|inlinks(e) \cap inlinks(e')|}{|inlinks(e) \cup inlinks(e')|}$$

where inlinks are the incoming links in the Wikipedia articles for the respective entity.

The semantic coherence enhances the coherent mapping of mentions to semantically related entities. For example, in the example text in Figure 1, the semantic coherence encourages to disambiguate the phrase "Ronaldo" as Portuguese footballer Cristiano Ronaldo rather than the famous Brazilian footballer Ronaldo.

Definition 4 (**Temporal coherence**). *$w\text{-}temp_{ff'}$ is the Jensen-Shannon Divergence between the histograms for*