

# Trust, but Verify!

## Better Entity Linking through Automatic Verification

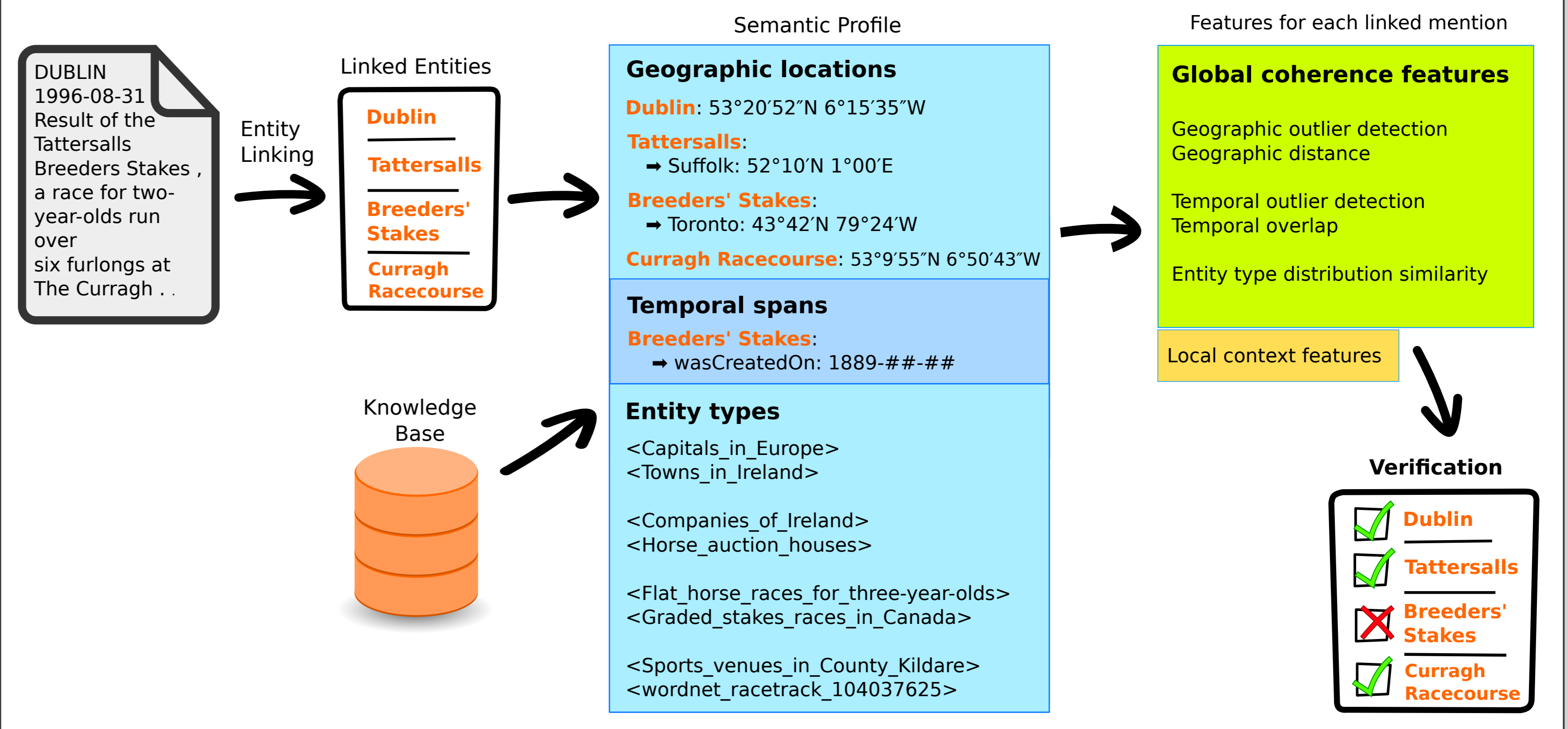
Benjamin Heinzerling<sup>1,2</sup> Michael Strube<sup>2</sup> Chin-Yew Lin<sup>3</sup>

<sup>1</sup>AIPHES <sup>2</sup>Heidelberg Institute for Theoretical Studies <sup>3</sup>Microsoft Research

### In a Nutshell

- ▶ Only one, generic measure of global coherence is commonly used in entity linking.
- ▶ We propose three new aspects of global coherence:
  - ▷ Geographic coherence
  - ▷ Temporal coherence
  - ▷ Entity type coherence
- ▶ Global disambiguation is expensive, adding these aspects even more so.
- ▶ Simple two-step disambiguation circumvents this problem:
  1. **Trust:** Extract rich semantic features for linked entities and assume these are mostly correct.
  2. **Verify:** Use these features to find likely entity linking mistakes.

### Automatic Verification



### Entity Linking

The screenshot shows an entity linking interface. The text snippet is: "DUBLIN (1996-08-31): Result of the Tattersalls Breeders Stakes, a race for two-year-olds at The Curragh...". Below the text, there are two entity cards:
 

- Tattersalls:** From Wikipedia, the free encyclopedia. For other uses, see Tattersall (disambiguation). Tattersalls (formerly spelled with an apostrophe) is the main auctioneer of race horses in the United Kingdom and Ireland.
- Breeders' Stakes:** Restricted race. Third leg in the Canadian Triple Crown. Location: Woodbine Racetrack, Toronto, Ontario, Canada. Inaugurated: 1889. Race type: Thoroughbred - Flat racing. Website: Woodbine Racetrack.

 There are also images of the racecourse and a horse race.

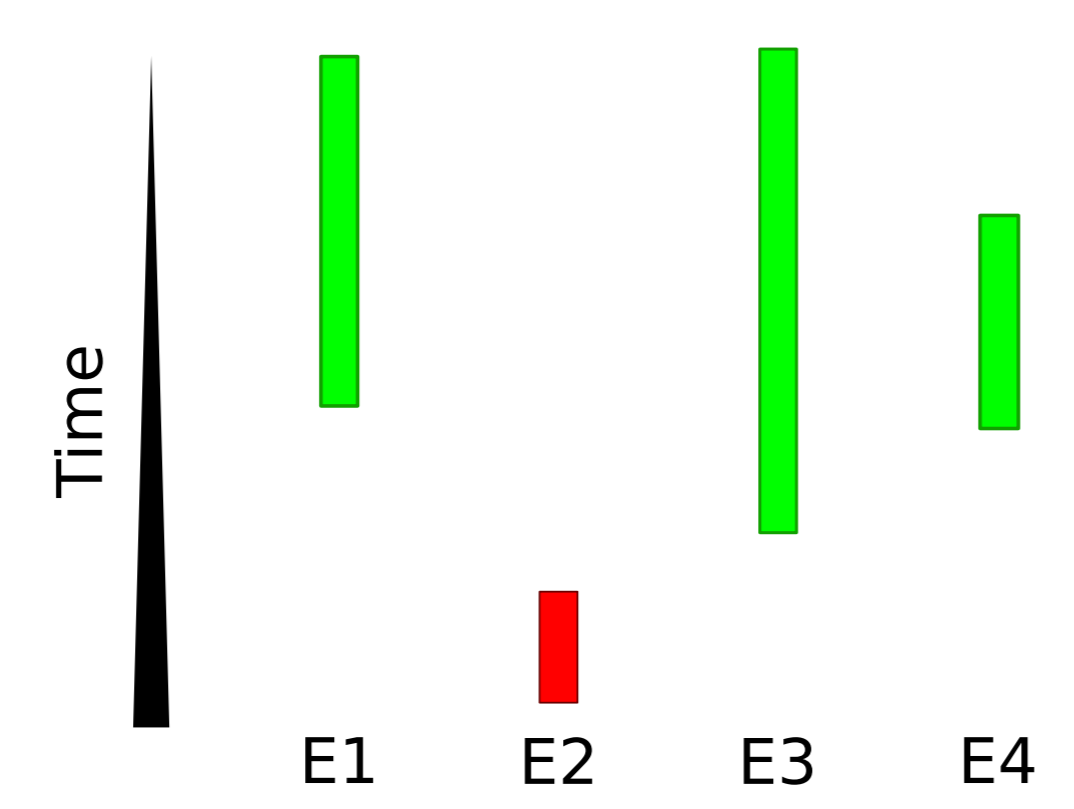
### Results



### Geographic Coherence



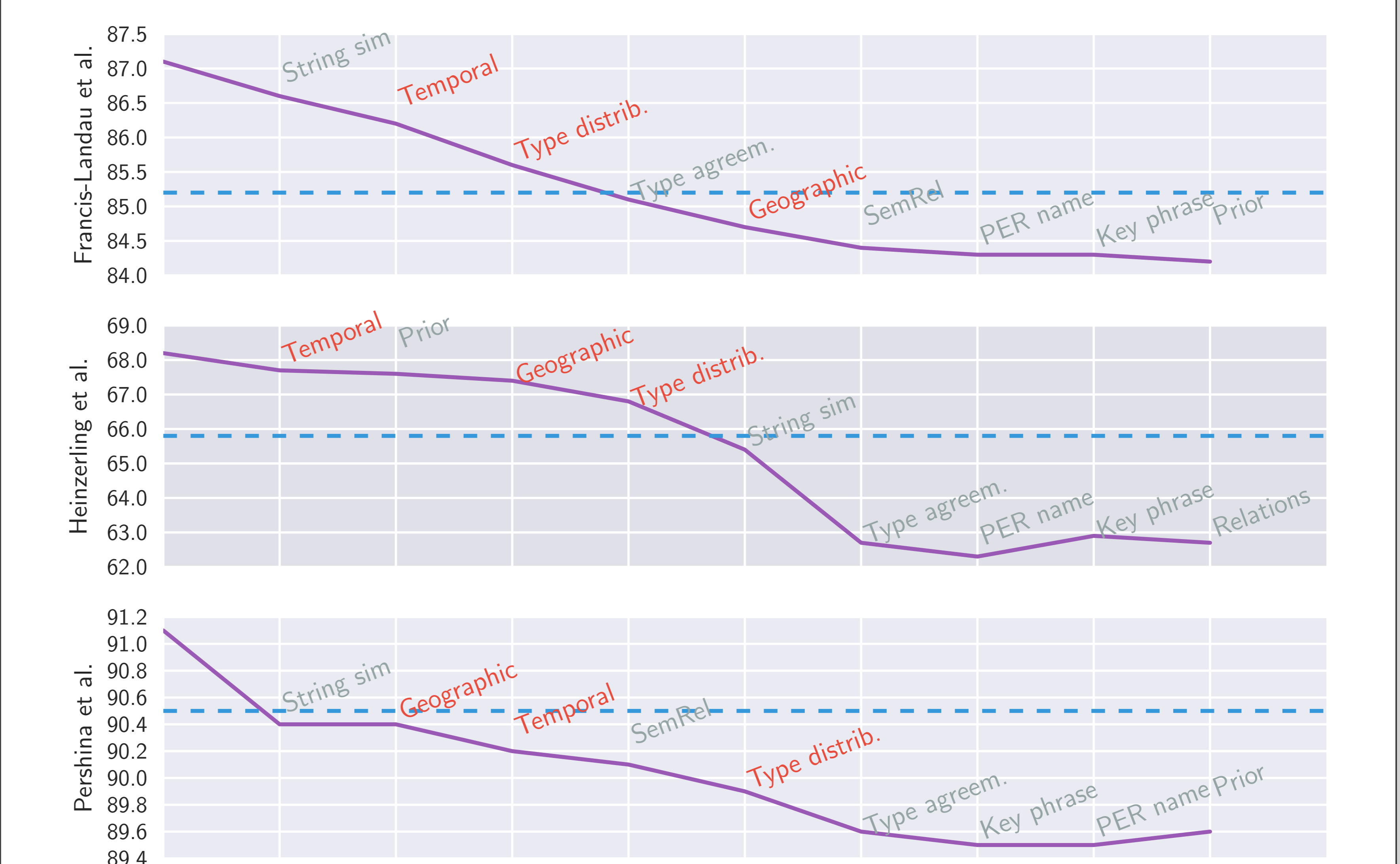
### Temporal Coherence



### Entity Type Coherence

TF-IDF Count	Type
1116	2 :base.rugby.rugby_club
244	3 :organization.organization
232	2 :base.schemastaging.sports_team_extra
184	2 :sports.sports_team
56	2 :base.tagit.concept

### Feature Set Ablation



benjamin.heinzerling@h-its.org

**Acknowledgements:** This work has been supported by the German Research Foundation as part of the Research Training Group "Adaptive Preparation of Information from Heterogeneous Sources" (AIPHES) under grant No. GRK 1994/1.