Nomen est omen

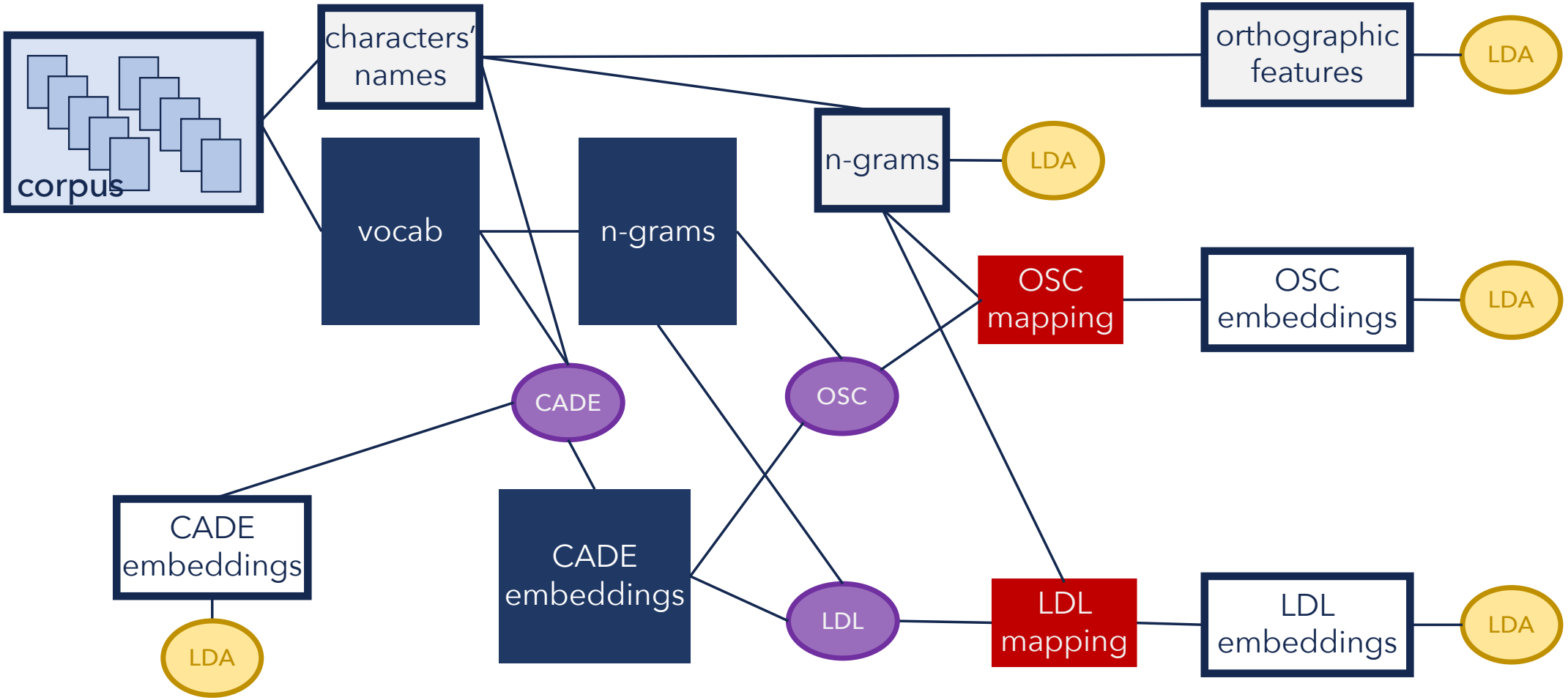
Fictional characters' names encode polarity, gender, and age

Fabiënne Reedijk, Stefano Scola, Niccolò Minetti, Niveditha Subramaniam, Giovanni Cassani

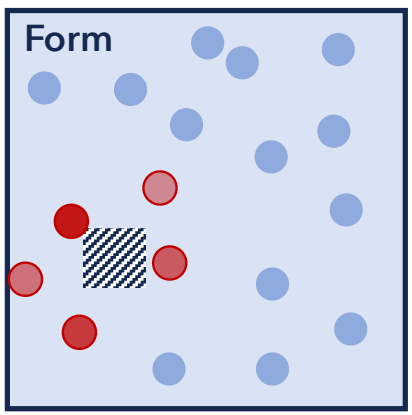
AMLaP 2021

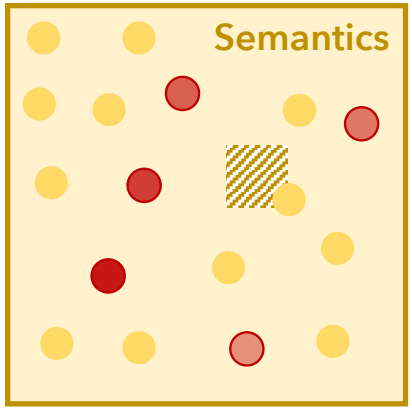
Tilburg University

The pipeline



Form to meaning





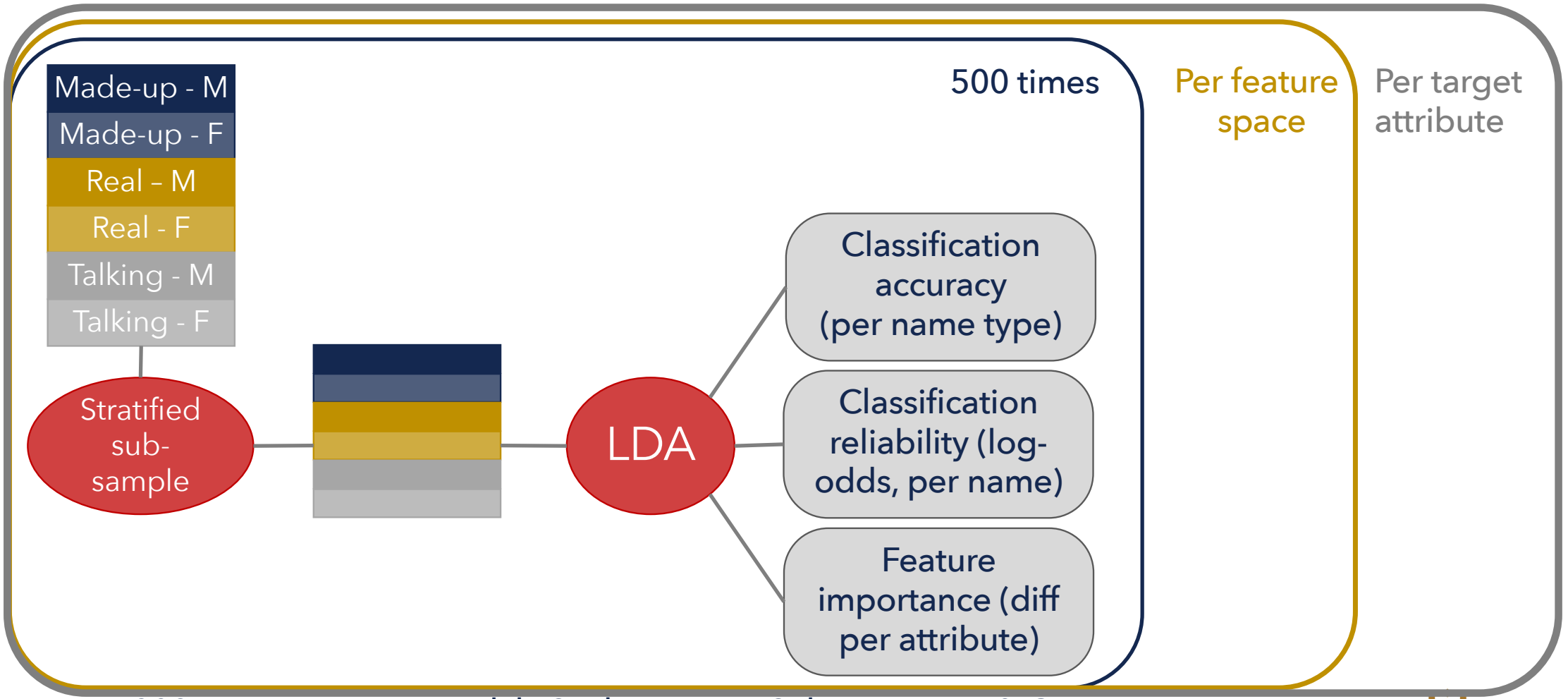
OSC: analogise in form space and average semantic neighbors

LDL: learn linear mapping from form to meaning and apply to novel words

Questions

- Can we discriminate characters' feature vectors in the different representational spaces?
- Is there a correlation between how difficult it is to discriminate characters in different feature spaces?
- Can we find reliable predictors?

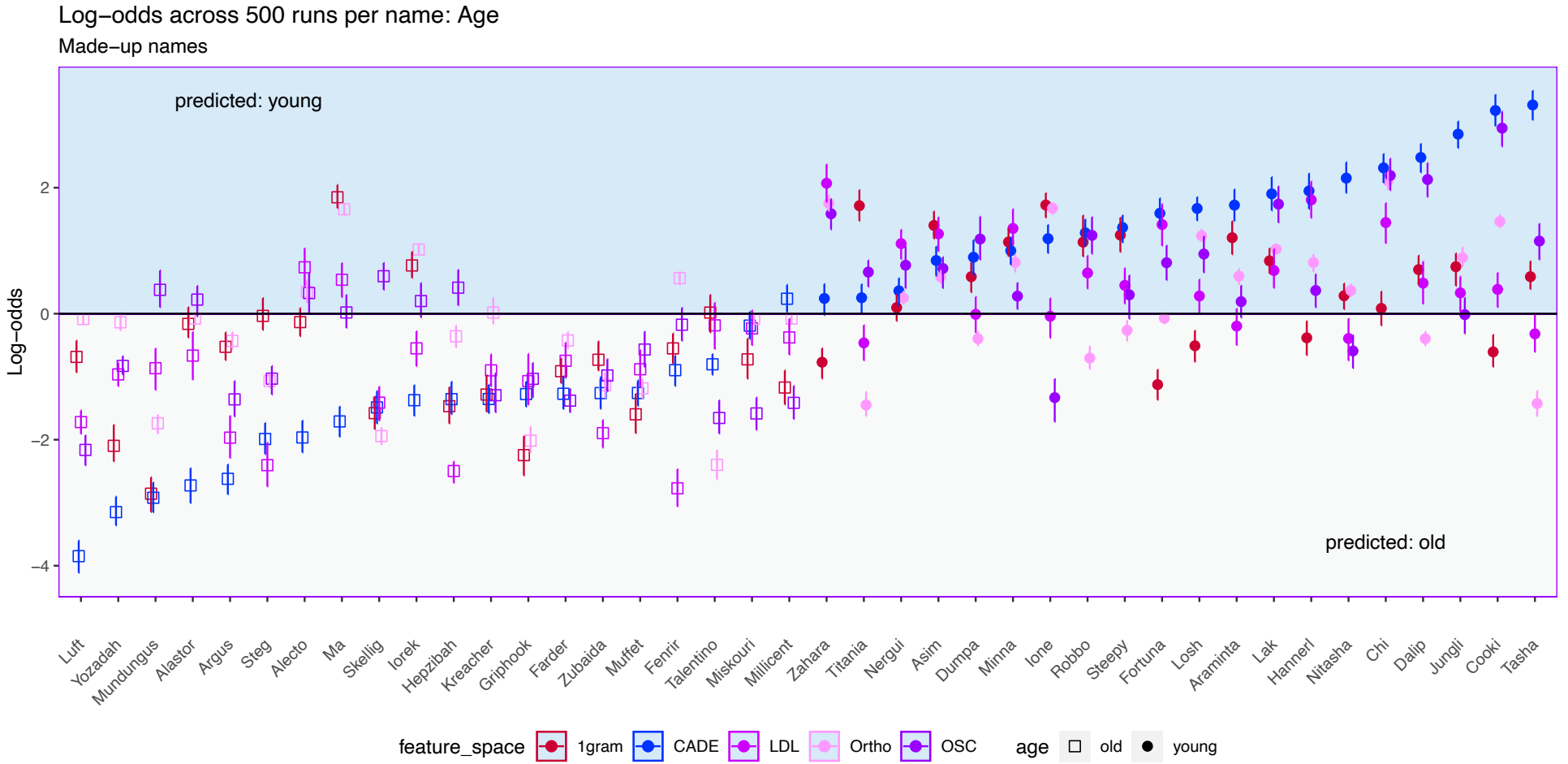
Bootstrapped LDAs



Classification accuracy

	Age		
	Real	Talking	Made-up
Unigrams	0.655 [0.651, 0.658]	0.606 [0.603, 0.610]	0.758 [0.754, 0.761]
Ortho features	0.690 [0.688, 0.693]	0.559 [0.556, 0.562]	0.688 [0.685, 0.691]
CADE	0.878 [0.875, 0.880]	0.956 [0.955, 0.958]	0.955 [0.953, 0.957]
LDL	0.847 [0.844, 0.850]	0.743 [0.739, 0.747]	0.775 [0.772, 0.779]
OSC	0.780 [0.776, 0.783]	0.801 [0.797, 0.804]	0.755 [0.751, 0.758]

Log-odds per name



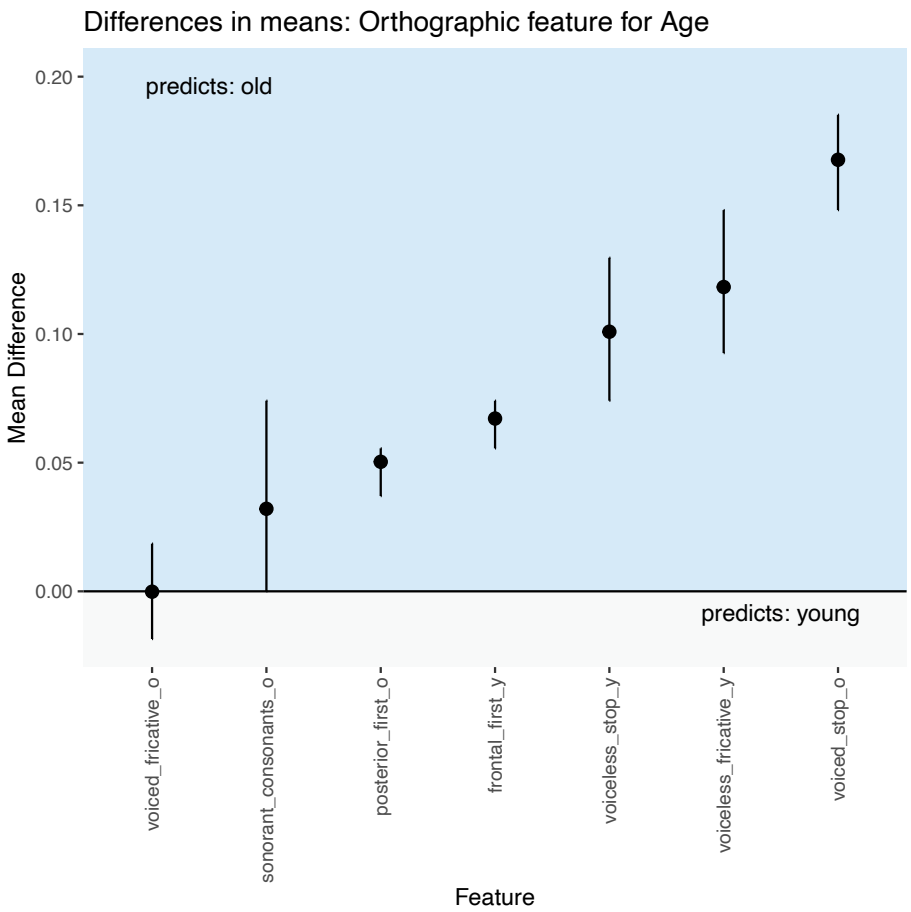
Reedijk, Scola, Minetti, Subramaniam, & Cassani

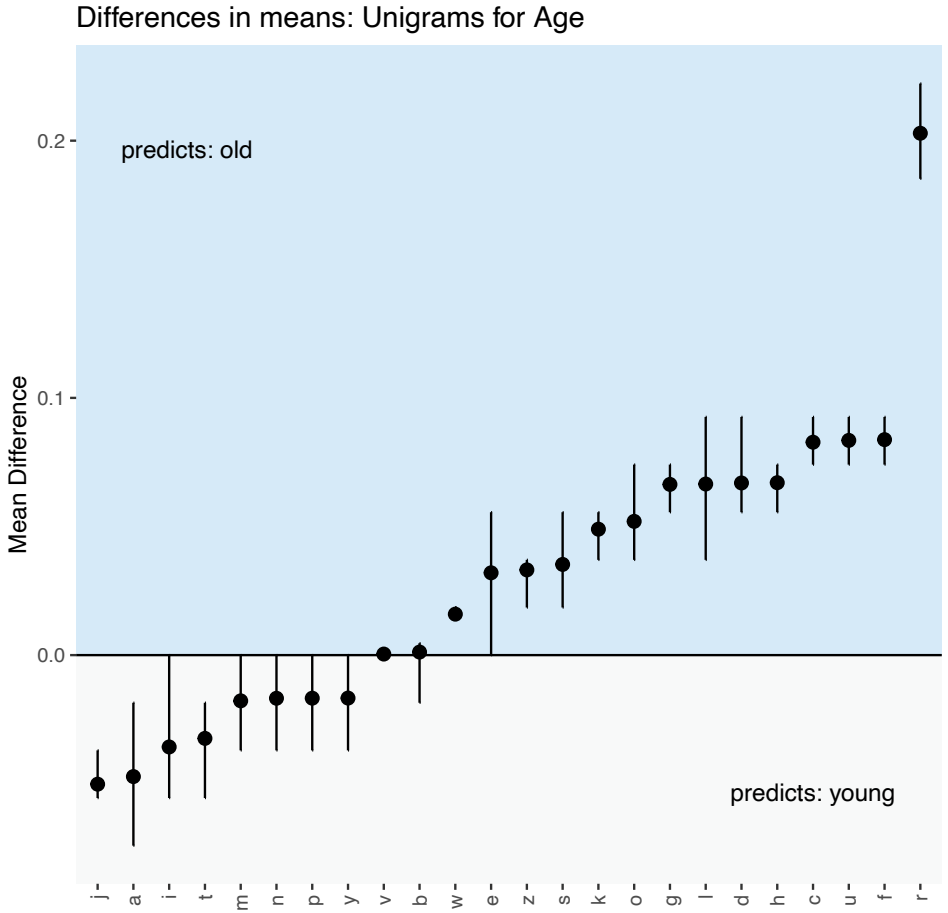
Tilburg University

Correlation Matrix

Age	CADE	OSC	LDL	Ortho
OSC	0.535			
LDL	0.466	0.412		
Ortho	0.198	0.228	0.313	
1grams	0.302	0.269	0.265	0.397

Feature discriminability: form

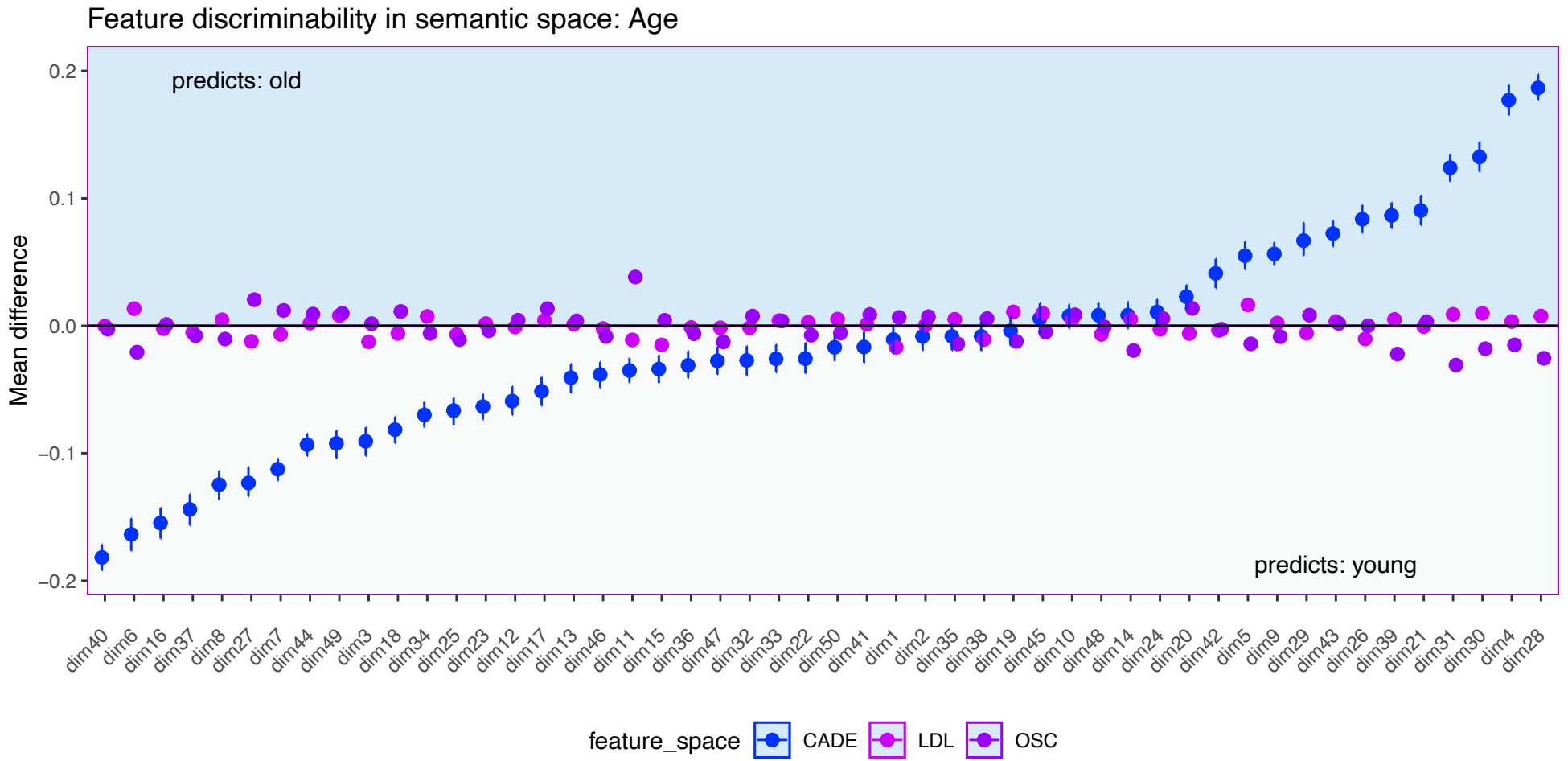




Reedijk, Scola, Minetti, Subramaniam, & Cassani

Tilburg University

Feature discriminability: semantics



Summary of results

- Very **good discriminability in semantic space**
- *Above chance discriminability in form space*
- **Very good discriminability in form-based semantic spaces**
- Moderate correlation between log-odds in different feature spaces
- Feature reliability makes sense (**r** for *male, old, evil*; CADE more discriminable than OSC and LDL)

Take-home

- Characters sharing a trait are described similarly across stories
- Characters sharing a trait are *named more similarly* than characters having a different trait
- It is possible to **infer semantic attributes from form alone**
- More **prototypical characters** along a trait are named with more **prototypical names** for that trait
- Sound-symbolic devices are consistent in form and consistently point to semantic space

Thank you!