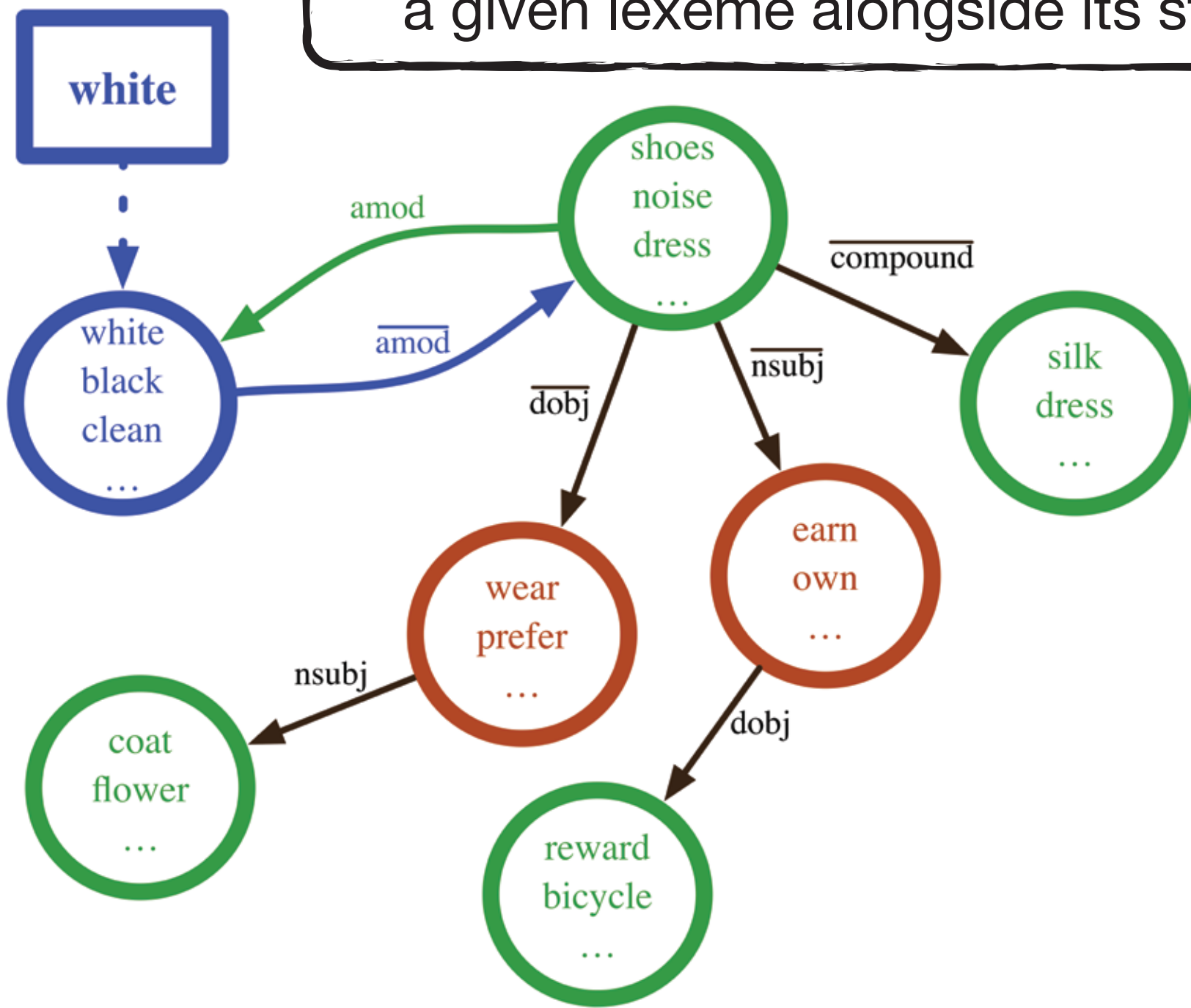


# Improving Semantic Composition with Offset Inference

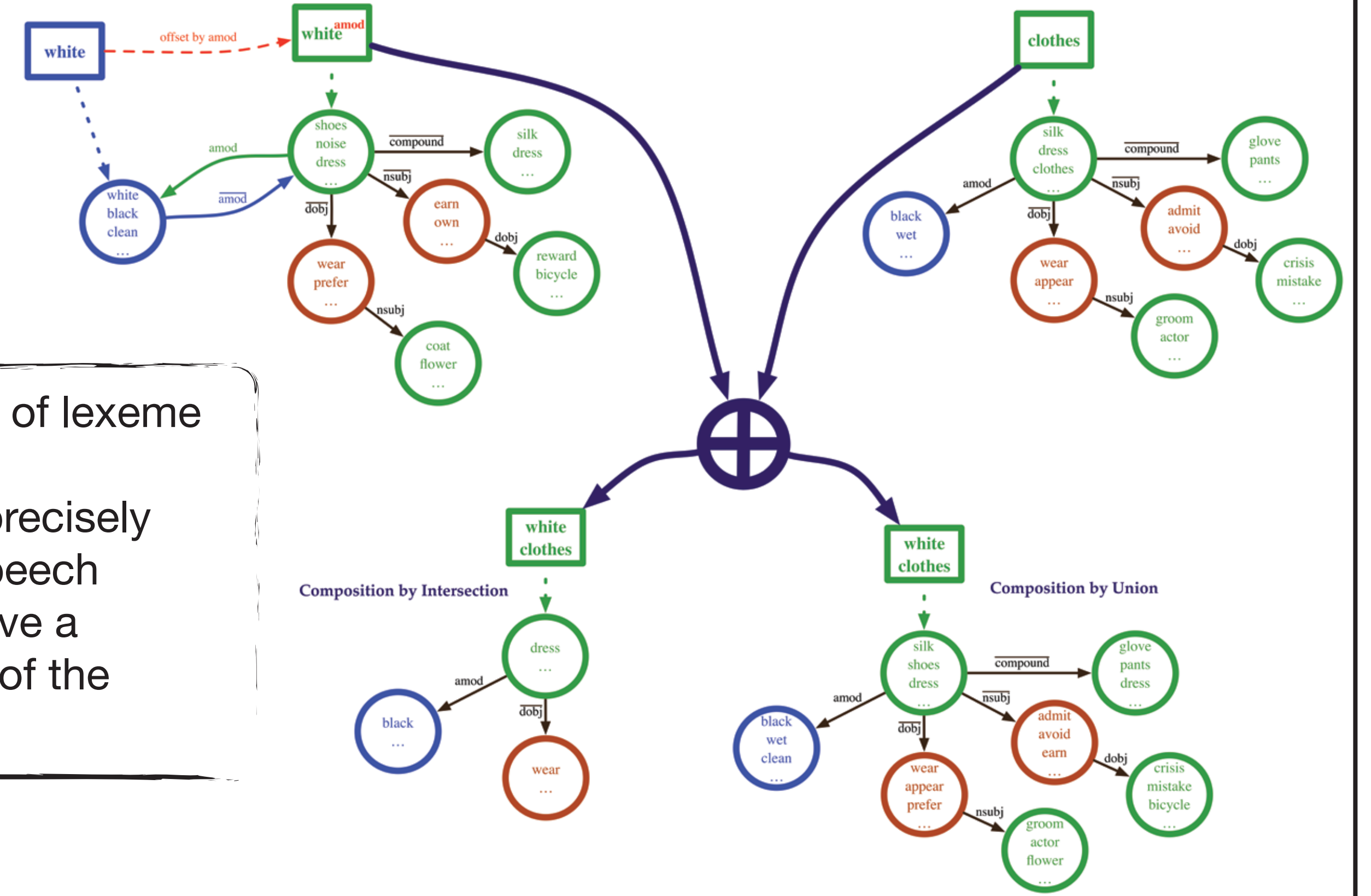
Thomas Kober, Julie Weeds, Jeremy Reffin and David Weir  
TAG Laboratory, University of Sussex

## Anchored Packed Trees (APTs)

- Lexemes are represented with a single higher-order dependency-typed structure
- APTs encode the distributional semantics of a given lexeme alongside its structure

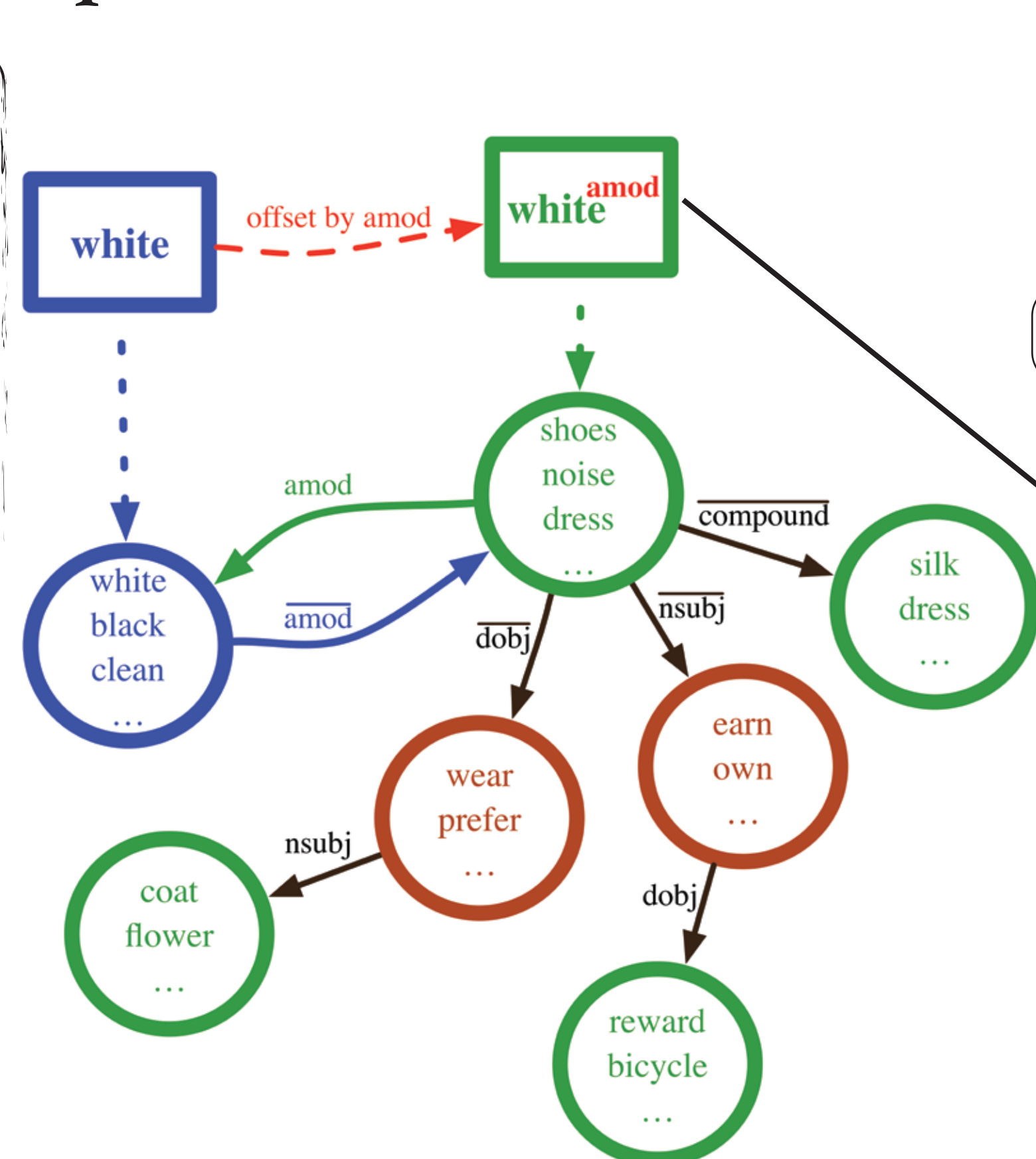


- Distributional composition is a process of lexeme contextualisation (Weir et al., 2016)
- The structure of APTs is important to precisely align lexemes with different parts of speech
- A composed sentence is a way to derive a contextualised representation of each of the lexemes in the sentence



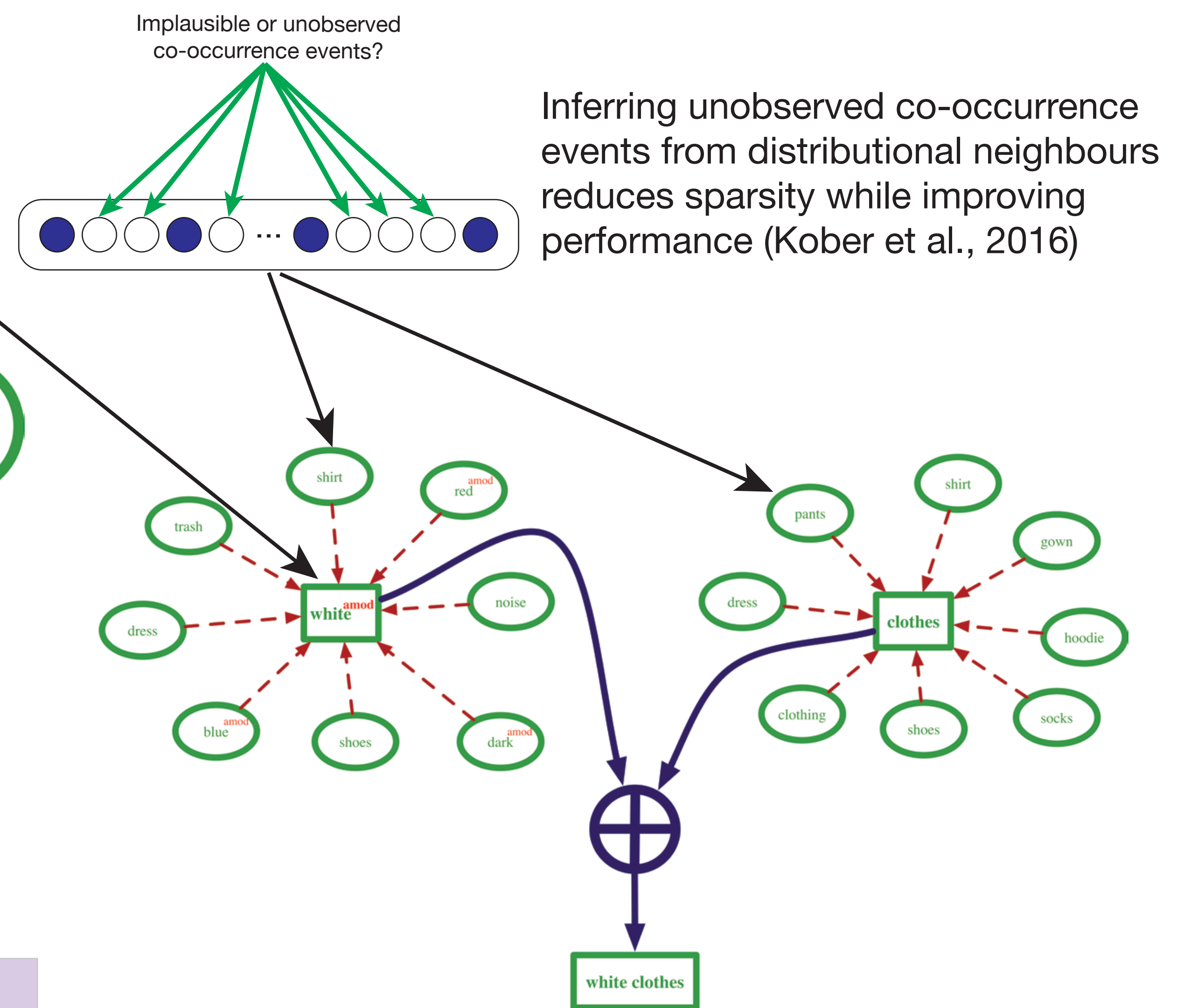
## Offset APT Representations

- Lexemes with different parts of speech live in different areas of the semantic space
- Offsetting is a mechanism to align the representations
- Central part of composition in APTs
- Represents a "things-that-can-be-white" structure when the adjective *white* is offset into noun space



white	white <sup>amod</sup>	clothes
:clean	amod:clean	amod:wet
amod:shoes	:shoes	:dress
amod.dobj:wear	dobj:wear	dobj:wear
amod.nsubj:earn	nsubj:earn	nsubj:admit

## Offset Inference



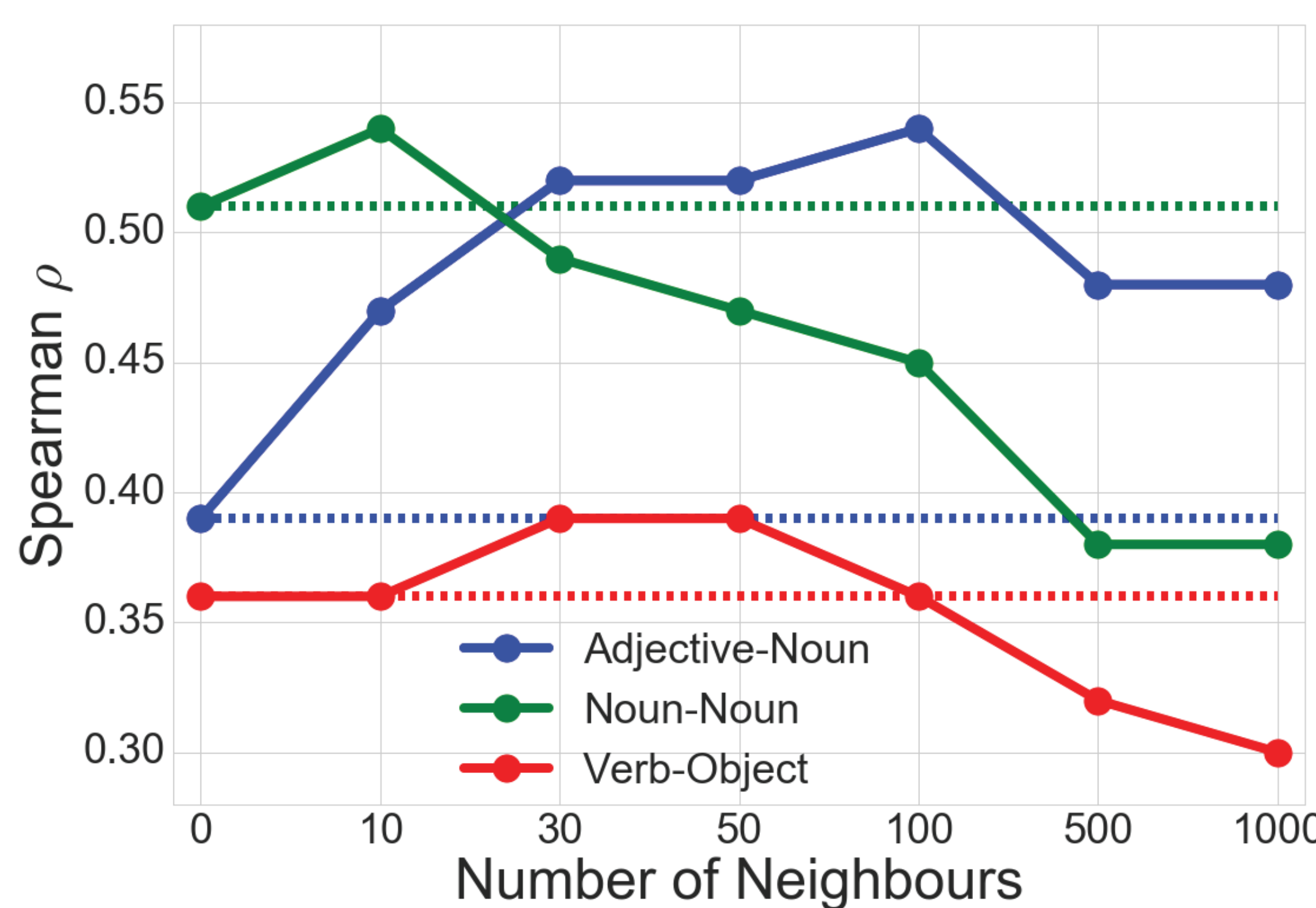
Inferring unobserved co-occurrence events from distributional neighbours reduces sparsity while improving performance (Kober et al., 2016)

	Representation	Nearest Neighbours
ancient vs. an "ancient thing"	ancient	medieval, greek, historic, modern, egyptian
	ancient <sup>amod</sup>	civilisation, mythology, tradition, ruin, monument
mother vs. actions done by a mother	mother	father, sister, wife, husband, daughter
	mother <sup>nsubj</sup>	wife <sup>nsubj</sup> , father <sup>nsubj</sup> , parent <sup>nsubj</sup> , woman <sup>nsubj</sup>
law vs. actions done by the law	law	legislation, policy, rule, practice, politics
	law <sup>dobj</sup>	violate, rule <sup>dobj</sup> , enact, repeal, principle
actions done to the law vs.	law <sup>nsubj</sup>	rule <sup>nsubj</sup> , principle <sup>nsubj</sup> , policy <sup>nsubj</sup> , criminalise

- Offset inference leverages the structure of APTs
- Infer unobserved co-occurrence events from contextualised representations
- Generalises the standard distributional inference algorithm in APTs, inference is based on offset APTs instead of standard distributional neighbours
- Offset inference and distributional composition realised by the same operation

## Results

Distributional Inference	ML10				ML08
	AN	NN	VN	Avg	VN
None	0.35	0.50	0.39	0.41	0.22
Standard	0.48	0.51	0.43	0.47	0.29
Offset	<b>0.49</b>	<b>0.52</b>	<b>0.44</b>	<b>0.48*</b>	<b>0.31*</b>



- Saturation effect of the number of neighbours added
- Up to the saturation point, DI is inferring useful information
- Beyond that DI degrades to just generic smoothing

- Compositionality benchmarks of Mitchell and Lapata (2008) (ML08), and Mitchell and Lapata (2010) (ML10)
- Substantial improvements over a baseline without an inference mechanism
- Small, but consistent and statistically significant improvements over standard distributional inference

## References

Thomas Kober, Julie Weeds, Jeremy Reffin and David Weir. 2016. *Improving Sparse Word Representations with Distributional Inference for Semantic Composition*. In Proceedings of EMNLP  
 Jeff Mitchell and Mirella Lapata. 2008. *Vector-based Models of Semantic Composition*. In Proceedings of ACL  
 Jeff Mitchell and Mirella Lapata. 2010. *Composition in distributional models of semantics*. Cognitive Science  
 David Weir, Julie Weeds, Jeremy Reffin and Thomas Kober. 2016. *Aligning Packed Dependency Trees: A theory of composition for distributional semantics*. Computational Linguistics