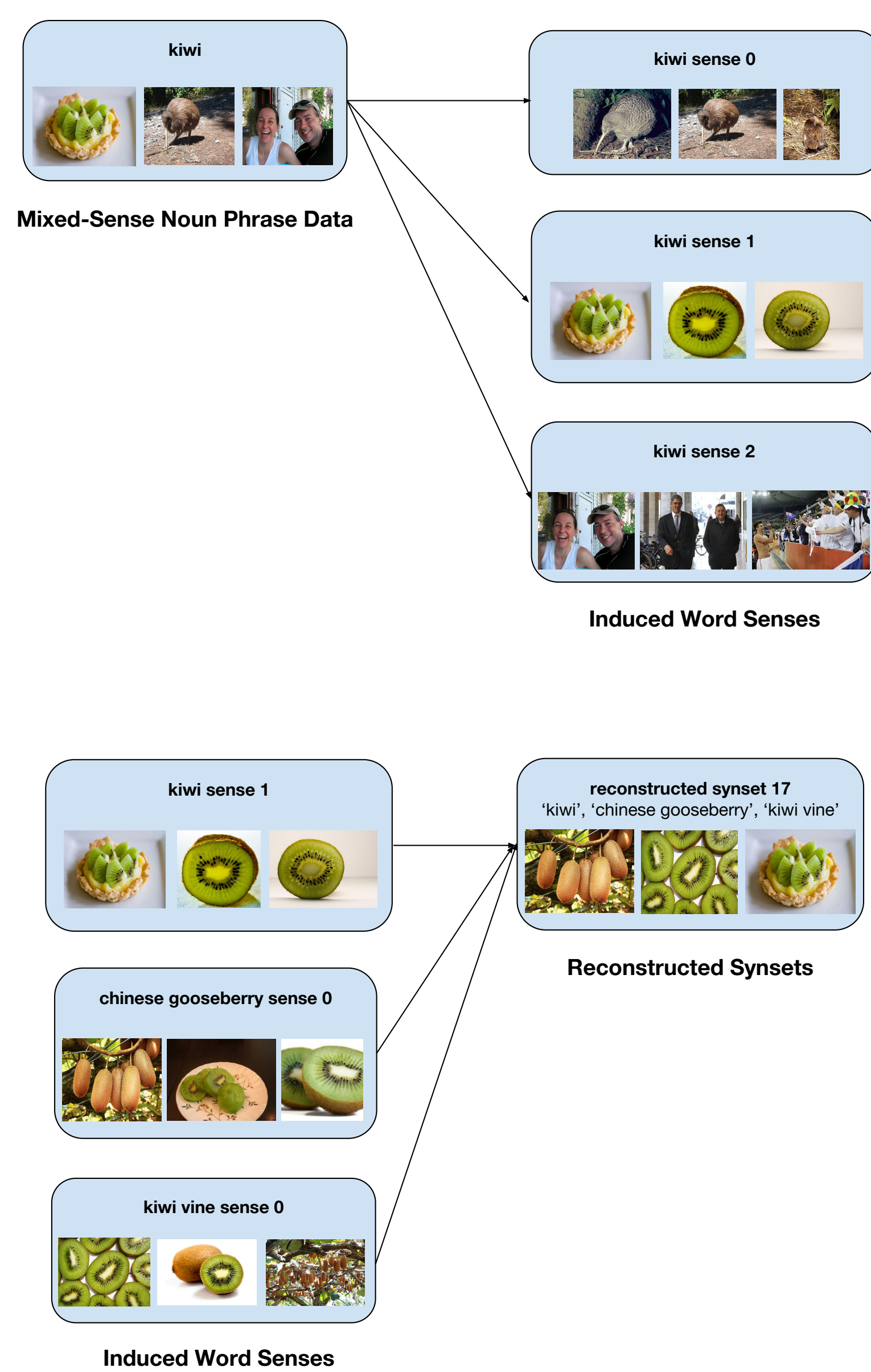


Synonymous Set of Word Senses (Synset) Induction

A word in natural language can be polysemous, having multiple meanings, as well as synonymous, meaning the same thing as other words. Word sense induction finds the senses of polysemous words. Synonymy detection finds when two words are interchangeable. We combine these tasks, first inducing word senses and then detecting similar senses to form word-sense synonym sets (*synsets*) in an unsupervised fashion using instances of noun phrases represented as both vectorized text (as an LSA embedding (Deerwester *et al.* 1990)) and by an image (as VGG's penultimate layer (Simonyan *et al.*, 2014)).

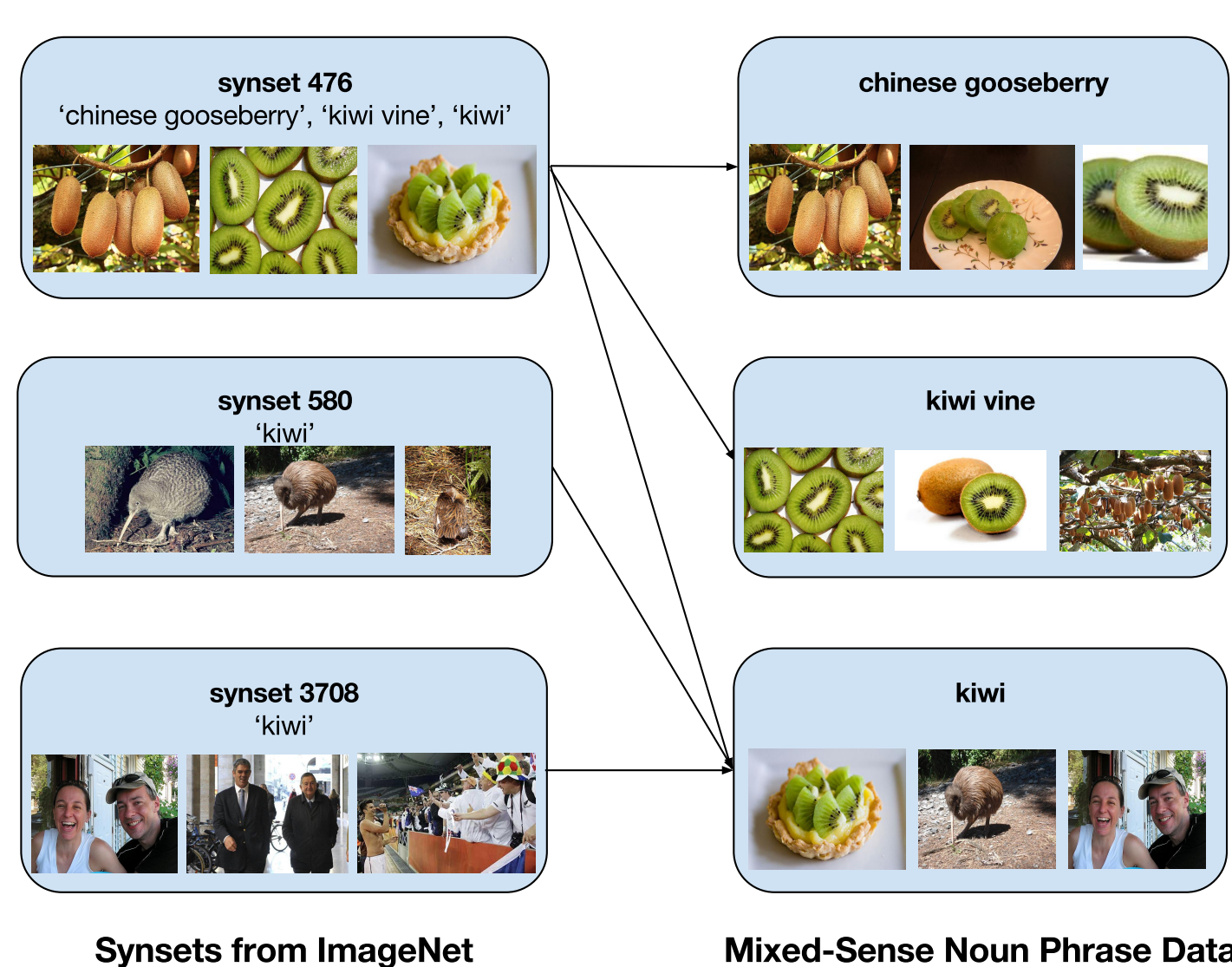


Clustering within each noun phrase's observations yields induced senses. Clustering is performed over these induced senses' means to construct synsets.

Multi-modal Dataset

Noun phrase relationships			
synonymous	polysemous	both	neither
4019	804	1017	2586

Noun phrases relationships for the ImageNet subset we consider.



Noun phrase observations are extracted from ImageNet synsets. Our task is to automatically induce synsets from these noun phrase's image and text observations in an unsupervised fashion. For each image, we use reverse image search to extract relevant text data.

Methodology

Cosine distance between observations is calculated in text-only, image-only, or multi-modal space (using a 0.5 weight between spaces).

Polysemy detection. We perform *k*-means clustering, selecting *k* by the gap statistic (Tibshirani *et al.*, 2001), for polysemy detection. The gap statistic selects the smallest number of clusters *k* that reduces within-dispersion compared to *k* - 1 by more than chance.

Synonymy detection. To form synsets from senses, we use a greedy merging approach. We compute a mean observation vector for each induced sense and greedily merge the nearest means to produce a final set of induced synsets.

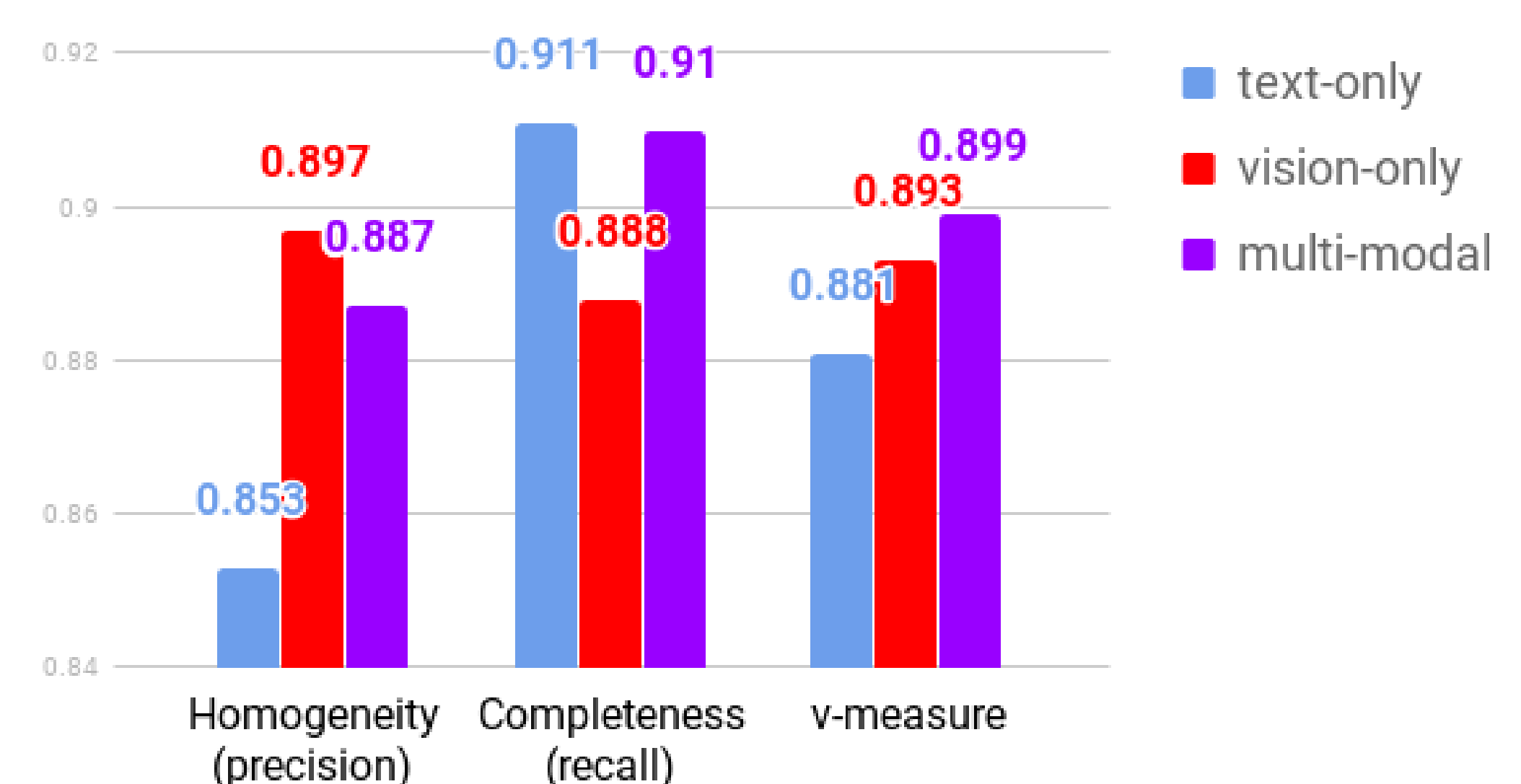
Results

Synsets containing the senses of the noun phrase "washboard" by different induction modalities:

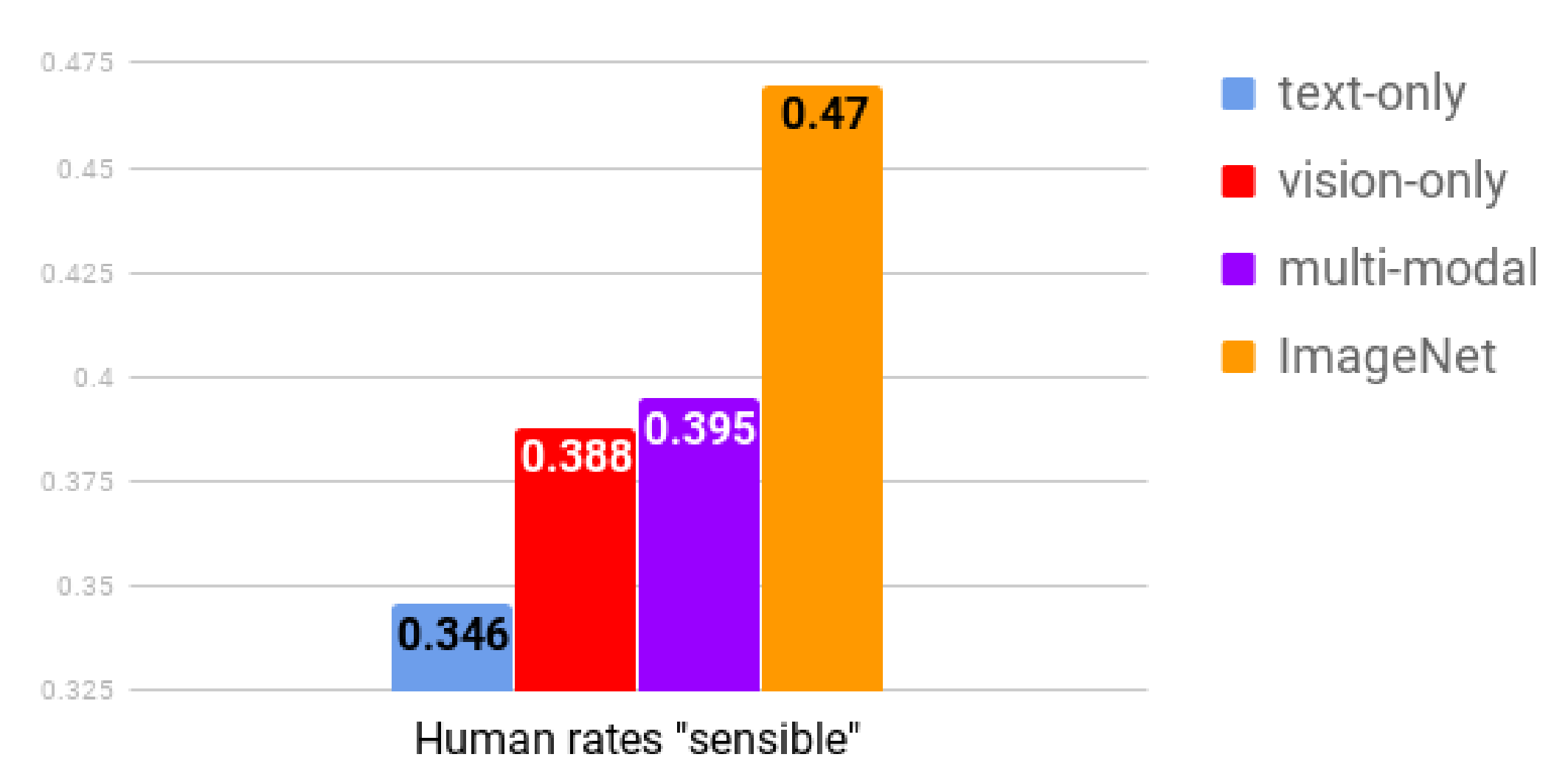


We use WSI metrics (Manandhar *et al.*, 2010) for an automated evaluation, and perform a Mechanical Turk study to evaluate the quality of induced synsets.

Synset Agreement with ImageNet



Human Evaluation



Humans rate our synsets from multi-modal induction as sensible about 84% as often as ImageNet's.