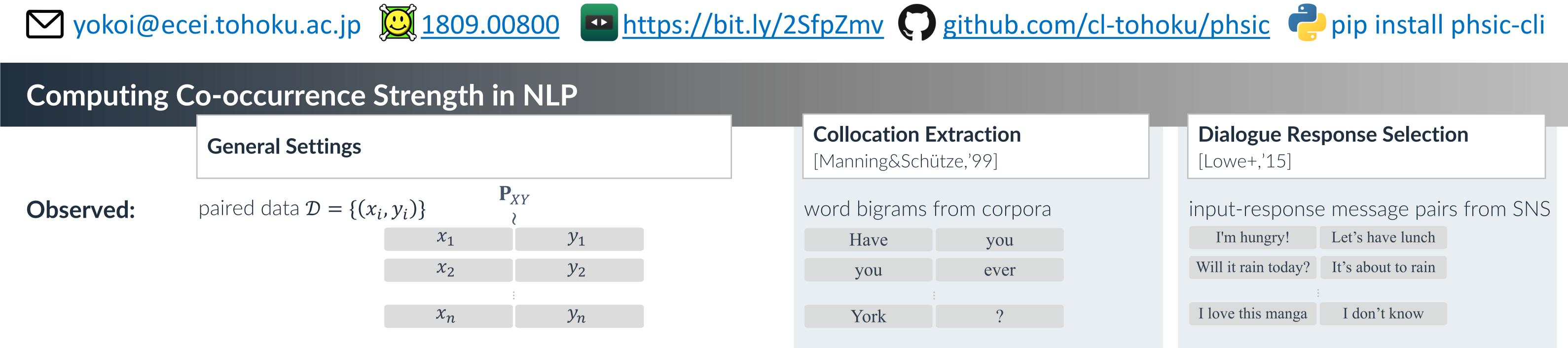
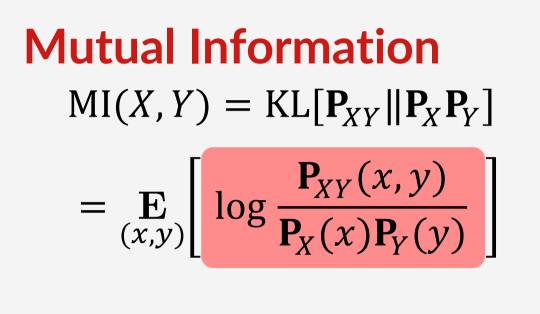
Pointwise HSIC: A Linear-Time Kernelized Co-occurrence Norm for Sparse Linguistic Expressions [EMNLP2018]

Sho Yokoi (Tohoku U./RIKEN AIP)*, Sosuke Kobayashi (PFN), Kenji Fukumizu (ISM), Jun Suzuki *, Kentaro Inui *

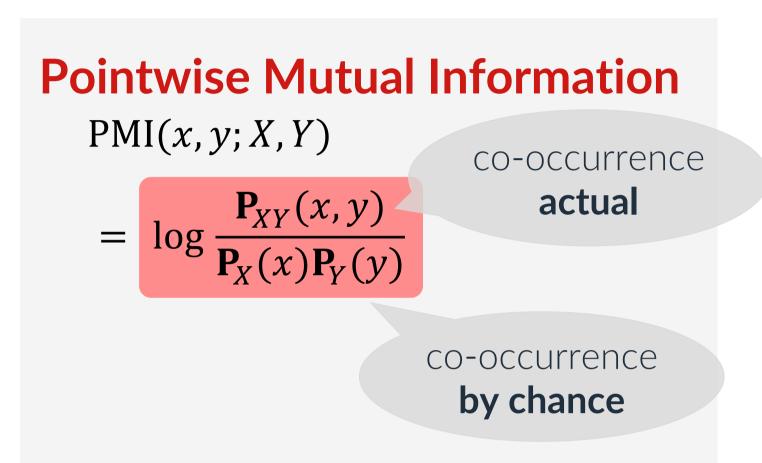


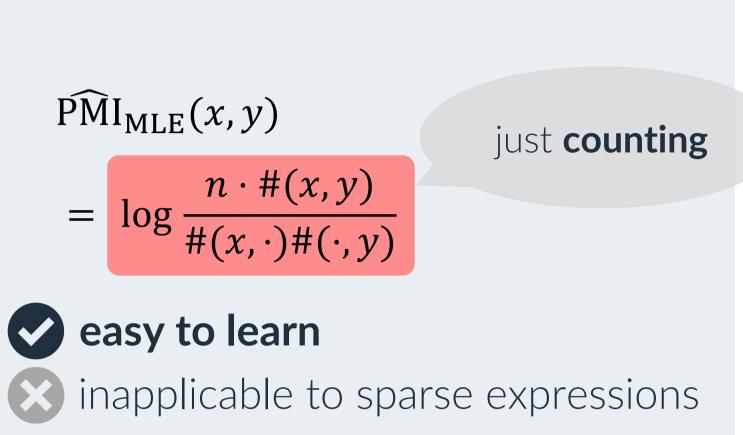
De facto Measure: Pointwise Mutual Information

compute co-occurrence strength of a pair (x, y)



Task:



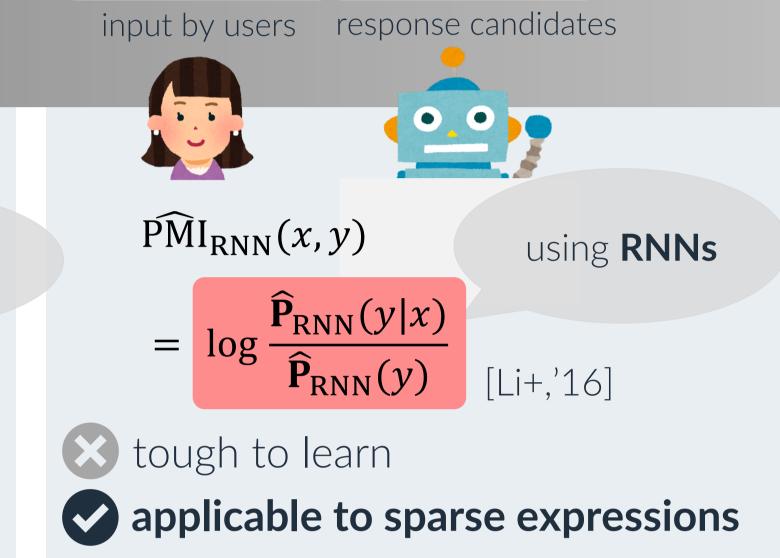


York

the

which pairs are collocation?

in



I'm so sleepy

I don't know

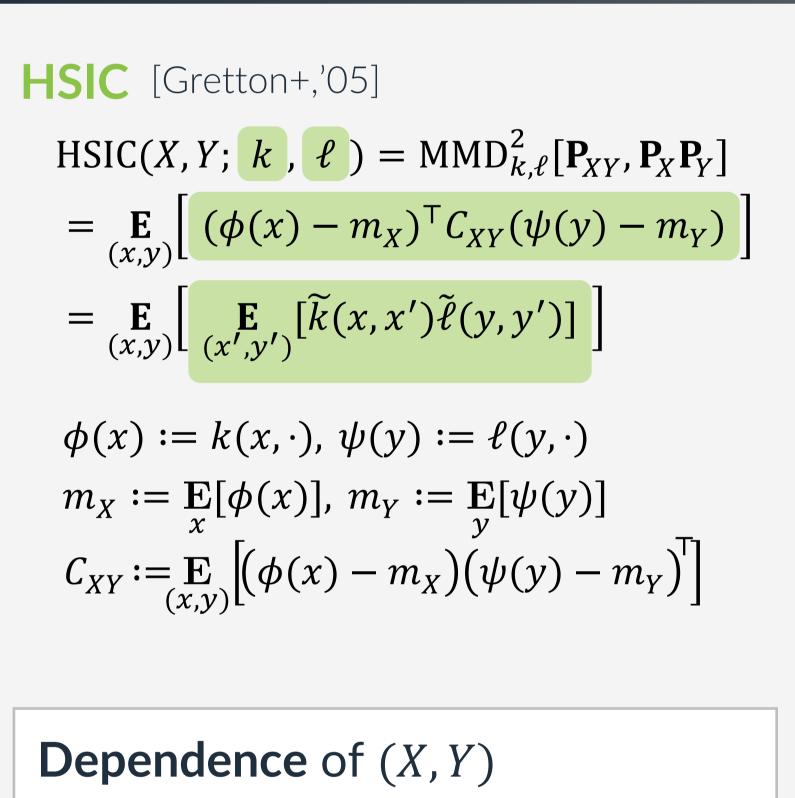
which response is the best?

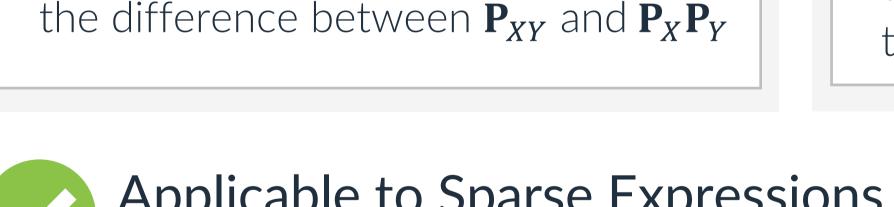
I've lost my wallet I saw it at the ...

I've lost my wallet

I've lost my wallet

Proposed Measure : Pointwise HSIC





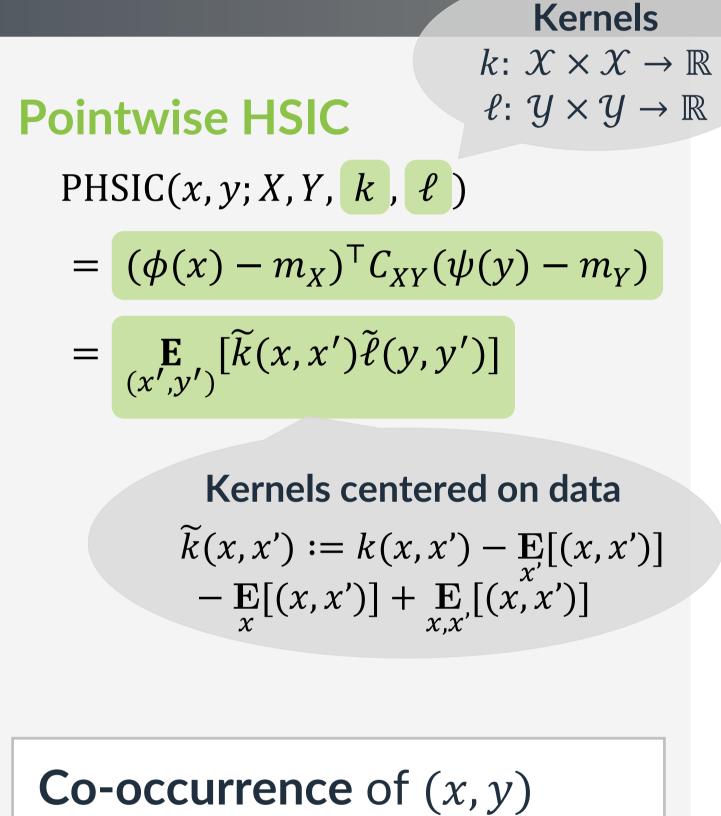
 $n = 5 \times 10^5$

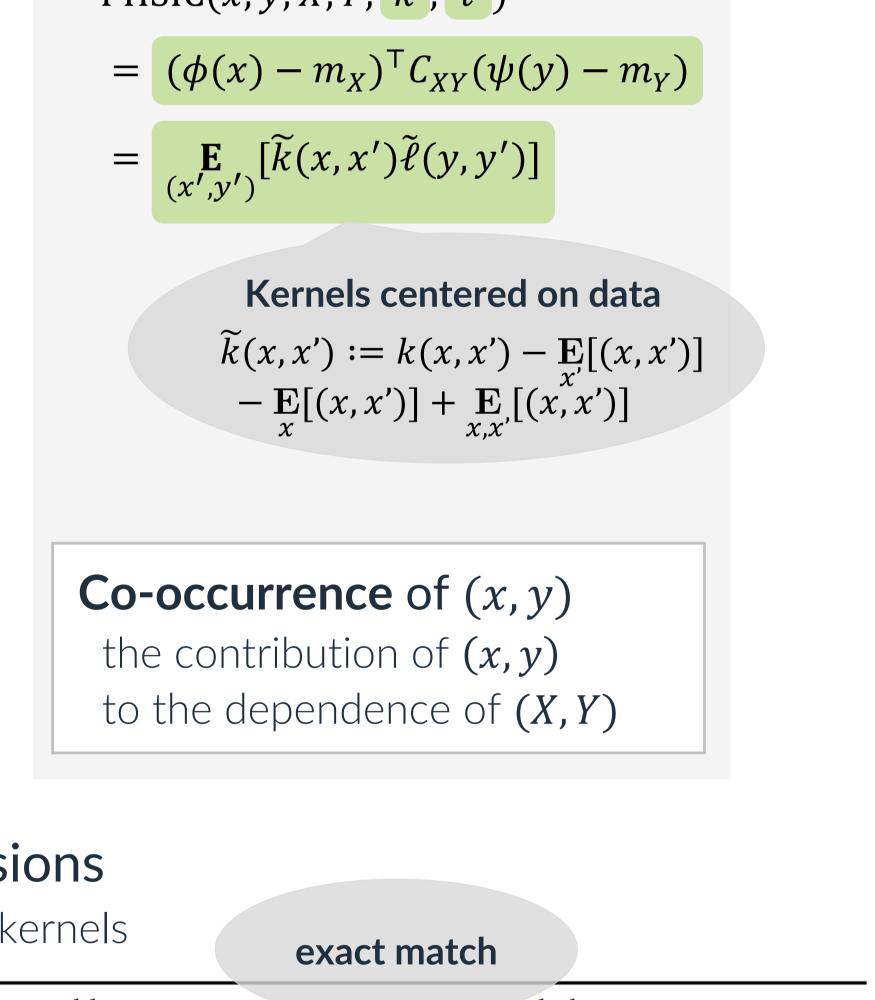
 $n = 10^3$

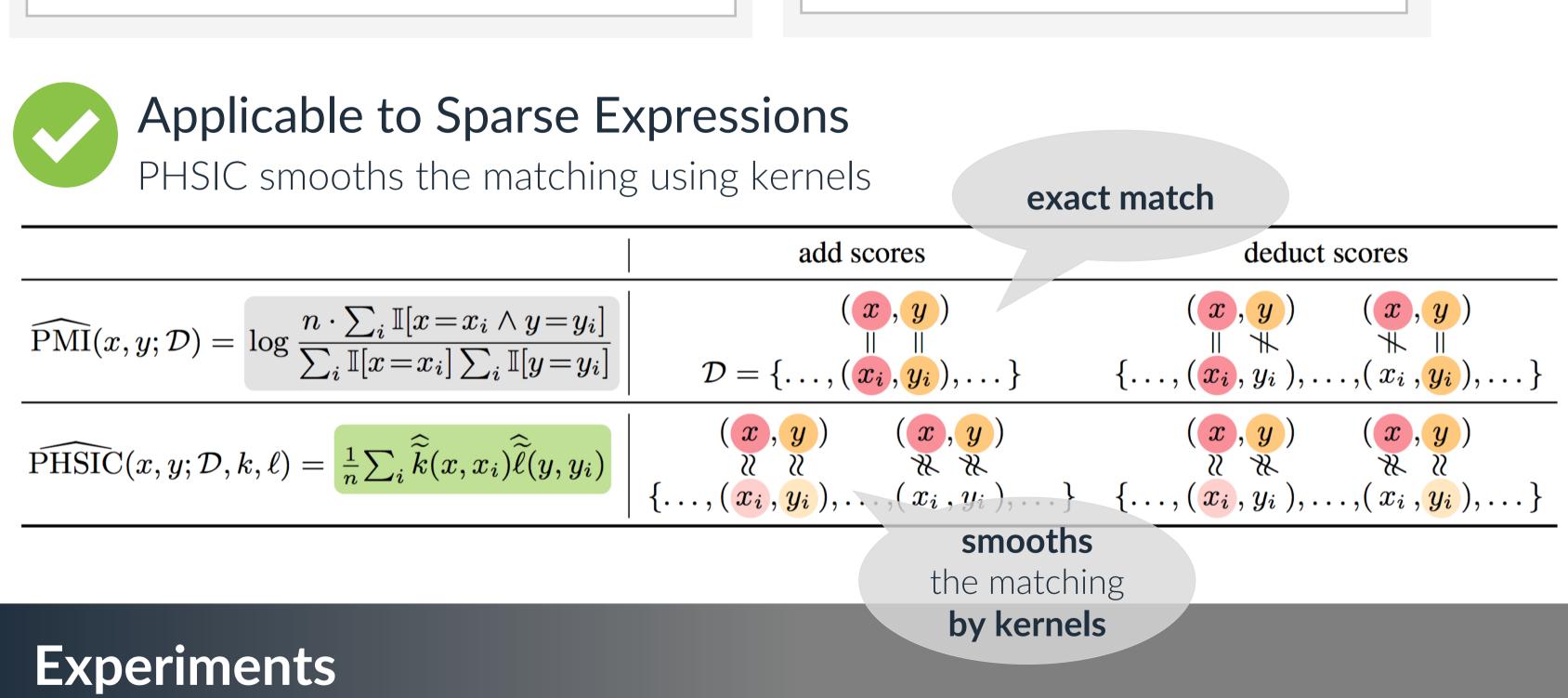
PMI (w/RNN)

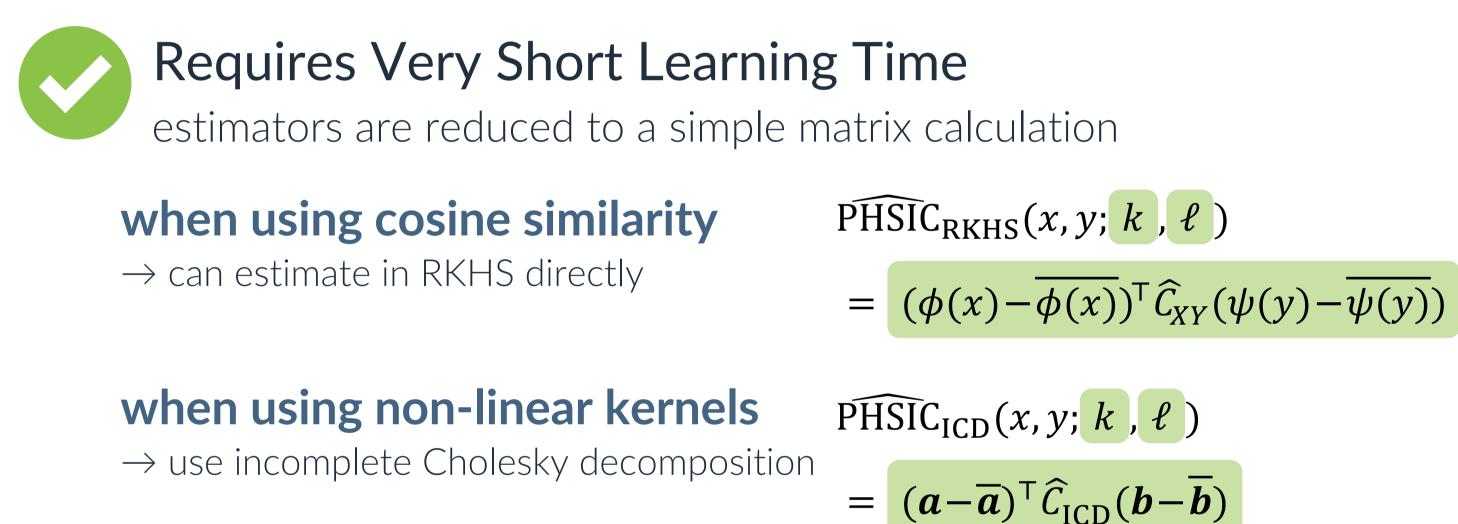
PMI (w/RNN)

PHSIC (cos of USE)









- Kernels are Available
 - Cosine Similarity between Sentence Vectors
 - Sentence vectors [Kiros+,'15; Dai&Le,'15; Iyyer+,'15; Hill+,'16; Cer+,'18]
 - Sum of word vectors [Mikolov+,'13; Pennington+,'14; Bojanowski+,'17]
 - Many pre-trained models are off-the-shelf!
 - Structure Kernels
 - o [Collins&Duffy,'02; Bunescu&Mooney,'06; Moschitti,'06]
 - Combinations

