An Empirical Investigation of Discounting in Cross-Domain Language Models

Berkeley

N L P

Greg Durrett and Dan Klein
UC Berkeley

Discounting

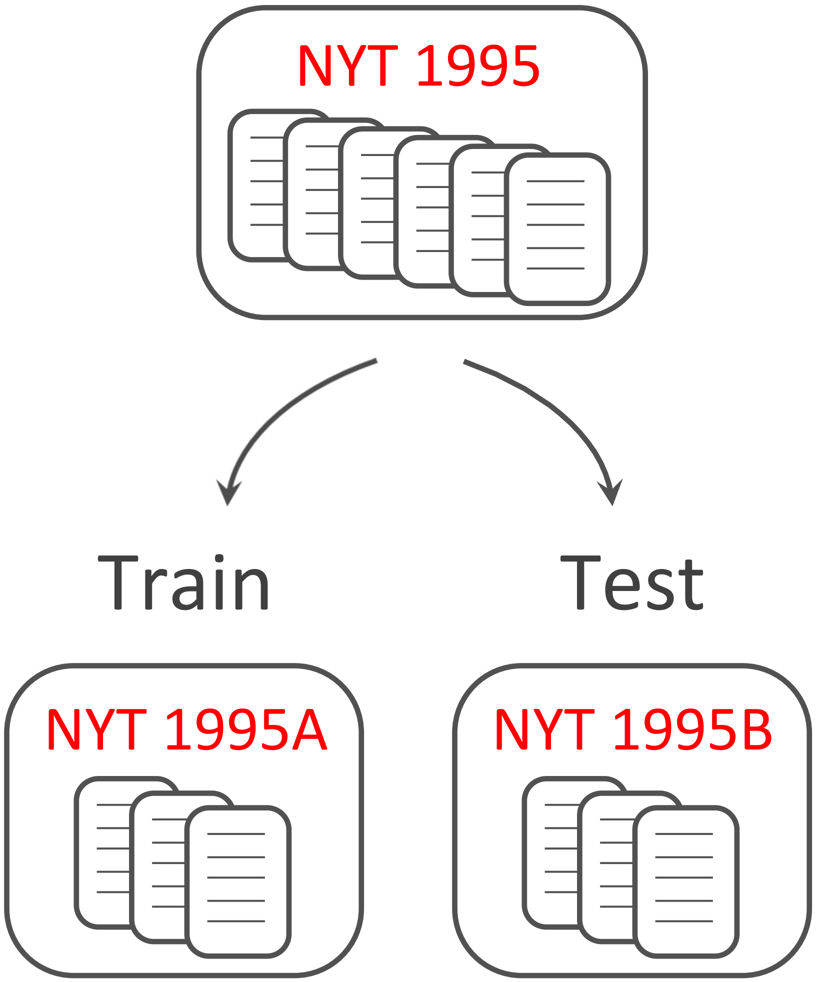
1

c(test) = 2

Discounting

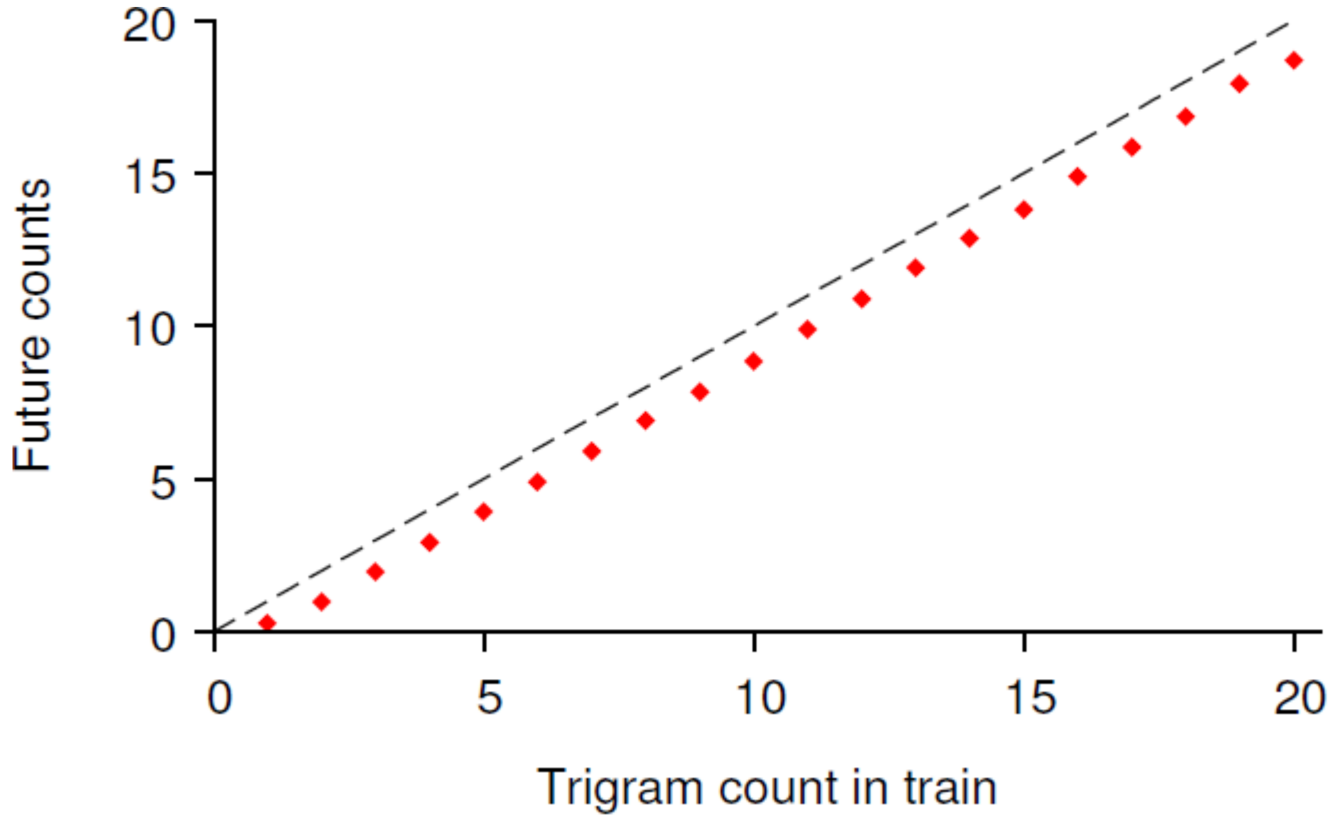
1	c*(1) = 0.28	"future count"
2	c*(2) = 0.97	
3	c*(3) = 1.96	

Experimental Setup

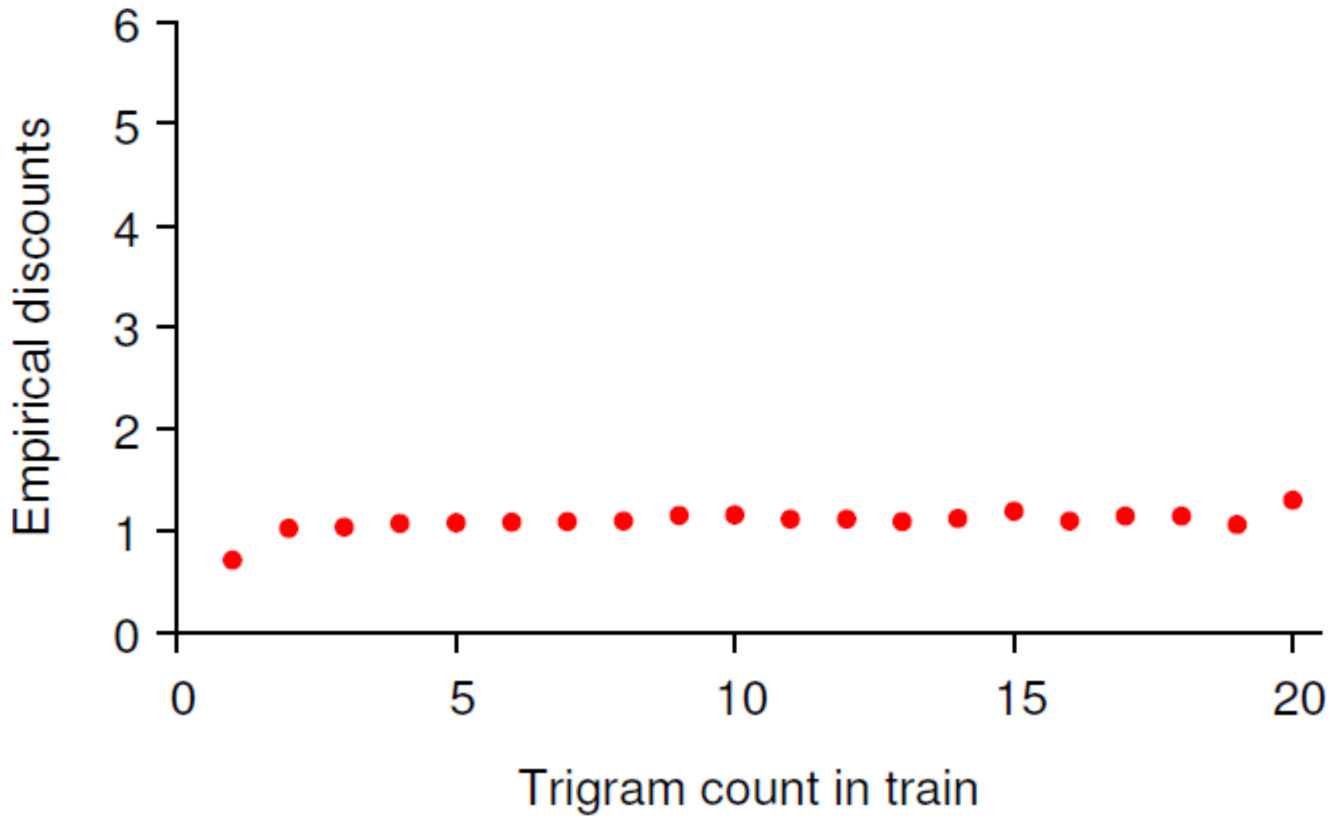


[Church and Gale, 1991]

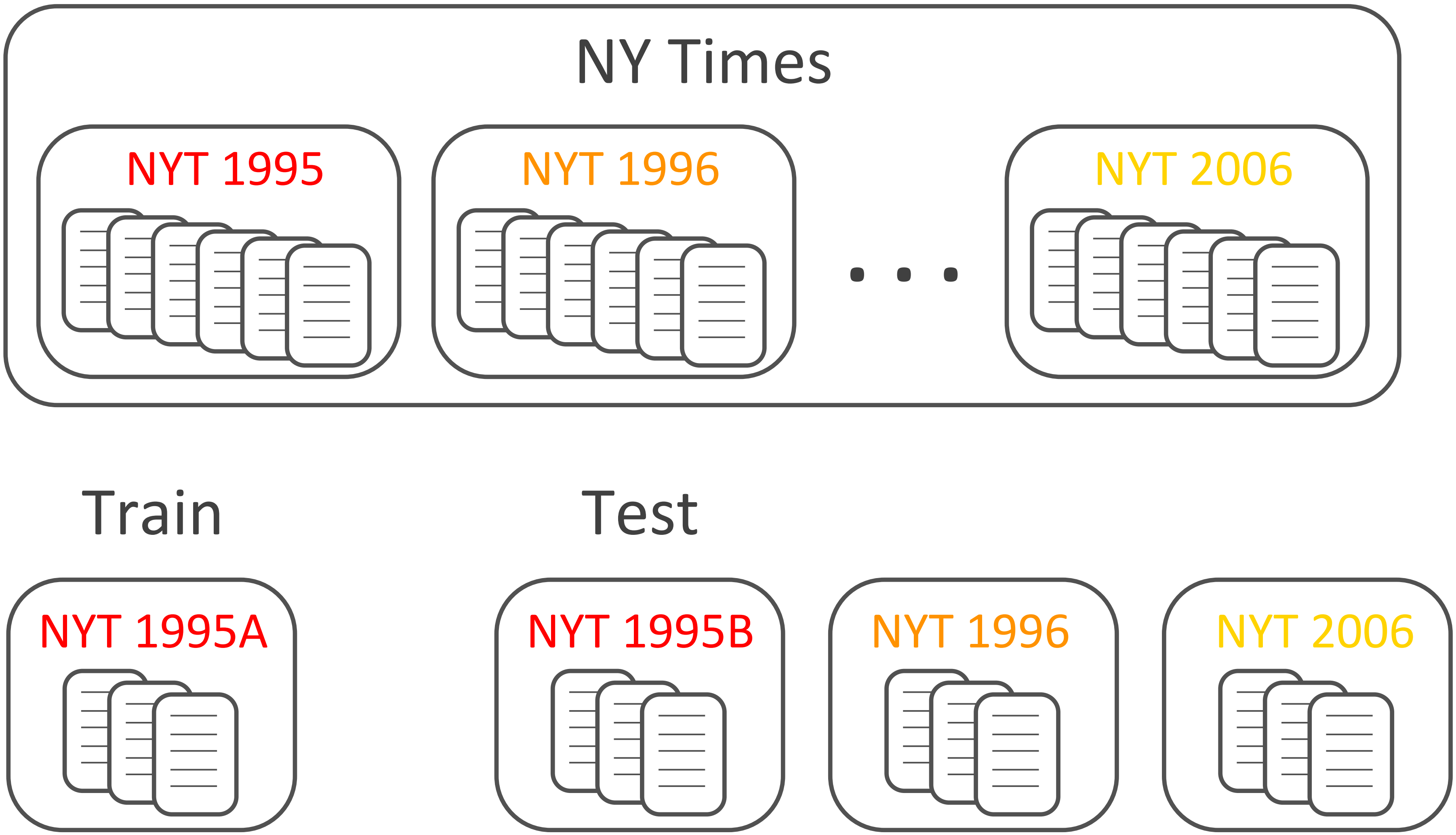
Future Counts



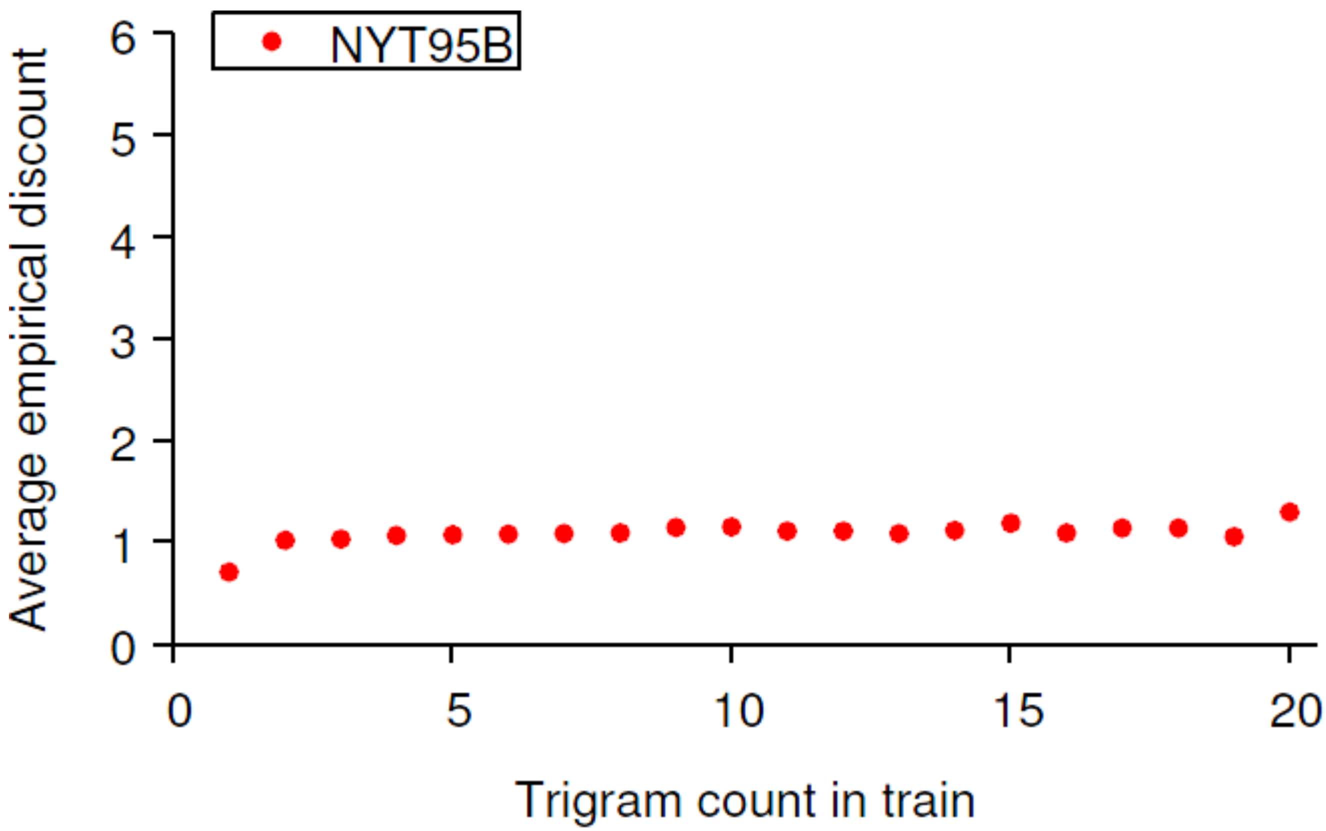
Empirical Discounts



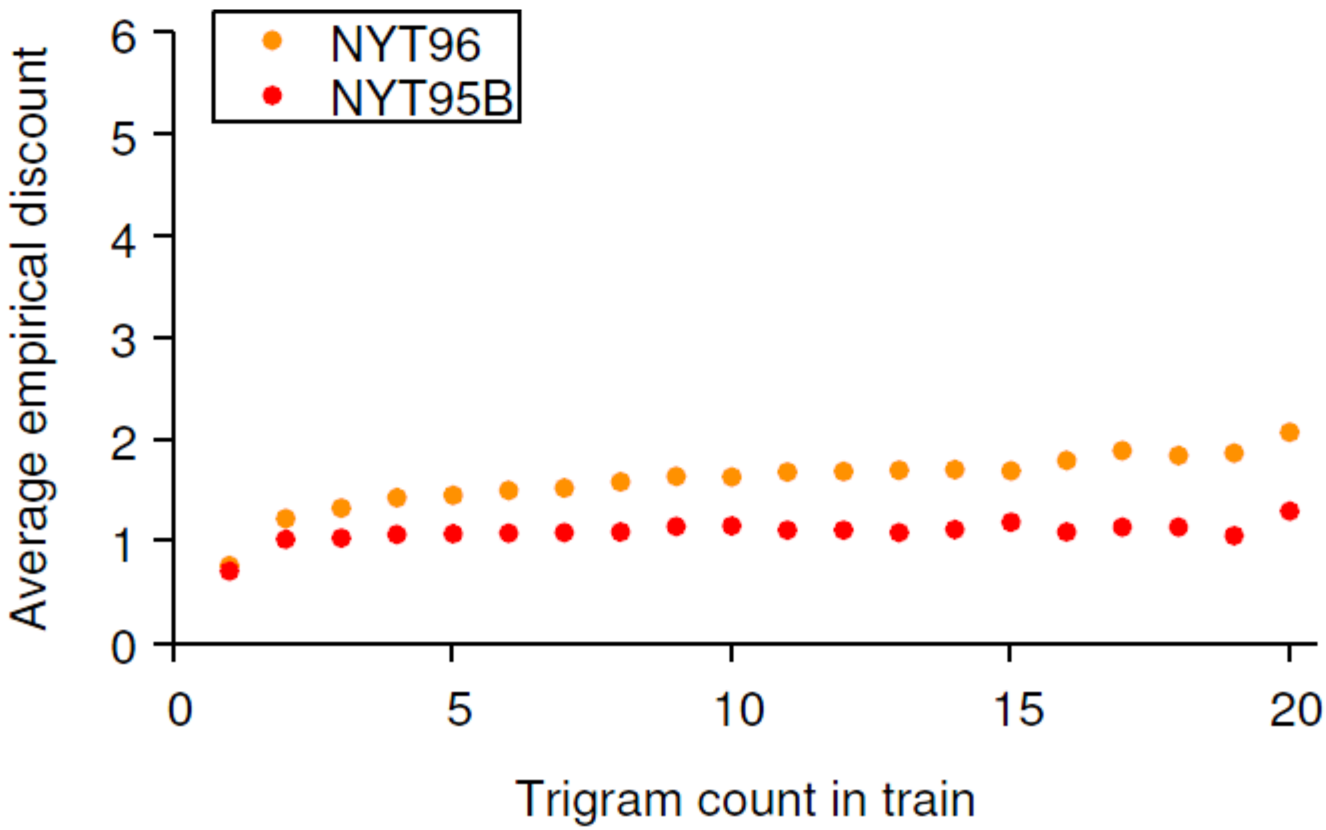
Experimental Setup



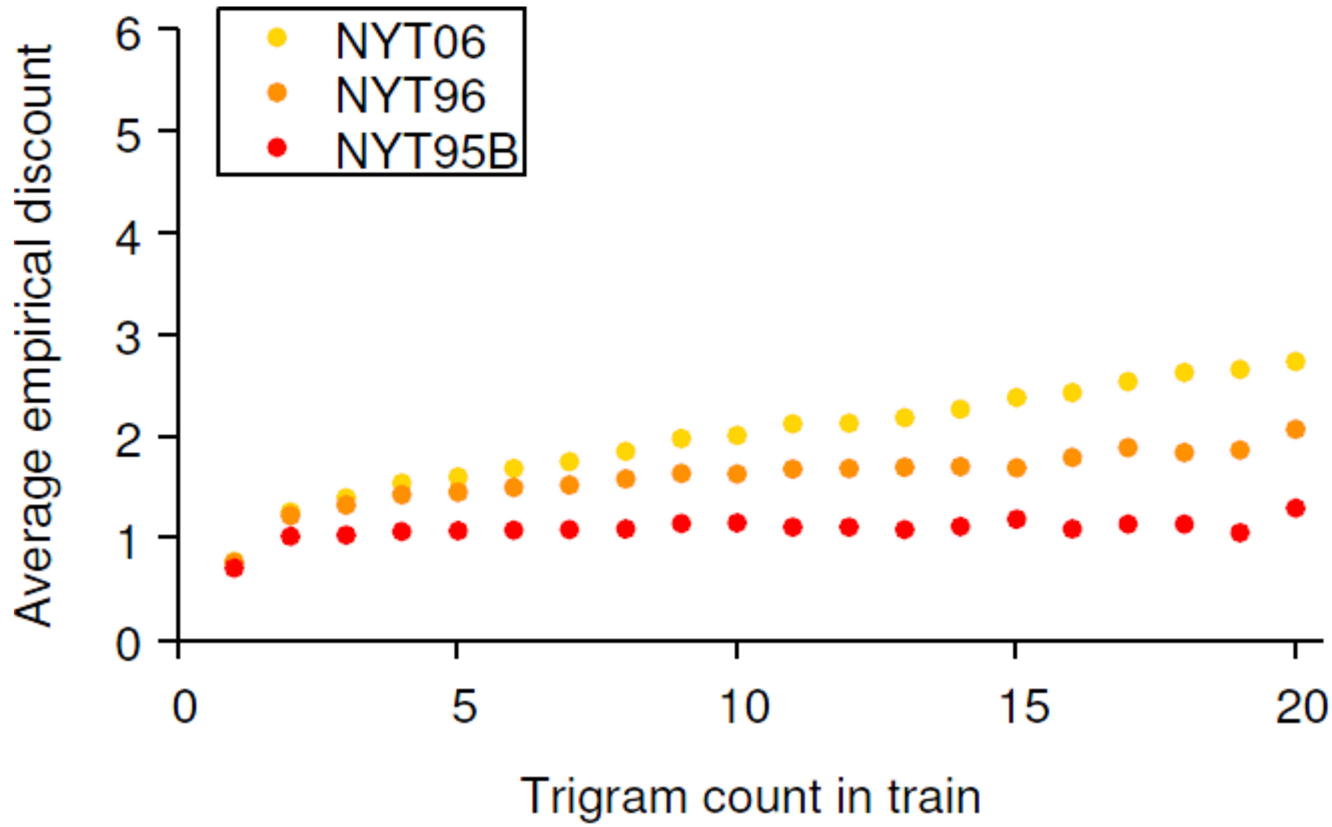
Empirical Discounts



Empirical Discounts

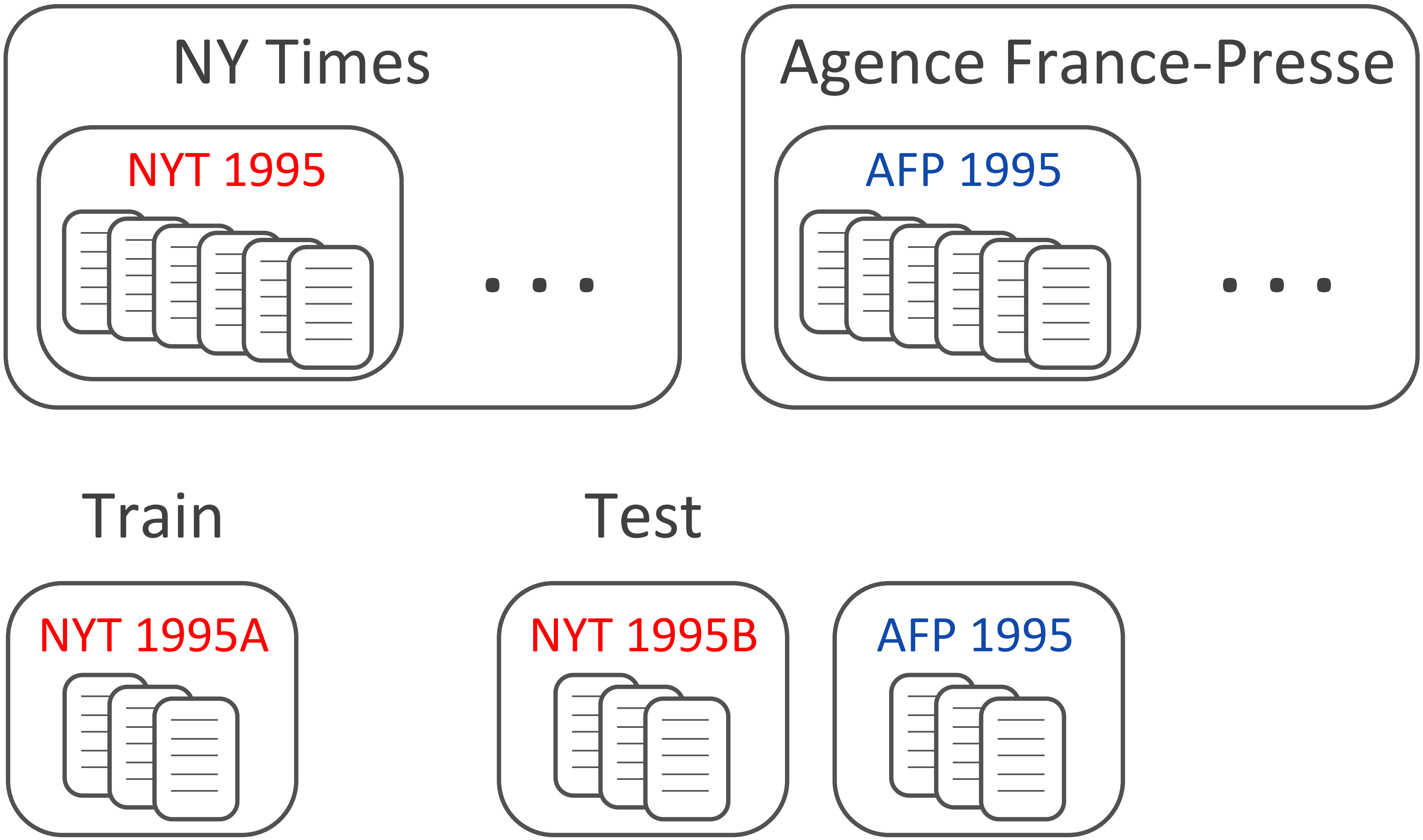


Empirical Discounts

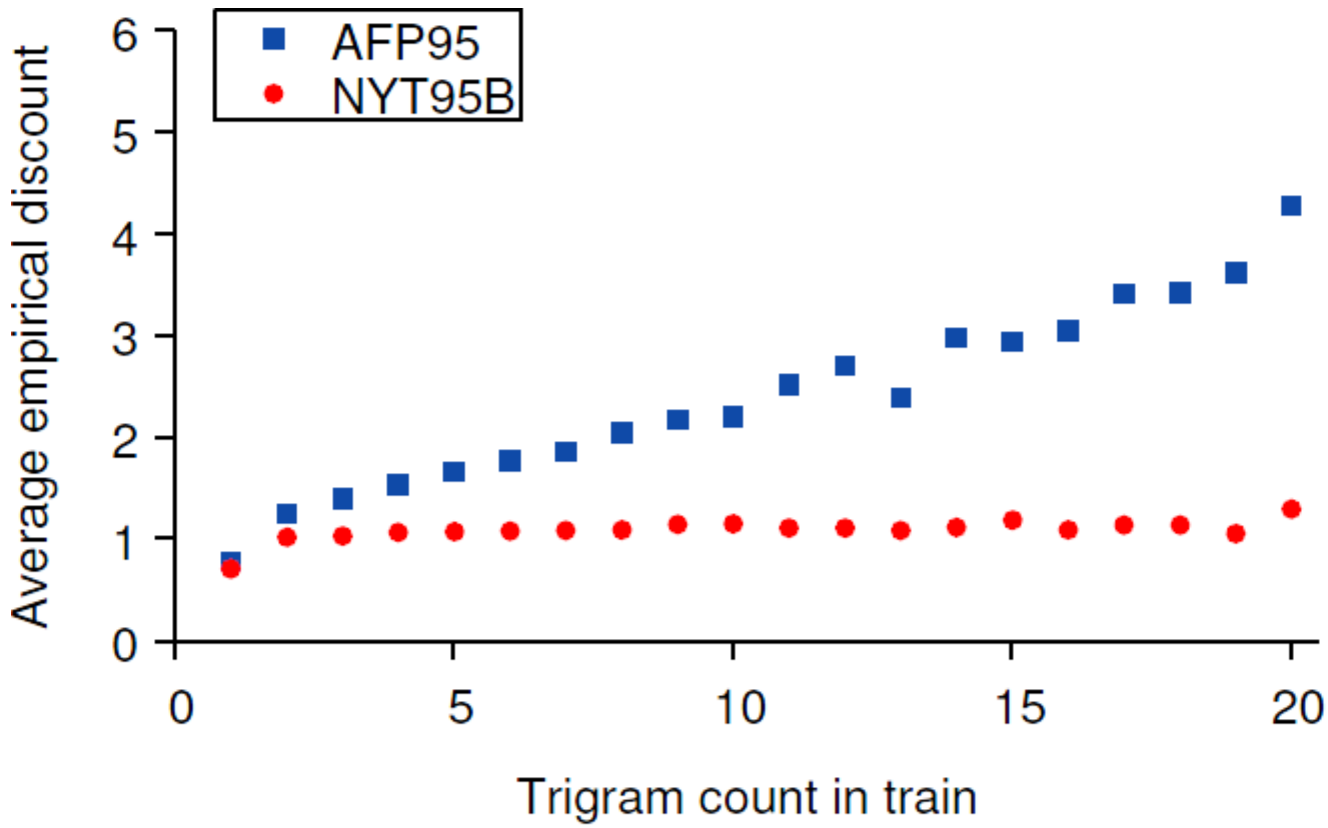


Changing the training year results in growing discounts

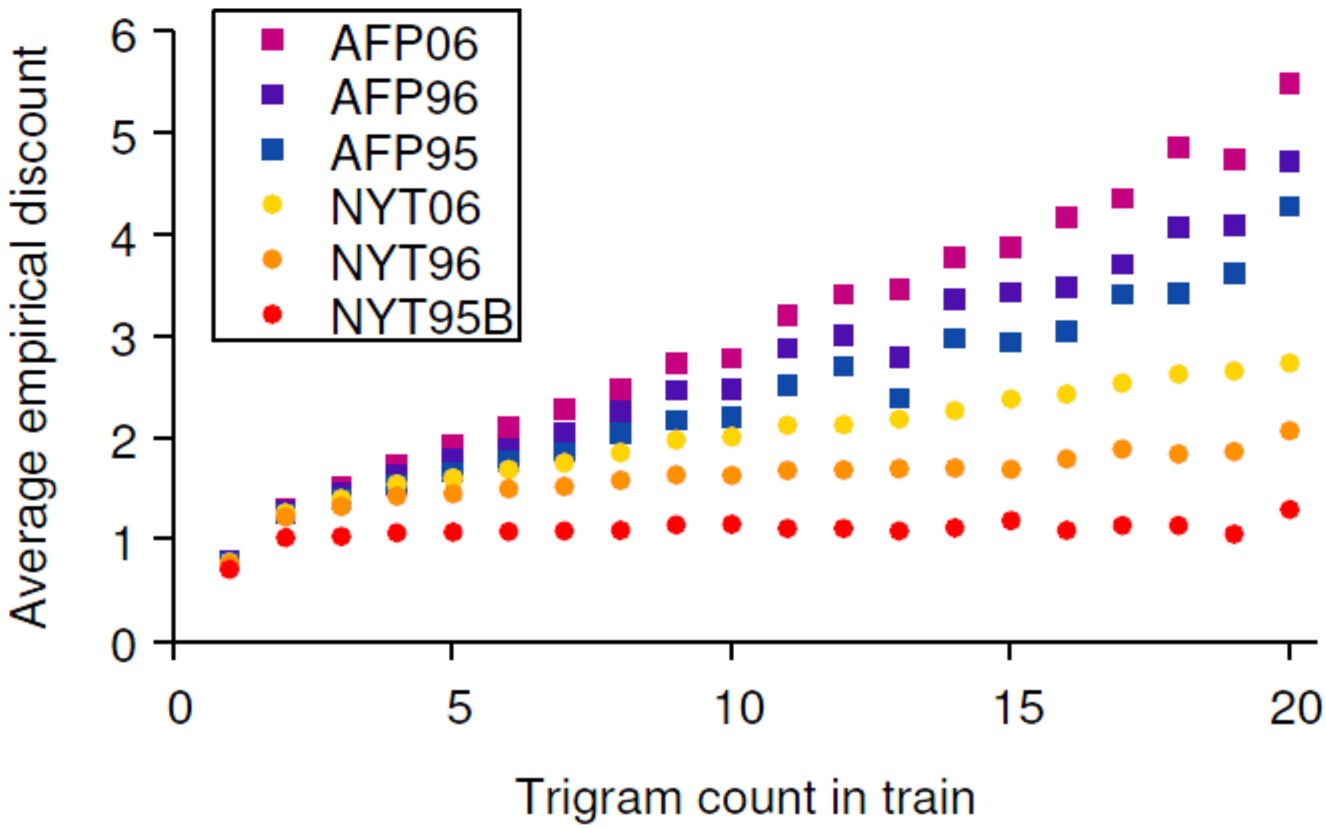
Experimental Setup



Empirical Discounts

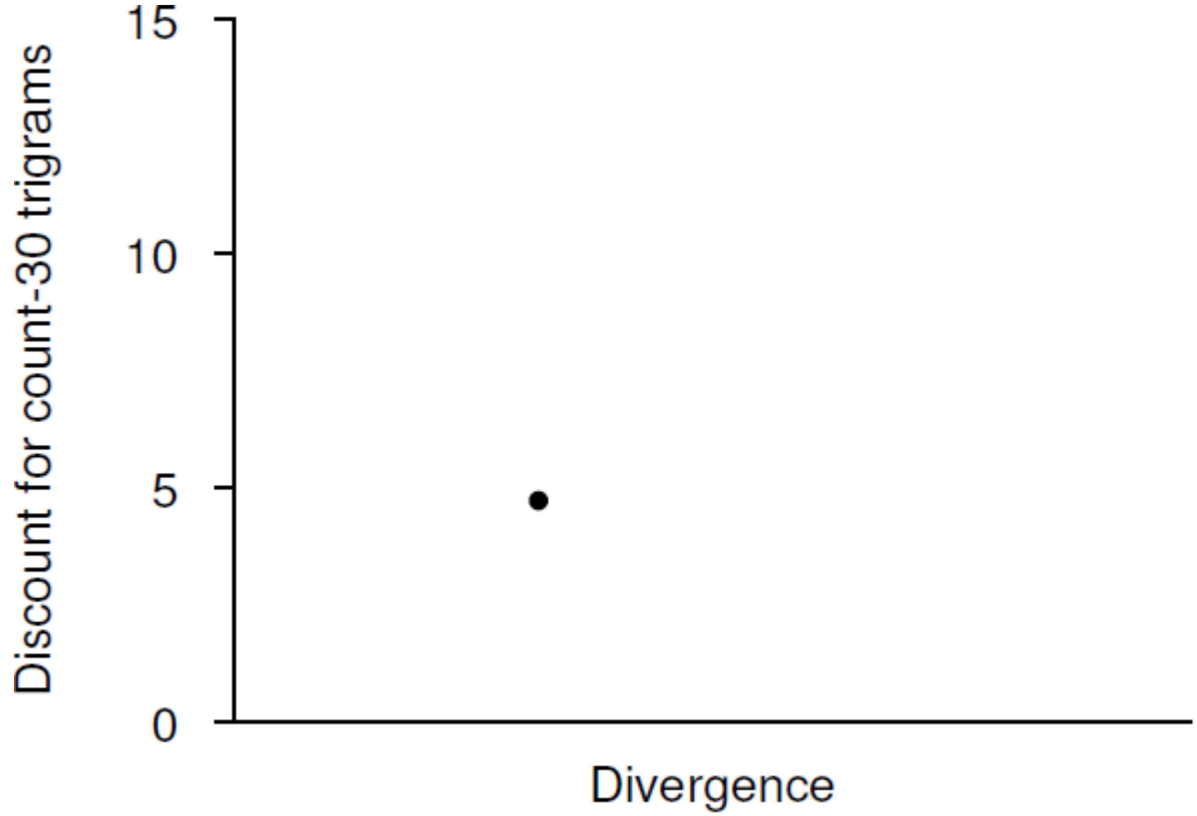


Empirical Discounts

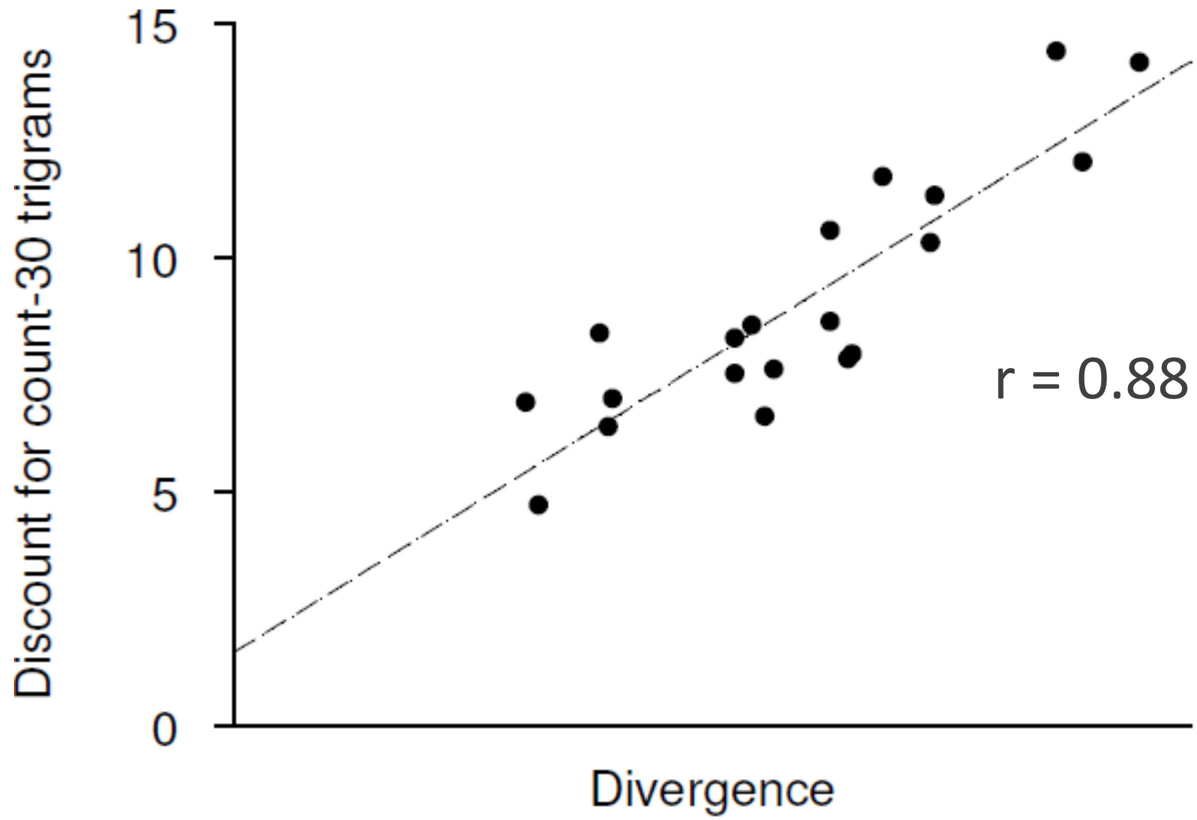


More discount growth as temporal divergence increases and source changes

Predicting Discount Growth



Predicting Discount Growth



More divergence yields more discounting

A Growing Discounts LM

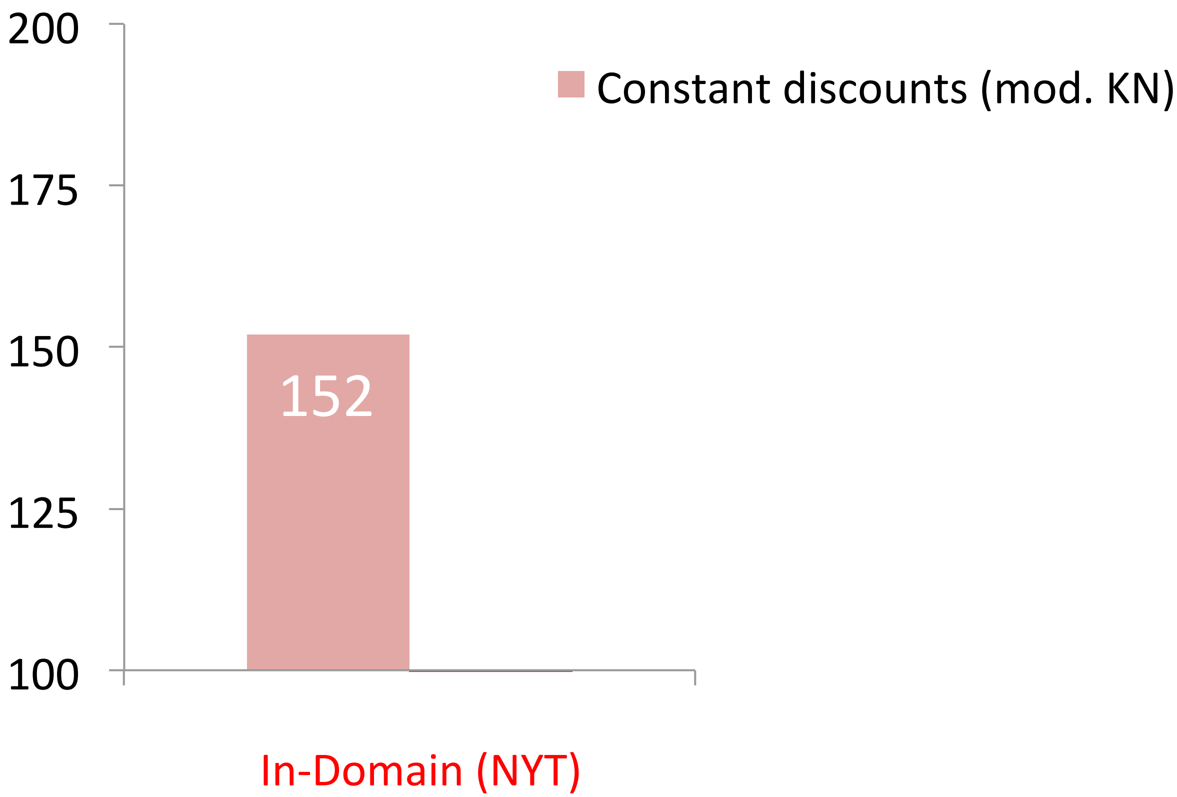
- Interpolated Kneser-Ney

$$p(n\text{-gram}) = \frac{c(n\text{-gram}) - d}{c(\text{context})} + \ldots$$

- Linearly Growing Discounts

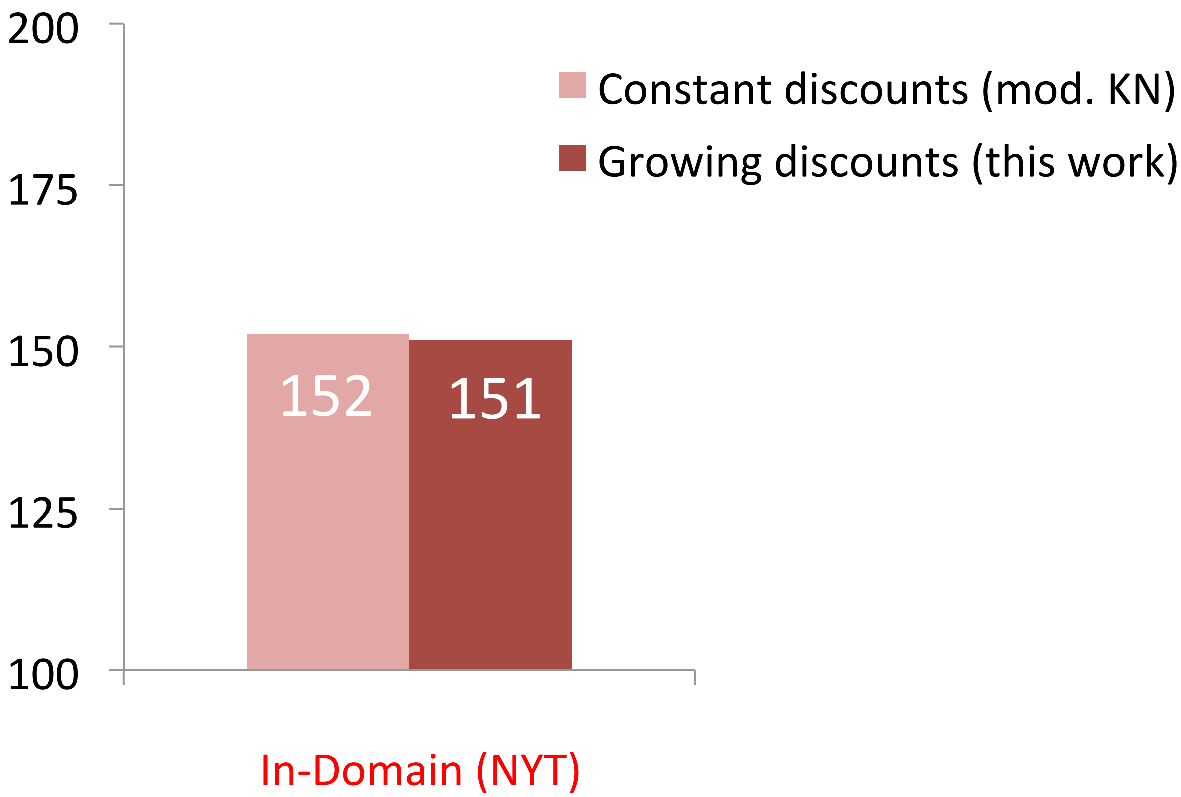
$$p(n\text{-gram}) = \frac{c(n\text{-gram}) - (d_1 + d_2 c(n\text{-gram}))}{c(\text{context})} + \ldots$$

Perplexity Results



(Parameters tuned to maximize held-out NYT perplexity)

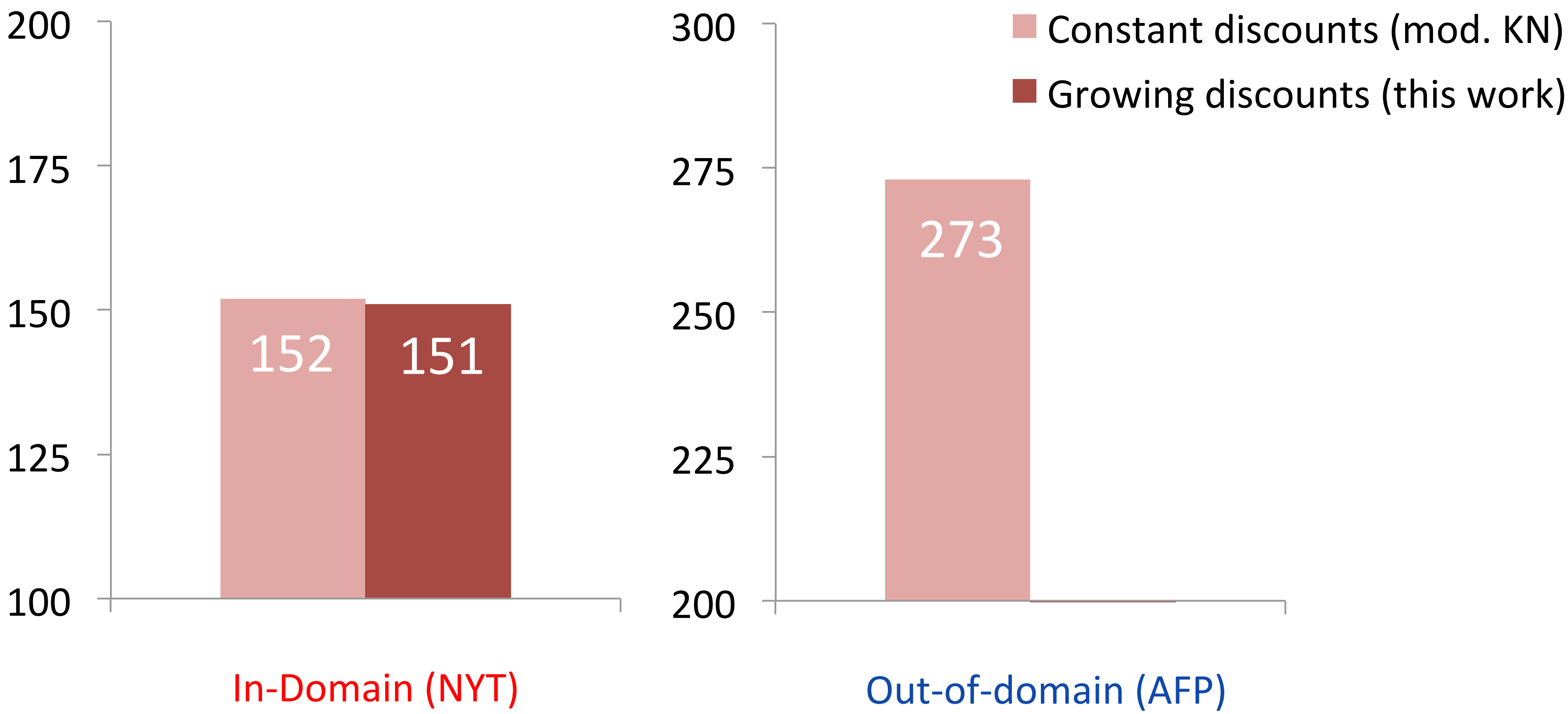
Perplexity Results



d_2 = 0.083

(Parameters tuned to maximize held-out NYT perplexity)

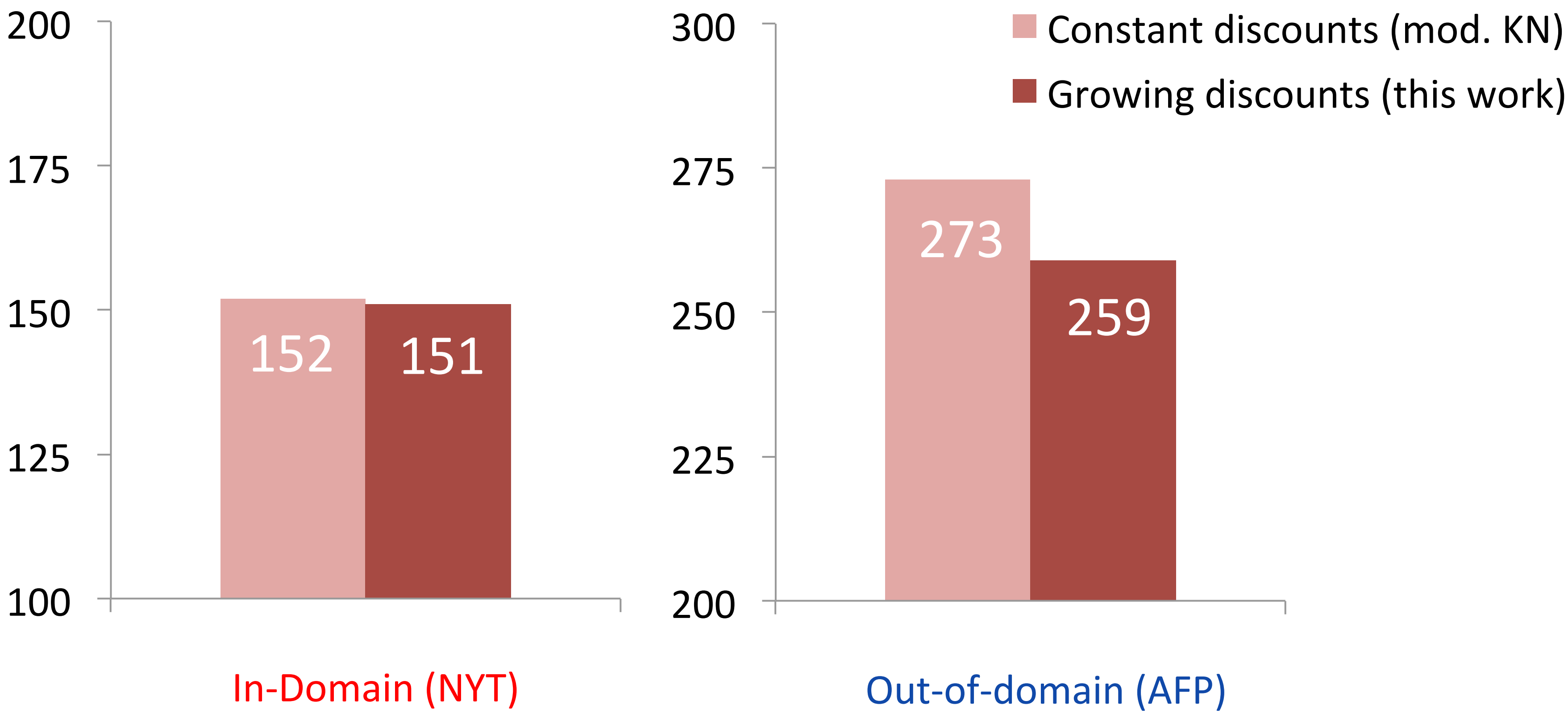
Perplexity Results



d_2 = 0.083

(Parameters tuned to maximize held-out NYT perplexity)

Perplexity Results



d_2 = 0.083

d_2 = 0.29

(Parameters tuned to maximize held-out NYT perplexity)

Conclusion

- Shape of discount must be changed (should grow with n-gram count) as corpora diverge
- Subtle cross-domain effects suggest using a qualitatively different model

Thank you!