

Appendix

Appendix A Full data on POS tags and percentage of improvement on the log likelihood score.

	Not in context, not in input	Not in context, in input	In context, not in input	In context, in input	Occurrence
NN	0.157771	0.698062	16.197878	5.578312	23639620
NNP	0.020258	0.527689	35.891013	8.141562	16840255
IN	-0.169446	-0.187711	3.308690	1.816983	13603903
JJ	0.400632	1.011774	13.077988	5.019096	12885333
DT	0.224373	0.196791	3.253798	2.453526	11128747
NNS	0.131760	0.594879	17.203110	5.242418	10088414
.	0.074150	-0.292448	7.026237	0.411417	6330229
RB	0.692010	0.897788	5.454908	2.711667	6174253
VB	0.384166	0.659601	9.219075	3.491015	5390427
,	0.731329	0.710795	3.420206	2.353697	5245713
PRP	-0.260621	-0.006360	3.305721	1.998650	5219071
CC	0.055603	0.277403	3.446472	2.634180	4463355
VBZ	0.355881	0.093804	5.872791	0.625349	3895160
VBG	0.588187	0.913552	11.268538	4.581329	3834853
VTB	0.260189	0.650080	12.332930	4.213668	3646850
VBD	0.312395	0.378780	6.878956	2.758350	3377739
VBP	0.333019	0.362408	5.038212	1.873432	3308882
TO	0.000127	-0.183880	3.921325	1.767642	3205519
CD	0.619997	1.148265	20.414486	5.950232	2723028
PRP\$	0.040851	0.269473	4.617825	2.683837	1942225
MD	-0.115006	0.047148	3.572827	2.559091	1590480
:	-0.729198	-0.690925	5.812855	3.404822	650535
WDT	0.455384	0.571508	3.444940	1.234093	576220
WRB	0.638024	0.921696	5.063341	3.163921	572072
JJR	1.010684	1.612511	7.880883	4.808698	483001
(0.226634	0.602507	3.785168	2.695342	450846
WP	1.296407	1.746595	5.018384	3.681695	437709
RP	0.347023	0.406551	5.269570	3.544124	416556
POS	-1.028733	0.472410	8.667241	2.427072	387925
)	0.200041	0.019679	4.084305	1.586972	369449
JJS	0.499507	0.790719	7.975327	4.348411	342354
NNPS	-0.206552	0.327515	30.412756	6.301600	233250
EX	0.793452	1.222533	4.670563	3.714079	201943
RBR	1.037655	2.046677	5.190104	4.050341	192250
PDT	0.604735	1.509266	3.752347	3.438381	84221
RBS	0.114155	0.759014	4.438339	4.140244	83746
FW	0.779163	0.560080	13.321369	4.776943	80379
\$	-0.071967	0.671289	18.502472	5.415212	63854
UH	0.154138	0.207683	7.501694	3.188253	45710
”	0.967397	1.077652	17.579383	2.770341	30181
WP\$	2.702986	3.293384	6.577276	3.215185	18004
#	0.767495	4.536762	31.027920	14.052459	14969
SYM	0.343766	-3.428490	20.566550	3.729796	5532
“	1.352541	4.927900	35.315082	16.644833	2998
LS	3.417315	-1.039910	57.893078	8.573192	836

Appendix B Visualization of context and change in token log likelihood

Context:

was humbly decorated but Grand Master Raphael Cotton^{er} ordered the renovation of the interior. Mattia Preti the artist from Calabria Italy was commissioned to embellish the interior with detailed carved stone wall designs finished with gold leaf gilding including the painting of the vaulted ceiling and side altars. They depicting the various scenes from the life of Saint John. The Knights of St. John were very decisive to spend and dedicate their resources for the best ornamentals for the cathedral. During 1748, Grand Master António Manoel de Vilhena built new adjoining buildings to enhance the requirements of the time. When Sir Walter Scott came to Malta in 1831 he said that the cathedral was a "magnificent church, the most striking interior ever seen." He spent three weeks in Malta during October 1831. The floor of the co-cathedral is covered with 400 marble carved tombstones with superbly intricate, marble-inlaid created by skilled tradesmen, tombstones create a formidable appearance to the large floors around all the church. All the buried knights are sons of Europe's noble families from the time of the 16th to the 18th century. Here lies the mastermind behind the building of Valletta, his name was Grand Master Jean Parisot de la Vall^{ette} among the Grand Masters buried within the crypt. St. John's Co-Cathedral always remained the central religious place for the Knights of Malta until the French occupied Malta in 1798. The French ransacked the church from its very rich belongings which were melted down into coins to pay their soldiers. The museum adjoining the cathedral contain several art objects. The most important are the Flemish 29 priceless tapestries according to designs by Peter Paul Rubens at the Flemish atelier of Judecos de Vos, famous as the court Weaver for King Louis XIV. They were donated by Aragonese Grand Master Ramon Perellos y Roccaful in 1697. It was the traditional custom that when a newly elected Grandmaster he is to bestow a gift^{or} gioi

Input:

nod to 21st century architecture – the new parliament designed by the eminent Italian architect Renzo Piano. The Knights of St. John's 268 year rule very much shaped Malta's built environment with miles of bastions, forts and palaces which still define the islands. But once inside the story is a very different one. It is here that one is likely to run out of superlatives in attempting to come to describe this magnificent building. The knights certainly went all out to create an interior quite unlike any other and no expense was spared in the embellishment of their main place of worship. The cathedral's walls are richly carved and the monuments to the various Grandmasters fill almost every available space. The church is the proud owner of two paintings by Caravaggio – including his largest and only signed canvas The Beheading of Saint John, commissioned for the church's Oratory where it hangs to this day. The Knights employed the Calabrian artist Mattia Preti for the most ambitious part of the church's décor – the magnificent cycle of ceiling paintings depicting the life of the saint. The renowned Maltese sculptor Melchiorre Gafa – at the time making a name for himself in Rome – was commissioned for the apse's sculpture of the Baptism of Christ. Sadly his early demise meant that the work was never completed and the project was eventually entrusted to Gafa's only pupil, the Italian Giuseppe Mazzuoli. Not to be overlooked is the cathedral's magnificent floor – made up of around 364 knights' tombstones in inlaid marble laid one next to the other. In the adjoining museum one finds a fine set of large Flemish tapestries which used to hang in the church on special feast days. The cathedral.

Targets:

a small crypt contains the ornate tombs of some of the prominent Grand Masters among them Jean Parisot de Vallette – the founder of the city. No visit to the islands is complete without a visit to this magnificent church – the apex of the Knights.

(a) Sampled from context-input-target triplet with top 1% of context utility.

Context:

just brings your outfit all together. So I am listing some of my favorite bangles that I hope to one day add to my collection. This first one that I would love to be able to own one day is a Cartier gold bangle bracelet. I just think these are so stylish and can be worn with any outfit. Okay I won't lie, Kourtney Kardashian was a total inspiration for it... yes Kourtney not Kylie... but I just think they look so cute and can really bring an outfit together. Let's just ignore the price though because unless I become a millionaire there is a good chance I will never be able to afford one of these bad boys. That's okay though, a girl can dream. The next bracelet is one that I actually already own a couple of. These Kendra Scott bracelets can make an outfit pop a little bit more. Even though these bracelets aren't technically bangles, they're cuffs, my wrist is so tiny they close all the way around and look like a bangle. Anyways, these Kendra Scott Erica Bracelets with stones can help an outfit stand out because of the different colored stones. They have so many different options and they match the stone-to-metal colors very nicely. Some of my go-to bracelets in my everyday outfits for sure. Now okay, I know you shouldn't own one of these bracelets until you actually "tie the knot," but I can't help it! I'm obsessed with the simplicity and such special meaning behind them. The Kate Spade knot bangles are definitely a gift that I want for when I get married. This next bangle is actually something that has been on my Christmas list for about three years now. The Coordinate bangle. I just think all of their jewelry is pretty flawless and can have such a special meaning behind it. For one of Omar's Christmas gifts last year (2014) I got us both matching leather bracelets. His bracelet has my home coordinates and my bracelet had his base in Japan's coordinates

Input:

We have our first big family vacation coming up this fall and it is a serious road trip...all the way to Texas. As each day brings us closer to our departure date I am already planning the packing in my head. This includes all of the making that will need to take place beforehand. There is nothing like a trip to speed up my progress on a garment or gift for Desmond. I saw these lush, silk wrapped bracelets in the latest Real Simple and they begged me to bring them along on our deep south adventure. Let the making begin! Many of you know that I am no strangers to wrapped bangles. Wrapped in wool, wrapped in cotton, I love a wrapped bangle so this project wasn't new territory. I even had some bangles lying around. These were hand-me-downs from my Mom. I've worn them for years and their shape made them great candidates for a silk-wrapped update. Did I mention I happened to have some hand-

Targets:

died silk ribbon in the studio? Inconceivable. I know I actually won these (I NEVER win anything) in a drawing from the Richmond Craft Mafia a couple of years ago. I'd been saving them for something glamorous and it was

(b) Randomly sampled.

Figure 1: Visualize the difference in log-likelihood for each target token with and without context.

Visualize the delta log-likelihood for each target token with and without context. The color of each target token is determined by $\log P(y_i|y_{<i}, \mathbf{x}, \mathbf{c}, \theta) - \log P(y_i|y_{<i}, \mathbf{x}, \theta')$. The color is green if the difference is positive (context helped). The color is red if the difference is negative (context harmed). We put a box around a target token if it only appeared in the context, but not input. We put a strike-through on a token if it appeared in the input. We show two examples here. Figure 1a is sampled from the context-input-target triplet with top 1% of context utility. Figure 1b is randomly sampled from the whole C4 train split. We provide more examples in the supplementary materials. They are interactive HTML files that you can mouse-over the tokens to see the difference in log-likelihood.

Appendix C Model training steps per second

Model Size	Steps/Sec	Training Devices	Number of Steps	Batch size
C4 Language Modeling				
113M	3.3	64 TPUv3	1,100,000	512
409M	2.2	64 TPUv3	1,100,000	512
1560M	1.5	128 TPUv3	1,100,000	512
Natural Question				
409M	212.5	64 TPUv3	40,000	64
1560M	74.6	64 TPUv3	40,000	64

Appendix D License for assets used in this work

Name	License	Link
C4 English Dataset	ODC-By-1.0	https://commoncrawl.org
Natural Question	CC BY-SA 3.0	https://github.com/google-research-datasets/natural-questions/tree/master/nq_open
T5X	Apache-2.0	https://github.com/google-research/t5x
mT5	Apache-2.0	https://github.com/google-research/multilingual-t5
T5 Retrieval	Apache-2.0	https://github.com/google-research/t5x_retrieval