

Summary of references

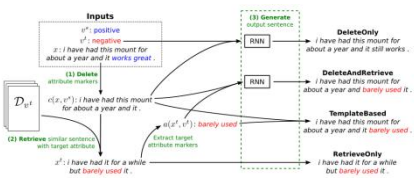
Stanza:

<https://stanfordnlp.github.io/stanza/depparse.html>

Pytextrank:

<https://derwen.ai/docs/ptr/overview/>

Title	Dataset	Method	Evaluation matrix																																																															
《Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer》	1.YELP	Baseline models:	<table><tr><th></th><th colspan="2">YELP</th><th colspan="2">CAPTIONS</th><th colspan="2">AMAZON</th></tr><tr><th></th><th>Classifier</th><th>BLEU</th><th>Classifier</th><th>BLEU</th><th>Classifier</th><th>BLEU</th></tr><tr><td>CROSSALIGNED</td><td>73.7%</td><td>3.1</td><td>74.3%</td><td>0.1</td><td>74.1%</td><td>0.4</td></tr><tr><td>STYLEEMBEDDING</td><td>8.7%</td><td>11.8</td><td>54.7%</td><td>6.7</td><td>43.3%</td><td>10.0</td></tr><tr><td>MULTIDECODER</td><td>47.0%</td><td>7.1</td><td>68.5%</td><td>4.6</td><td>68.3%</td><td>5.0</td></tr><tr><td></td><td>81.7%</td><td>11.8</td><td>92.5%</td><td>17.1</td><td>68.7%</td><td>27.1</td></tr><tr><td></td><td>95.4%</td><td>0.4</td><td>95.5%</td><td>0.7</td><td>70.3%</td><td>0.9</td></tr><tr><td></td><td>85.7%</td><td>7.5</td><td>83.0%</td><td>9.0</td><td>45.6%</td><td>24.6</td></tr><tr><td></td><td>88.7%</td><td>8.4</td><td>96.8%</td><td>7.3</td><td>48.0%</td><td>22.8</td></tr></table>		YELP		CAPTIONS		AMAZON			Classifier	BLEU	Classifier	BLEU	Classifier	BLEU	CROSSALIGNED	73.7%	3.1	74.3%	0.1	74.1%	0.4	STYLEEMBEDDING	8.7%	11.8	54.7%	6.7	43.3%	10.0	MULTIDECODER	47.0%	7.1	68.5%	4.6	68.3%	5.0		81.7%	11.8	92.5%	17.1	68.7%	27.1		95.4%	0.4	95.5%	0.7	70.3%	0.9		85.7%	7.5	83.0%	9.0	45.6%	24.6		88.7%	8.4	96.8%	7.3	48.0%	22.8
		YELP		CAPTIONS		AMAZON																																																												
		Classifier		BLEU	Classifier	BLEU	Classifier	BLEU																																																										
	CROSSALIGNED	73.7%		3.1	74.3%	0.1	74.1%	0.4																																																										
	STYLEEMBEDDING	8.7%		11.8	54.7%	6.7	43.3%	10.0																																																										
MULTIDECODER	47.0%	7.1	68.5%	4.6	68.3%	5.0																																																												
	81.7%	11.8	92.5%	17.1	68.7%	27.1																																																												
	95.4%	0.4	95.5%	0.7	70.3%	0.9																																																												
	85.7%	7.5	83.0%	9.0	45.6%	24.6																																																												
	88.7%	8.4	96.8%	7.3	48.0%	22.8																																																												
2.Reviews on Amazon		RETRIEVEONLY: Returns, word by word, sentences retrieved that are similar to the original sentence with the deleted attribute. Guaranteed to produce a grammatically correct sentence with the target attributes, but its content may differ from the original sentence.																																																																
3.Image captions		TEMPLATEBASED: Replace the deleted attribute in the original sentence with the target attribute																																																																
		Neural modes:																																																																
	<table><tr><th>Dataset</th><th>Attributes</th><th>Train</th><th>Dev</th><th>Test</th></tr><tr><td rowspan="2">YELP</td><td>Positive</td><td>270K</td><td>2000</td><td>500</td></tr><tr><td>Negative</td><td>180K</td><td>2000</td><td>500</td></tr><tr><td rowspan="3">CAPTIONS</td><td>Romantic</td><td>6000</td><td>300</td><td>0</td></tr><tr><td>Humorous</td><td>6000</td><td>300</td><td>0</td></tr><tr><td>Factual</td><td>0</td><td>0</td><td>300</td></tr><tr><td rowspan="2">AMAZON</td><td>Positive</td><td>277K</td><td>985</td><td>500</td></tr><tr><td>Negative</td><td>278K</td><td>1015</td><td>500</td></tr></table>	Dataset	Attributes	Train	Dev	Test	YELP	Positive	270K	2000	500	Negative	180K	2000	500	CAPTIONS	Romantic	6000	300	0	Humorous	6000	300	0	Factual	0	0	300	AMAZON	Positive	277K	985	500	Negative	278K	1015	500	DELETEONLY: An RNN is used to embed the sentences with the attributes removed into a vector. It then connects the final hidden state to the learned embedding of the target attribute and feeds it to the RNN decoder to generate the final sentence. The decoder attempts to generate words that represent the source content and target attributes while maintaining fluency.																												
Dataset	Attributes	Train	Dev	Test																																																														
YELP	Positive	270K	2000	500																																																														
	Negative	180K	2000	500																																																														
CAPTIONS	Romantic	6000	300	0																																																														
	Humorous	6000	300	0																																																														
	Factual	0	0	300																																																														
AMAZON	Positive	277K	985	500																																																														
	Negative	278K	1015	500																																																														
		DELETEANDRETRIEVE: Similar to DELETEONLY, but uses the attributes of the retrieved sentence rather than the target attributes. As with DELETEONLY, it encodes the sentence with the original deleted attribute using an RNN. The RNN decoder uses the splicing and content embedding of																																																																

		<p>this vector to generate the final sentence.</p> 																																																																																																																																																																																																																																																																					
Text Style Transfer: A Review and Experimental Evaluation	<div>1. Yelp a corpus of restaurant reviews</div> <div>2. Amazon product review datasets</div> <div>3. IMDb movie reviews datasets</div>	<p>Creating a taxonomy to organize the TST models, and provide a comprehensive summary of the state of the art. We review the existing evaluation methodologies for TST tasks, and conduct a large-scale reproducibility study where we experimentally benchmark 19 state-of-the-art TST algorithms on two publicly available datasets. Finally, we expand on current trends and provide new perspectives on the new and exciting developments in the TST field.</p>	<p>Table 10: Example outputs Yelp datasets.</p> <table><tr><th>Model</th><th>From negative to positive (Yelp)</th><th>From positive to negative (Yelp)</th></tr><tr><td>Source</td><td>Food was cold and lacking of flavors.</td><td>Excellent service, and food, and nice atmosphere.</td></tr><tr><td>Deletions</td><td>They were great, with a wide variety of flavors.</td><td>I would give, zero, stars if I could.</td></tr><tr><td>Template</td><td>Parameters of flavors and textures.</td><td>Excellent service, good food, and nice atmosphere.</td></tr><tr><td>Deletions</td><td>Of a large list of flavors.</td><td>A waste of time and money.</td></tr><tr><td>Deletions</td><td>Food was bland and of no flavors.</td><td>Great service, good food, and great atmosphere.</td></tr><tr><td>Deletions</td><td>The food was cold and lacking of flavors.</td><td>The service, nice food, and atmosphere.</td></tr><tr><td>Deletions</td><td>Food was fresh and delicious of flavors.</td><td>Horrible service, horrible food, and horrible atmosphere.</td></tr><tr><td>Deletions</td><td></td><td>Slow service, good food, and nice atmosphere.</td></tr><tr><td>Deletions</td><td></td><td>Fast service, poor food, and poor atmosphere.</td></tr><tr><td>Deletions</td><td>Food was delicious and fresh.</td><td>Service was ok but the food was not hot.</td></tr><tr><td>Deletions</td><td>Food was cold and boring.</td><td>However, food was good, and poor, poor atmosphere.</td></tr><tr><td>Deletions</td><td>Food was nicely and surprising of flavors.</td><td>Excellent service, good food, nice atmosphere and nice atmosphere.</td></tr><tr><td>Deletions</td><td>Food was prepared, surprising and service.</td><td>Terrible service, not food, and the atmosphere.</td></tr><tr><td>Deletions</td><td>Food was cold and authentic which has.</td><td>Disappointing service, both nice and the atmosphere.</td></tr><tr><td>Deletions</td><td>Food was fresh and great of flavors.</td><td>Excellent service, good food, and nice atmosphere.</td></tr><tr><td>Deletions</td><td>Food was great and clean of flavors.</td><td>Terrible service, bad food, and nice atmosphere.</td></tr><tr><td>Deletions</td><td>Food was delicious and authentic of flavors.</td><td>Terrible service, horrible food, and bland food.</td></tr><tr><td>Deletions</td><td>Food was fresh and exceptional of flavors.</td><td>Terrible service, bad food, and terrible atmosphere.</td></tr><tr><td>Deletions</td><td></td><td>Horrible service, bad food, and poor service.</td></tr></table> <p>Table 9: TST results on Yelp dataset (sentiment transfer tasks).</p> <table><tr><th>Model</th><th>ACC (%)</th><th>BLEU</th><th>CS</th><th>WO</th><th>F1</th><th>G-Score</th><th>F1-Score</th><th>#Params</th><th>FLOPs</th></tr><tr><td>DeletionsOnly</td><td>84.2</td><td>28.7</td><td>0.893</td><td>0.301</td><td>115</td><td>1.80</td><td>0.034</td><td>9.14M</td><td>27.9K</td></tr><tr><td>Template</td><td>78.2</td><td>48.1</td><td>0.850</td><td>0.463</td><td>1959</td><td>1.04</td><td>0.002</td><td>-</td><td>-</td></tr><tr><td>Deletions</td><td>88.1</td><td>30.0</td><td>0.897</td><td>0.464</td><td>101</td><td>1.87</td><td>0.039</td><td>9.88M</td><td>29.7K</td></tr><tr><td>B-GST</td><td>89.2</td><td>46.5</td><td>0.939</td><td>0.449</td><td>216</td><td>1.88</td><td>0.018</td><td>-</td><td>-</td></tr><tr><td>G-GST</td><td>72.7</td><td>52.0</td><td>0.967</td><td>0.417</td><td>407</td><td>1.55</td><td>0.019</td><td>-</td><td>-</td></tr><tr><td>PTD</td><td>82.3</td><td>37.4</td><td>0.982</td><td>0.737</td><td>245</td><td>1.94</td><td>0.016</td><td>-</td><td>-</td></tr><tr><td>UST</td><td>74.0</td><td>41.0</td><td>0.929</td><td>0.448</td><td>394</td><td>1.36</td><td>0.010</td><td>79.87M</td><td>2386.4G</td></tr><tr><td>SMILE</td><td>84.4</td><td>14.8</td><td>0.907</td><td>0.294</td><td>210</td><td>1.313</td><td>0.019</td><td>76.27M</td><td>4418.5G</td></tr><tr><td>DLST</td><td>91.2</td><td>7.6</td><td>0.904</td><td>0.484</td><td>86</td><td>1.41</td><td>0.045</td><td>64.0M</td><td>6.14K</td></tr><tr><td>BST</td><td>83.1</td><td>2.9</td><td>0.827</td><td>0.076</td><td>201</td><td>0.49</td><td>0.013</td><td>-</td><td>-</td></tr><tr><td>CAAE</td><td>82.7</td><td>11.2</td><td>0.901</td><td>0.277</td><td>145</td><td>1.15</td><td>0.027</td><td>12.92M</td><td>0.43K</td></tr><tr><td>ARAE</td><td>83.2</td><td>18.0</td><td>0.874</td><td>0.270</td><td>138</td><td>1.31</td><td>0.028</td><td>4.20M</td><td>5.61K</td></tr><tr><td>Cut-Gem</td><td>69.6</td><td>49.5</td><td>0.933</td><td>0.707</td><td>384</td><td>1.69</td><td>0.010</td><td>13.7M</td><td>6.13K</td></tr><tr><td>Multi-Dec</td><td>69.6</td><td>17.2</td><td>0.887</td><td>0.244</td><td>299</td><td>0.99</td><td>0.013</td><td>-</td><td>-</td></tr><tr><td>StyleEmb</td><td>47.5</td><td>51.4</td><td>0.926</td><td>0.433</td><td>217</td><td>1.31</td><td>0.018</td><td>-</td><td>-</td></tr><tr><td>Deletions</td><td>79.0</td><td>58.3</td><td>0.97</td><td>0.401</td><td>134</td><td>2.29</td><td>0.030</td><td>44M</td><td>0.862K</td></tr><tr><td>DAST</td><td>90.7</td><td>49.7</td><td>0.961</td><td>0.705</td><td>323</td><td>1.77</td><td>0.012</td><td>26.68M</td><td>0.822K</td></tr><tr><td>DAST-C</td><td>89.6</td><td>41.2</td><td>0.933</td><td>0.540</td><td>450</td><td>1.48</td><td>0.009</td><td>24.31M</td><td>0.50K</td></tr><tr><td>PPST</td><td>85.3</td><td>41.7</td><td>0.902</td><td>0.527</td><td>104</td><td>2.06</td><td>0.038</td><td>34.32M</td><td>76.1K</td></tr></table>	Model	From negative to positive (Yelp)	From positive to negative (Yelp)	Source	Food was cold and lacking of flavors.	Excellent service, and food, and nice atmosphere.	Deletions	They were great, with a wide variety of flavors.	I would give, zero, stars if I could.	Template	Parameters of flavors and textures.	Excellent service, good food, and nice atmosphere.	Deletions	Of a large list of flavors.	A waste of time and money.	Deletions	Food was bland and of no flavors.	Great service, good food, and great atmosphere.	Deletions	The food was cold and lacking of flavors.	The service, nice food, and atmosphere.	Deletions	Food was fresh and delicious of flavors.	Horrible service, horrible food, and horrible atmosphere.	Deletions		Slow service, good food, and nice atmosphere.	Deletions		Fast service, poor food, and poor atmosphere.	Deletions	Food was delicious and fresh.	Service was ok but the food was not hot.	Deletions	Food was cold and boring.	However, food was good, and poor, poor atmosphere.	Deletions	Food was nicely and surprising of flavors.	Excellent service, good food, nice atmosphere and nice atmosphere.	Deletions	Food was prepared, surprising and service.	Terrible service, not food, and the atmosphere.	Deletions	Food was cold and authentic which has.	Disappointing service, both nice and the atmosphere.	Deletions	Food was fresh and great of flavors.	Excellent service, good food, and nice atmosphere.	Deletions	Food was great and clean of flavors.	Terrible service, bad food, and nice atmosphere.	Deletions	Food was delicious and authentic of flavors.	Terrible service, horrible food, and bland food.	Deletions	Food was fresh and exceptional of flavors.	Terrible service, bad food, and terrible atmosphere.	Deletions		Horrible service, bad food, and poor service.	Model	ACC (%)	BLEU	CS	WO	F1	G-Score	F1-Score	#Params	FLOPs	DeletionsOnly	84.2	28.7	0.893	0.301	115	1.80	0.034	9.14M	27.9K	Template	78.2	48.1	0.850	0.463	1959	1.04	0.002	-	-	Deletions	88.1	30.0	0.897	0.464	101	1.87	0.039	9.88M	29.7K	B-GST	89.2	46.5	0.939	0.449	216	1.88	0.018	-	-	G-GST	72.7	52.0	0.967	0.417	407	1.55	0.019	-	-	PTD	82.3	37.4	0.982	0.737	245	1.94	0.016	-	-	UST	74.0	41.0	0.929	0.448	394	1.36	0.010	79.87M	2386.4G	SMILE	84.4	14.8	0.907	0.294	210	1.313	0.019	76.27M	4418.5G	DLST	91.2	7.6	0.904	0.484	86	1.41	0.045	64.0M	6.14K	BST	83.1	2.9	0.827	0.076	201	0.49	0.013	-	-	CAAE	82.7	11.2	0.901	0.277	145	1.15	0.027	12.92M	0.43K	ARAE	83.2	18.0	0.874	0.270	138	1.31	0.028	4.20M	5.61K	Cut-Gem	69.6	49.5	0.933	0.707	384	1.69	0.010	13.7M	6.13K	Multi-Dec	69.6	17.2	0.887	0.244	299	0.99	0.013	-	-	StyleEmb	47.5	51.4	0.926	0.433	217	1.31	0.018	-	-	Deletions	79.0	58.3	0.97	0.401	134	2.29	0.030	44M	0.862K	DAST	90.7	49.7	0.961	0.705	323	1.77	0.012	26.68M	0.822K	DAST-C	89.6	41.2	0.933	0.540	450	1.48	0.009	24.31M	0.50K	PPST	85.3	41.7	0.902	0.527	104	2.06	0.038	34.32M	76.1K
Model	From negative to positive (Yelp)	From positive to negative (Yelp)																																																																																																																																																																																																																																																																					
Source	Food was cold and lacking of flavors.	Excellent service, and food, and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions	They were great, with a wide variety of flavors.	I would give, zero, stars if I could.																																																																																																																																																																																																																																																																					
Template	Parameters of flavors and textures.	Excellent service, good food, and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Of a large list of flavors.	A waste of time and money.																																																																																																																																																																																																																																																																					
Deletions	Food was bland and of no flavors.	Great service, good food, and great atmosphere.																																																																																																																																																																																																																																																																					
Deletions	The food was cold and lacking of flavors.	The service, nice food, and atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was fresh and delicious of flavors.	Horrible service, horrible food, and horrible atmosphere.																																																																																																																																																																																																																																																																					
Deletions		Slow service, good food, and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions		Fast service, poor food, and poor atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was delicious and fresh.	Service was ok but the food was not hot.																																																																																																																																																																																																																																																																					
Deletions	Food was cold and boring.	However, food was good, and poor, poor atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was nicely and surprising of flavors.	Excellent service, good food, nice atmosphere and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was prepared, surprising and service.	Terrible service, not food, and the atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was cold and authentic which has.	Disappointing service, both nice and the atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was fresh and great of flavors.	Excellent service, good food, and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was great and clean of flavors.	Terrible service, bad food, and nice atmosphere.																																																																																																																																																																																																																																																																					
Deletions	Food was delicious and authentic of flavors.	Terrible service, horrible food, and bland food.																																																																																																																																																																																																																																																																					
Deletions	Food was fresh and exceptional of flavors.	Terrible service, bad food, and terrible atmosphere.																																																																																																																																																																																																																																																																					
Deletions		Horrible service, bad food, and poor service.																																																																																																																																																																																																																																																																					
Model	ACC (%)	BLEU	CS	WO	F1	G-Score	F1-Score	#Params	FLOPs																																																																																																																																																																																																																																																														
DeletionsOnly	84.2	28.7	0.893	0.301	115	1.80	0.034	9.14M	27.9K																																																																																																																																																																																																																																																														
Template	78.2	48.1	0.850	0.463	1959	1.04	0.002	-	-																																																																																																																																																																																																																																																														
Deletions	88.1	30.0	0.897	0.464	101	1.87	0.039	9.88M	29.7K																																																																																																																																																																																																																																																														
B-GST	89.2	46.5	0.939	0.449	216	1.88	0.018	-	-																																																																																																																																																																																																																																																														
G-GST	72.7	52.0	0.967	0.417	407	1.55	0.019	-	-																																																																																																																																																																																																																																																														
PTD	82.3	37.4	0.982	0.737	245	1.94	0.016	-	-																																																																																																																																																																																																																																																														
UST	74.0	41.0	0.929	0.448	394	1.36	0.010	79.87M	2386.4G																																																																																																																																																																																																																																																														
SMILE	84.4	14.8	0.907	0.294	210	1.313	0.019	76.27M	4418.5G																																																																																																																																																																																																																																																														
DLST	91.2	7.6	0.904	0.484	86	1.41	0.045	64.0M	6.14K																																																																																																																																																																																																																																																														
BST	83.1	2.9	0.827	0.076	201	0.49	0.013	-	-																																																																																																																																																																																																																																																														
CAAE	82.7	11.2	0.901	0.277	145	1.15	0.027	12.92M	0.43K																																																																																																																																																																																																																																																														
ARAE	83.2	18.0	0.874	0.270	138	1.31	0.028	4.20M	5.61K																																																																																																																																																																																																																																																														
Cut-Gem	69.6	49.5	0.933	0.707	384	1.69	0.010	13.7M	6.13K																																																																																																																																																																																																																																																														
Multi-Dec	69.6	17.2	0.887	0.244	299	0.99	0.013	-	-																																																																																																																																																																																																																																																														
StyleEmb	47.5	51.4	0.926	0.433	217	1.31	0.018	-	-																																																																																																																																																																																																																																																														
Deletions	79.0	58.3	0.97	0.401	134	2.29	0.030	44M	0.862K																																																																																																																																																																																																																																																														
DAST	90.7	49.7	0.961	0.705	323	1.77	0.012	26.68M	0.822K																																																																																																																																																																																																																																																														
DAST-C	89.6	41.2	0.933	0.540	450	1.48	0.009	24.31M	0.50K																																																																																																																																																																																																																																																														
PPST	85.3	41.7	0.902	0.527	104	2.06	0.038	34.32M	76.1K																																																																																																																																																																																																																																																														
Style Transfer Through Back-Translation	<div>Gender</div> <div>1. Reddy and Knight’ s (2016) dataset of reviews from Yelp annotated for two genders corresponding to markers of sex.</div> <div>Political slant</div> <div>2. Top-level comments on Facebook posts from all 412 current members of the United States Senate and House who have public Facebook pages.</div> <div>Sentiment.</div> <div>3. Yelp a corpus of restaurant reviews.</div>	<p>Focus on transferring author attributes: gender and political slant, and on sentiment modification.</p> <p>The second task is novel: given a sentence by an author with a particular political leaning, rephrase the sentence to preserve its meaning but to confound classifiers of political slant.</p> 	<p>Figure 1: Style transfer pipeline: to rephrase a sentence and reduce its stylistic characteristics, the sentence is back-translated. Then, separate style-specific generators are used for style transfer.</p> <table><tr><th>Style</th><th>class</th><th>train</th><th>dev</th><th>test</th></tr><tr><td>gender</td><td>2.57M</td><td>2.67M</td><td>4.5K</td><td>535K</td></tr><tr><td>political</td><td>80K</td><td>540K</td><td>4K</td><td>56K</td></tr><tr><td>sentiment</td><td>2M</td><td>444K</td><td>63.5K</td><td>127K</td></tr></table>	Style	class	train	dev	test	gender	2.57M	2.67M	4.5K	535K	political	80K	540K	4K	56K	sentiment	2M	444K	63.5K	127K																																																																																																																																																																																																																																																
Style	class	train	dev	test																																																																																																																																																																																																																																																																			
gender	2.57M	2.67M	4.5K	535K																																																																																																																																																																																																																																																																			
political	80K	540K	4K	56K																																																																																																																																																																																																																																																																			
sentiment	2M	444K	63.5K	127K																																																																																																																																																																																																																																																																			

Dear Sir or Madam, May I
Introduce the GY AFC Dataset:
Corpus, Benchmarks and Metrics
for Formality Style Transfer

Yahoo Answers,a question
answering forum,contains
a large number of informal
sentences and allows
redistribution of data.
Hence, we use the
Yahoo Answers L6 corpus⁵
to create our GYAFC
dataset of informal and
formal sentence pairs.

Domain	Total	Informal	Formal
All Yahoo Answers	40M	24M	16M
Entertainment & Music	3.8M	2.7M	700K
Family & Relationships	7.8M	5.6M	1.8M

Table 2: Yahoo Answers corpus statistics

	Informal to Formal			Formal to Informal	
	Train	Tune	Test	Tune	Test
E&M	52,595	2,877	1,416	2,356	1,082
F&R	51,967	2,788	1,332	2,247	1,019

Table 3: GYAFC dataset statistics

Informal: <i>I'd say it is punk though.</i>
Formal: <i>However, I do believe it to be punk.</i>
Informal: <i>Gotta see both sides of the story.</i>
Formal: <i>You have to consider both sides of the story.</i>

Table 1: Informal sentences with formal rewrites.

1.Rule-based Approach

Develop a set of rules to
automatically make an informal
sentence more formal where we
capitalize first word and proper
nouns, remove repeated
punctuations, handcraft a list of
expansion for contractions etc.

2.Phrase-based Machine Translation

Use a combination of training
regimes to develop our model.
Train on the output of the rule based
approach when applied to GYAFC.
This is meant to force the PBMT
model to learn generalizations
outside the rules.

3.Neural Machine Translation
Experiment with three NMT models

NMT baseline: Our baseline model is
a bidirectional LSTM
encoder-decoder model with
attention.

NMT Copy: Jhamtani et al., (2017)
introduce a copy-enriched NMT
model for style transfer to better
handle stretches of text which
should not be changed. We
incorporate this mechanism into our
NMT Baseline.

NMT Combined:We augment the
data used to train NMT Copy via two
techniques:

1) we run the PBMT model on

Human-based Evaluation:

Formality,fluency,meaning
preservation,overall Ranking

Automatic Metrics:

Formality,fluency,meaning
preservation,overall Ranking

Model	Formality		Fluency		Meaning		Combined		Overall	
	Human	PT16	Human	H14	Human	H15	Human	Auto	BLEU	PINC
Original Informal	7.23	1.00	7.60	2.80	7.50	2.80	7.50	2.80	50.00	0.00
Formal Reference	0.38	0.17	4.45	1.52	4.57	3.64	5.68	4.67	100.0	0.57
Rule-based	0.30	0.34	4.00	3.00	4.88	4.41	5.54	4.00	61.38	0.21
PBMT	0.10*	0.00*	3.96	3.24*	4.64*	4.19*	5.27	4.82*	67.20*	0.26
NMT Baseline	0.80*	0.07*	4.05	3.82*	3.55*	3.89*	4.96*	4.84*	56.61	0.30*
NMT Copy	0.02*	0.00*	4.07	3.45*	3.88*	3.87*	4.97*	4.81*	58.01	0.30*
NMT Combined	0.10*	0.00*	4.09*	3.77*	4.46*	4.20*	5.32*	4.82*	67.87*	0.26

Table 13: Results of models on 500 test sentences for informal to formal task evaluated using human judgments and automatic metrics for three criteria of evaluation: formality, fluency and meaning preservation on the F&R domain. Scores marked with * are significantly different from the rule-based scores with $p < 0.001$.

		additional source data, and 2) we use back-translation (Sennrich et al., 2016c) of the PBMT model to translate the large number of in-domain target style sentences from GYAFC.	

There is a delicate balance between retaining the original content and removing the original attributes, and existing adversarial models tend to sacrifice one or the other.

It introduces a simple approach to text attribute transfer whose main advantage comes from induction bias, i.e. attributes are usually represented as locally recognisable phrases, which allows it to outperform previous models based on adversarial training. It also reflects some problems in that content and attributes cannot be so cleanly separated by word boundaries.

A fruitful direction would be to develop a concept of attributes that is more general than N-grams, more general than N-grams, but with more inductive bias than arbitrary potential vectors.