



# Microsoft Translator at WMT 2019

## Towards Large-Scale Document-Level Neural Machine Translation



Marcin Junczys-Dowmunt

We train large-scale document level translation models. Our approach to document-level MT is inspired by recent work in language modeling – BERT by Devlin et al. (2018) and GPT-2 by Radford et al. (2019) – rather than previous document-level MT work. Instead of explicitly modeling document-level context, we just train on and translate sequences of up to 1024 sub-words.

Our document-level models drastically outperform comparable sentence-level models in terms of human evaluation.

## Data Augmentation

- 1 Data-filtering via Dual-Crossentropy Filtering and language identification;
- 2 Noised back-translation of WMT monolingual German data with document boundaries;
- 3 Oversampling document-level parallel data via randomly selected sub-documents;  $\frac{1}{2}$  size of back-translations;
- 4 Oversampling of parallel data via randomly assembled fake documents;  $\frac{1}{2}$  size of back-translations;

## Models and Training

We train 6 and 12-block transformer models on sequences of up to 1024 subwords with relatively standard hyper-parameters.

## Human Evaluation

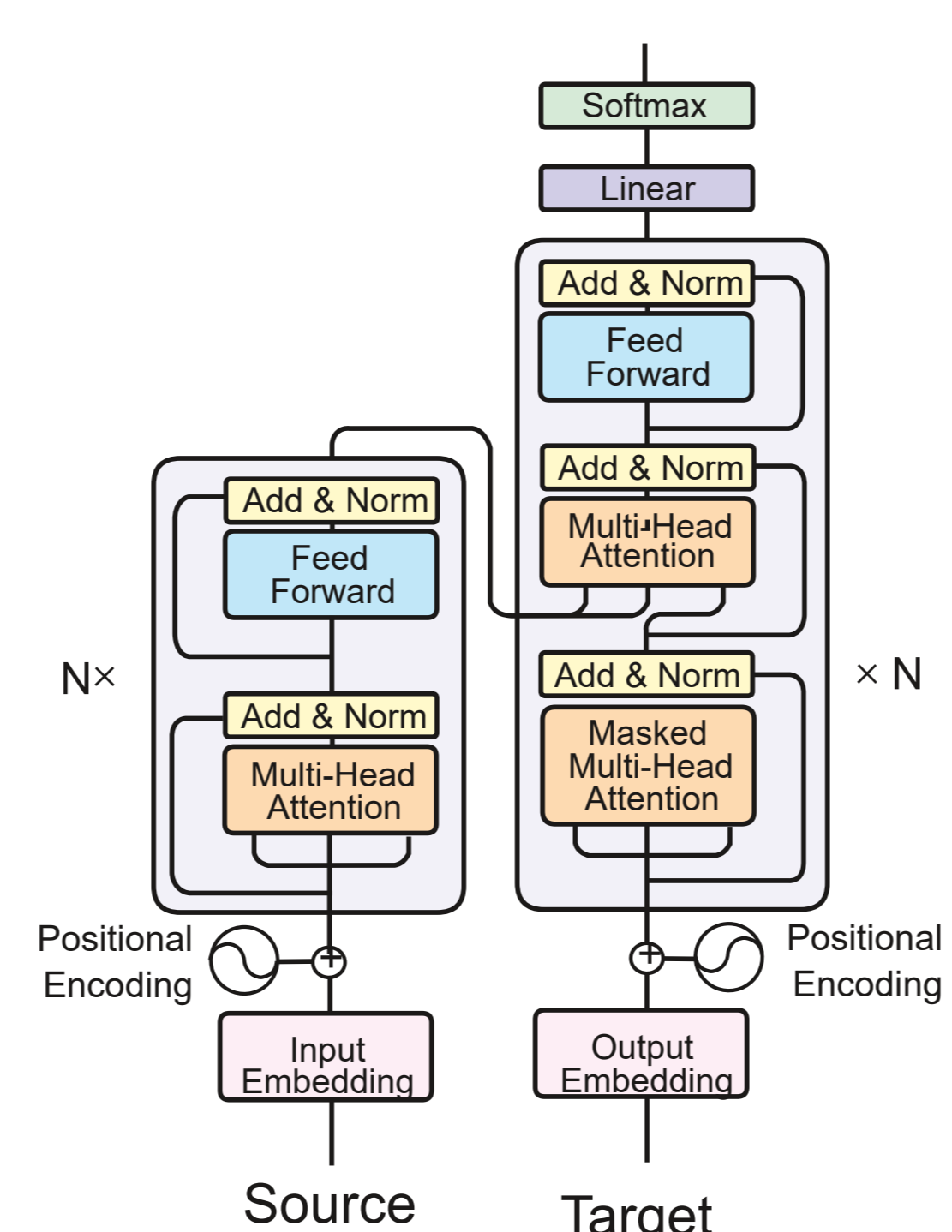
<https://www.overleaf.com/project/5d323c94d273>

Ave.	Ave. z	System
90.3	0.347	Facebook-FAIR
93.0	0.311	Microsoft-WMT19-sent-doc
92.6	0.296	Microsoft-WMT19-doc-level
90.3	0.240	HUMAN
87.6	0.214	MSRA-MADL
...		
84.2	0.094	online-B
86.6	0.094	Microsoft-WMT19-sent-level
87.3	0.081	JHU
...		
82.4	-0.132	TartuNLP-c
76.3	-0.400	online-X
43.3	-1.769	en-de-task

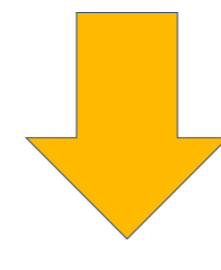
In 911 Call, Professor Admits to Shooting Girlfriend  
 In a 911 call, his voice only slightly shaky, college professor Shannon Lamb told police he had shot his girlfriend and officers needed to get over to their house.  
 Lamb made a point to say his "sweet dog" was there alive and probably upset, and said the dead woman's family contacts could be found on her phone.  
 Inside the home, officers found Amy Prentiss' body and a hand-written note scribbled on a white legal pad: "I am so very sorry I wish I could take it back I loved Amy and she is the only woman who ever loved me", read the letter authorities say was signed by Lamb.  
 There was no indication that Lamb, who was teaching two online classes for Delta State University in Cleveland, Mississippi, had already traveled 300 miles to the school's campus, where police believe he shot and killed a well-liked history professor, Ethan Schmidt, in the doorway to his office.  
 Delta State University police chief Lynn Buford said university officials heard about the shooting at 10:18 a.m.  
 He said Lamb made the fateful 911 call sometime after that.  
 By the end of the day, there would be one more death: Lamb took his own life as police closed in on him.  
 A day after the school shooting forced students and faculty to hide behind locked doors, authorities were still trying to piece together what motivated Lamb.  
 The details released by investigators at both ends of the state as well as students and staff who knew him helped paint a picture of a talented but possibly troubled teacher.  
 Students said they looked forward to his class.  
 Police in Gautier, where Prentiss died, said he had no history of violence or criminal record.  
 Schmidt himself had included Lamb in a book he wrote where he acknowledged the "wonderful people" he shared his academic life with.  
 Both taught in the Division of Social Sciences and History, which lists 17 faculty members, and many students took courses from both.  
 At the same time, there were some inclinations of problems.



<BEG> In 911 Call, Professor Admits to Shooting Girlfriend <SEP> In a 911 call, his voice only slightly shaky, college professor Shannon Lamb told police he had shot his girlfriend and officers needed to get over to their house. <SEP> Lamb made a point to say his "sweet dog" was there alive and probably upset, and said the dead woman's family contacts could be found on her phone. <SEP> Inside the home, officers found Amy Prentiss' body and a hand-written note scribbled on a white legal pad: "I am so very sorry I wish I could take it back I loved Amy and she is the only woman who ever loved me", read the letter authorities say was signed by Lamb. <SEP> There was no indication that Lamb, who was teaching two online classes for Delta State University in Cleveland, Mississippi, had already traveled 300 miles to the school's campus, where police believe he shot and killed a well-liked history professor, Ethan Schmidt, in the doorway to his office. <SEP> Delta State University police chief Lynn Buford said university officials heard about the shooting at 10:18 a.m. <SEP> He said Lamb made the fateful 911 call sometime after that. <SEP> By the end of the day, there would be one more death: Lamb took his own life as police closed in on him. <SEP> A day after the school shooting forced students and faculty to hide behind locked doors, authorities were still trying to piece together what motivated Lamb. <SEP> The details released by investigators at both ends of the state as well as students and staff who knew him helped paint a picture of a talented but possibly troubled teacher. <SEP> Students said they looked forward to his class. <SEP> Police in Gautier, where Prentiss died, said he had no history of violence or criminal record. <SEP> Schmidt himself had included Lamb in a book he wrote where he acknowledged the "wonderful people" he shared his academic life with. <SEP> Both taught in the Division of Social Sciences and History, which lists 17 faculty members, and many students took courses from both. <SEP> At the same time, there were some inclinations of problems. <SEP> <BRK>



<BEG> In 911-Anruf gibt Professor zu, Freundin erschossen zu haben <SEP> In einem 911-Anruf sei seine Stimme nur leicht wackelig gewesen, sagte College-Professor Shannon Lamb der Polizei, er habe seine Freundin erschossen und Beamte hätten in ihr Haus eindringen müssen. <SEP> Lamb gab zu Protokoll, sein "süßer Hund" sei lebend da gewesen und vermutlich aufgebracht gewesen und sagte, die Familienkontakte der toten Frau seien auf ihrem Handy gefunden worden. <SEP> Im Inneren des Hauses fanden Beamte die Leiche von Amy Prentiss und einen handschriftlichen Zettel, der auf ein weißes Postler gekritzelt war: "Es tut mir so leid, dass ich wünsche, ich könnte sie zurücknehmen, ich liebte Amy und sie ist die einzige Frau, die mich jemals geliebt hat", hieß es in dem Brief, den die Behörden unterschrieben hatten. <SEP> Es gab keine Hinweise darauf, dass Lamb, der zwei Online-Kurse für die Delta State University in Cleveland, Mississippi, unterrichtete, bereits 300 Meilen zum Campus der Schule gereist war, wo er nach Polizeiangaben einen beliebten Geschichtsprofessor, Ethan Schmidt, in der Tür zu seinem Büro erschossen haben soll. <SEP> Der Polizeichef der Delta State University, Lynn Buford, sagte, die Beamten der Universität hätten von der Schießerei um 10:18 Uhr gehört. <SEP> Er sagte, Lamb habe den schicksalhaften 911-Anruf irgendwann danach gemacht. <SEP> Am Ende des Tages werde es einen weiteren Tod geben: Lamb habe sich das Leben genommen, als die Polizei ihn eingeschlossen habe. <SEP> Einen Tag nach der Schießerei in der Schule, die Studenten und Dozenten gezwungen hatte, sich hinter verschlossenen Türen zu verstecken, versuchten die Behörden noch immer zusammenzufügen, was Lamb motivierte. <SEP> Die von den Ermittlern an beiden Enden des Staates veröffentlichten Details sowie Studenten und Angestellte, die ihn kannten, halfen dabei, ein Bild eines talentierten, aber möglicherweise aufgewühlten Lehrers zu malen. <SEP> Studenten sagten, sie freuten sich auf seine Klasse. <SEP> Die Polizei in Gautier, wo Prentiss starb, sagte, er habe keine Geschichte von Gewalt oder Vorstrafen gehabt. <SEP> Schmidt selbst hatte Lamb in ein von ihm verfasstes Buch aufgenommen, in dem er die "wunderbaren Menschen" würdigte, mit denen er sein akademisches Leben teilte. <SEP> Beide lehrten in der Abteilung für Sozialwissenschaften und Geschichte, die 17 Fakultätsmitglieder auflistet, und viele Studenten nahmen Kurse von beiden. <SEP> Gleichzeitig gab es einige Neigungen zu Problemen. <SEP> <BRK>



In 911-Anruf gibt Professor zu, Freundin erschossen zu haben  
 In einem 911-Anruf sei seine Stimme nur leicht wackelig gewesen, sagte College-Professor Shannon Lamb der Polizei, er habe seine Freundin erschossen und Beamte hätten in ihr Haus eindringen müssen.  
 Lamb gab zu Protokoll, sein "süßer Hund" sei lebend da gewesen und vermutlich aufgebracht gewesen und sagte, die Familienkontakte der toten Frau seien auf ihrem Handy gefunden worden.  
 Im Inneren des Hauses fanden Beamte die Leiche von Amy Prentiss und einen handschriftlichen Zettel, der auf ein weißes Postler gekritzelt war: "Es tut mir so leid, dass ich wünsche, ich könnte sie zurücknehmen, ich liebte Amy und sie ist die einzige Frau, die mich jemals geliebt hat", hieß es in dem Brief, den die Behörden unterschrieben hatten.  
 Es gab keine Hinweise darauf, dass Lamb, der zwei Online-Kurse für die Delta State University in Cleveland, Mississippi, unterrichtete, bereits 300 Meilen zum Campus der Schule gereist war, wo er nach Polizeiangaben einen beliebten Geschichtsprofessor, Ethan Schmidt, in der Tür zu seinem Büro erschossen haben soll.  
 Der Polizeichef der Delta State University, Lynn Buford, sagte, die Beamten der Universität hätten von der Schießerei um 10:18 Uhr gehört.  
 Er sagte, Lamb habe den schicksalhaften 911-Anruf irgendwann danach gemacht.  
 Am Ende des Tages werde es einen weiteren Tod geben: Lamb habe sich das Leben genommen, als die Polizei ihn eingeschlossen habe.  
 Einen Tag nach der Schießerei in der Schule, die Studenten und Dozenten gezwungen hatte, sich hinter verschlossenen Türen zu verstecken, versuchten die Behörden noch immer zusammenzufügen, was Lamb motivierte.  
 Die von den Ermittlern an beiden Enden des Staates veröffentlichten Details sowie Studenten und Angestellte, die ihn kannten, halfen dabei, ein Bild eines talentierten, aber möglicherweise aufgewühlten Lehrers zu malen.  
 Studenten sagten, sie freuten sich auf seine Klasse.  
 Die Polizei in Gautier, wo Prentiss starb, sagte, er habe keine Geschichte von Gewalt oder Vorstrafen gehabt.  
 Schmidt selbst hatte Lamb in ein von ihm verfasstes Buch aufgenommen, in dem er die "wunderbaren Menschen" würdigte, mit denen er sein akademisches Leben teilte.  
 Beide lehrten in der Abteilung für Sozialwissenschaften und Geschichte, die 17 Fakultätsmitglieder auflistet, und viele Studenten nahmen Kurse von beiden.  
 Gleichzeitig gab es einige Neigungen zu Problemen.

## Bonus Material: ContraPro Challenge

The ContraPro challenge test set by Müller et al. (2018) measures document-level MT performance for anaphora resolution.

Previous state of the art: ca. 60%  
Ensemble of our models with 10-10 context: 88.5%

## How much context?

		Left context (sentences, src+trg)										
		0	1	2	3	4	5	6	7	8	9	10
Right context (sentences, src-only)	0	0.4920	0.6800	0.7314	0.7728	0.8031	0.8214	0.8379	0.8476	0.8513	0.8573	0.8611
	1	0.5001	0.7179	0.7724	0.8046	0.8297	0.8384	0.8486	0.8556	0.8605	0.8638	0.8644
	2	0.5029	0.7509	0.8033	0.8258	0.8423	0.8476	0.8564	0.8608	0.8639	0.8655	0.8675
	3	0.5090	0.7809	0.8215	0.8399	0.8513	0.8552	0.8612	0.8633	0.8667	0.8674	0.8695
	4	0.5129	0.7988	0.8366	0.8500	0.8574	0.8604	0.8646	0.8664	0.8680	0.8690	0.8708
	5	0.5153	0.8139	0.8459	0.8583	0.8628	0.8651	0.8672	0.8689	0.8707	0.8699	0.8708
	6	0.5191	0.8238	0.8528	0.8631	0.8667	0.8673	0.8705	0.8698	0.8717	0.8702	0.8726
	7	0.5219	0.8313	0.8568	0.8662	0.8697	0.8700	0.8710	0.8698	0.8717	0.8723	0.8731
	8	0.5206	0.8342	0.8623	0.8694	0.8710	0.8703	0.8707	0.8703	0.8717	0.8728	0.8735
	9	0.5212	0.8375	0.8659	0.8709	0.8714	0.8706	0.8716	0.8724	0.8717	0.8721	0.8753
	10	0.5228	0.8408	0.8668	0.8711	0.8723	0.8705	0.8723	0.8725	0.8717	0.8724	0.8742

## Visualization

