Syntax-Informed Interactive Neural Machine Translation

Kamal Kumar Gupta[1], Rejwanul Haque[2], Asif Ekbal[1], Pushpak Bhattacharyya[1], and Andy Way[2]

[1]Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India
[2]ADAPT Centre, School of Computing, Dublin City University, Dublin, Ireland
[1]{kamal.pcs17, asif, pb}@iitp.ac.in
[2]firstname.lastname@adaptcentre.ie

***Abstract*—In interactive machine translation (MT), human translators correct errors in automatic translations in collaboration with the MT systems, and this is an effective way to improve productivity gain in translation. Phrase-based statistical MT (PB-SMT) has been the mainstream approach to MT for the past 30 years, both in academia and industry. Neural MT (NMT), an end-to-end learning approach to MT, represents the current state-of-the-art in MT research. The recent studies on interactive MT have indicated that NMT can significantly outperform PB-SMT.**

In this work, first we investigate the possibility of integrating lexical syntactic descriptions in the form of supertags into the state-of-the-art NMT model, Transformer. Then, we explore whether integration of supertags into Transformer could indeed reduce human efforts in translation in an interactive-predictive platform. From our investigation we found that our syntax-aware interactive NMT (INMT) framework significantly reduces simulated human efforts in the French–to–English and Hindi–to–English translation tasks, achieving a 2.65 point absolute corresponding to 5.65% relative improvement and a 6.55 point absolute corresponding to 19.1% relative improvement, respectively, in terms of word prediction accuracy (WPA) over the respective baselines.

***Index Terms*—machine translation, neural machine translation, interactive neural machine translation**

I. Introduction

Translation service providers (TSPs) who use MT in their production exploit human translators for correcting erroneous automatic translations, and by this, they produce high quality translations for their corporate customers. Interactive MT, a promising use-case of the industrial MT services and an active field of MT research, aims to reduce human efforts in automatic translation workflows (TWs) with employing an iterative collaborative strategy with its two most important components: human translators and MT engine. Figure 1 represents the interactive-predictive protocol.

The recent studies on interactive MT [1], [2] have shown that NMT [3] can significantly surpass PB-SMT [4]. The MT researchers have also investigated integration of advance machine learning features into the interactive MT models in order to further minimise human efforts in translation [5], [6]. In a different MT research context, Nadejde et al. [7] have integrated CCG (combinatory categorical grammar) syntactic categories [8] into the target-side of the then state-of-the-art attentional recurrent neural network (RNN) MT models [3], and they found that supertags can help improve translation quality in a German-to-English translation task, a high-resource language-pair, and a Romanian-to-English translation task, a low-resource language-pairs.

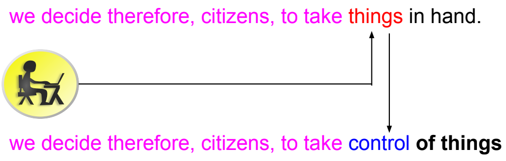

we decide therefore, citizens, to take things in hand.
we decide therefore, citizens, to take control of things

Fig. 1: Interactive protocol in collaboration with an MT system and a user. The user wants to translate the French sentence ‘Nous décidons donc, citoyens, de prendre les choses en main.’ to English. The reference translation is ‘we decide therefore, citizens, to take control of things’ which is used here to simulate the user. The user corrects the first wrong word (*things*) from the hypothesis. The validated prefix (magenta phrase) and the last modified word (*control*) are fed back to the NMT system which generates a correct suffix (*of things*).

In this work, we first investigate the possibility of integrating supertags into the current state-of-the-art NMT model, Transformer [9], and then explore whether integration of the supertags into Transformer could indeed reduce human efforts in translation in an interactive-predictive scenario. To summarise, our main contributions in this paper are as follows: (i) to the best of our knowledge, this is the first study that investigates the possibility of integrating syntactic knowledge sources into an interactive MT model, (ii) we explore the possibility of integrating CCG supertags into the current state-of-the-art MT system, Transformer, (iii) we test our syntax-informed interactive Transformer models on French-to-English, a high resource language pair, and Hindi-to-English, a low-resource language pair, and present our results with a thorough and comparative analysis on translations produced by our syntax-informed and baseline interactive MT systems.

The remainder of the paper is organised as follows. In Section II, we discuss related work. Section III provides a short description on our syntax-informed interactive NMT, and Section IV explains our motivation for considering supertags in our experiments. In Section V, we present our experimental setups. Section VI describes our evaluation plan, experimental results and analysis, while Section VII concludes and provides avenues for further work.

II. Related Work

Foster et al. [10] were the first to introduce the idea of interactive-predictive MT, as an alternative to pure post-editing MT. There have been a number of papers that explored this strategy in order to minimise human efforts in translation and cover many use-cases involving SMT: e.g. applying online [11] and active [12] learning techniques, use of translation memories [13], [14], predicting the partially typed words and prefix matching [15], word-graphs for reducing response time [16], segment-based approaches [2], suggesting more than one suffix [17], and exploring multimodal interaction [18]. Since the introduction of NMT to the MT community, researchers have been investigating interactive-predictive protocol with the RNN-based MT models, with a focus on reducing human efforts in translation, e.g. [19], [1], [5], [6]. As of yet, to the best of our knowledge, no one has investigated the interactive-predictive protocol with the current state-of-the-art Transformer model [9].

The strategy of exploiting syntactic knowledge from the source and/or target languages for the betterment of translation is not new in MT research; it was successfully applied in the era of classical MT, e.g. [20], [21], [22], [23], and is continually being applied to improve current state-of-the-art NMT models, e.g. [7], [24], [25], [26]. Nadejde et al. [7] have exploited CCG syntactic categories [8] from target language in order to improve a RNN MT model [3]. In this paper, we investigate the possibility of integrating CCG supertags into the current state-of-the-art MT system, Transformer, with an aim to minimise human efforts in translation in an interactive-predictive scenario.

III. Syntax-informed Interactive NMT

This section presents our syntax-informed interactive NMT model. In NMT, at time step *i*, the conditional probability of predicting output token y_i given a source sentence x and the previously generated output token $y_1, ..., y_{i-1}$ is modelled as $\mathrm{p}(y_i|\{y_1, ..., y_{i-1}\}, x)$.

In interactive protocol, the user corrects the left-most wrong word of translation produced by the MT system. The feedback is returned back to the MT system in the form of $\hat{y}_1^{i-1}$ which is the validated prefix together with the corrected word. Thus, in interactive NMT, the conditional context becomes $\hat{y}_1^{i-1}$, and the conditional probability of predicting output token y_i is modelled as $\mathrm{p}(y_i|\{\hat{y}_1, ..., \hat{y}_{i-1}\}, x)$.

In our work, we adopted the best-performing strategy (i.e. *interleaving*) of Nadejde et al. [7], which first predicts the CCG supertag ($\hat{s}_i$) of the word (y_i) to be predicted next. As a result, the length of conditional context becomes twice the number of words in context plus one. As far as the syntax-informed interactive NMT is concerned, the conditional probability of predicting output token y_i is modelled as $\mathrm{p}(y_i|\{\hat{s}_1, \hat{y}_1, ..., \hat{s}_{i-1}, \hat{y}_{i-1}, \hat{s}_i\}, x)$ where $\hat{s}_1^{i-1}$ is the CCG sequence of the validated prefix $\hat{y}_1^{i-1}$ and $\hat{s}_i$ is the supertag of the word (y_i) to be predicted next.

IV. Syntactic Context Features

This section explains why we consider a rich and complex syntactic feature, supertags, as context in our experiments. Supertagging, a kind of syntactic parsing (e.g. lexicalised tree adjoining grammar [27], [28], combinatory categorial grammar [8], [29], [30]), assigns rich and complex lexical syntactic descriptions (i.e. supertags) to words. In other words, supertags include information such as the POS tag and local subcategorisation information of a word and the hierarchy of phrase categories that the word projects upwards. They are known to be context sensitive tags that preserve the global syntactic information at local lexical level. Having this property, supertags resolve ambiguity in short- and long-distance dependencies by capturing the previous and next syntactic dependencies of a lexical term. For example, it signifies whether a particular lexical term is expecting a preposition in the sentence or denoting which among other lexical terms a particular term corresponds to. The interactive neural MT model predicts new hypothesis primarily based on a validated context (prefix) including the left-most modified word by the user. In case of our syntax-informed model, the prediction of the next words is also conditioned on the CCG supertags [8] of the user validated prefix and the word to be predicted next. Our intuition underpinning modelling supertags in this work is that such complex and rich syntactic knowledge sources, which inherently capture long-distance word-to-word dependencies in a sentence, may be useful to improve the subsequent predictions in interactive NMT, especially for the longer sentences.

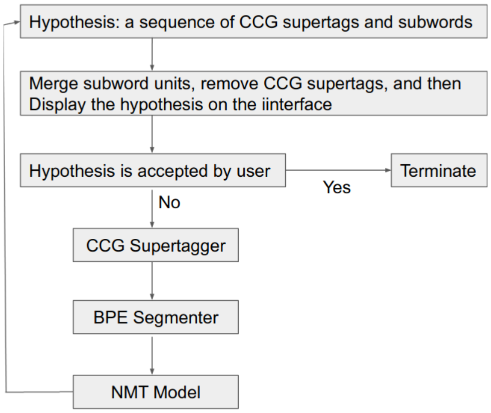

Hypothesis: a sequence of CCG supertags and subwords
Merge subword units, remove CCG supertags, and then Display the hypothesis on the iinterface
Hypothesis is accepted by user
Yes
Terminate
No
CCG Supertagger
BPE Segmenter
NMT Model

Fig. 2: Flowchart of hypothesis generation process in our syntax-informed INMT platform.

V. Experimental Setups

A. Methods of forming conditional syntactic context

We carried out our experiments following two setups, which are described as follows. In our first setup, we directly use the supertags that are predicted by Transformer as a part

TABLE I: An example of an English sentence with its words associated with CCG supertags. A: a French (source) sentence, B: subword form of source sentence, C: an English (target) sentence, D: subword form of target sentence,E: CCG supertags are distributed to each of the subwords of a word. As an example: CCG supertag *N/N* of word *themed* is distributed over all its subword units, i.e. them@@ and ed.

A	J'ai jamais joué à un jeu à boire basé sur le thème Nazi cela dit .
B	J@@ '@@ ai jamais joué à un jeu à b@@ oire basé sur le thème N@@ azi cela dit .
C	never played a Nazi themed drinking game though .
D	never played a Nazi them@@ ed drinking game though .
E	NP never (S[dcl]\NP)/NP played NP[nb]/N a N/N Nazi N/N them@@ N/N ed N/N drinking N game (NP\NP)/NP though N .

TABLE II: An example showing applying *On the fly CCG supertagger* on hypothesis. As can be seen from rows 5 and 6, the user replaces incorrect token *play* with correct token *drinking*. The new token *drinking* gets the CCG supertag of the incorrect token *play*, *(S[dcl]\NP)/NP*, which is also incorrect. In second setup, *On the fly CCG supertagger* is applied on hypothesis (validated prefix and suffix). As can be seen from row 7, a new CCG supertag sequence is generated for the hypothesis, and we see that the CCG supertag (*N*) is assigned to the new token *drinking*.

Input sentence (subword)	J@@ '@@ ai jamais joué à un jeu à b@@ oire basé sur le thème N@@ azi cela dit .
Reference	never played a Nazi themed drinking game though .
Initial Hypothesis	never played a Nazi drinking play there .
Hypothesis after several iterations	NP never S[pss]\NP played NP/NP a N/N Nazi N them@@ N ed **(S[dcl]\NP)/NP play** (NP\NP)/NP though N .
INMT interface	never played a Nazi themed **play** though .
Correction by user	never played a Nazi themed **drinking** though .
Applying on the fly CCG supertagger	NP never S[pss]\NP played NP/NP a N/N Nazi N them@@ N ed **N drinking** (S\NP)\(S\NP) though N/N .
New hypothesis	never played a Nazi themed **drinking game** though .

of conditional context for the prediction of the remaining hypothesis. This means this setup follows the interleaving technique of Nadejde et al. [7], in which the CCG supertag of a token is kept before the token as shown in Table I. For an example, $word_i$ is produced by the decoder in a hypothesis having ccg_i as its CCG supertag that has been predicted in the previous time step. In the interface, the user sees that $word_i$ is not appropriate in the context, i.e. incorrectly predicted by the MT system, and replaces $word_i$ with a new token $word_{new}$. Now, when the modified context (i.e. validated prefix) is fed back to the NMT model, $word_{new}$ will have the tag of $word_i$, i.e. ccg_i. In other words, the final two tokens of the conditional context would be ccg_i $word_{new}$. We carried out an analysis to see how closely these supertags are related to the new words that have been added by the user (cf. Section VI-D). In this regard, since we followed standard practice for NMT training, we applied the byte-pair encoding (BPE) segmentation[1] [31] to the tokens. The subword units of a word inherit the CCG supertag of the word. As an example, we show an English sentence with supertags in Table I. We see from row E of Table I that CCG supertag 'N/N' of a word 'themed' is distributed over its subwords (i.e. them@@ and ed).

In this context, Akoury et al. [24] showed that integrating target-side ground-truth syntactic information into Transformer at decoding time significantly improved their system's translation quality, and their syntax-based model outperformed the baseline Transformer model with a large margin in terms of BLEU [32]. However, in reality, there is no way of obtaining the target-side ground-truth syntactic information at decoding time. But, in interactive-predictive mode, we got a way to obtain a slightly better CCG sequence for the partial translation (i.e. validated prefix) and inject them into the model at runtime, which we believe can positively impact the model's subsequent predictions. In other words, in our second setup, we integrate a CCG supertagger in our INMT framework, and apply that on validated prefix and unchecked suffix on the fly. The supertagger is invoked when the user makes a correction. In other words, when user inserts a new token $word_{new}$ in the place of an incorrectly predicted token ($word_i$), the CCG supertagger is invoked and applied to the validated prefix and unchecked suffix on the fly. As above, Section VI-D shows statistics in relation to the quality of such supertag sequences. In Table II, we show how we apply *On the fly CCG supertagger* on modified hypothesis via an example.

B. MT systems

We carried out our experiments with a high-resource language-pair, French-to-English. This is regarded as an important language-pair in the translation industry. In addition to this, we tested our method on a low-resource and less-explored language pair, Hindi-to-English. For French-to-English we used UN corpus[2] [33], and the training and development sets contain 12,238,995 and 1,500 sentences, respectively. For Hindi-to-English we used the IIT Bombay English-Hindi

[1] https://github.com/rsennrich/subword-nmt

[2] https://www.statmt.org/wmt13/training-parallel-un.tgz

parallel corpus[3] [34] that is compiled from a variety of existing sources, e.g. OPUS[4] [35], and the training and development sets contain 1,513,548 and 520 sentences, respectively.

As for the French-to-English task, we used 1,500 sentences from the WMT15 news test set *newstest2015* as our test set. For the Hindi-to-English task we considered the WMT14 news test set *newstest2014* as our test set.

In order to build our MT systems, we used Sockeye[5] [36] toolkit. Our training set-up is described below. The tokens of the training, evaluation and validation sets are segmented into subword units using the BPE technique [37] proposed by [31]. We performed 32,000 join operations. We use 6 layers at encoder and decoder sides each, 8-head attention, hidden layer of size 512, embedding vector of size 512, learning rate 0.0002, minimum batch size of 1800 tokens. Easyccg[6] [38] tool is used for generating CCG supertag sequences for English sentences.

Table III shows the performance of the baseline and our syntax-informed NMT systems in terms of BLEU for both the French-to-English and Hindi-to-English translation tasks. We see from Table III that the BLEU scores of the baseline and syntax-informed MT systems are comparable in both cases. Additionally, we performed statistical significance test using bootstrap resampling methods [39]. We found that the difference in BLEU scores of the MT systems (baseline and syntax-informed) are not statistically significant. Surprisingly, this finding contradicts with the findings of [7] who found supertags helpful in their case and their systax-based NMT systems significantly surpassed their baseline RNN MT systems.

TABLE III: BLEU scores for baseline and syntax-informed NMT systems

	Fr→En	Hi→En
Baseline	26.9	18.12
Syntax-Informed NMT system	27.1	18.81

VI. Results and Discussion

In this section first we explain the strategy that we adopted for evaluating the interactive-predictive MT systems. Then, we present our evaluation results with some discussions and analysis.

A. Evaluation Strategy for INMT

We evaluate the performance of the INMT systems using two evaluation metrics, word stroke ratio (WSR) and word prediction accuracy (WPA). WSR denotes the total number of token replacements required to obtain the desired hypothesis [2]. Word prediction accuracy (WPA) is the percentage of words that the INMT system predicted correctly, given a prefix of all the previous translator-produced words [1]. The process of evaluating translations in interactive scenarios is expensive as it requires human evaluators. As an alternative, we adopted the reference-simulated evaluation strategy as in [2], where instead of taking feedback from the real user, reference sentence is used as the feedback. In other words, the reference sentences are used to simulate the user. Note that the supertag sequences of the reference sentences are not considered for evaluation. As shown in Figure 2, each time the NMT system generates a hypothesis it is compared with the reference sentence from left to right.

TABLE IV: WSR and WPA scores of the syntax-informed and baseline INMT systems. **A.** Fr→En and **B.** Hi→En

		Baseline	CCG supertags by Transformer	On the fly CCG supertagger	CCG Tagged (GT)
A	WSR	53.77	**51.70**	**50.61**	**29.44**
	WPA	46.82	**48.29**	**49.47**	**70.53**
B	WSR	65.68	**61.58**	**59.12**	**36.89**
	WPA	34.32	**38.41**	**40.87**	**63.10**

B. Evaluation Results

We obtain WPA and WSR scores to evaluate the French-to-English and Hindi-to-English INMT systems on the test sets, which are reported in Table IV. Note that WSR is an error metric, which means that lower scores are better. The top- and bottom-half of the table represents the French-to-English and Hindi-to-English translation tasks, respectively. We can see from the table that our the supertag-based INMT systems outperform the respective baselines regardless of the translation tasks.

We obtained best WSR and WPA scores when *on the fly CCG supertagger* is applied on the modified hypothesis (see second experimental setup; cf. Section V-A). As for the French-to-English task, we achieve a 2.65 point absolute corresponding to 5.65% relative improvement in terms of WPA and a 3.16 point absolute corresponding to 5.87% relative reduction in terms of WSR over the baseline. As far as the Hindi-to-English translation task is concerned, we achieve a 6.55 point absolute corresponding to 19.1% relative improvement in terms of WPA and a 6.65 point absolute corresponding to 9.98% relative reduction in terms of WSR over the baseline. We found that these gains are statistically significant [40]. For comparison, we also report the WPA and WSR scores of our syntax-informed INMT systems on an ideal setup, i.e. when we feed Transformer with the ground-truth CCG supertags instead of those predicted by the model or generated by the *on the fly CCG supertagger*. As expected, in this setup, the syntax-informed INMT systems surpassed their respective baselines with a large-margin (cf. last column of Table IV).

C. Evaluation based on test set sentence lengths

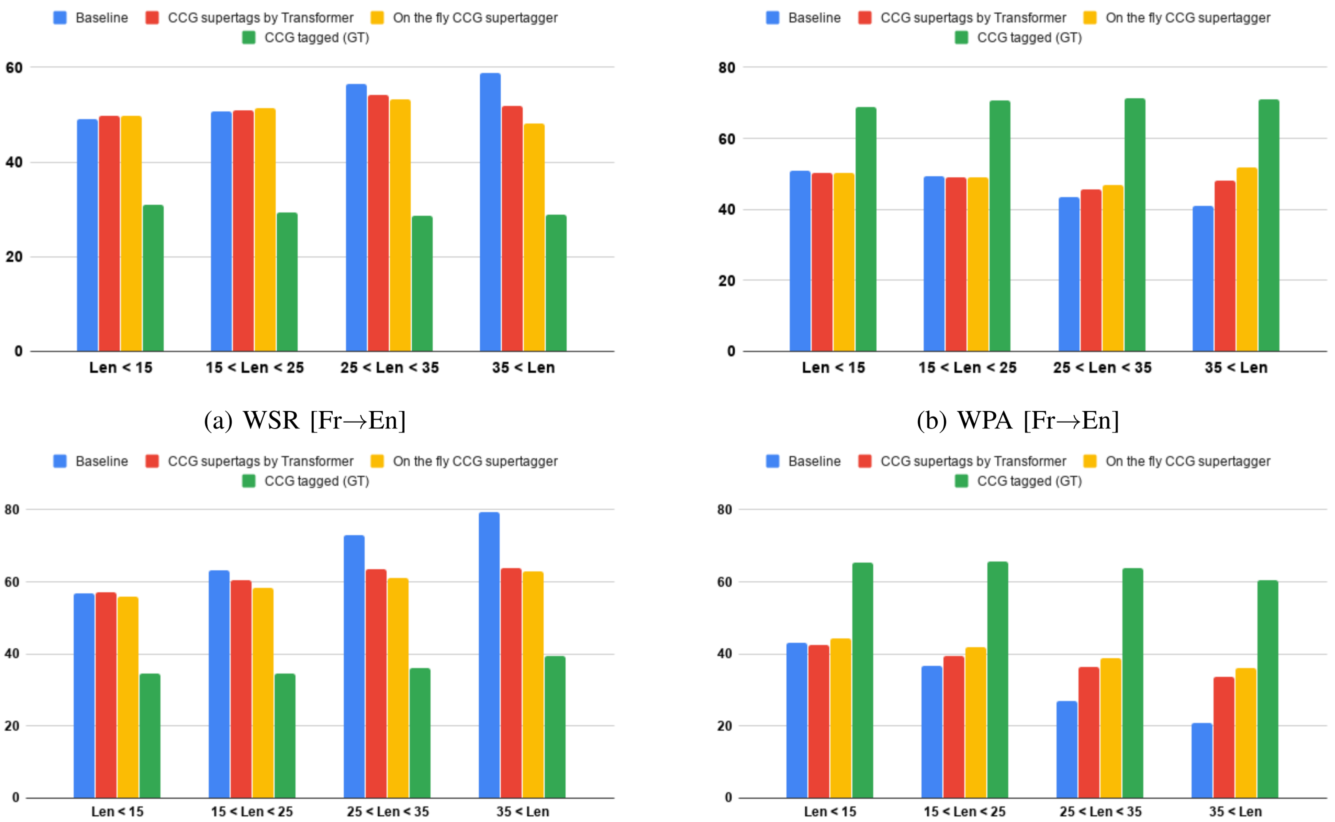
For further analysis, we place the sentences of the test set into four sets (cf. Figure 3) as per the sentence length measures, i.e. the first set contains those sentences whose

[3] http://www.cfilt.iitb.ac.in/iitb_parallel/

[4] http://opus.lingfil.uu.se/

[5] https://github.com/awslabs/sockeye

[6] https://github.com/mikelewis0/easyccg



(a) WSR [Fr→En]

(b) WPA [Fr→En]

(c) WSR [Hi→En]

(d) WPA [Hi→En]

Fig. 3: WSR and WPA scores of the syntax-informed and baseline INMT systems with respect to the sentence-length based test sets.

lengths are less than and equal to 15, the second set contains those sentences whose lengths are above 15 and less than and equal to 25, the third set contains those sentences whose lengths are above 25 and less than and equal to 35, and the fourth set contains those sentences whose lengths are above 35. This division was made based on the lengths of reference sentences. In Figure 3, we plot the distributions of WPA and WSR scores over the sentence-length based sets. This figure provides us a better insights on the performance of the baseline and syntax-informed MT systems, especially how they will perform on the varying length of sentences that they would have to predict. As can be seen from Figure 3, our syntax-based INMT systems produce increasingly better WSR and WPA scores as the length of the reference sentences increases. As discussed above, supertags encode wider context of a sentence, which could help the decoder to capture long-range word-to-word dependencies at generation time. Hence, the integration of the rich syntactic feature (CCG supertags) from the target-side into the interactive NMT system can play an important role for the translation of longer sentences. The next section provides further analysis and discusses the impact of incorporating CCG supertags on interactive predictions.

TABLE V: % of CCG supertags that becomes incorrect when the user replaces the incorrectly predicted token in hypothesis with the token of his choice.

	Fr–>En		Hi–>En	
	% wrong tags (CCG supertags by transformer)	% wrong tags (On the fly CCG supertagger)	% wrong tags (CCG supertags by transformer)	% wrong tags (On the fly CCG supertagger)
Whole testset	41.07	23.95	45.79	24.51
Len <15	40.64	23.88	44.58	23.72
15 <Len <25	40.84	23.04	45.47	24.14
25 <Len <35	42.80	25.28	47.25	25.96
35 <Len	39.32	24.33	46.94	24.95

D. *CCG supertags of the words of the user choice*

As mentioned in Section V-A, we came up with two different ways to use CCG supertags as the conditional context for the predictions in INMT. First, in *CCG supertags by Transformer* setup, if the user makes a correction, the user's choice of word inherits the CCG supertag of the word that the user has just corrected, which is, in fact, predicted by the INMT system. The new word and the incorrect word that the user has just corrected could be syntactically or semantically different. As a result, the supertag that the new word inherits could be incorrect. We calculate percentage of CCG supertags that become incorrect for the new words when the predicted words were wrong and edited by the user. This shows us how much correct or incorrect contextual information for supertags is returned back to the decoder for the prediction of the remaining hypothesis. We also produced such statistics for the second experimental setup, *On the fly CCG supertagger*. In Table V, we show the percentage of CCG supertags those were incorrectly assigned to new words on both the experimental setups. We clearly see from the table that the second setup (*On the fly CCG supertagger*) is far better than the first setup (*CCG supertags by Transformer*) in terms of assigning correct CCG supertags to the new words that the user has just corrected, i.e. better by 17.12% and 21.28% for the French-to-English and Hindi-to-English translation tasks, respectively.

E. Incorrect predictions versus time-steps

We carry out another analysis to see how the MT systems' prediction accuracy varies over the time of a translation. In Section VI-C, we plot the distributions of the WPA and WSR scores over the sentence-length based test sets. As above, we also consider the sentence-length based test sets for this analysis. This time, we detect the number of incorrect predictions by the INMT systems over the translation and plot those numbers over the time steps.

Figure 4 presents eight graphs, and the top four and bottom four graphs represent the French-to-English and Hindi-to-English translation tasks, respectively. Four graphs represent four sentence-length based test sets which we defined above in Section VI-C. The x-axis of the graphs represents time steps, i.e. word positions in translation. The y-axis of the graphs represents the average number of incorrectly predicted words. For comparison we plot curves for the baseline and syntax-informed INMT systems considering two setups (*CCG supertags by Transformer* and *on the fly CCG supertagger*; cf. Section V). We also show curves for the ideal setup, i.e. when we feed Transformer with the ground-truth CCG supertags instead of the supertags predicted by Transformer or generated by *on the fly CCG supertagger*. These graphs show a clear picture in terms of interactive predictions by the baseline and syntax-informed INMT systems. We see the all curves go downward over time (i.e. increasing positions of translation) for both baseline and syntax-informed INMT systems. We also see from the figure that in most cases supertags play an important role for predicting correct tokens, especially in latter stages of translation. In other words, integration of supertags into the interactive-predictive platform has positively impacted human effort in translation. When we see the graphs for the sets of longer sentences, we see that the supertag features have even more impact on predicting correct tokens in translation. As above, we clearly see the *on the fly CCG supertagger* setup is again more productive than the *CCG supertags by Transformer* setup most of the cases.

VII. Conclusion

In this paper, we integrated a rich and complex syntactic feature (supertags) into the current state-of-the-art neural MT model, Transformer. Furthermore, we test whether the integration of such knowledge sources into Transformer could indeed reduce human effort in translation in an interactive-predictive scenario. We carried out our experiments with French-to-English, a high resource language pair, and Hindi-to-English, a low-resource and less-explored language pair, and present our results with a comparative error analysis.

From our evaluation results we found that our syntax-aware Transformer models outperform the baseline transformer models with small gains in terms of BLEU, and the gains are not statistically significant. This finding contradicts to the findings of Nadejde et al. [7] who found supertags effective in significantly improving their RNN MT models.

We compared our syntax-informed and baseline Transformer models on an interactive-predictive setup. We integrated supertags into Transformer in two different ways, and both setups were found to be effective in reducing human efforts in translation. Most importantly, although our best-performing syntax-informed and the baseline Transformer models are comparable in terms of BLEU in both the French-to-English and Hindi-to-English translation tasks, we found that the best-performing syntax-informed interactive NMT framework significantly reduces human efforts in translation in the French–to–English and Hindi–to–English translation tasks, achieving 2.65 and 6.55 point absolutes corresponding to 5.65% and 19.1% relative improvements, respectively, in terms of WPA over the respective baselines.

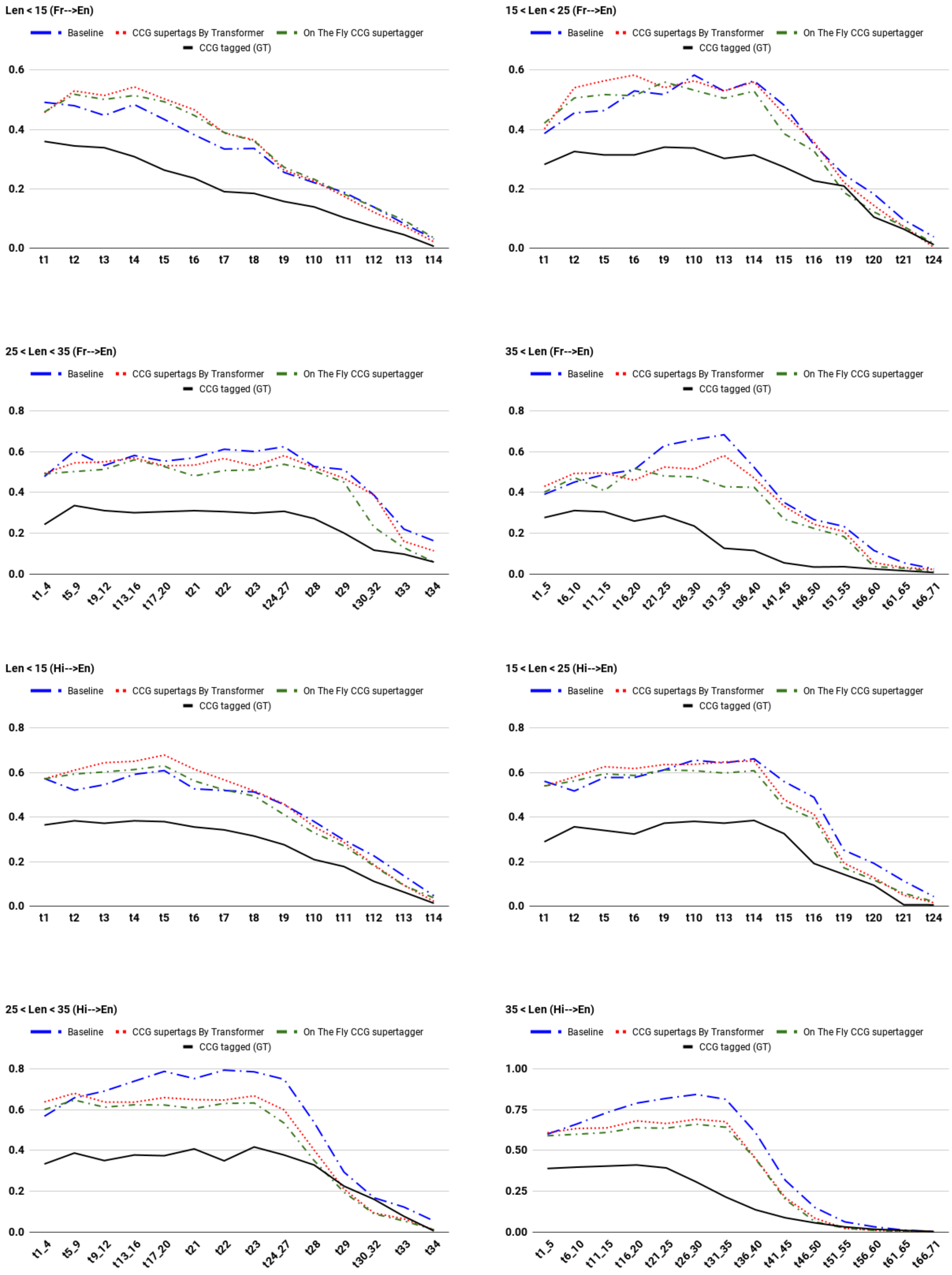
We carried out an extensive error analysis with a variety of criteria. Our analysis unraveled many sides of our syntax-aware models in an interactive-predictive environment. For an example, we particularly found that our syntax-informed interactive-predictive models have positively impacted more for the translation of the longer sentences.

Given the importance of interactive MT in translation industry, the findings of this work can be crucial for their production as our methods can positively impact their productivity gain in translation.

In future, we plan to evaluate our interactive MT systems on a real translation project with human translators. We also plan to integrate language-independent contextual knowledge into the interactive-predictive NMT systems.

Acknowledgement

Authors gratefully acknowledge the generous grant of TDIL, MeiTY, Govt. of India for the project ”Hindi to English Machine Translation for Judicial Domain [11(3)/2015-HCC(TDIL]” to carry out this research. Asif Ekbal acknowledges Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly MediaLab Asia). Rejwanul Haque and Andy Way acknowledge the ADAPT Centre for Digital Content Technology, funded under the Science Foundation Ireland (SFI) Research Centres Programme (Grant No. 13/RC/2106) and is co-funded under the European Regional Development Fund.



Len < 15 (Fr-->En)

15 < Len < 25 (Fr-->En)

25 < Len < 35 (Fr-->En)

35 < Len (Fr-->En)

Len < 15 (Hi-->En)

15 < Len < 25 (Hi-->En)

25 < Len < 35 (Hi-->En)

35 < Len (Hi-->En)

Fig. 4: Average word replacement required at each token position; Len is length of reference sentence

References

[1] R. Knowles and P. Koehn, "Neural interactive translation prediction," in *Proceedings of the Association for Machine Translation in the Americas*, Austin, TX, 2016, pp. 107–120.

[2] Á. Peris, M. Domingo, and F. Casacuberta, "Interactive neural machine translation," *Computer Speech & Language*, vol. 45, pp. 201–220, 2017.

[3] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, 2015.

[4] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation," in *HLT-NAACL 2003: conference combining Human Language Technology conference series and the North American Chapter of the Association for Computational Linguistics conference series*, Edmonton, AB, 2003, pp. 48–54.

[5] Á. Peris and F. Casacuberta, "Active learning for interactive neural machine translation of data streams," in *Proceedings of the 22nd Conference on Computational Natural Language Learning*, Brussels, Belgium, 2018, pp. 151–160.

[6] T. K. Lam, S. Schamoni, and S. Riezler, "Interactive-predictive neural machine translation through reinforcement and imitation," in *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*, Dublin, Ireland, 2019, pp. 96–106.

[7] M. Nădejde, S. Reddy, R. Sennrich, T. Dwojak, M. Junczys-Dowmunt, P. Koehn, and A. Birch, "Predicting target language CCG supertags improves neural machine translation," in *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark, 2017, pp. 68–79.

[8] M. Steedman, "The syntactic process," *MIT Press*, vol. 24, 2000.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[10] G. Foster, P. Isabelle, and P. Plamondon, "Target-text mediated interactive machine translation," *Machine Translation*, vol. 12, no. 1-2, pp. 175–194, 1997.

[11] D. Ortiz-Martínez, "Online learning for statistical machine translation," *Computational Linguistics*, vol. 42, no. 1, pp. 121–161, 2016.

[12] J. González-Rubio, D. Ortiz-Martínez, and F. Casacuberta, "Active learning for interactive machine translation," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France, 2012, pp. 245–254.

[13] S. Green, J. Chuang, J. Heer, and C. D. Manning, "Predictive translation memory: A mixed-initiative system for human language translation," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 2014, pp. 177–187.

[14] S. Barrachina, O. Bender, F. Casacuberta, J. Civera, E. Cubel, S. Khadivi, A. Lagarda, H. Ney, J. Tomás, E. Vidal *et al.*, "Statistical approaches to computer-assisted translation," *Computational Linguistics*, vol. 35, no. 1, pp. 3–28, 2009.

[15] P. Koehn, C. Tsoukala, and H. Saint-Amand, "Refinements to interactive translation prediction based on search graphs," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimore, MD, 2014, pp. 574–578.

[16] G. Sanchis-Trilles, D. Ortiz-Martínez, and F. Casacuberta, "Efficient wordgraph pruning for interactive translation prediction," in *Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT)*, Prague, Czech Republic, 2014, pp. 27–34.

[17] P. Koehn, "A process study of computer-aided translation," *Machine Translation*, vol. 23, no. 4, pp. 241–263, 2009.

[18] V. Alabau, A. Sanchis, and F. Casacuberta, "Improving on-line handwritten recognition in interactive machine translation," *Pattern Recognition*, vol. 47, no. 3, pp. 1217–1228, 2014.

[19] J. Wuebker, S. Green, J. DeNero, S. Hasan, and M.-T. Luong, "Models and inference for prefix-constrained machine translation," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, 2016, pp. 66–75.

[20] H. Hassan, K. Sima'an, and A. Way, "Supertagged phrase-based statistical machine translation." Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 288–295.

[21] Y. Marton and P. Resnik, "Soft syntactic constraints for hierarchical phrased-based translation," in *Proceedings of ACL-08: HLT*, Columbus, Ohio, 2008, pp. 1003–1011.

[22] R. Haque, S. Kumar Naskar, A. Van Den Bosch, and A. Way, "Supertags as source language context in hierarchical phrase-based smt," in *Proceedings of AMTA 2010: The Ninth Conference of the Association for Machine Translation in the Americas*, Denver, CO., 2010, pp. 210–219.

[23] R. Haque, S. K. Naskar, A. van den Bosch, and A. Way, "Integrating source-language context into phrase-based statistical machine translation," *Machine Translation*, vol. 25, no. 3, pp. 239–285, 2011.

[24] N. Akoury, K. Krishna, and M. Iyyer, "Syntactically supervised transformers for faster neural machine translation," in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Volume 1: Long Papers*, Florence, Italy, 2019, pp. 1269–1281.

[25] A. Eriguchi, Y. Tsuruoka, and K. Cho, "Learning to parse and translate improves neural machine translation," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Volume 2: Short Papers*, Vancouver, BC, 2017, pp. 72–78.

[26] R. Aharoni and Y. Goldberg, "Towards string-to-tree neural machine translation," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Volume 2: Short Papers*, Vancouver, BC, 2017, pp. 132–140.

[27] S. Bangalore and A. K. Joshi, "Supertagging: An approach to almost parsing." *Computational Linguistics*, vol. 25, no. 2, pp. 237–265, 1999.

[28] J. Chen, S. Bangalore, and K. Vijay-Shanker, "Automated extraction of tree-adjoining grammars from treebanks," *Natural Language Engineering*, vol. 12, no. 3, pp. 251–299, 2006.

[29] J. Hockenmaier, "Data and models for statistical parsing with combinatory categorial grammar," 2003.

[30] S. Clark and J. R. Curran, "The importance of supertagging for wide-coverage CCG parsing," in *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland, 2004, pp. 282–288.

[31] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, August 2016, pp. 1715–1725.

[32] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: a method for automatic evaluation of machine translation." in *ACL-2002: 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA: ACL, 2002, pp. 311–318.

[33] M. Ziemski, M. Junczys-Dowmunt, and B. Pouliquen, "The united nations parallel corpus v1. 0," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, 2016, pp. 3530–3534.

[34] A. Kunchukuttan, P. Mehta, and P. Bhattacharyya, "The IIT Bombay English-Hindi parallel corpus," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018.

[35] J. Tiedemann, "Parallel data, tools and interfaces in OPUS," in *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*, Istanbul, Turkey, 2012, pp. 2214–2218.

[36] F. Hieber, T. Domhan, M. Denkowski, D. Vilar, A. Sokolov, A. Clifton, and M. Post, "The sockeye neural machine translation toolkit at AMTA 2018," in *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, Boston, MA, 2018, pp. 200–207.

[37] P. Gage, "A new algorithm for data compression," *C Users Journal*, vol. 12, no. 2, pp. 23–38, 1994.

[38] M. Lewis and M. Steedman, "A* ccg parsing with a supertag-factored model," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 990–1000.

[39] P. Koehn, "Statistical significance tests for machine translation evaluation," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, D. Lin and D. Wu, Eds., Barcelona, Spain, July 2004, pp. 388–395.

[40] A. Yeh, "More accurate tests for the statistical significance of result differences," in *Proceedings of the 18th conference on Computational linguistics - Volume 2, COLING 2000*, Saarbrücken, Germany, 2000, pp. 947–953.