Syntax-Driven Machine Translation as a Model of ESL Revision

Huichao Xue, Rebecca Hwa
Department of Computer Science, University of Pittsburgh

Introduction

Motivation

When learning a second language, English as a Second Language(ESL) students make mistakes along the way.

The problem here is that they come to the US like illegal. → (But he(she) actually means) → The problem here is that they come to the US illegally.

Can we help them?

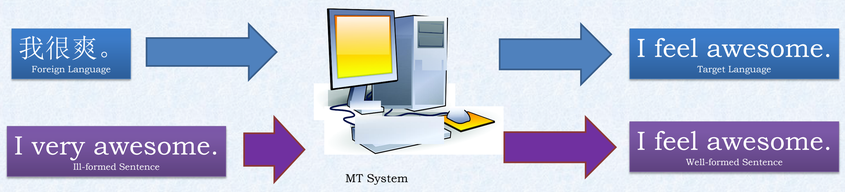
Our ultimate goal is to build an automatic error correction system.

Background -- Two Research Directions

I. Correcting specific types of errors
1. Determiners (Yi et al., 2008; Gamon et al., 2008)
2. Prepositions (Chodorow et al., 2007; Gamon et al., 2008; Hermet et al., 2008)
3. Mass versus Count nouns (Nagata et al., 2006; Brockett et al. (2006).)

II. Approaches that deal with more general errors
They try to correct all kinds of errors at once.

Machine Translation (MT) is one way to achieve this purpose.



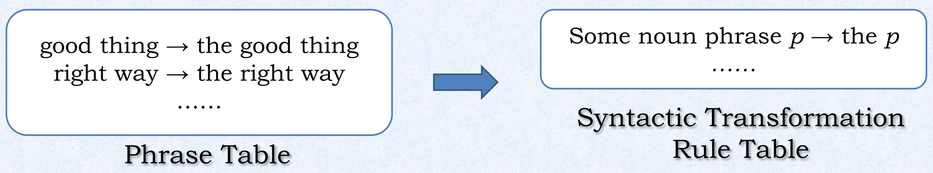
Rather than transforming between languages, we want MT systems to transform from ill-formed sentences to well-formed sentences.

MT has also been used in correcting Specific types of errors in Brockett et al. (2006).

Our Approach -- Syntax-Driven MT in error correction

Why Syntax-Driven (as compared to Phrase-Based)?

1. It can be less data-intensive because the mapping is formed at a structural level rather than the surface word level.



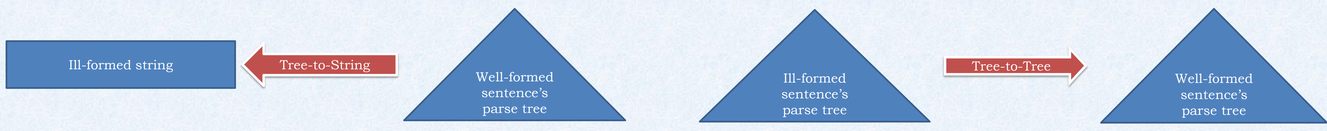
2. Syntactic transformations provide an intuitive description of error corrections, especially for syntactic errors.

Our Research Question

The question we ask: **What kind of Syntax-Driven MT model is better suited in Error Correction?**

Our scenario is different from MT in that *sometimes it is not possible to assign a sensible syntactic structure to an ungrammatical sentence.*

We focus on two particular Syntax-Driven MT formalisms that reflects different perspectives on the role of syntax, and compare.

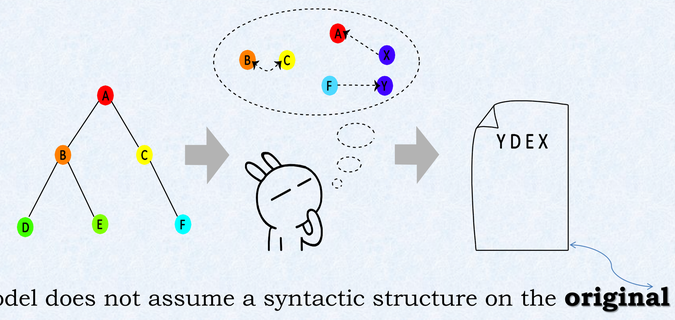


The Corpus

Students are often asked to review their previously written essays, we use that data as the resource.

Syntax-Driven Model 1 –Tree-To-String Model

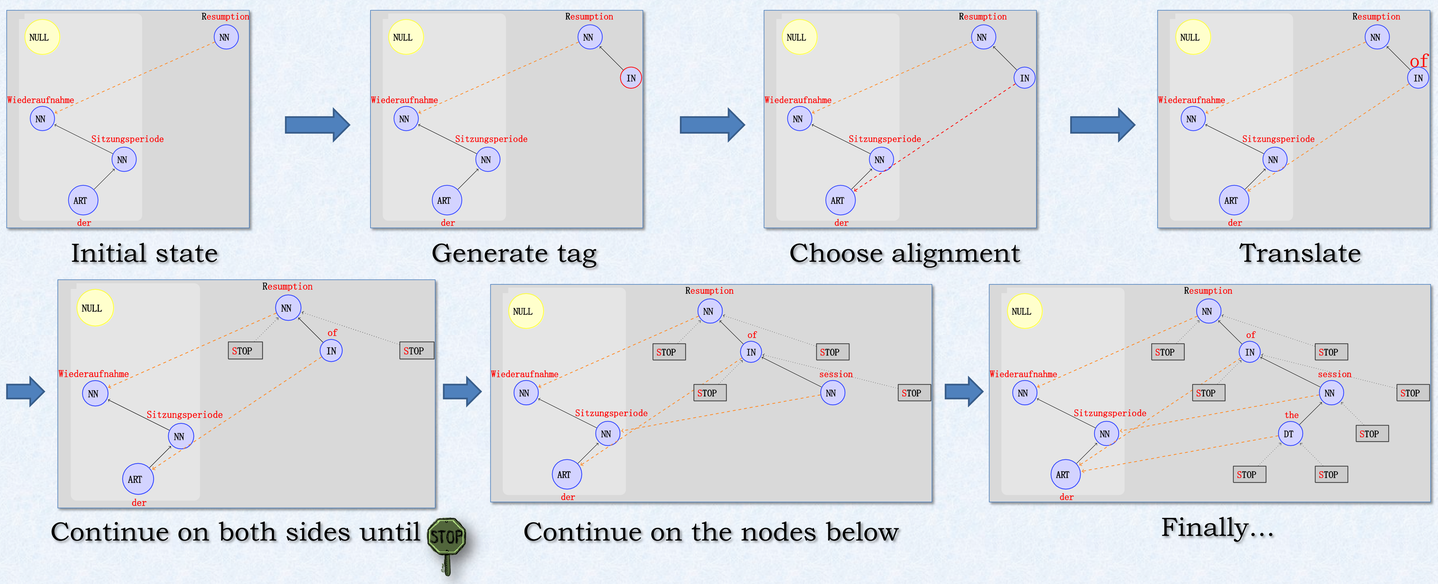
We pick the Yamada & Knight Tree-To-String Model. It assumes that the observed source sentence is the result of transformation performed on the parse tree of the intended target sentence due to a noisy communication channel.



This Tree-To-String model does not assume a syntactic structure on the **original draft**.

Syntax-Driven Model 2 –Tree-To-Tree Model

We use Quasi-Synchronous Grammar (Smith and Eisner, 2006), which is a generative model that aims to produce the most likely target tree for a given source tree.



Experiments

-- We compare the two models above.

Model Comparison

We use Cross Entropy as a metric that measures the distance between the learned probabilistic model and the real data.

$$-\frac{1}{|C|}\sum_{(O,R)\in C}\log \Pr_{M}(O,R) \Rightarrow \begin{array}{l}\Pr_{YK}(O,R)\approx \Pr_{YK}(O\mid \hat{\tau}_R)\Pr(\hat{\tau}_R)\\ \Pr_{QG}(O,R)\approx \Pr_{QG}(\hat{\tau}_R\mid \hat{\tau}_O)\Pr(\hat{\tau}_O)\end{array}$$

Here we use Viterbi approximations to make the computation tractable.

Corpus

•Ideally, we would like to analyze a parallel corpus containing ill-formed sentences and their revised versions.
•However, such "Gold standard data" is hard to obtain.
•We choose an alternative: the essay revisions made by ESL speakers.
•This "less than perfect" data still offers us an opportunity to model how ESL speakers make mistakes.

Dataset

We randomly split the dataset into a training corpus of 4566 sentence pairs and a test corpus of 100 pairs.

Experiment 1

How well can the models describe the ESL revision domain?
•How are the two models' coverage relate?
•What about cross entropy?
•Both models have difficulties interpreting some transformations. Are these pairs harder? We show that the edit distances and sentence lengths are **higher with 90% confidence**.
•Sometimes models have trouble with **sentence pairs that require no change**.

Experiment 1

	Neither	$D_{QG}\cap D_{YK}$	$D_{QG}-D_{YK}$	$D_{YK}-D_{QG}$
Number of instances	38	33	26	3
Average edit distance	2.42	1.88	2.08	1
Average *O* Length	14.63	12.36	12.58	6.67
Average *R* Length	13.87	12.06	12.62	6.67
QG cross entropy	N/A	127.95	138.9	N/A
YK cross entropy	N/A	78.76	N/A	43.84
% of identical pairs	53%	48%	58%	67%

While YK model has a more limited coverage, it models those transformations with a greater certainty.

Experiment 2

We also expect better models to fit worse to bad sentence pairs. This means a drop of coverage or a raise of cross entropy.

We construct such negative scenarios by reversing the original testing corpus.

Coverage drops: QG from 59 to 49 pairs, YK from 36 to 20.
Non-identical pairs: 16/49 in QG, 1/20 in YK.

Experiment 2

	Neither	$D_{QG}\cap D_{YK}$	$D_{QG}-D_{YK}$	$D_{YK}-D_{QG}$
Number of instances	50	19	30	1
Average edit distance	2.88	0.05	2.17	1
Average *O* Length	14.18	9.00	12.53	17
Average *R* Length	14.98	9.05	12.47	16
QG cross entropy	N/A	81.85	139.36	N/A
YK cross entropy	N/A	51.2	N/A	103.75
% of identical pairs	40%	95%	50%	0%

The trained YK model is a worse fit to the negative corpus, compared to QG.

Conclusions

This work investigates the suitability of syntax-driven MT approaches for modeling the revision writing process of ESL learners.

We have considered a tree-to-string model (YK model), as well as a tree-to-tree model (QG model). We evaluate the trained models by measuring their cross entropies on two corpora: a normal corpus and a negative corpus.

The results suggest that:

•QG has a better coverage of the transformations, while YK has a lower entropy.
•On the negative corpus, YK was more perplexed.
•YK might be a promising approach for automatic grammar error correction.