Lost in Machine Translation: A Method to Reduce Meaning Loss

Reuben Cohn-Gordon | Noah Goodman | Stanford University

2019

The problem

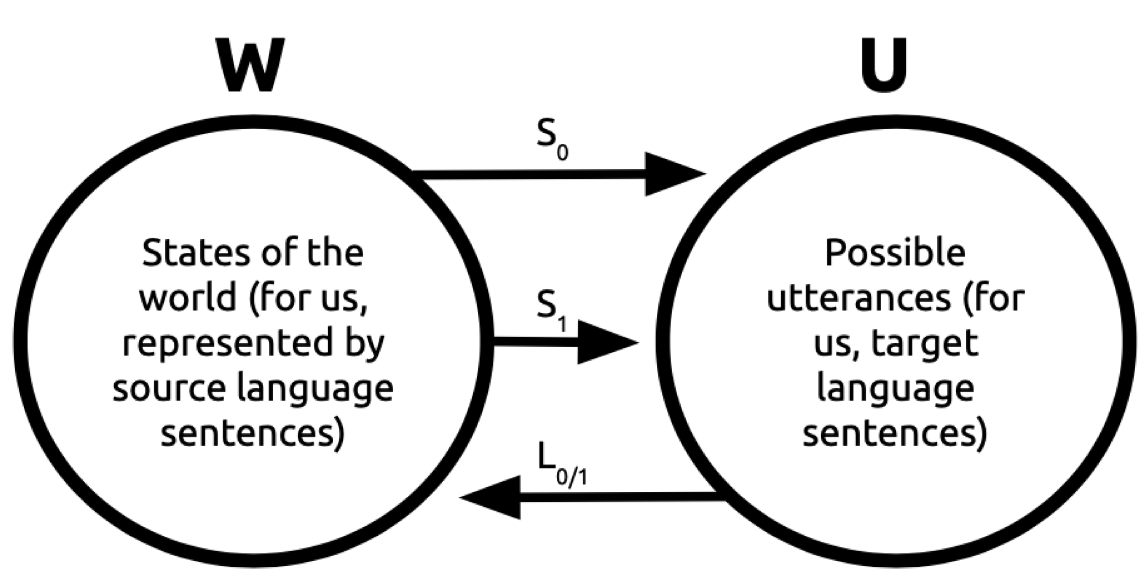
I cut my finger

$\neq$

I cut my finger off

But state-of-the-art systems map both to a single sentence in French: ***Je me suis coupé le doigt.***

Bayesian model of pragmatics



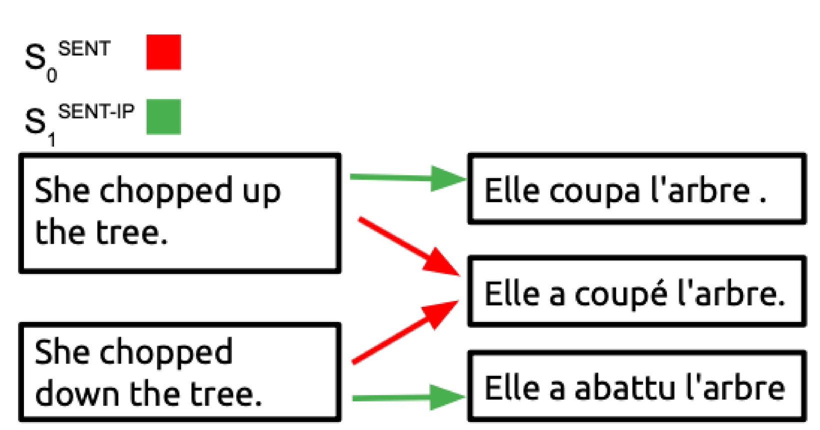
- The *Rational Speech Acts* (RSA) framework [1] models pragmatic inferences.
- E.g. the inference that *Some apples are red* means that not all are (because if all were, speaker would have said *All apples are red*)
- Let W be a set of states , and U be a set of possible utterances
- Given a state w, S_1 prefers utterances u which are good for S_0 but also communicate w to L_0

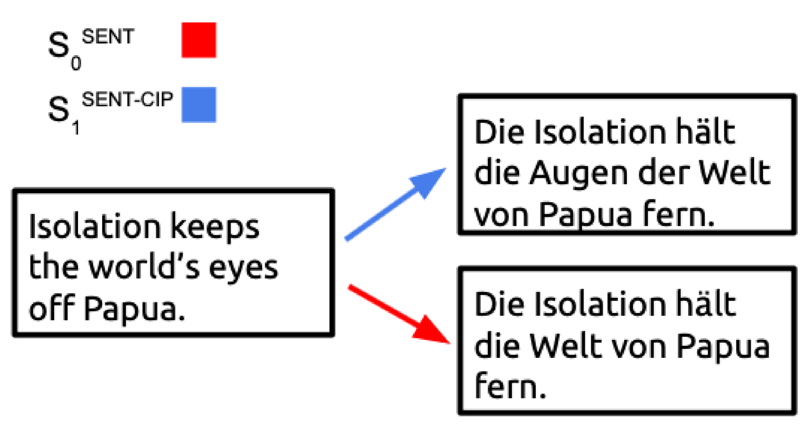
$S_0(u|w) \propto [\![u]\!](w) \cdot P(u)$ (literal speaker)
$L_0(w|u) \propto [\![u]\!](w) \cdot P(w)$ (literal listener)
$S_1(u|w) \propto S_0(u|w) \cdot L_0(w|u)$ (**informative speaker**)

Informative Translation

- RSA in the domain of translation: a source language sentences corresponds to a world state w. A target language sentence corresponds to a possible message u, which a translation model decodes from w.
- **Intuition: $\mathbf{S_1}$ tries to maximize the one-to-one nature of the mapping.**
- For a set of target language sentences W and some $w \in W$, the S_1 utility says: pick the best translation for w according to S_0 which also allows L_0 to best guess the original sentence w.
- One version ($\mathrm{S}_1^{\text{SNT-IP}}$) for explicitly selected W, one ($\mathrm{S}_1^{\text{SNT-CIP}}$) for unbounded W (all sequences of words).

Examples





Evaluation

- **Eval 1**: Translate to target language with model. Translate back (with separate system). Do you get back what you started with? (distance measured in BLEU)
- **Eval 2**: On an aligned corpus, measure translation quality of S_1 vs. S_0 by BLEU score.

Model	Cycle	Translate
S_0^{SNT}	43.35	37.42
$S_1^{\text{SNT-CIP}}$	**47.34**	**38.29**

Figure: Scores for the non-pragmatic and pragmatic models, on 750 English-German WMT pairs.

Model and Inference

- A trained neural model $S_0^{WD}(wd|w,c)$ is a distribution over the next word given a source sentence and a partial translation. Likewise L_0^{WD}, but from target language to source.
- We use pretrained neural transformer models [2]
- Because U and W are infinite, we need to approximate S_1. We extend the approach of [3], with a model $S_1^{\text{SNT-CIP}}$, in terms of $S_1^{\text{WD-C}}$:

$$S_1^{\text{WD-C}}(wd|w,c) \propto S_0^{\text{WD}}(wd|w,c) \cdot \Sigma_k(L_0^{\text{SNT}}(w|c+wd+k) \cdot S_0^{\text{SNT}}(k|w,c+wd)) \quad (1)$$

$$S_1^{\text{SNT-CIP}}(u|w,c) = \Pi_t S_1^{\text{WD-C}}(u[t]|w,c+u[:t])$$

Conclusions

- Meaning distinctions in the source language should be preserved in the target language.
- An explicit utility function for informativity (as in S_1) is a simple solution to meaning loss in translation, which improves quality generally

References

[1] Michael C. Frank and Noah D. Goodman. "Predicting Pragmatic Reasoning in Language Games". In: *Science* 336.6084 (2012), p. 998.

[2] Myle Ott et al. "fairseq: A Fast, Extensible Toolkit for Sequence Modeling". In: *arXiv preprint arXiv:1904.01038* (2019).

[3] Ramakrishna Vedantam et al. "Context-aware captions from context-agnostic supervision". In: *Computer Vision and Pattern Recognition (CVPR)*. Vol. 3. 2017.