

Exploring Grammatical Error Correction with Not-So- Crummy Machine[†] Translation

Nitin Madnani
Joel Tetreault
Martin Chodorow



[†]*Good Applications for Crummy Machine Translation.*
Ken Church & Ed Hovy. *Mach. Trans.* 8(4). 1993

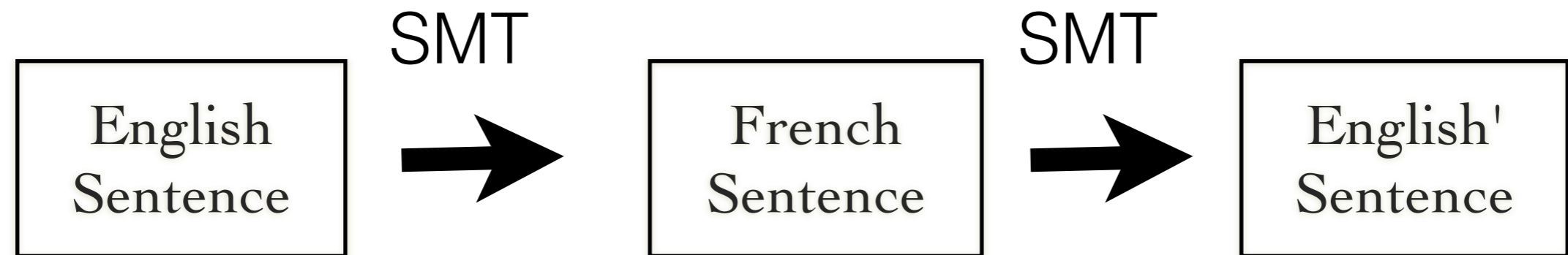
Whole Sentence Error Correction

- Most work in correction targets specific error types such as prepositions and determiners
- Large variety of grammatical errors in L2 writing
- "*Less holistic, error-type specific approaches are unable to disentangle compounded errors of style and grammar*" (Anonymous Reviewer)
- Can we use statistical machine translation (SMT) to do whole sentence error correction *without requiring error detection*?
 - Useful for feedback and content scoring

Related Work

- Brockett et al. (2006) apply the SMT framework to an engineered *parallel* corpus to correct mass noun errors.
- Park and Levy (2011) attempt whole sentence correction using a noisy channel model but do not use any SMT.
- Connections to work in MT system combination and paraphrase generation (more in paper)

Round-trip Machine Translation (RTMT)



- Leverage a Statistical MT (SMT) system's large phrase tables to generate alternative renderings of original sentence
- Each rendering should retain the meaning of the original sentence – to some (crummy) extent

RTMT for Preposition Errors

- Hermet et al. (2009) used RTMT to fix preposition errors

E: Most people don't want to go study to college

F: La plupart des gens ne veulent pas aller étudier au collège

E': Most people don't want to go study at college

- Performance on 150 sentences with pre-marked errors

System	Performance
RTMT	66.4%
Classifier	68.7%
Hybrid	81.2%

Using first and second language models to correct preposition errors in second language authoring.
Matthieu Hermet & Alain Désilets. Proceedings of BEA. 2009

Our RTMT Methodology

- Focus on all the errors in the sentence without performing detection first
- Use *multiple* pivot languages to generate several round-trip translations (via the Google Translate API)
- Use alignment algorithm to combine round-trip translations into a lattice
- Explore several different methods to generate full-sentence corrections from the lattice

Multiple Round-trip Translations

Original	Both experience and books are very important <u>about living</u> .
Swedish	Both experience and books are very important in live.
Italian	Both books are very important experience and life.
Russian	And the experience, and a very important book about life.
French	Both experience and the books are very important in life.
German	Both experience and books are very important about life.
Chinese	Related to life experiences and the books are very important.
Spanish	Both experience and the books are very important about life.
Arabic	Both experience and books are very important for life.

Multiple Round-trip Translations

Original	Both experience and books are very important <u>about living</u> .
Swedish	Both experience and books are very important in live.
Italian	Both books are very important experience and life.
Russian	And the experience, and a very important book about life.
French	Both experience and the books are very important in life.
German	Both experience and books are very important about life.
Chinese	Related to life experiences and the books are very important.
Spanish	Both experience and the books are very important about life.
Arabic	Both experience and books are very important for life.

Why not use an LM to pick the most "fluent"?

Multiple Round-trip Translations

Original	Both experience and books are very important <u>about living</u> .
Swedish	Both experience and books are very important in live.
Italian	Both books are very important experience and life.
Russian	And the experience, and a very important book about life.
French	Both experience and the books are very important in life.
German	Both experience and books are very important about life.
Chinese	Related to life experiences and the books are very important.
Spanish	Both experience and the books are very important about life.
Arabic	Both experience and books are very important for life.

Why not use an LM to pick the most "fluent"?

Combining Round-trip Translations

- No single round-trip translation always fixes every error; yet may fix ones that no other translation does
- Better alternative: **combine** translations to produce a better overall correction; Also increases likelihood of meaning preservation
- Develop heuristic alignment algorithm using the TERp MT metric (Snover et al., 2009)
 - TERp computes minimal number of edit operations (insertion, deletion, substitution, synonymy etc.) needed to turn one sentence into another. Generates alignment as by-product.

Fluency, Adequacy, or HTER? Exploring different human judgments with a tunable MT metric.
Matthew Snover, Nitin Madnani, Richard Schwartz & Bonnie Dorr. Proceedings of WMT. 2009.

TERp alignment

both experience and books are very important about living

and the experience , and a very important book about life

TERp alignment

both experience and books are very important about living

and the experience , and a very important book about life

TERp alignment

both experience and books are very important about living

and the experience , and book a very important about life

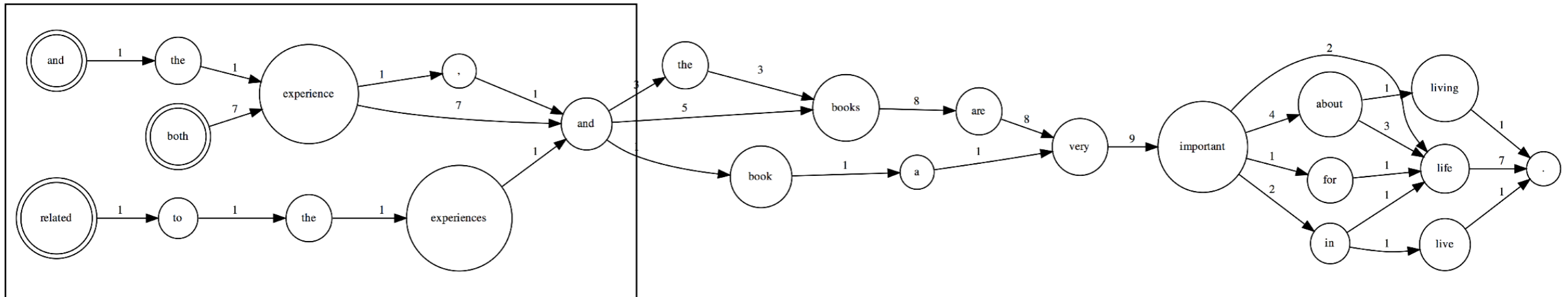
Shift before computing alignment, if needed

TERp alignment

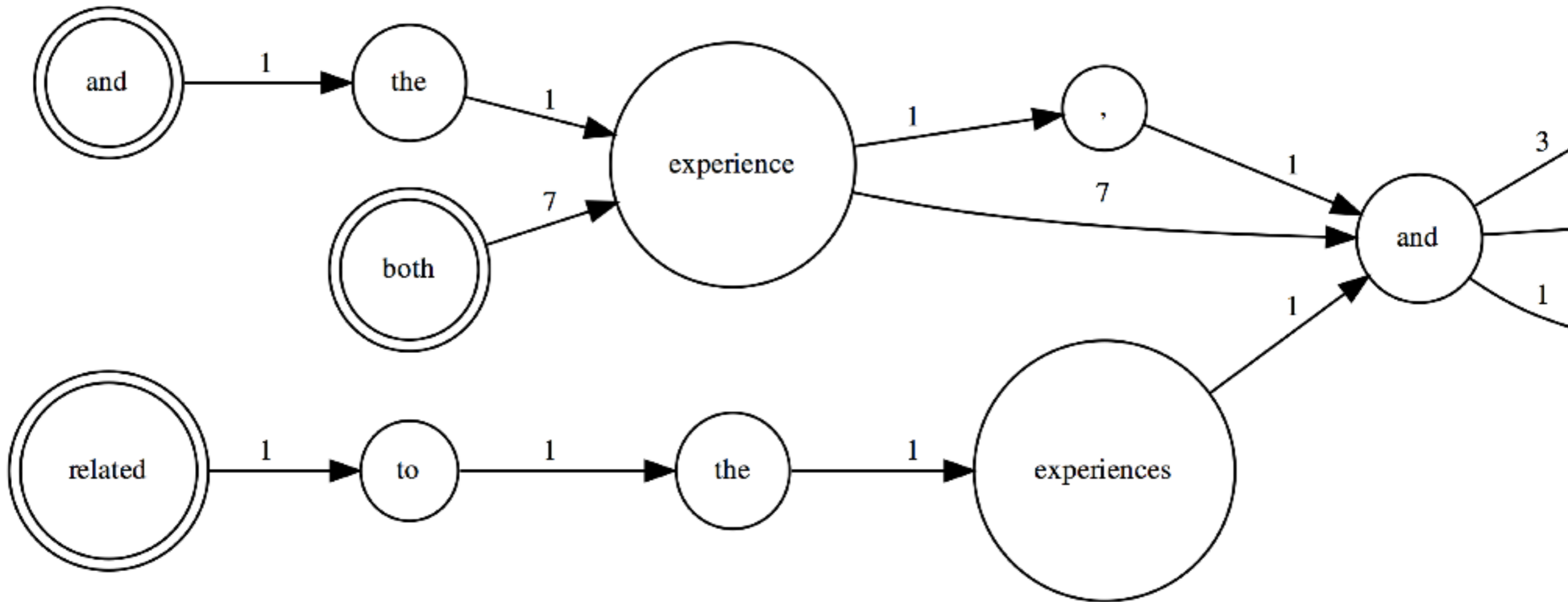
both experience		and books are very important about living						
S	M	M	T	S	M	M	M	Y
and	the experience	,	and book	a	very important	about	life	

(I)nsertion, (S)ubstitution, (M)atch, s(T)emming, s(Y)nonymy

Example TERp Lattice



Example TERp Lattice



Generating Corrections

- **Baseline LM (B).** The round-trip translation with the lowest perplexity as measured by a 5-gram Gigaword LM
- **Greedy (G).** Path extracted from lattice using greedy best-first strategy at each node
- **1-Best (1).** Shortest path extracted using OpenFST toolkit
- **LM Re-ranked (L).** Top item from 20-best list extracted using OpenFST and reranked with 5-gram LM
- **Product Re-ranked (P).** Same as L except reranked using product of LM score and path cost
- **Full LM composition (C).** Lattice composed with trigram finite state language model and shortest path extracted

Evaluation

- 200 manually selected sentences from TOEFL essays
 - 162 sentences with at least one error; 38 perfectly grammatical
 - Sentences chosen on the basis of length, and number & types of errors

# of Errors	Sentences	Avg. Length
1	61	14.4
2	45	19.9
3	29	24.2
4	14	29.4
> 4	13	38.0

Evaluation

- For each sentence, rater shown original sentence alongside one of the 6 corrections and asked to rate:
 - **Grammaticality**: Correction > Original, Correction = Original, Correction < Original
 - **Meaning**: Correction preserves meaning, Correction changes meaning
- Correction successful *iff* $C > O$ and meaning preserved

	C > O	C = O	C < O
Meaning Preserved	Success	Draw	Failure
Meaning Changed	Failure	Failure	Failure

Results

162 sentences with grammatical errors

Method	Success	Draw	Failure
Baseline LM (B)	21%	9%	70%
Greedy (G)	36%	33%	31%
1-Best (1)	32%	30%	38%
LM Re-ranked (L)	30%	17%	54%
Product Re-ranked (P)	23%	38%	40%
LM Composition (C)	19%	12%	69%
<i>Oracle</i>	<i>60%</i>	<i>40%</i>	-

Results

38 sentences with NO grammatical errors

Method	Success	Draw	Failure
Baseline LM (B)			
Greedy (G)			
1-Best (1)			
LM Re-ranked (L)			
Product Re-ranked (P)			
LM Composition (C)			
<i>Oracle</i>			

Results

38 sentences with NO grammatical errors

Method	Success	Draw	Failure
Baseline LM (B)		-	
Greedy (G)		-	
1-Best (1)		-	
LM Re-ranked (L)		-	
Product Re-ranked (P)		-	
LM Composition (C)		-	
<i>Oracle</i>	-	-	-

Results

38 sentences with NO grammatical errors

Method	Success	Draw	Failure
Baseline LM (B)	26%	-	74%
Greedy (G)	79%	-	21%
1-Best (1)	61%	-	39%
LM Re-ranked (L)	34%	-	66%
Product Re-ranked (P)	42%	-	58%
LM Composition (C)	29%	-	71%
<i>Oracle</i>	-	-	-

Example Successes (Greedy)

That's why I like to make travel by using my own car.

That's why I like to travel using my own car.

Other important thing can be the way you make your work, because working hard is not enough.

Another important thing may be the way you do your work, because working hard is not enough.

Facing on this problem, hard work and careful calculation maybe be useless.

Faced with this problem, hard work and careful calculation may be useless.

Example Draws (Greedy)

I think science and mathematics will need to future.

I think that science and mathematics will need to the future.

Chinese people use a different written words.

The Chinese use a different written words.

Book is not the only way to learn something.

The book is not the only way to learn something.

Example Failures (Greedy)

And simply there is fantastic for everyone.

All magic and simply there is fantastic for all.

I hope that share a room with she can be certainly kindle, because she is likely me and so will not be a problem with she.

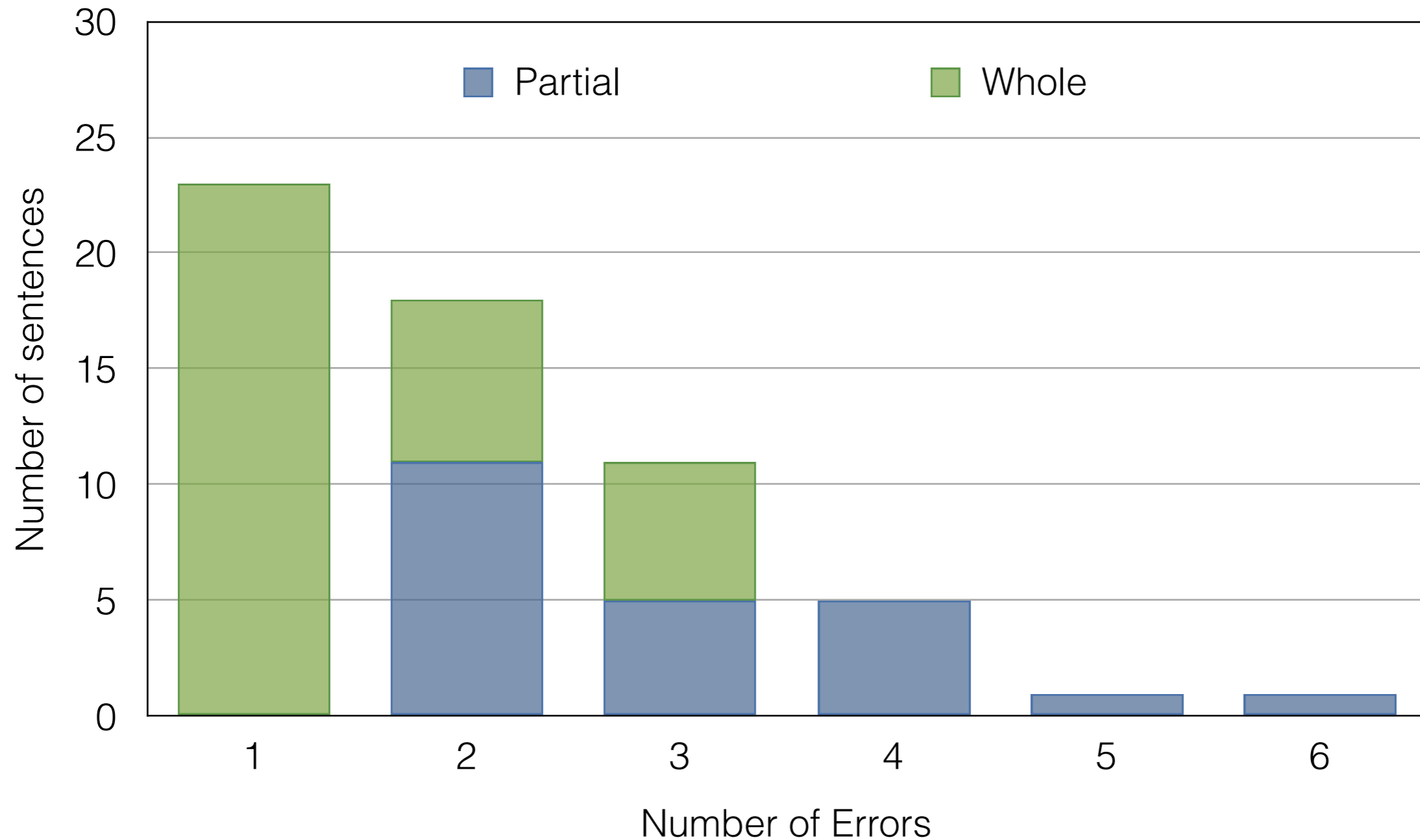
I hope that sharing a room with her can be certainly kindle, because it is likely that I and so there will be no problems with it.

We can go to the shopping place, have a good food, meet some new friend, watch a movie.

We can go to the shopping place of, have a good food, meet some new friends, watch a movie.

Whole Sentence Corrections

Greedy Successes (59)



Conclusions

- Combining several round-trip translations has potential for whole sentence correction
- Most failures due to new errors from Google Translate
- Reduce SMT noise by identifying errors first and then focusing translation process only on errorful spans?
- Exploit n -best round-trip translations instead of 1-best?
- Use crowd-sourcing for quicker annotation and evaluation?

Thanks!

Original	I will take your questions now.
Swedish	I will take your questions now.
Italian	I will take your questions now.
Russian	I will take your questions now.
French	I will take your questions now.
German	I will take your questions now.
Chinese	Now I am ready to answer your questions.
Spanish	I will answer your questions now.
Arabic	I take your questions now.