MT – Final thoughts

David Kauchak
CS159 – Spring 2023

Some slides adapted from

Philipp Koehn
School of Informatics
University of Edinburgh

Kevin Knight
USC/Information Sciences Institute
USC/Computer Science Department

Dan Klein
Computer Science Department
UC Berkeley

1

Admin

Assignment 6

2

Language translation

https://www.youtube.com/watch?v=Q6jzl_Oy2IQ
https://www.youtube.com/watch?v=vV1SkTdizZI

3

Word-alignment Evaluation

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

How good of an alignment is this?
How can we quantify this?

4

Word-alignment Evaluation

System:

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

Human

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

How can we quantify this?

5

Word-alignment Evaluation

System:

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

Human

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

Precision and recall!

6

Word-alignment Evaluation

System:

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

Human

The old man is happy. He has fished many times.

El viejo está feliz porque ha pescado muchos veces.

Precision: $\frac{6}{7}$ Recall: $\frac{6}{10}$

7

What kind of Translation Model?

Mary did not slap the green witch

Word-level models

Phrasal models

Syntactic models

Semantic models

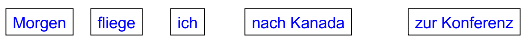
Maria no dió una botefada a la bruja verde

8

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$



1. Sentence is divided into phrases

9

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$

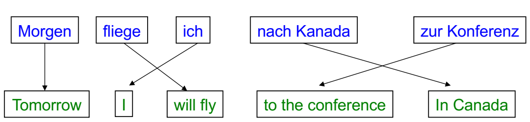
1. Sentence is divided into phrases
2. Phrases are translated (avoids a lot of weirdness from word-level model)

10

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$



1. Sentence is divided into phrases
2. Phrase are translated (avoids a lot of weirdness from word-level model)
3. Phrases are reordered

11

Phrase table

natuerlich

Translation	Probability
of course	0.5
naturally	0.3
of course ,	0.15
, of course ,	0.05

12

Phrase table

den Vorschlag

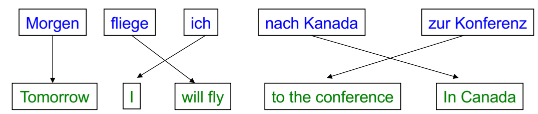
Translation	Probability
the proposal	0.6227
's proposal	0.1068
a proposal	0.0341
the idea	0.0250
this proposal	0.0227
proposal	0.0205
of the proposal	0.0159
the proposals	0.0159
the suggestions	0.0114
...	

13

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$



Advantages?

14

Advantages of Phrase-Based

Many-to-many mappings can handle non-compositional phrases

Easy to understand

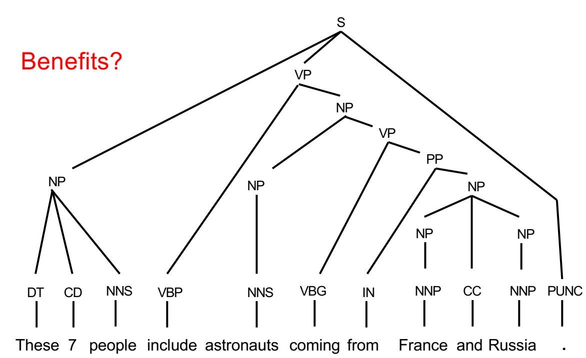
Local context is very useful for disambiguating
- "Interest rate" → …
- "Interest in" → …

The more data, the longer the learned phrases
- Sometimes whole sentences!

15

Syntax-based models

Benefits?



16

Syntax-based models

Benefits

- Can use syntax to motivate word/phrase movement
- Could ensure grammaticality

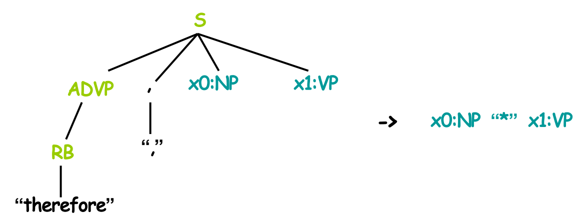
Two main types:

- p(foreign *string* | English parse tree)
- p(foreign *parse tree* | English parse tree)

Why always English parse tree?

17

Tree to string rule



18

Tree to string rules examples

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航，员
9. PUNC(.) → .
10. NP(x0:DT, CD(7), NNS(people) → x0 , 7人
11. VP(VBG(coming), PP(IN(from), x0:NP)) → 来自 ,x0
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1, x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT(“the”), x0:JJ, x1:NN) → x0 , x1

Contiguous phrase pair substitution rules (1–12)

Higher-level rules (13–17)

19

Tree to string rules examples

Both VBP(“include”) and VBP(“includes”) will translate to “中包括” in Chinese.

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航，员
9. PUNC(.) → .
10. NP(x0:DT, CD(7), NNS(people) → x0 , 7人
11. VP(VBG(coming), PP(IN(from), x0:NP)) → 来自 ,x0
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1, x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT(“the”), x0:JJ, x1:NN) → x0 , x1

Contiguous phrase pair substitution rules (6–12)

Higher-level rules (13–17)

20

Tree Transformations

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航，员
9. PUNC(.) → .
10. NP(x0:DT, CD(7), NNS(people) → x0 , 7人
11. VP(VBG(coming), PP(IN(from), x0:NP)) → 来自 ,x0
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1, x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT("the"), x0:JJ, x1:NN) → x0 , x1

The phrase "coming from" translates to "来自" only if followed by an NP (whose translation is then placed to the right of "来自").

ase pair
es
plates)

Higher-level rules

21

Tree Transformations

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航，员
9. PUNC(.) → .
10. NP(x0:DT, CD(7), NNS(people) → x0 , 7人
11. VP(VBG(coming), PP(IN(from), x0:NP)) → 来自 ,x0
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1, x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT("the"), x0:JJ, x1:NN) → x0 , x1

Translate an English NP ("astronauts") modified by a gerund VP ("coming from France and Russia") as follows:
(1) translate the VP,
(2) type the Chinese word "的",
(3) translate the NP.

hrase pair
Substitution rules
(alignment templates)

Higher-level rules

22

Tree Transformations

1. DT(these) → 这
2. VBP(include) → 中包括
3. VBP(includes) → 中包括
4. NNP(France) → 法国
5. CC(and) → 和
6. NNP(Russia) → 俄罗斯
7. IN(of) → 的
8. NP(NNS(astronauts)) → 宇航，员
9. PUNC(.) → .
10. NP(x0:DT, CD(7), NNS(people) → x0 , 7
11. VP(VBG(coming), PP(IN(from), x0:NP)) →
12. IN(from) → 来自
13. NP(x0:NNP, x1:CC, x2:NNP) → x0 , x1 , x2
14. VP(x0:VBP, x1:NP) → x0 , x1
15. S(x0:NP, x1:VP, x2:PUNC) → x0 , x1, x2
16. NP(x0:NP, x1:VP) → x1 , 的 , x0
17. NP(DT("the"), x0:JJ, x1:NN) → x0 , x1

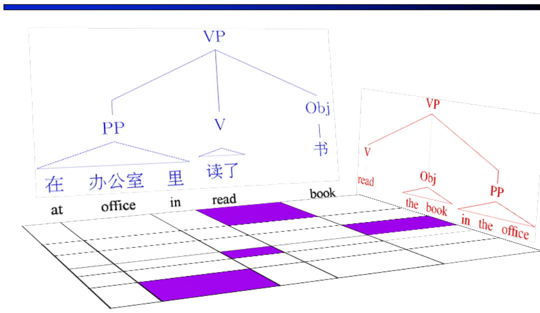
Contiguous phrase pair
Substitution rules
(alignment templates)

To translate "the JJ NN", translate the JJ and NN (and drop "the").

Higher-level rules

23

Tree to tree example



24

MT Evaluation

How do we do it?

What data might be useful?

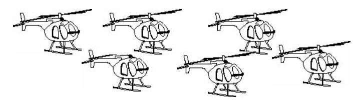
25

MT Evaluation

Source only

Manual:
- SSER (subjective sentence error rate)
- Correct/Incorrect
- Error categorization

Extrinsic:
Objective usage testing



Automatic:
- WER (word error rate)
- BLEU (Bilingual Evaluation Understudy)
- NIST

26

MT Evaluation exercise

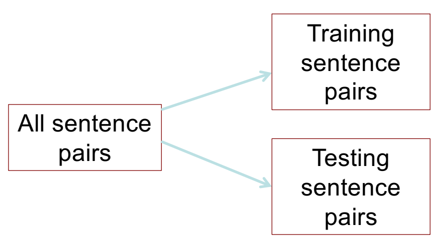
Play with an MT system

1. Find a few examples of the system doing interesting (surprising?) ”good” translations.
2. Find some examples of the system making mistakes (consider, idioms and common expressions)

27

Automatic Evaluation

Common NLP/machine learning/AI approach



28

Automatic Evaluation

Reference (human) translation:
The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

Machine translation:
The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

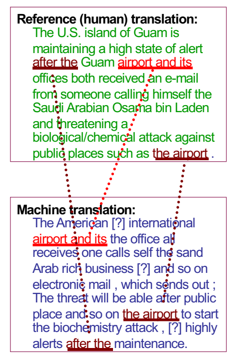
Machine translation 2:
United States Office of the Guam International Airport and were received by a man claiming to be Saudi Arabian businessman Osama bin Laden, sent emails, threats to airports and other public places will launch a biological or chemical attack, remain on high alert in Guam.

Ideas?

29

BLEU Evaluation Metric

(Papineni et al, ACL-2002)



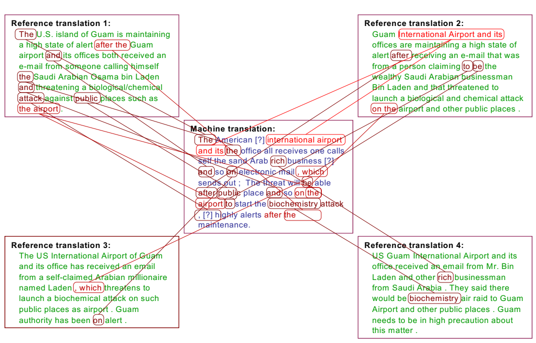
Basic idea:

Combination of n-gram precisions of varying size

What percentage of machine n-grams can be found in the reference translation?

30

Multiple Reference Translations



31

N-gram precision example

Candidate 1: *It is a guide to action which ensures that the military always obey the commands of the party.*

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*

Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Reference 3: *It is the practical guide for the army always to heed directions of the party.*

What percentage of machine n-grams can be found in the reference translations? Do unigrams, bigrams and trigrams.

32

N-gram precision example

Candidate 1: *It is a guide to action which ensures that the military always* obey *the commands of the party.*

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*
Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*
Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 17/18

33

N-gram precision example

Candidate 1: *It is a guide to action which ensures that the military*

always obey the commands of the party.

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*
Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*
Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 17/18
Bigrams: 10/17

34

N-gram precision example

Candidate 1: *It is a guide to action which ensures that the military*

always obey the commands of the party.

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*
Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*
Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 17/18
Bigrams: 10/17
Trigrams: 7/16

35

N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions guide that party commands.*

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*
Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*
Reference 3: *It is the practical guide for the army always to heed directions of the party.*

36

N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions guide that party commands.*

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*

Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 12/14

37

N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions*

guide that party commands.

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*

Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 12/14
Bigrams: 4/13

38

N-gram precision example 2

Candidate 2: *It is to ensure the army forever hearing the directions*

guide that party commands.

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*

Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*

Reference 3: *It is the practical guide for the army always to heed directions of the party.*

Unigrams: 12/14
Bigrams: 4/13
Trigrams: 1/12

39

N-gram precision

Candidate 1: *It is a guide to action which ensures that the military always obey the commands of the party.*

Unigrams: 17/18
Bigrams: 10/17
Trigrams: 7/16

Candidate 2: *It is to ensure the army forever hearing the directions guide that party commands.*

Unigrams: 12/14
Bigrams: 4/13
Trigrams: 1/12

Any problems/concerns?

40

N-gram precision example

Candidate 3: the
Candidate 4: It is a

Reference 1: *It is a guide to action that ensures that the military will forever heed Party commands.*
Reference 2: *It is the guiding principle which guarantees the military forces always being under the command of the Party.*
Reference 3: *It is the practical guide for the army always to heed directions of the party.*

What percentage of machine n-grams can be found in the reference translations? Do unigrams, bigrams and trigrams.

41

BLEU Evaluation Metric

(Papineni et al, ACL-2002)

Reference (human) translation:
The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

Machine translation:
The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

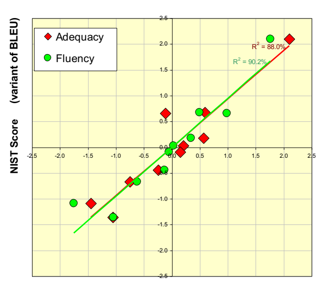
N-gram precision (score is between 0 & 1)
- What percentage of machine n-grams can be found in the reference translation?
- Not allowed to use same portion of reference translation twice (can't cheat by typing out "the the the the the")

Brevity penalty
- Can't just type out single word "the" (precision 1.0!)

*** Amazingly hard to "game" the system (i.e., find a way to change machine output so that BLEU goes up, but quality doesn't)

42

BLEU Tends to Predict Human Judgments



slide from G. Doddington (NIST)

43

BLEU: Problems?

Doesn't care if an incorrectly translated word is a name or a preposition
- *gave it to Albright* (reference)
- *gave it at Albright* (translation #1)
- *gave it to altar* (translation #2)

What happens when a program reaches human level performance in BLEU but the translations are still bad?
- maybe sooner than you think …

47

Appendix A

Input: corpus of English/Foreign sentence pairs (no alignment)

```
for some number of iterations:
    for (E, F) in corpus:
        for e in E:
            for f in F:
```
$p(f \to e) = p(f|e) / \sum_{e\ in\ E} p(f|e)$

count(e,f) += $p(f \to e)$

count(e) += $p(f \to e)$

for all (e,f) in count:

p(f|e) = count(e,f) / count(e)

48

Appendix A

for (E, F) in corpus:

for e in E:

for f in F:

$$p(f \to e) = \frac{p(f|e)}{\sum_{e\ in\ E} p(f|e)}$$

count(e,f) += $p(f \to e)$

count(e) += $p(f \to e)$

Pair 1: E: green house
F: casa verde

Pair 2: E: the house
F: la casa

Step 1: calculate p(f -> e) for all pairs of words in the two sentences (assume p(f|e) is a constant for all f,e)

49

Appendix A

for (E, F) in corpus:

for e in E:

for f in F:

$$p(f \to e) = \frac{p(f|e)}{\sum_{e\ in\ E} p(f|e)}$$

count(e,f) += $p(f \to e)$

count(e) += $p(f \to e)$

Pair 1: E: green house
F: casa verde

Pair 2: E: the house
F: la casa

Step 2: aggregate the counts

50

Appendix A

Pair 1: E: green house
F: casa verde

Pair 2: E: the house
F: la casa

for all (e,f) in count:

p(f|e) = count(e,f) / count(e)

Step 3: recalculate p(e|f)

51

Appendix A

Input: corpus of English/Foreign sentence pairs (no alignment)

```
for some number of iterations:
  for (E, F) in corpus:
    for e in E:
      for f in F:
        p(f → e) = p(f|e)/ Σ_{e in E} p(f|e)
        count(e,f) += p(f → e)
        count(e) += p(f → e)

      for all (e,f) in count:
            p(f|e) = count(e,f) / count(e)
```

52

Worksheet

Pair 1:
p(casa → green) =
p(casa → house) =
p(verde → green) =
p(verde → house =

Pair 2:
p(la → the) =
p(la → house) =
p(casa → the) =
p(casa → house =

---

count(green, casa) =
count(green, verde) =

count(house, casa) =
count(house, verde) =
count(house, la) =

count(the, casa) =
count(the, la) =

count(green) =
count(house) =
count(the) =

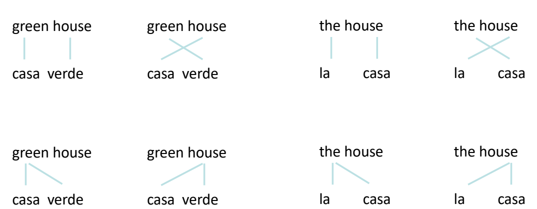
---

p(casa | green) =
p(verde | green) =

p(casa | house) =
p(verde | house) =
p(la | house) =

p(casa | the) =
p(la | the) =

53



p(casa | green)	1/2
p(verde | green)	1/2
p(la | green)	0

p(casa | house)	1/2
p(verde | house)	1/4
p(la | house)	1/4

p(casa | the)	1/2
p(verde | the)	0
p(la | the)	1/2

c(casa,green) = 1/4+1/4 = 1/2
c(verde,green) = 1/4+1/4 = 1/2
c(la, green) = 0

c(casa,house) = 1/4+1/4+1/4+1/4 = 1
c(verde,house) = 1/4+1/4 = 1/2
c(la,house) = 1/4+1/4 = 1/2

c(casa,the) = 1/4+1/4 = 1/2
c(verde,the) = 0
c(la,the) = 1/4+1/4 = 1/2

Then, calculate the probabilities by normalizing the counts

54