Machine Translation Concluded

David Kauchak

CS159 – Fall 2020

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

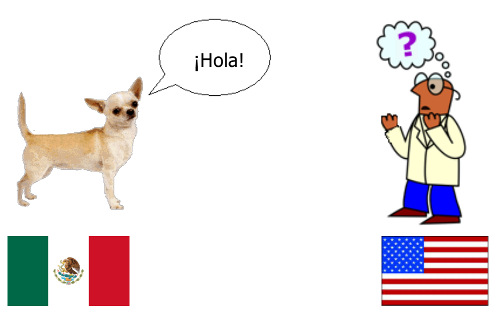
Assignment 5b

Assignment 6 available

Quiz 3: 11/10

2

Language translation



3

Word models: IBM Model 1

NULL Mary did not slap the green witch

Maria no dió una botefada a la bruja verde

p(verde | green)

Each foreign word is aligned to exactly one English word

This is the **ONLY** thing we model!

$$p(f_1 f_2 \ldots f_{|F|}, a_1 a_2 \ldots a_{|F|} \mid e_1 e_2 \ldots e_{|E|}) = \prod_{i=1}^{|F|} p(f_i \mid e_{a_i})$$

4

Training without alignments

Initially assume a p(f|e) are equally probable

Repeat:

- Enumerate all possible alignments
- Calculate how probable the alignments are under the current model (i.e. p(f|e))
- Recalculate p(f|e) using counts from **all** alignments, **weighted** by how probable they are

(Note: theoretical algorithm)

5

EM alignment

E-step

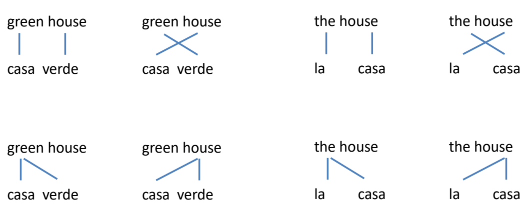
- Enumerate all possible alignments
- Calculate how probable the alignments are under the current model (i.e. p(f|e))

M-step

- Recalculate p(f|e) using counts from **all** alignments, **weighted** by how probable they are

(Note: theoretical algorithm)

6



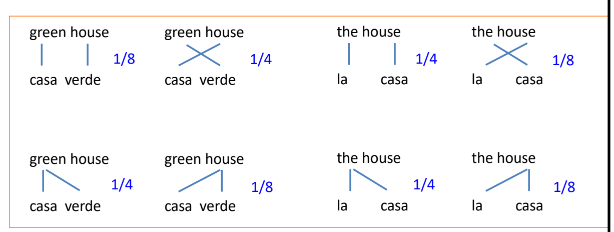
p(casa | green)	1/2
p(verde | green)	1/2
p(la | green)	0

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

E-step: Given p(F|E), calculate p(A,F|E)

7



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

E-step: Given p(F|E), calculate p(A,F|E)

Calculate unnormalized counts

8

green house / casa verde 1/8
green house / casa verde 1/4
the house / la casa 1/4
the house / la casa 1/8

green house / casa verde 1/4
green house / casa verde 1/8
the house / la casa 1/4
the house / la casa 1/8

sum = (3/4)　　**sum = (3/4)**

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

$$p(a_i|E,F) = \frac{p(F,ai|E)}{\sum_{j=1}^{4} p(F,aj|E)}$$

E-step: Given p(F|E), calculate p(A,F|E)

Normalize by the sum

9

green house / casa verde 1/6
green house / casa verde 1/3
the house / la casa 1/3
the house / la casa 1/6

green house / casa verde 1/3
green house / casa verde 1/6
the house / la casa 1/3
the house / la casa 1/6

sum = (3/4)　　**sum = (3/4)**

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

$$p(a_i|E,F) = \frac{p(F,ai|E)}{\sum_{j=1}^{4} p(F,aj|E)}$$

E-step: Given p(F|E), calculate p(A,F|E)

Normalize by the sum

10

green house / casa verde 1/6
green house / casa verde 1/3
the house / la casa 1/3
the house / la casa 1/6

green house / casa verde 1/3
green house / casa verde 1/6
the house / la casa 1/3
the house / la casa 1/6

M-step: calculate unnormalized counts for p(f|e) given the alignments

p(casa | green)	
p(verde | green)	
p(la | green)	

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

p(casa | the)	
p(verde | the)	
p(la | the)	

c(casa,green) = 1/6+1/3 = 3/6
c(verde,green) = 1/3+1/3 = 4/6
c(la, green) = 0

c(casa,house) = 1/3+1/6+ 1/3+1/6 = 6/6
c(verde,house) = 1/6+1/6 = 2/6
c(la,house) = 1/6+1/6 = 2/6

c(casa,the) = 1/6+1/3 = 3/6
c(verde,the) = 0
c(la,the) = 1/3+1/3 = 4/6

11

green house / casa verde 1/6
green house / casa verde 1/3
the house / la casa 1/3
the house / la casa 1/6

green house / casa verde 1/3
green house / casa verde 1/6
the house / la casa 1/3
the house / la casa 1/6

M-step: normalize

p(casa | green)	3/7
p(verde | green)	4/7
p(la | green)	0

p(casa | house)	3/5
p(verde | house)	1/5
p(la | house)	1/5

p(casa | the)	3/7
p(verde | the)	0
p(la | the)	4/7

c(casa,green) = 1/6+1/3 = 3/6
c(verde,green) = 1/3+1/3 = 4/6
c(la, green) = 0

c(casa,house) = 1/3+1/6+ 1/3+1/6 = 6/6
c(verde,house) = 1/6+1/6 = 2/6
c(la,house) = 1/6+1/6 = 2/6

c(casa,the) = 1/6+1/3 = 3/6
c(verde,the) = 0
c(la,the) = 1/3+1/3 = 4/6

Then, calculate the probabilities by normalizing the counts

12

Implementation details

For |E| English words and |F| foreign words, how many alignments are there?

Repeat:

E-step

- Enumerate all possible alignments
- Calculate how probable the alignments are under the current model (i.e. p(f|e))

M-step

- Recalculate p(f|e) using counts from **all** alignments, **weighted** by how probable they are

13

Implementation details

Each foreign word can be aligned to any of the English words (or NULL)

$(|E|+1)^{|F|}$

Repeat:

E-step

- Enumerate all possible alignments
- Calculate how probable the alignments are under the current model (i.e. p(f|e))

M-step

- Recalculate p(f|e) using counts from **all** alignments, **weighted** by how probable they are

14

Thought experiment

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

El viejo está feliz porque ha pescado muchos veces.

His wife talks to him.

Su mujer habla con él.

The sharks await.

Los tiburones esperan.

$$p(f_i \mid e_{a_i}) = \frac{count(f\ aligned\text{-}to\ e)}{count(e)}$$

p(el | the) = 0.5
p(Los | the) = 0.5

15

If we had the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

```
for (E, F) in corpus:
    for aligned words (e, f) in pair (E,F):
        count(e,f) += 1
        count(e) += 1

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

16

If we had the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

```
for (E, F) in corpus:
    for e in E:
        for f in F:
            if f aligned-to e:
                count(e,f) += 1
                count(e) += 1

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


If we had the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

```
for (E, F) in corpus:
    for aligned words (e, f) in pair (E,F):
        count(e,f) += 1
        count(e) += 1
```

```
for (E, F) in corpus
    for e in E:
        for f in F:
            if f aligned-to e:
                count(e,f) += 1
                count(e) += 1
```

Are these equivalent?

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


Thought experiment #2

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

El viejo está feliz porque ha pescado muchos veces.

80 annotators

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

El viejo está feliz porque ha pescado muchos veces.

20 annotators

$$p(f_i \mid e_{a_i}) = \frac{count(f\ aligned\text{-}to\ e)}{count(e)}$$

Use partial counts:
- count(viejo | man) 0.8
- count(viejo | old) 0.2

Without the alignments

```
if f aligned-to e:
        count(e,f) += 1
        count(e) += 1
```

$p(f \rightarrow e)$: probability that f is aligned to e *in this pair*

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

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

Key: use ***expected*** counts (i.e., how likely based on the current model), rather than actual counts

Without alignments

$p(f \to \mathrm{e})$: probability that f is aligned to e *in this pair*

a b c

y z

What is $p(y \to \mathrm{a})$?

Put another way, of all things that y could align to in this sentence, how likely is it to be a?

21

Without alignments

$p(f \to \mathrm{e})$: probability that f is aligned to e *in this pair*

a b c

y z

Of all things that *y* could align to, how likely is it to be *a*:

p(y | a)

Does that do it?

No! p(y | a) is how likely y is to align to a over the whole data set.

22

Without alignments

$p(f \to \mathrm{e})$: probability that f is aligned to e *in this pair*

a b c

y z

Of all things that *y* could align to, how likely is it to be *a*:

$$\frac{\text{p(y | a)}}{\text{p(y | a) + p(y | b) + p(y | c)}}$$

23

Without the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

```
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)
```

24

EM: without the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

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

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

25

EM: without the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

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

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

26

EM: without the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

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

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

Where are the E and M steps?

27

EM: without the alignments

Input: corpus of English/Foreign sentence pairs along with alignment

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

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

Calculate how probable the alignments are under the current model (i.e. p(f|e))

28

EM: without the alignments

Input: corpus of English/Foreign sentence pairs along with 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)
```

Recalculate p(f|e) using counts from **all** alignments, **weighted** by how probable they are

29

NULL

Sometimes foreign words don't have a direct correspondence to an English word

Adding a NULL word allows for p(f | NULL), i.e. words that appear, but are not associated explicitly with an English word

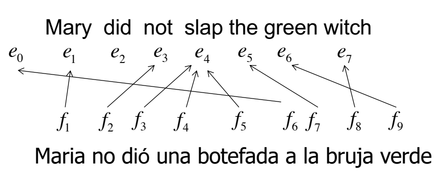
Implementation: add "NULL" (**or some unique string representing NULL**) to each of the English sentences, often at the beginning of the sentence

p(casa | NULL)	1/3
p(verde | NULL)	1/3
p(la | NULL)	1/3

30

Benefits of word-level model

Rarely used in practice for modern MT system



Two key side effects of training a word-level model:

- Word-level alignment
- p(f | e): translation dictionary

How do I get this?

31

Word alignment

100 iterations

p(casa | green)	0.005
p(verde | green)	0.995
p(la | green)	0

p(casa | house)	~1.0
p(verde | house)	~0.0
p(la | house)	~0.0

p(casa | the)	0.005
p(verde | the)	0
p(la | the)	0.995

green house

casa verde

How should these be aligned?

the house

la casa

32

Word alignment

100 iterations

p(casa | green)	0.005
p(verde | green)	0.995
p(la | green)	0

p(casa | house)	~1.0
p(verde | house)	~0.0
p(la | house)	~0.0

p(casa | the)	0.005
p(verde | the)	0
p(la | the)	0.995

green house

casa verde

Why?

the house

la casa

33

Word-level alignment

$$alignment(E,F) = \arg_A \max p(A,F \mid E)$$

Which for IBM model 1 is:

$$alignment(E,F) = \arg_A \max \prod_{i=1}^{|F|} p(f_i \mid e_{a_i})$$

Given a model (i.e. trained p(f|e)), how do we find this?

Align each foreign word (f in F) to the English word (e in E) with highest p(f|e)

$$a_i = \arg_{j:1-|E|} \max p(f_i \mid e_j)$$

34

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?

35

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?

36

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!

37

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: $\frac{6}{10}$

38

Problems for Statistical MT

Preprocessing

Language modeling

Translation modeling

Decoding

Parameter optimization

Evaluation

39

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

40

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$

Morgen | fliege | ich | nach Kanada | zur Konferenz

1. Sentence is divided into phrases

41

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$

Morgen | fliege | ich | nach Kanada | zur Konferenz

Tomorrow | will fly | I | In Canada | to the conference

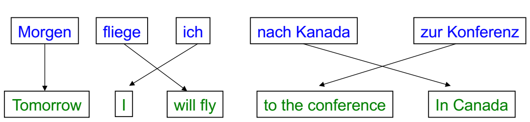
1. Sentence is divided into phrases
2. Phrases are translated (avoids a lot of weirdness from word-level model)

42

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

43

Phrase table

natuerlich

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

44

Phrase table

den Vorschlag

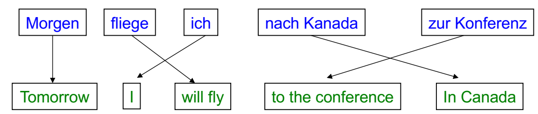
Translation	Probability
the proposal	0.6227
‘s proposal	0.1068
a proposal	0.0341
the idea	0.0250
this proposal	0.0227
proposal	0.0205
of the proposal	0.0159
the proposals	0.0159
the suggestions	0.0114
…	

45

Phrasal translation model

The models define probabilities over inputs

$$p(f \mid e)$$



Advantages?

46

Advantages of Phrase-Based

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

Easy to understand

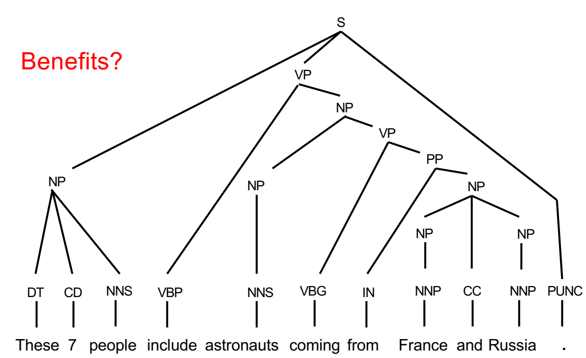
Local context is very useful for disambiguating
- “Interest rate” → …
- “Interest in” → …

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

47

Syntax-based models

Benefits?



48

Syntax-based models

Benefits

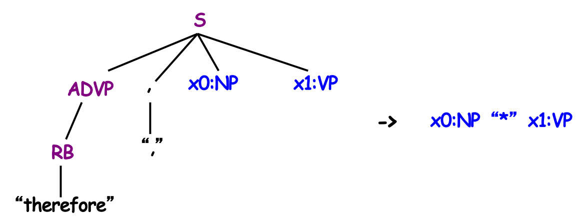
- Can use syntax to motivate word/phrase movement
- Could ensure grammaticality

Two main types:

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

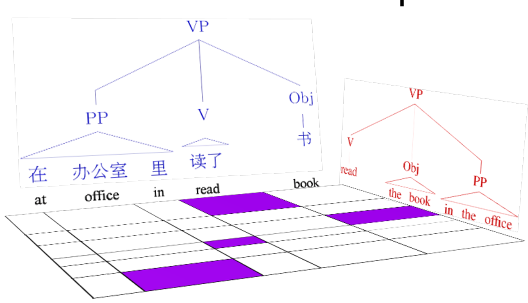
49

Tree to string rule



50

Tree to tree example



51

Problems for Statistical MT

Preprocessing

Language modeling

Translation modeling

Decoding

Parameter optimization

Evaluation

52

MT Evaluation

How do we do it?

What data might be useful?

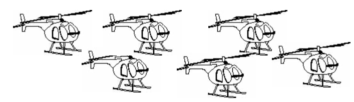
53

MT Evaluation

Source only

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

Extrinsic:
Objective usage testing

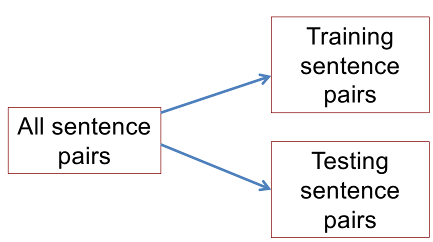


Automatic:
- WER (word error rate)
- BLEU (Bilingual Evaluation Understudy)
- NIST

54

Automatic Evaluation

Common NLP/machine learning/AI approach



55

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?

56

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.

Basic idea:

Combination of n-gram precisions of varying size

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

57

Multiple Reference Translations

Reference translation 1:
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 .

Reference translation 2:
Guam International Airport and its offices are maintaining a high state of alert after receiving an e-mail that was from a person claiming to be the wealthy Saudi Arabian businessman Bin Laden and that threatened to launch a biological and chemical attack on the airport and other public places .

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.

Reference translation 3:
The US International Airport of Guam and its office has received an email from a self-claimed Arabian millionaire named Laden , which threatens to launch a biochemical attack on such public places as airport . Guam authority has been on alert .

Reference translation 4:
US Guam International Airport and its office received an email from Mr. Bin Laden and other rich businessman from Saudi Arabia . They said there would be biochemistry air raid to Guam Airport and other public places . Guam needs to be in high precaution about this matter .

58

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.

59

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

60

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

61

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

62

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.*

63

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

64

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

65

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

66

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?

67

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.

68

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")

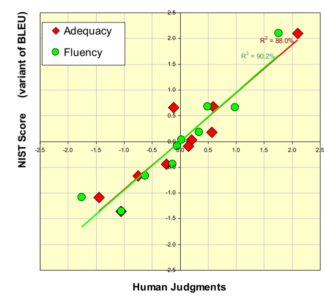
Brevity penalty

- 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)

69

BLEU Tends to Predict Human Judgments



slide from G. Doddington (NIST)

70

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 …

71