CS626: Speech, NLP and the Web

Machine Translation cntd., MT Evaluation

Pushpak Bhattacharyya

Computer Science and Engineering Department

IIT Bombay

Week 14 of 24h October, 2022

Phrase Based MT

An Example

English--> Hindi	*People*	*Of*	*Mumbai*
Mumbai			X
Ke		X	
Log	X		

Table X.5: creation of phrase alignments from word alignments through grow-dia algorithm

Phrase Alignment Process

- Run IBM model 3 in both directions- source to target and target to source- to create what are called *alignment sets*. There are two alignment sets: one in each direction.

- Then apply a process called *symmetrisation* to obtain phrase alignments.

Alignments: “People of Mumbai”←→”Mumbai ke logoM”

- *A1: {<Mumbai, Mumbai>, <of, ke>, <people, log>}*

- *A2: {< mumbai, Mumbai>, < ke, of>, < log, people>}*

Grow Diagonal Process

- *People of --> ke log* (blue square)
- *of Mumbai --> mumbai ke* (yellow square)
- *People of Mumbai --> mumbai ke log* (red square)

Illustration with “people of…”

English--> Hindi ↓	*People*	*Of*	*Mumbai*
Mumbai			X
Ke		X	
Log	X		

Linguistic and Non-linguistic Phrases

- '*People of Mumbai*' → '*mumbai ke log*'
  - noun phrase (NP) alignment
- '*of Mumbai*' → '*mumbai ke*'
  - preposition phrase (PP),
- '*people of*' → '*ke log*' is not a linguistic phrase (headedness property violated)

Case of Null Alignment

English--> Hindi ↓	*Meet*	*The*	*People*	*Of*	*Mumbai*
Mumbai					X
Ke				X	
LogoM			X		
Se					
Miliye	X				

Table X.6: phrase alignments in case of null alignment

Growing Bigger Alignments

English--> Hindi ↓	*Meet*	*The*	*People*	*Of*	*Mumbai*
Mumbai					X
Ke				X	
LogoM			X		
Se					
Miliye	X				

Table X.6: phrase alignments in case of null alignment

Grow-Diag with null alignment

- The red box will expand into the cell <*se*, *the*> and create the alignment '*the people of Mumbai*'-->'*mumbai ke logoM se*'.
- The alignment '*the people'-->'logoM*' too will be created, merging the <*ke*, *the*> cell with <*logoM*, *people*>.
- Cells from null rows or null columns can be merged upwards or downwards, thereby associating the row-word/column-word with the next phrase or the previous phrase.

Consequence of Grow-Diag with null alignment

- Due to the null alignment of ‘*the*’, all of the following phrase alignments are possible:

‘*meet the*’--> ‘*se miliye*’

‘*the people*’ <-> ’*logoM*’

‘*the people*’ --> ’*logoM se*’

‘the’ and ‘se’ both can be *both* prefix and suffix of phrases.

Influence of Data (1/2)

- Q: Which phrase amongst the above will be retained?
- A: ALL! But with different probabilities.
- '*the people*'--> '*logoM*' should have higher probability than '*the people*'--> '*logoM se*',
- Because '*people*' is likely to seen more in the company of '*logoM*' than '*logoM se*'
- as in *'tell the people'--> 'logoM ko bolo'*, '*have faith in the people*'--> '*logoM pe viswaas rakho*' and so on

Mathematics of PBSMT

Follows (1) Koehn (2010) and (2) upcoming text book “Natural Language Processing” Bhattacharyya and Joshi (2022)

Starting equation

$$e_{best} = \arg\max_e P(e \mid f) = \arg\max_e [P(f \mid e) P_{LM}(e)]$$

where *e* and *f* have their usual meaning of output and input respectively; the translation with the highest score is e_{best}. $P(f|e)$ and $P_{LM}(e)$ are the translation model and language model, respectively

Modelling P(f|e)

$$P(\overline{f}_1^I \mid \overline{e}_1^I) = P(\overline{f_1}, \overline{f_2}, \ldots, \overline{f_I} \mid \overline{e_1}, \overline{e_2}, \ldots, \overline{e_I})$$

$$= \prod_{i=1}^{I} \Phi(\overline{f_i} \mid \overline{e_i})\, d(start_i - end_{i-1} - 1)$$

LHS is the probability of sequence of *I* phrases in the sentence *f*, given *I* phrases in sentence *e*. *Φ* is called the *phrase translation probability* and *d(.)* is the *distortion probability.*

Distortion: $d(start_i - end_{i-1} - 1)$

$$P(\overline{f}_1^I \mid \overline{e}_1^I) = P(\overline{f}_1, \overline{f}_2, \ldots, \overline{f}_I \mid \overline{e}_1, \overline{e}_2, \ldots, \overline{e}_I)$$

$$= \prod_{i=1}^{I} \Phi(\overline{f}_i \mid \overline{e}_i)\, d(start_i - end_{i-1} - 1)$$

- $start_i$: starting position of the translation of the i^{th} phrase of *e* in *f*
- end_{i-1}: end position of the translation of the $(i-1)^{th}$ phrase of *e* in *f.*

Reordering of Phrases

- $start_i$-end_{i-1}-1 is a measure of the distance between the translation of i^{th} phrase and the translation of the $(i-1)^{th}$ phrase of e as they appear as the j^{th} and k^{th} phrase in f.

- It is, thus, also a measure of the *reordering* of phrases induced by the translation.

Illustration of PBSMT

Hindi: *आज जल्दी आना*

Hindi Transliteration: *aaj jaldii aanaa*

English gloss: *today soon come*

The expected output is:

English: *come soon today*

Phrases

HP1: "*aaj*"

HP2: "*jaldii aanaa*"

Now, from the phrase table, we find the translation units for these two phrases as:

EP1: "*today*" (with probability p_1; there are other translations too)

EP2: "*come soon*" (with probability p_2; there are other translations too)

There could be multiple translations

C1: “*come soon today*”

C2: “*today come soon*”

e generates *f* and is scored according to:

- 1. Language model probability $P_{LM}(e)$ (the prior)
- 2. Product of probabilities of translations of phrases of *e*
- 3. Product of probabilities of distance by which phrases have moved around

Computing distortion (1/2)

- For candidate C1, the distance by which EP1 has moved is found as follows:

  $Start_i$ for *translation(“come soon”)= “jaldii aao”* is 2.

  End_{i-1}= 0 (we assume a null before the starting phrase whose translation occupies position 0).

Computing distortion (2/2)

- Distance for "*come soon*" = 2-0-1=1.
- For *today*
- $Start_i = 1$
- $End_{i-1} = 3$
- So the distance for *today* = 1-3-1= -3

- This means *translation("today")*= "*aaj*" has moved left three words to occupy the starting place of the Hindi sentence.

Distortion probability from data (1/2)

- To compute the distortion probability of "*come soon*", we will observe from the data how many times "*come soon*" has travelled a particular distance compared to other distances it has travelled.

- The ratio of these two numbers is the required probability.

Distortion probability from data (2/2)

- We can get the distortion probability of *today* in a similar manner.

- Spatiotemporal adverbs are normally found at the end of the sentence for English. Hence, "*come soon today*" should be scored over "*today come soon*", provided there is enough evidence of spatiotemporal adverbs in the parallel corpora.

Indian Language SMT *(LREC 2014)*

	hi	ur	pa	bn	gu	mr	kK	ta	te	ml	en
hi		61.28	68.21	34.96	51.31	39.12	37.81	14.43	21.38	10.98	29.23
ur	61.42		52.02	29.59	39.00	27.57	28.29	11.95	16.61	8.65	22.46
pa	73.31	56.00		29.89	43.85	30.87	30.72	10.75	18.81	9.11	23.83
bn	37.69	32.08	31.38		28.14	22.09	23.47	10.94	13.40	8.10	18.76
gu	55.66	44.12	45.14	28.50		32.06	30.48	12.57	17.22	8.01	19.78
mr	45.11	32.60	33.28	23.73	32.42		27.81	10.74	12.89	7.65	17.62
kK	41.92	34.00	34.31	24.59	31.07	27.52		10.36	14.80	7.89	17.07
ta	20.48	18.12	15.57	13.21	16.53	11.60	11.87		8.48	6.31	11.79
te	28.88	25.07	25.56	16.57	20.96	14.94	17.27	8.68		6.68	12.34
ml	14.74	13.39	12.97	10.67	9.76	8.39	9.18	5.90	5.94		8.61
en	28.94	22.96	22.33	15.33	15.44	12.11	13.66	6.43	6.55	4.65	

Baseline PBSMT - % BLEU scores (S1)

- **Clear partitioning of translation pairs by language family pairs**, based on translation accuracy.
  - Shared characteristics within language families make translation simpler
  - Divergences among language families make translation difficult

(Anoop Kunchukuttan, Abhijit Mishra, Pushpak Bhattacharyya, LREC 2014)

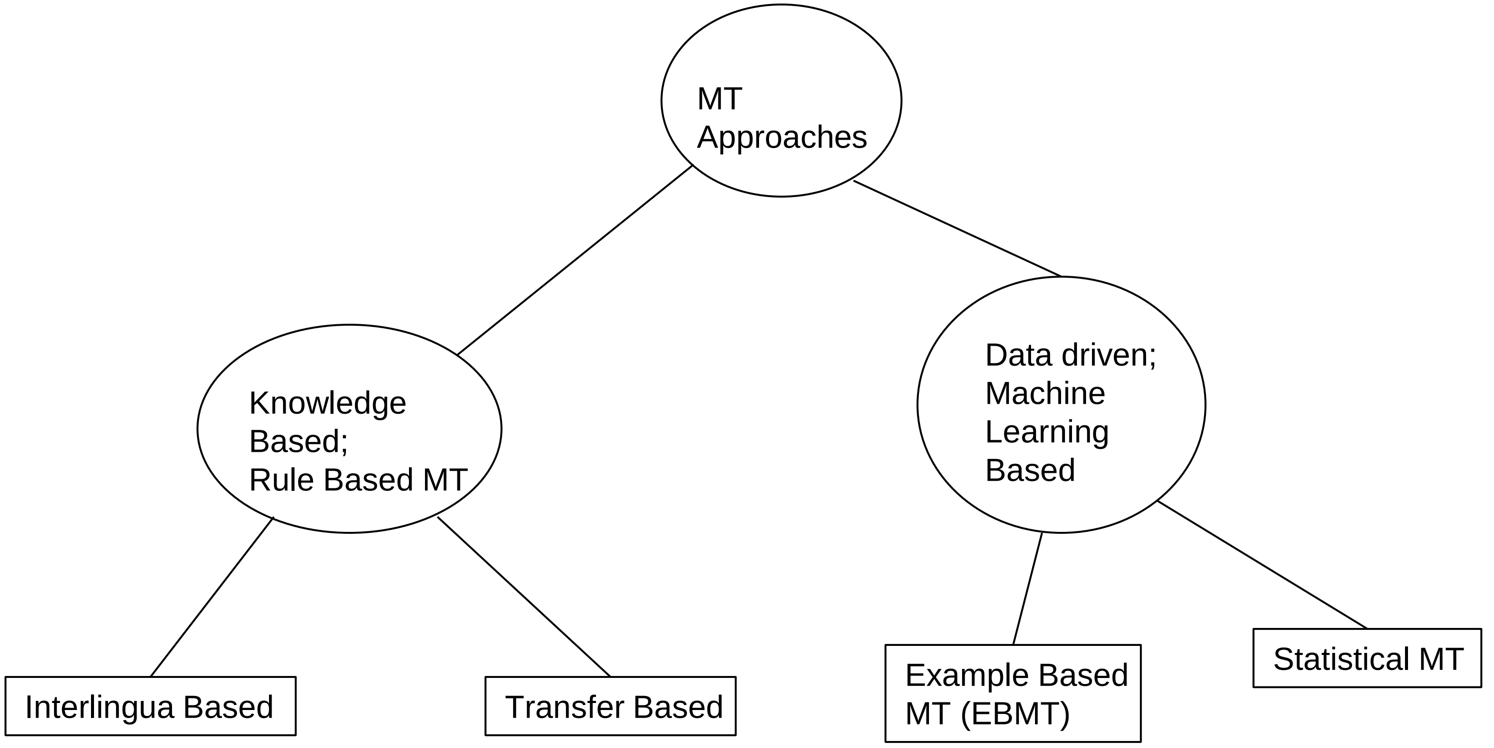
Observations on the Bleu scores

- Clear partitioning of Indo-Aryan and Dravidian languages.
- Rows: source languages, Cols: Target languages
- Take Hindi row: bleu scores good for Punjabi and urdu which are very close to Hindi.

What does “closeness” of languages mean

- Vocabulary overlap
- Word order same
- Morphology same (suffix vs. post position)
- Tonality same

Taxonomy of MT systems



RBMT-EBMT-SMT spectrum: knowledge (rules) intensive to data (learning) intensive

RBMT EBMT SMT

The tricky case of ‘have’ translation

- *Peter has a house*
- *Peter has a brother*
- *This hotel has a museum*

The tricky case of ‘have’ translation

English

- *Peter has a house*
- *Peter has a brother*
- *This hotel has a museum*

Marathi

- पीटरकडे एक घर आहे/ piitar kade ek ghar aahe
- पीटरला एक भाऊ आहे/ piitar laa ek bhaauu aahe
- ह्या हॉटेलमध्ये एक संग्रहालय आहे/ hyaa hotel madhye ek saMgrahaalay aahe

RBMT

If

syntactic subject is animate AND syntactic object is owned by subject

Then

“have” should translate to “kade … aahe”

If

syntactic subject is animate AND syntactic object denotes kinship with subject

Then

“have” should translate to “laa … aahe”

If

syntactic subject is inanimate

Then

“have” should translate to “madhye … aahe”

EBMT

X have Y →

Xkade Y aahe /

Xlaa Y aahe /

Xmadhye Y aahe

SMT

- *has a house ←→ kade ek ghar aahe*
- *has a car ←→ kade ek gaadii aahe*
- *has a brother ←→ laa ek bhaau aahe*
- *has a sister ←→ laa ek bahiin aahe*
- *hotel has ←→ hotel madhye*
- *hospital has ←→ hospital madhye*

SMT: new sentence

"This hospital has 100 beds"

- *n*-grams (*n=1, 2, 3, 4, 5*) like the following will be formed:
  - *"This", "hospital",… (unigrams)*
  - *"This hospital", "hospital has", "has 100",… (bigrams)*
  - *"This hospital has", "hospital has 100", … (trigrams)*

DECODING !!!

Why is MT difficult?

Language divergence

Why is MT difficult: Language Divergence

- Languages have different ways of expressing meaning
  - Lexico-Semantic Divergence
  - Structural Divergence

Our work on English-IL Language Divergence with illustrations from Hindi
(Dave, Parikh, Bhattacharyya, Journal of MT, 2002)

Languages differ in expressing thoughts: Agglutination

Finnish: “istahtaisinkohan”

English: "I wonder if I should sit down for a while“

Analysis:

- ist + "sit", verb stem
- ahta + verb derivation morpheme, "to do something for a while"
- isi + conditional affix
- n + 1st person singular suffix
- ko + question particle
- han a particle for things like reminder (with declaratives) or "softening" (with questions and imperatives)

Language Divergence Theory:
Lexico-Semantic Divergences (few examples)

- Conflational divergence
  - F: vomir; E: to be sick
  - E: *stab;* H: *chure se maaranaa (knife-with hit)*
  - S: *Utrymningsplan;* E: *escape plan*

- Categorial divergence
  - Change is in POS category:
  - *The play is on_PREP (vs. The play is Sunday)*
  - *Khel chal_rahaa_haai_VM (vs. khel ravivaar ko haai)*

Language Divergence Theory:

Structural Divergences

- SVO→SOV
  - E: *Peter plays basketball*
  - H: *piitar basketball kheltaa haai*

- Head swapping divergence
  - E: *Prime Minister of India*
  - H: *bhaarat ke pradhaan mantrii (India-of Prime Minister)*

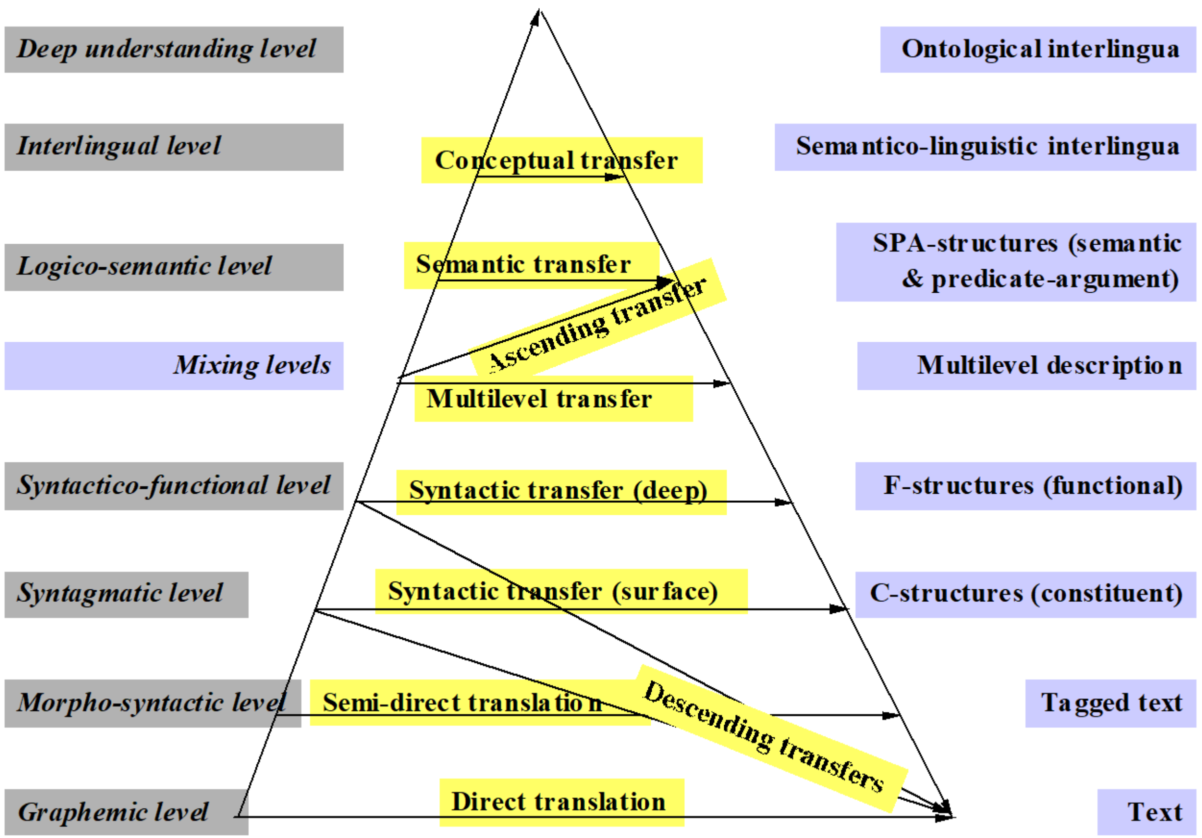
Language Divergence Theory: *Syntactic Divergences* (few examples)

- Constituent Order divergence
  - E: *Singh, the PM of India, will address the nation today*
  - H: *bhaarat ke pradhaan mantrii, singh, … (India-of PM, Singh…)*
- Adjunction Divergence
  - E: *She will visit here in the summer*
  - H: *vah yahaa garmii meM aayegii (she here summer-in will come)*
- Preposition-Stranding divergence
  - E: *Who do you want to go with?*
  - H: *kisake saath aap jaanaa chaahate ho? (who with…)*

Vauquois Triangle

Kinds of MT Systems

(point of entry from source to the target text)



Simplified Vauquois Triangle

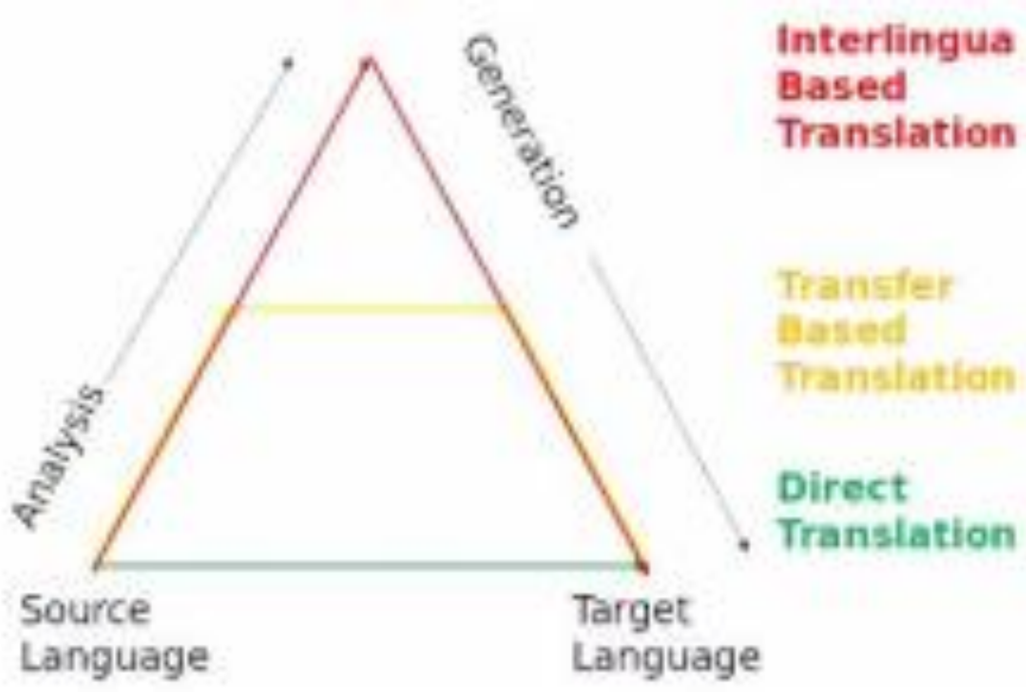
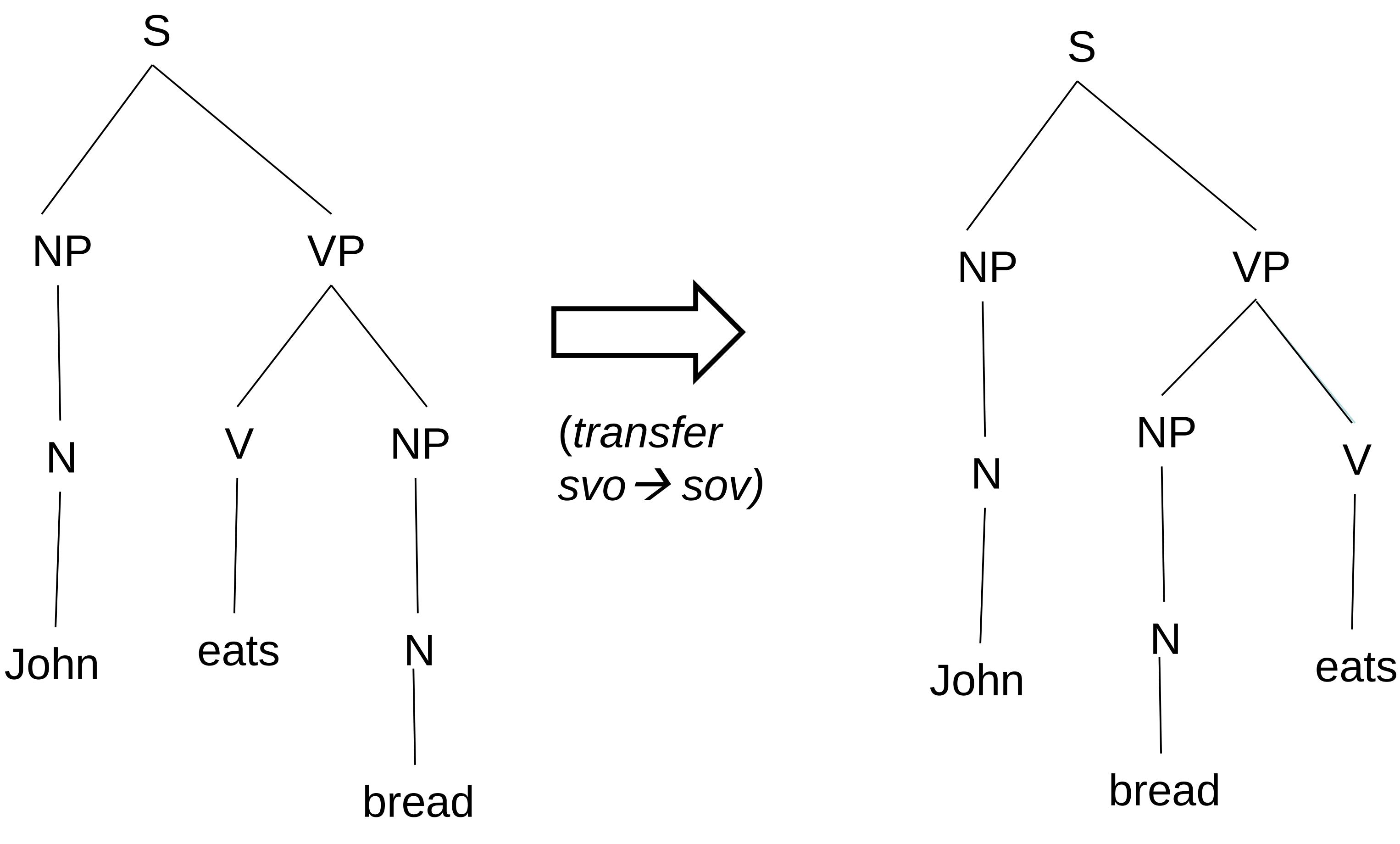


Figure X.6: Abridged Vauquois Triangle

Illustration of transfer SVO→SOV



Universality hypothesis

Universality hypothesis: At the level of “deep meaning”, all texts are the “same”, whatever the language.

NLP evaluation

Focus on MT evaluation

(Credit: Aditya Joshi, Kashyap Popat, Shubham Gautam)

Precision/Recall

Precision:

How many results returned were correct?

Recall:

What portion of correct results were returned?

Adapting precision/recall to NLP tasks

Document Classification: Taxonomy

- Labels form a taxonomy
- E.g.
  - Financial
    - Stocks
    - Tradings
    - Merger and acquisition, etc.
  - Sports
  - Cultural
  - Literature

Document Retrieval and Classification

- **Document Retrieval**

Precision =

$$\frac{|\text{Documents relevant and retrieved}|}{|\text{Documents retrieved}|}$$

Recall=

$$\frac{|\text{Documents relevant and retrieved}|}{|\text{Documents relevant}|}$$

- **Classification**

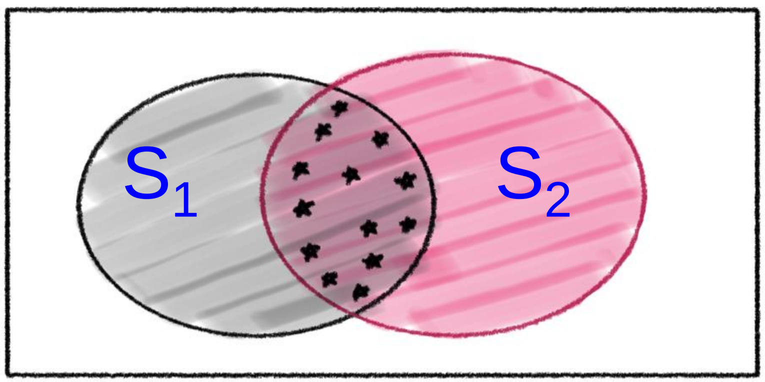
Precision =

$$\frac{|\text{True Positives}|}{|\text{True Positives + False Positives}|}$$

Recall=

$$\frac{|\text{True Positives}|}{|\text{True Positives + False Negatives}|}$$

Venn Diagram illustrating “Actual” vs “Obtained”



SET S_1 ··· OBTAINED

SET S_2 ··· ACTUAL

$S_1 \cap S_2$ ··· TRUE POSITIVES

$S_1 - (S_1 \cap S_2)$ ·· FALSE POSITIVES

$S_2 - (S_1 \cap S_2)$ ··· FALSE NEGATIVES

$(S_1 \cup S_2)^c$ -- TRUE NEGATIVES

$$Precision = \frac{|S_1 \bigcap S_2|}{|S_1|}$$

$$Recall = \frac{|S_1 \bigcap S_2|}{|S_2|}$$

Evaluation in MT

- Operational evaluation
  - "Is MT system A operationally better than MT system B? Does MT system A cost less?"
- Typological evaluation
  - "Have you ensured which linguistic phenomena the MT system covers?"
- Declarative evaluation
  - "How does quality of output of system A fare with respect to that of B?"

Adequacy (also called comprehensibility, fidelity, faithfulness) and Fluency

- Assign scores to specific qualities of output
  - Fluency: How good the output is as a well-formed target language entity
  - Adequacy: How good the output is in terms of preserving content of the source text

Form Content Dichotomy

- Ancient philosophical concept
- Consider a pot of milk: milk has the form of pot
- Pot has the content as milk.
- Adequacy refers to content, fluency refers to form

Adequacy and Fluency cntd.

For example, I am attending a lecture

मैं एक व्याख्यान बैठा हूँ

Main ek vyaakhyan baitha hoon

I a lecture sit (Present-first person)

I sit a lecture : Adequate but not fluent

मैं व्याख्यान हूँ

Main vyakhyan hoon

I lecture am

I am lecture: fluent but not adequate.

ADEQUACY AND FLUENCY SCALE

Adequacy and Fluency are measured in the scale of 1 to 5.

1: BAD !

2: MEDIOCRE !

3: GOOD !

4: VERY GOOD !

5: EXCELLENT !

What are human evaluators most sensitive to?

Native speakers are particularly keen on the correct usage of morphological variations and function words in the language.

e.g. “Rahul ka behen” instead of “Rahul ki behen” would be critically penalized.

Similarly, “Mary kitab padta hai” rather than “Mary kitab padti hai” would get a much lower score.

BLEU

Used in any kind of natural language generation situation: QA, Summarization, MT, Paraphrasing and so on

Foundational Point

- Human evaluation is the ultimate yardstick
- Any automatic evaluation MUST correlate well with human evaluation
- BLEU for last 20 years has satisfied reasonably this requirement
- Except in case of high morphological complexity, in which case we have to use subword based BLEU

Allied point: IAA

- Human evaluation is the skyline
- But human evaluation is subjective
- We must have multiple evaluators and compute inter-annotator agreement

How is translation performance measured?

The closer a machine translation is to a professional human translation, the better it is.

- A corpus of good quality human reference translations

- A numerical “translation closeness” metric

Suggested Papers

K. Papineni, S. Roukos, T. Ward, and W. Zhu. *Bleu: a method for automatic evaluation of machine translation,* ACL 2002.

Chris Callison-Burch, Miles Osborne, Phillipp Koehn, *Re-evaluating the role of Bleu in Machine Translation Research, European ACL (EACL) 2006, 2006.*

R. Ananthakrishnan, Pushpak Bhattacharyya, M. Sasikumar and Ritesh M. Shah, *Some Issues in Automatic Evaluation of English-Hindi MT: More Blues for BLEU*, **ICON 2007**, Hyderabad, India, Jan, 2007.

Cntd.

Preliminaries

- **Candidate Translation(s):** Translation returned by an MT system
- **Reference Translation(s):** ‘Perfect’ translation by humans

Goal of BLEU: To correlate with human judgment

Formulating BLEU (Step 1): Precision

I had lunch now.

Reference 1: मैने अभी खाना खाया

maine abhi khana khaya

I now food ate

I ate food now.

Reference 2 : मैने अभी भोजन किया

maine abhi bhojan kiyaa

I now meal did

I did meal now

Candidate 1: **मैने** अब **खाना खाया**

maine ab khana khaya

I now food ate

I ate food now

matching unigrams: 3,

matching bigrams: 1

Unigram precision: Candidate 1: 3/4 = 0.75, Similarly, bigram precision: Candidate 1: 0.33

Formulating BLEU (Step 1): Precision

I had lunch now.

Reference 1: मैने अभी खाना खाया
maine abhi khana khaya
I now food ate
I ate food now.

Reference 2 : मैने अभी भोजन किया
maine abhi bhojan kiyaa
I now meal did
I did meal now

Candidate 2: **मैने अभी** लंच एट
maine abhi lunch ate matching unigrams: 2,
I now lunch ate
I ate lunch (OOV) now(OOV) matching bigrams: 1

Unigram precision: Candidate 2: 2/4 = 0.5

Similarly, bigram precision: Candidate 2 = 0.33

Precision: Not good enough

Reference: *aapkii badii meharbaanii hogii*

I will be very thankful to you

Candidate 1: *aap badii meharbaanii hogii*

matching unigram: 3

Candidate 2: ***aapkii aapkii aapkii meharbaanii***

matching unigrams: 4

Unigram precision: Candidate 1: 3/4 = 0.75,

Candidate 2: 4/4 = 1

Formulating BLEU (Step 2): Modified Precision

- Clip the total count of each candidate word with its maximum reference count

- Countclip(n-gram) = min (count, max_ref_count)

  Reference: *aapkii badii meharbaanii hogii*

  Candidate 2: : ***aapkii aapkii aapkii meharbaanii***

  matching unigrams:
  (aapkii : min(3, 1) = 1)
  (meharbaaniii: min (1, 1) = 1)
  Modified unigram precision: 2/4 = 0.5

Modified n-gram precision

For entire test corpus, for a given n,

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}$$

n-gram: Matching n-grams in C

n-gram’: All n-grams in C

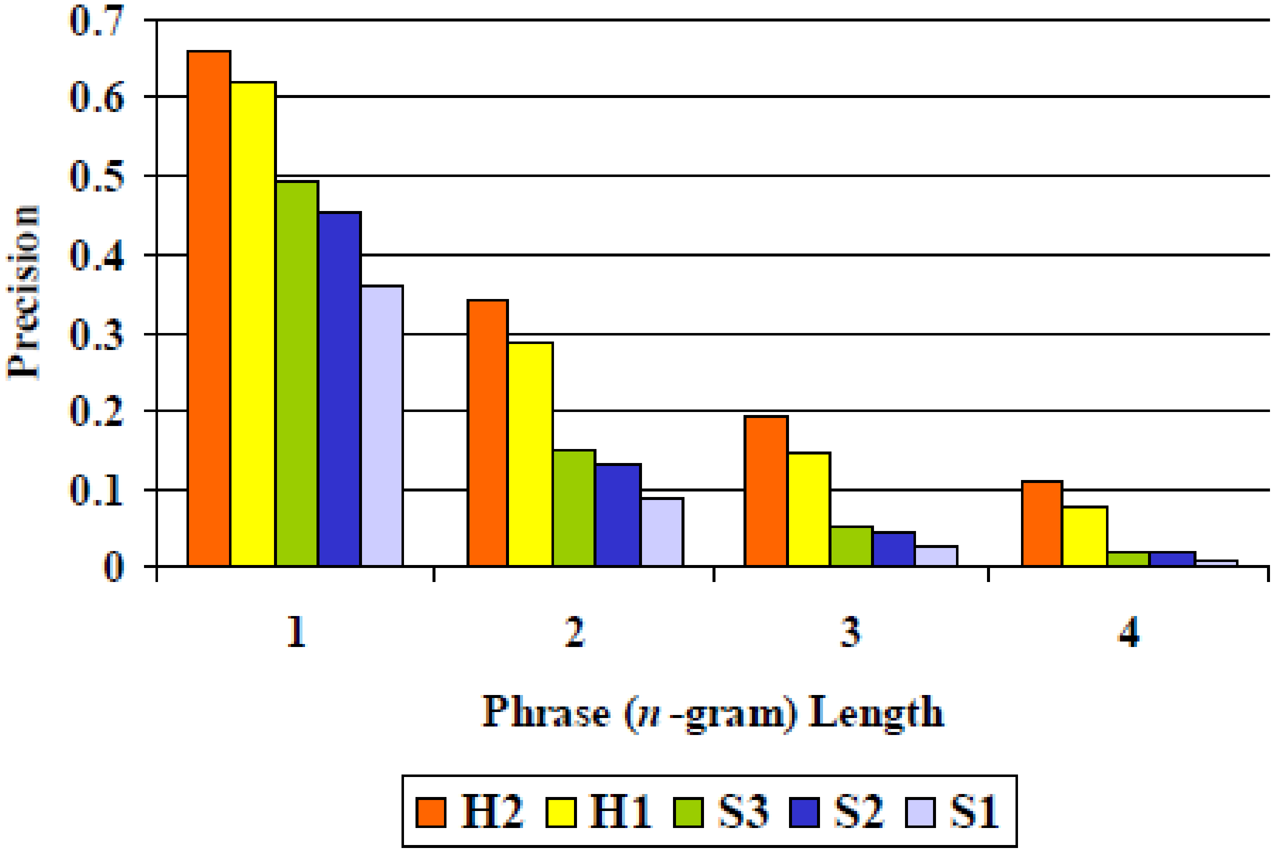
Modified precision for n-grams

Overall candidates of test corpus

Calculating modified n-gram precision (1/2)

- From the original BLEU paper (Papineni et al. 2002)
- 127 source sentences were translated by two human translators and three MT systems
- Translated sentences evaluated against professional reference translations using modified n-gram precision

Calculating modified n-gram precision (2/2)



Decaying precision with increasing n

Comparative ranking of the five

Combining precision for different values of n-grams?

A point about length of n-grams

- 1 and 2-grams stress vocabulary match or lexical goodness

- 3-4 and higher n-grams stress structural match or syntactic goodness

Formulation of BLEU: Recap

- Precision cannot be used as is
- Modified precision considers ‘clipped word count’

‘Recall’ for MT (1/2)

- Candidates shorter than references
- Reference: क्या ब्लू लंबे वाक्य की गुणवत्ता को समझ पाएगा?

  kya blue lambe vaakya ki guNvatta ko samajh paaega?

  will blue long sentence-of quality (case-marker) understand able(III-person-male-singular)?

  Will blue be able to understand quality of long sentence?

Candidate: **लंबे वाक्य**

lambe vaakya

long sentence

long sentence

modified unigram precision: 2/2 = 1

modified bigram precision: 1/1 = 1

Recall for MT (2/2)

Reference 1: मैने खाना खाया

maine khaana khaaya

I food ate

I ate food

Candidate 2: मैने खाना खाया

maine khaana khaaya

I food ate

I ate food

Modified unigram precision: 1

Candidate longer than references

Reference 2: मैने भोजन किया

maine bhojan kiyaa

I meal did

I had meal

Candidate 1: मैने खाना भोजन किया

maine khaana bhojan kiya

I food meal did

I had food meal

Modified unigram precision: 1

Formulating BLEU (Step 3): Incorporating recall

- Sentence length indicator of ‘good match’
- Brevity penalty (BP):
  - Multiplicative factor
  - Candidate translations that match reference translations in length must be ranked higher

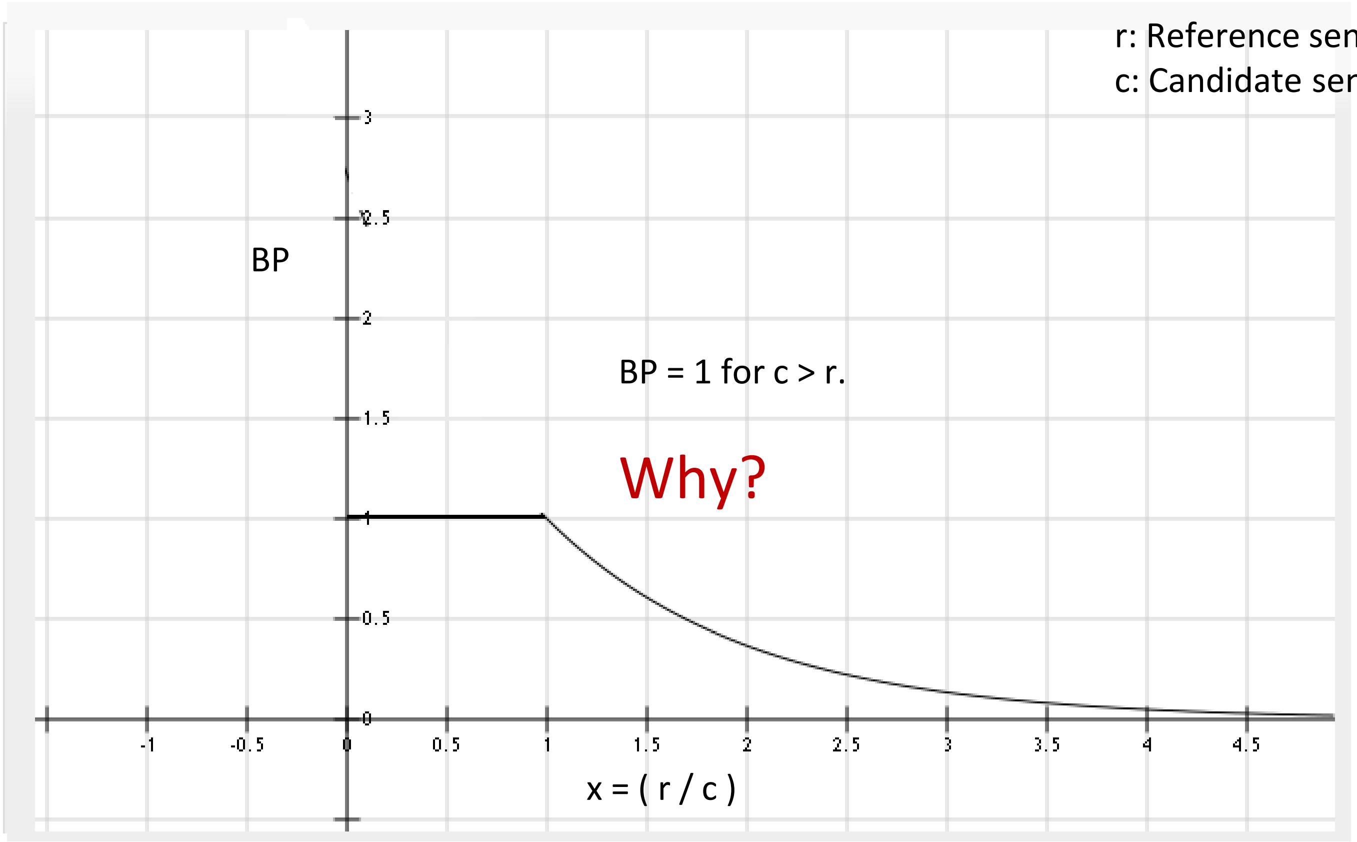
Candidate 1: लंबे वाक्य

Candidate 2: क्या ब्लू लंबे वाक्य की गुणवत्ता समझ पाएगा?

Formulating BLEU (Step 3): Brevity Penalty

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

r: Reference sentence length
c: Candidate sentence length



Graph drawn using www.fooplot.com

BP does not penalize translations longer than reference

Why?

Translations longer than reference are already penalized by modified precision

Validating the claim:

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}$$

Final BLEU Score Formula

Recall -> Brevity Penalty

Precision→ Modified n-gram precision

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}$$

$$\mathrm{B{\scriptstyle LEU}} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

Giving importance to Recall: Ref n-grams

ROUGE

- **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
- ROUGE is a package of metrics: ROUGE-N, ROUGE-L, ROUGE-W and ROUGE-S

ROUGE-N

ROUGE-N

$$= \frac{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\boxed{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)}}$$

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n\text{-}gram \in C} Count_{clip}(n\text{-}gram)}{\boxed{\sum_{C' \in \{Candidates\}} \sum_{n\text{-}gram' \in C'} Count(n\text{-}gram')}}$$

ROUGE-N incorporates Recall

Will BLEU be able to understand quality of long sentences?

Reference translation:
क्या ब्लू लंबे वाक्य की गुणवत्ता को समझ पाएगा?
Kya bloo lambe waakya ki guNvatta ko samajh paaega?

Candidate translation:
लंबे वाक्य
Lambe vaakya

ROUGE-N: 1 / 8
Modified n-gram Precision: 1

Other ROUGEs

- ROUGE-L
  - Considers longest common subsequence
- ROUGE-W
  - Weighted ROUGE-L: All common subsequences are considered with weight based on length
- ROUGE-S
  - Precision/Recall by matching skip bigrams

ROUGE v/s BLEU

	ROUGE	BLEU
Handling incorrect words	Skip bigrams, ROUGE-N	N-gram mismatch
Handling incorrect word order	Longest common sub-sequence	N-gram mismatch
Handling recall	ROUGE-N incorporates missing words	Precision cannot detect ‘missing’ words. Hence, brevity penalty!

$$\text{ROUGE-N} = \frac{\sum_{S \in \{ReferemceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)}$$

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n \right)$$

Language Typology

Proto-Language *(Wikipedia)*

Meaning:	Sanskrit	Latin:
"three"	*trayas*	*tres*
"seven"	*sapta*	*septem*
"eight"	*ashta*	*octo*
"nine"	*nava*	*novem*
"snake"	*sarpa*	*serpens*
"king"	*raja*	*regem*
"god"	*devas*	*divus* ("divine")

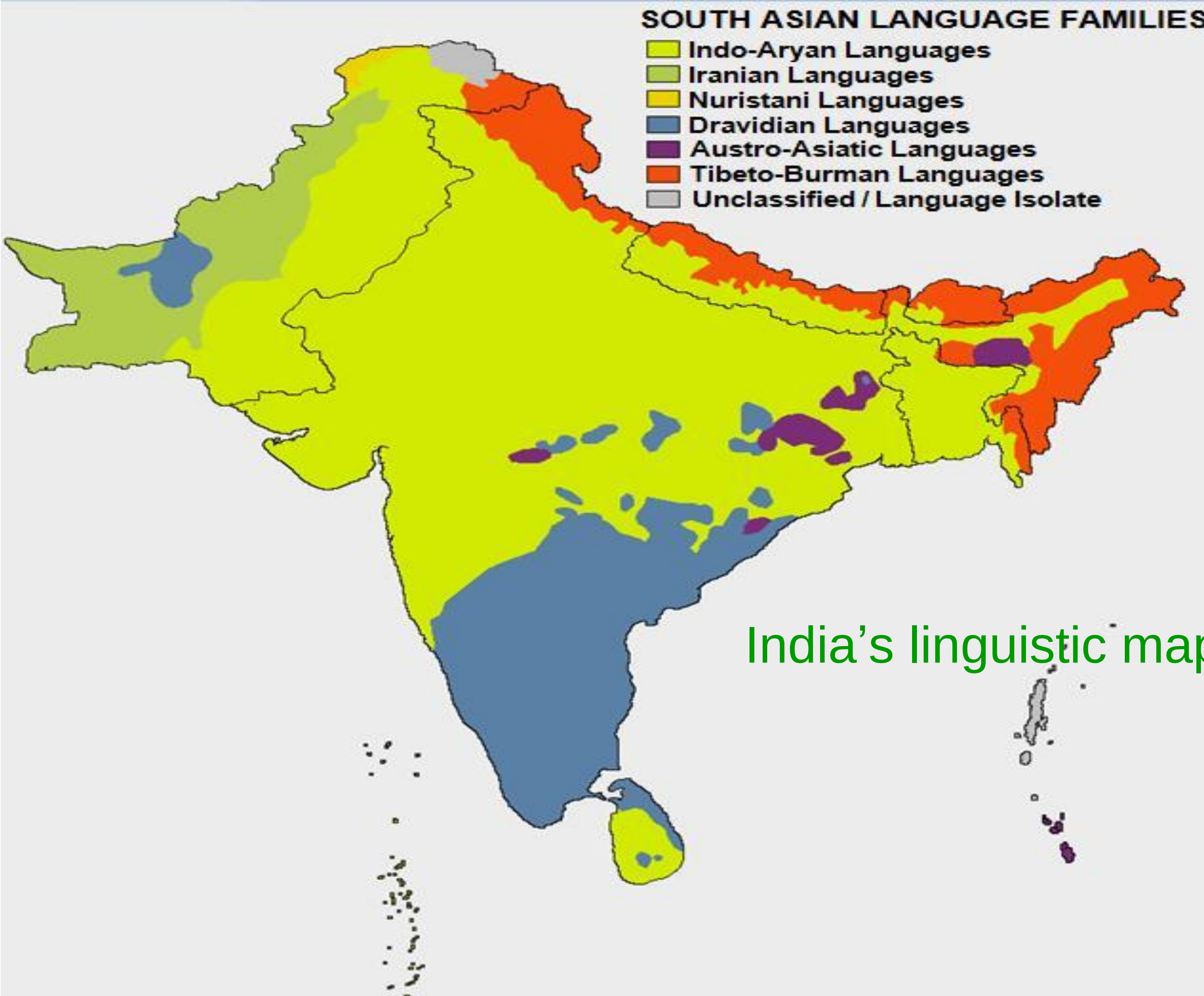
One of the indications that languages descended from a single source

Word order based

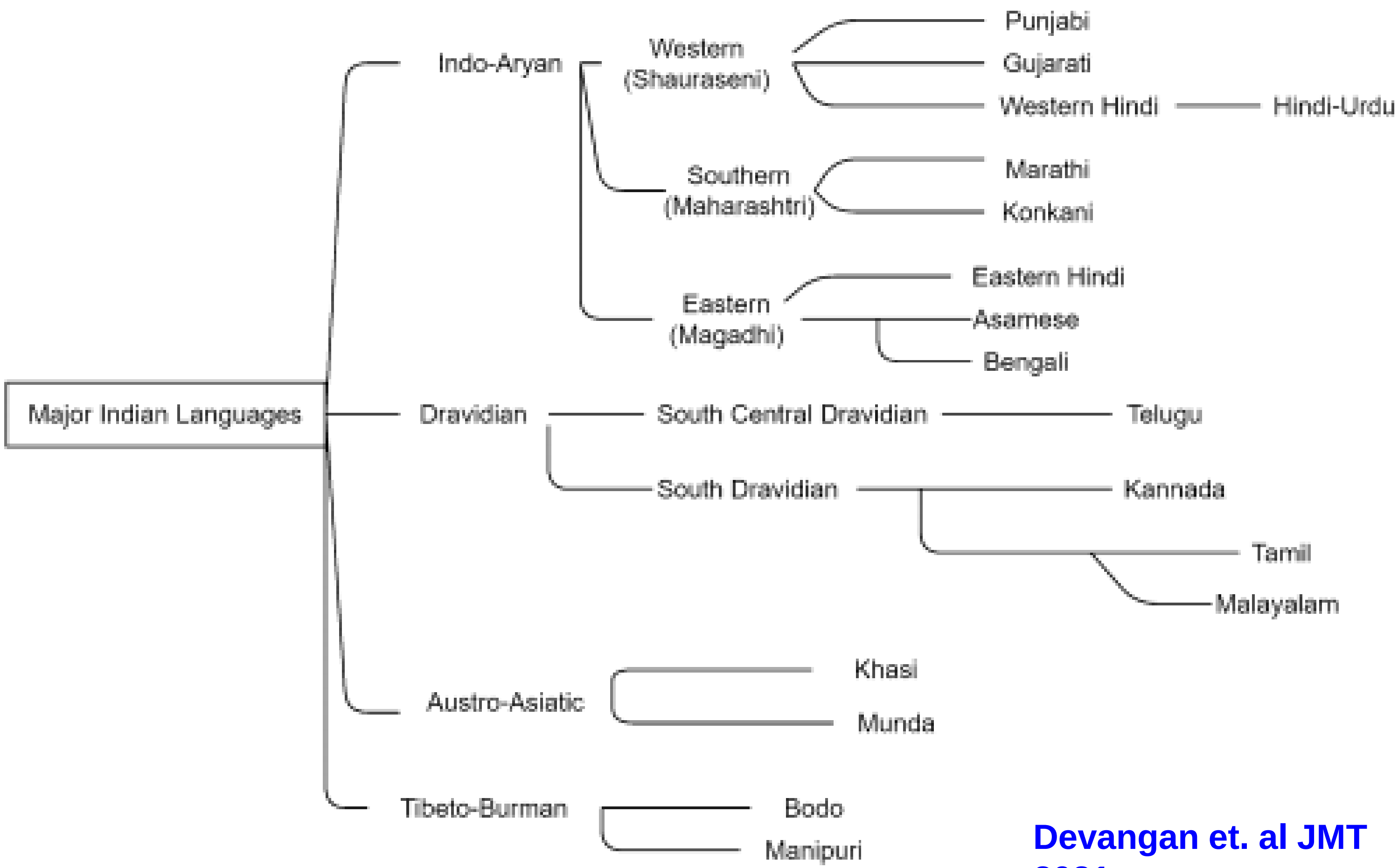
- Object–subject–verb (OSV)
- Object–verb–subject (OVS)
- Subject–verb–object (SVO): English
- Subject–object–verb (SOV): Most Indian Languages
- Verb–subject–object (VSO)
- Verb–object–subject (VOS)

Dominant Word Order Distribution Across Languages (Wikipedia)

Type	Languages	%	Families	%
SOV (Hindi)	2,275	43,3%	239	65.3%
SVO (English)	2,117	40.3%	55	15%
VSO (tagalog in phillipines)	503	9.5%	27	7.4%
VOS (Malagasy in Madagaskar)	174	3.3%	15	4.1%
NODOM (Sanskrit)	124	2.3%	26	7.1%
OVS (Korean and Japanese, many times)	40	0.7%	3	0.8%
OSV (Warao in Venezuela)	19	0.3%	1	0.3%



India’s linguistic map



Devangan et. al JMT 2021

Fig. 1 Tree diagram to illustrate the language closeness of major Indian languages

Languages in the virtual world (distribution in wiki)

Language	Language (local)	Wiki	Articles	Total pages	Edits	Admins	Users	Active users	Images	Depth
English	English	en	**6,410,964**	54,656,272	1,051,463,707	1,073	42,581,314	125,590	896,934	1,089
German	Deutsch	de	**2,633,864**	7,283,550	215,991,122	188	3,812,385	18,226	129,654	92
French	français	fr	**2,375,292**	11,619,832	187,687,175	160	4,235,175	18,926	65,803	244
Spanish	español	es	**1,732,256**	7,542,247	139,381,422	65	6,387,142	15,684	0	207
Japanese	日本語	ja	**1,301,239**	3,832,919	86,415,483	40	1,860,415	14,924	37,243	85
Chinese	中文	zh	**1,241,883**	6,814,699	68,427,675	65	3,156,759	8,120	57,853	202
Arabic	العربية	ar	**1,143,719**	7,520,121	55,818,627	27	2,168,623	7,936	47,915	230

Language wise distribution of wiki pages (ILS)

Language	Language (local)	Wiki	Articles	Total pages	Edits	Admins	Users	Active users	Images	Depth
Urdu	اردو	ur	**165,859**	**955,369**	4,788,257	11	137,648	253	11,553	113
Hindi	हिन्दी	hi	**150,620**	**1,141,451**	5,376,777	7	662,647	1,132	3,493	203
Tamil	தமிழ்	ta	**142,394**	**448,258**	3,310,745	34	197,483	298	8,048	34
Marathi	मराठी	mr	**81,812**	**271,480**	1,968,126	10	137,989	186	19,142	38
Malayalam	മലയാളം	ml	**76,428**	**473,218**	3,630,635	15	155,617	289	6,752	206
Telugu	తెలుగు	te	**73,722**	**295,532**	3,388,038	13	108,941	170	15,744	103
Nepali	नेपाली	ne	**32,051**	**101,538**	1,039,004	7	55,209	95	1,256	48
Gujarati	ગુજરાતી	gu	**29,750**	**112,438**	800,940	3	65,032	68	0	55
Kannada	ಕನ್ನಡ	kn	**27,457**	**126,761**	1,081,877	4	71,656	124	2,474	111
Odia	ଓଡ଼ିଆ	or	**15,907**	**71,467**	436,189	5	29,485	60	125	