BBN's Systems for the Chinese-English Sub-task of the NTCIR-10 PatentMT Evaluation

Zhongqiang Huang, Jacob Devlin, Spyros Matsoukas, Rich Schwartz
{zhuang, jdevlin, smatsouk, schwartz}@bbn.com

Speech, Language, and Multimedia
Raytheon BBN Technologies
Cambridge, MA, U.S.A.

Overview

Raytheon BBN Technologies

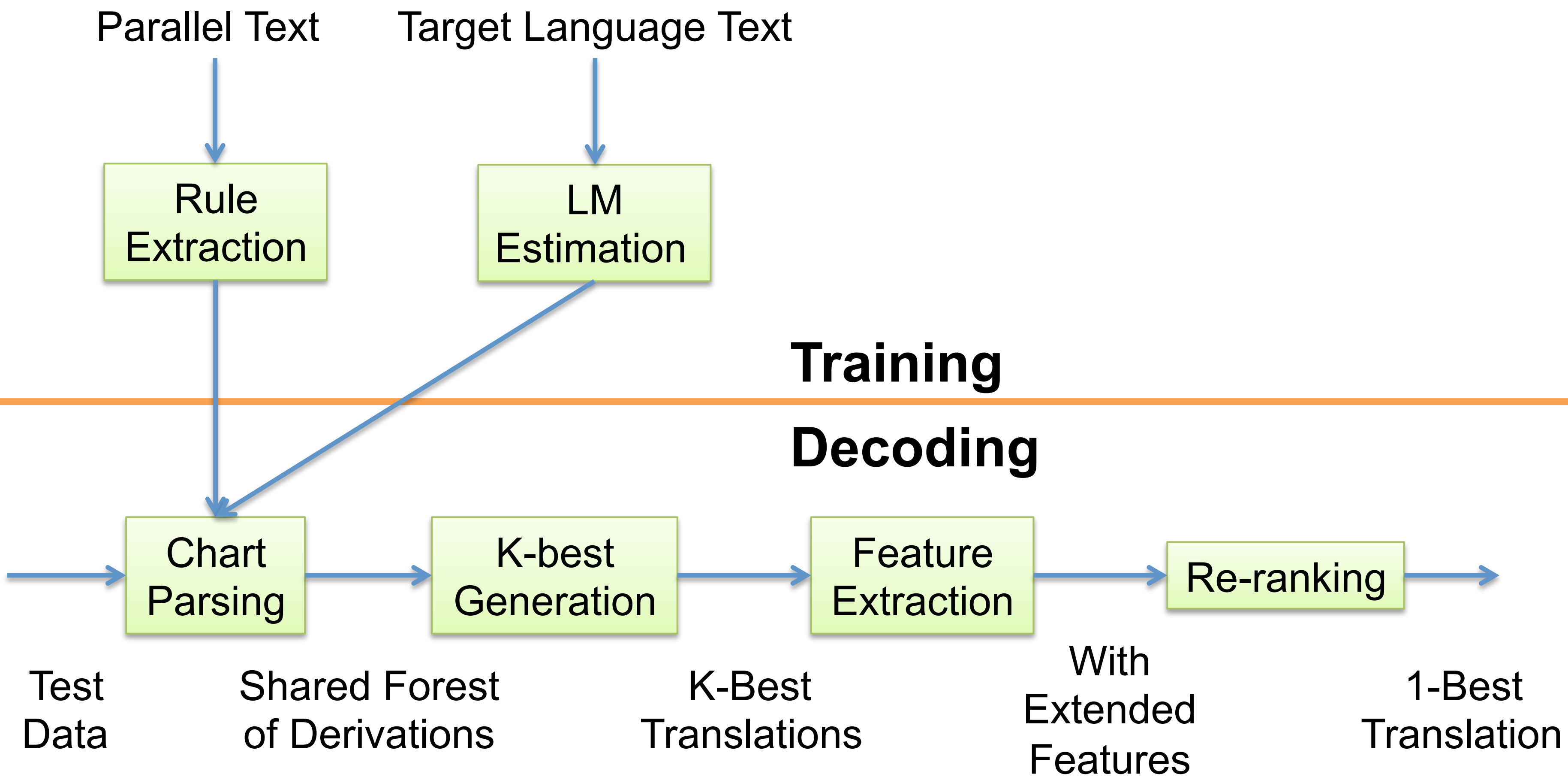
- Statistical machine translation framework
- Building patent machine translation systems
- Official evaluation results
- Summary

Raytheon
BBN Technologies

Part I:
Statistical Machine Translation Framework

Statistical Machine Translation (MT) Framework

Raytheon
BBN Technologies



String-to-Dependency Translation Model

Raytheon
BBN Technologies

- Modified version of Chiang's Hiero algorithm
- Extract hierarchical rules with well-formed dependencies on the target side
  - Well-formed dependency structure:
    - Single rooted tree, with each child being a complete sub-tree
    - Sequence of siblings, each being a complete sub-tree
  - Use POS tag of head word as non-terminal labels on the target side

X : X_1 出发 去 X_2 → VB : NR_1 leaves for NN_2

- Extract all phrasal rules, ignoring dependency
- Features:
  - 10+ core features
  - ~50K sparse binary features

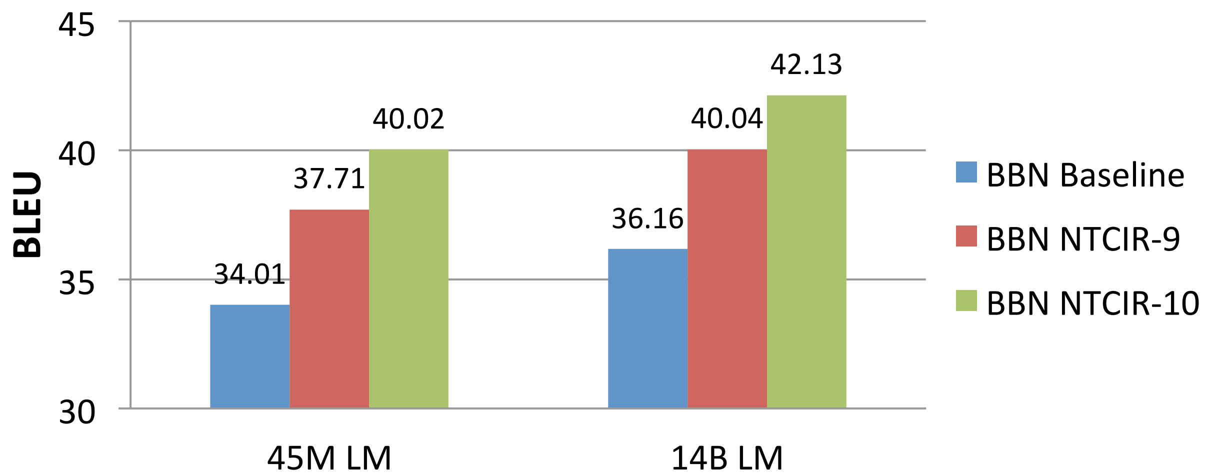
Raytheon
BBN Technologies

Part II:
Building Patent Machine Translation Systems

BBN Patent MT systems - Overview

Raytheon
BBN Technologies

- Data released by the NTCIR-10 organizers
  - Parallel data: 45M words of Chinese-English sentence pairs
  - Extra LM data: 14B words of US patents in English
  - Development data: 2K Chinese-English sentence pairs, split into tuning and test set
- Model training
  - Translation Model: trained on the 45M parallel corpus
  - Language Models:
    - 45M LM: trained on the target side of the 45M parallel corpus
    - 14B LM: trained on the 45M words plus the 14B US patent words
- Summary of results on the test set (development)



Review of Work for BBN NTCIR-9

Raytheon BBN Technologies

- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”

Review of Work for BBN NTCIR-9

Raytheon
BBN Technologies

- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”
- Special token sharing
  - Replace special tokens with a common token for each type in translation and language model
    - Numbers: e.g., 2,596, -123.321
    - Patent IDs: e.g., No.5,400,788, No. 5,405,753
    - Math expressions: e.g., p=0.004, Sine(45)=0.7071
    - Material names: e.g., C15H23N2O5P, LiEt3BH
    - Labeled names: e.g., 3.05kg, 200ml

Review of Work for BBN NTCIR-9

Raytheon
BBN Technologies

- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”
- Special token sharing
  - Replace special tokens with a common token for each type in translation and language model
    - Numbers: e.g., 2,596, -123.321
    - Patent IDs: e.g., No.5,400,788, No. 5,405,753
    - Math expressions: e.g., p=0.004, Sine(45)=0.7071
    - Material names: e.g., C15H23N2O5P, LiEt3BH
    - Labeled names: e.g., 3.05kg, 200ml
- Patent case-LM
  - Re-trained on the 45M LM data

Review of Work for BBN NTCIR-9

- Consistent tokenization
    - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”
- Special token sharing
    - Replace special tokens with a common token for each type in translation and language model
        - Numbers: e.g., 2,596, -123.321
        - Patent IDs: e.g., No.5,400,788, No. 5,405,753
        - Math expressions: e.g., p=0.004, Sine(45)=0.7071
        - Material names: e.g., C15H23N2O5P, LiEt3BH
        - Labeled names: e.g., 3.05kg, 200ml
- Patent case-LM
    - Re-trained on the 45M LM data
- Optimized word segmenter
    - Re-optimized for patent translation

Review of Work for BBN NTCIR-9

Raytheon BBN Technologies

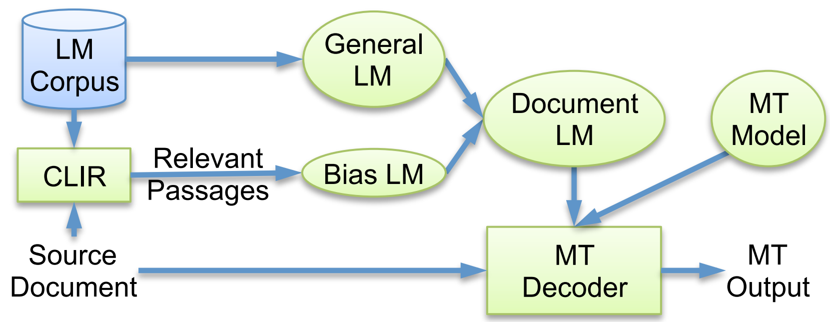
- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., "IS-1000" vs. "IS – 1000"
- Special token sharing
  - Replace special tokens with a common token for each type in translation and language model
    - Numbers: e.g., 2,596, -123.321
    - Patent IDs: e.g., No.5,400,788, No. 5,405,753
    - Math expressions: e.g., p=0.004, Sine(45)=0.7071
    - Material names: e.g., C15H23N2O5P, LiEt3BH
    - Labeled names: e.g., 3.05kg, 200ml
- Patent case-LM
  - Re-trained on the 45M LM data
- Optimized word segmenter
  - Re-optimized for patent translation
- Top 100 sparse features
  - Due to the smaller tuning set, we use only the top 100 features of the highest weights in each category of the 50K sparse features

Review of Work for BBN NTCIR-9

Raytheon BBN Technologies

- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”
- Special token sharing
  - Replace special tokens with a common token for each type in translation and language model
    - Numbers: e.g., 2,596, -123.321
    - Patent IDs: e.g., No.5,400,788, No. 5,405,753
    - Math expressions: e.g., p=0.004, Sine(45)=0.7071
    - Material names: e.g., C15H23N2O5P, LiEt3BH
    - Labeled names: e.g., 3.05kg, 200ml
- Patent case-LM
  - Re-trained on the 45M LM data
- Optimized word segmenter
  - Re-optimized for patent translation
- Top 100 sparse features
  - Due to the smaller tuning set, we use only the top 100 features of the highest weights in each category of the 50K sparse features

- Document-level LM adaptation
  - Find documents in monolingual English patent corpus that are similar to test document
  - Estimate a separate LM and interpolate with the general LM



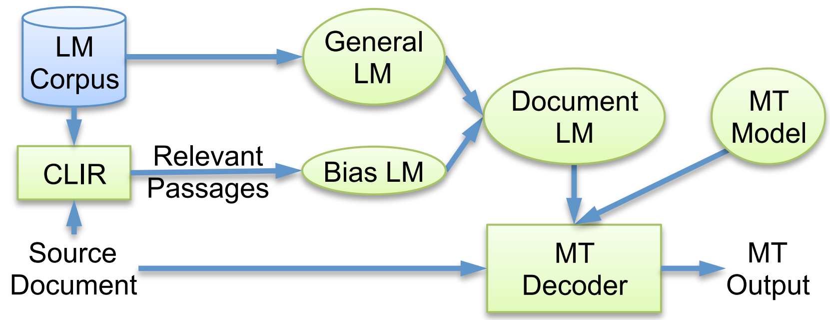
$$P_{LM}(s) = (1-\alpha)P_{generalLM}(s) + \alpha P_{biasLM}(s)$$

Review of Work for BBN NTCIR-9

Raytheon BBN Technologies

- Consistent tokenization
  - Fixed inconsistent tokenization of ASCII strings in the source and target sides, e.g., “IS-1000” vs. “IS – 1000”
- Special token sharing
  - Replace special tokens with a common token for each type in translation and language model
    - Numbers: e.g., 2,596, -123.321
    - Patent IDs: e.g., No.5,400,788, No. 5,405,753
    - Math expressions: e.g., p=0.004, Sine(45)=0.7071
    - Material names: e.g., C15H23N2O5P, LiEt3BH
    - Labeled names: e.g., 3.05kg, 200ml
- Patent case-LM
  - Re-trained on the 45M LM data
- Optimized word segmenter
  - Re-optimized for patent translation
- Top 100 sparse features
  - Due to the smaller tuning set, we use only the top 100 features of the highest weights in each category of the 50K sparse features

- Document-level LM adaptation
  - Find documents in monolingual English patent corpus that are similar to test document
  - Estimate a separate LM and interpolate with the general LM



$$P_{LM}(s) = (1-\alpha)P_{generalLM}(s) + \alpha P_{biasLM}(s)$$

System	BLEU
BBN Baseline with 45M LM	**34.01**
+ consistent tokenization	34.56
+ more token sharing	34.97
+ patent case-LM	36.47
+ optimized word segmenter	36.95
+ top 100 features	**37.71**
+ 14B LM	39.14
+ document-level LM adaptation	**40.04**

Development for NTCIR-10 Evaluation

Raytheon
BBN Technologies

- Overview
  - Miscellaneous additional features
  - Sentence-level LM adaptation
  - Robust context dependent translation
  - Recurrent neural network LM
  - Translation-based caser

Miscellaneous Additional Features

Raytheon BBN Technologies

- Bigram lexical translation model
  - Extension of context-based lexical probabilities to model joint likelihood of target bigrams given source context

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Apply chain rule and use simple back-off smoothing
  - Similarly for the backward direction

Miscellaneous Additional Features

Raytheon
BBN Technologies

- Bigram lexical translation model
  - Extension of context-based lexical probabilities to model joint likelihood of target bigrams given source context

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Apply chain rule and use simple back-off smoothing
  - Similarly for the backward direction
- Trait features, e.g.,
  - Percent of NULL source content words
  - Percent of words that re-order
  - Percent of low-frequency n-grams
  - Source-to-target length ratio

Miscellaneous Additional Features

Raytheon BBN Technologies

- Bigram lexical translation model
  - Extension of context-based lexical probabilities to model joint likelihood of target bigrams given source context

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Apply chain rule and use simple back-off smoothing
  - Similarly for the backward direction
- Trait features, e.g.,
  - Percent of NULL source content words
  - Percent of words that re-order
  - Percent of low-frequency n-grams
  - Source-to-target length ratio
- Disable feature normalization

Miscellaneous Additional Features

Raytheon BBN Technologies

- Bigram lexical translation model
  - Extension of context-based lexical probabilities to model joint likelihood of target bigrams given source context

$$P\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Apply chain rule and use simple back-off smoothing
  - Similarly for the backward direction
- Trait features, e.g.,
  - Percent of NULL source content words
  - Percent of words that re-order
  - Percent of low-frequency n-grams
  - Source-to-target length ratio
- Disable feature normalization

System	Test
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	**38.06**
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	**39.51**

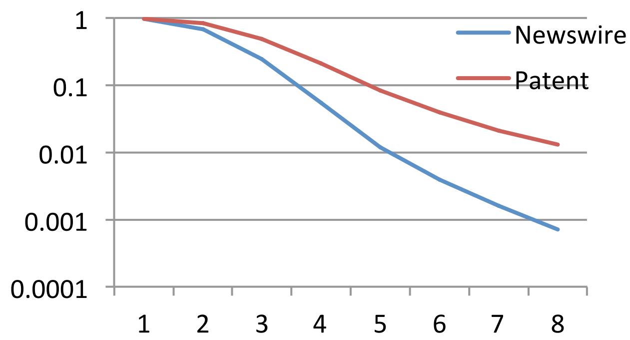
Sentence-level LM adaptation

- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents

Sentence-level LM adaptation

Raytheon
BBN Technologies

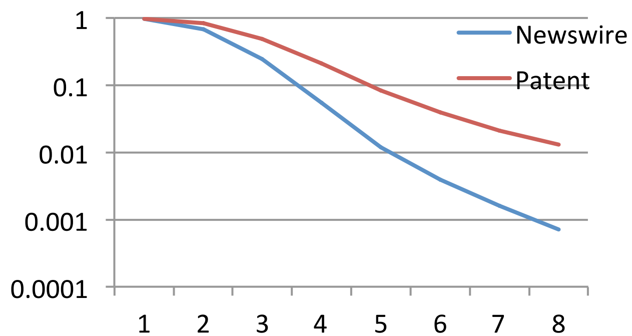
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



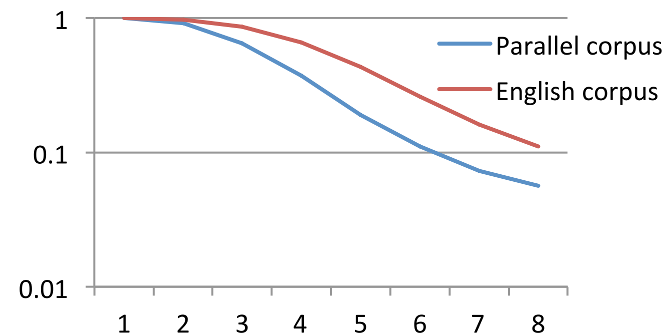
Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)

Sentence-level LM adaptation

- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)

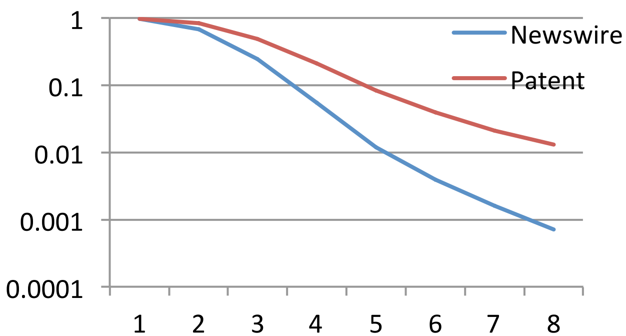


Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus

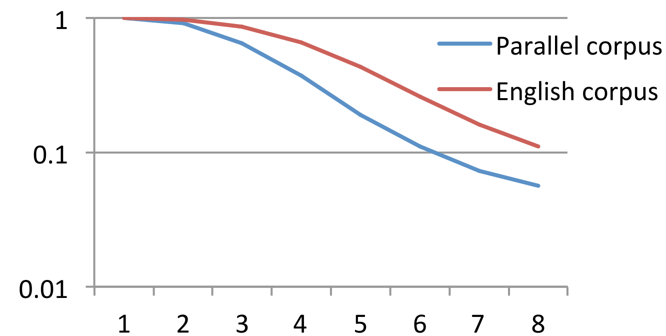
Sentence-level LM adaptation

Raytheon BBN Technologies

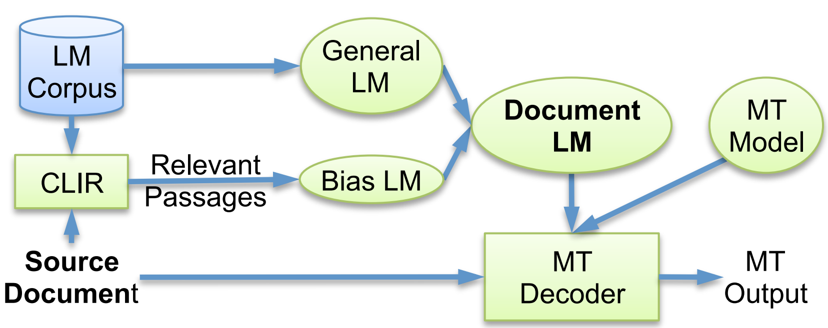
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)



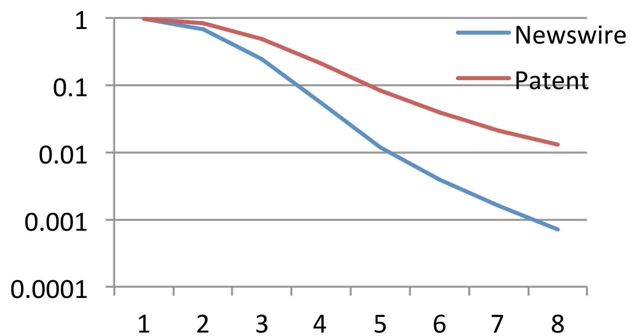
Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus



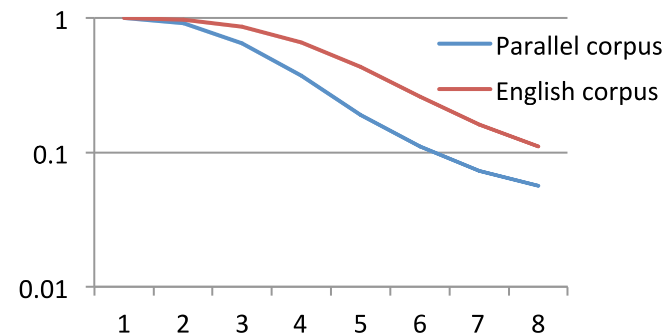
Sentence-level LM adaptation

Raytheon BBN Technologies

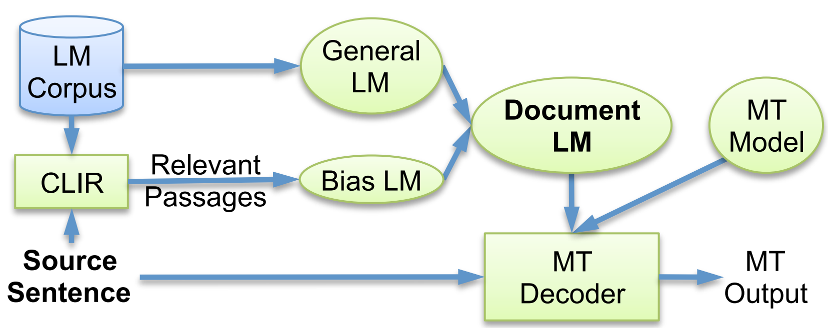
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)



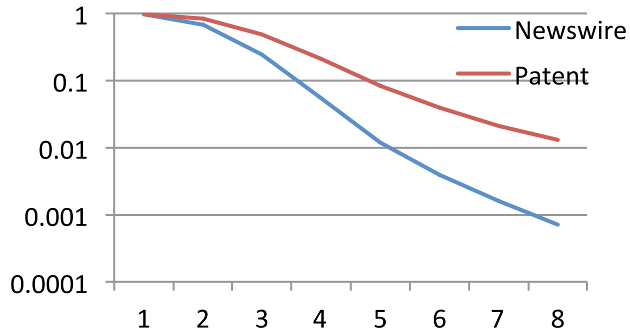
Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus



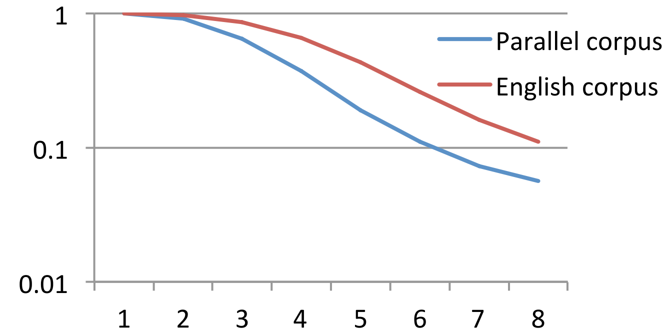
Sentence-level LM adaptation

Raytheon
BBN Technologies

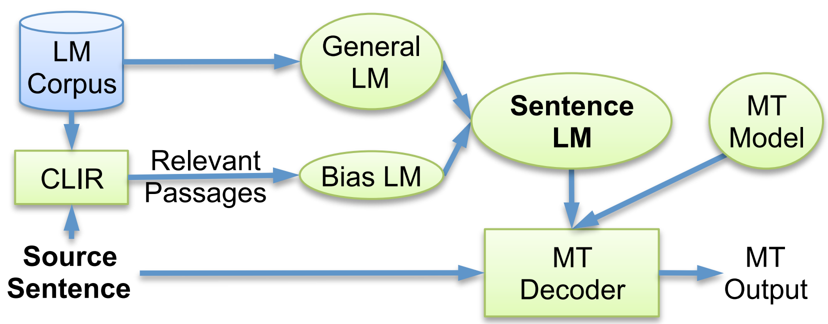
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)



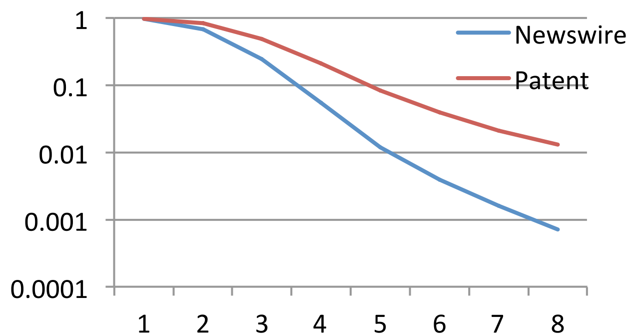
Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus



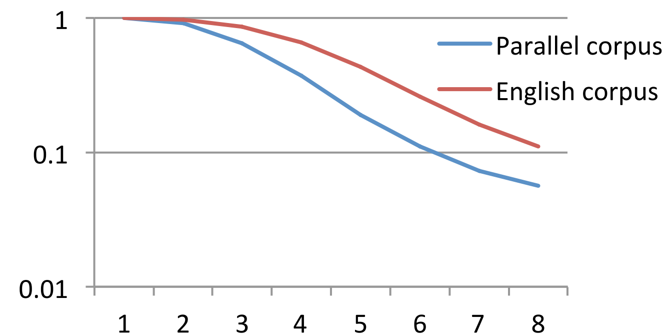
Sentence-level LM adaptation

Raytheon BBN Technologies

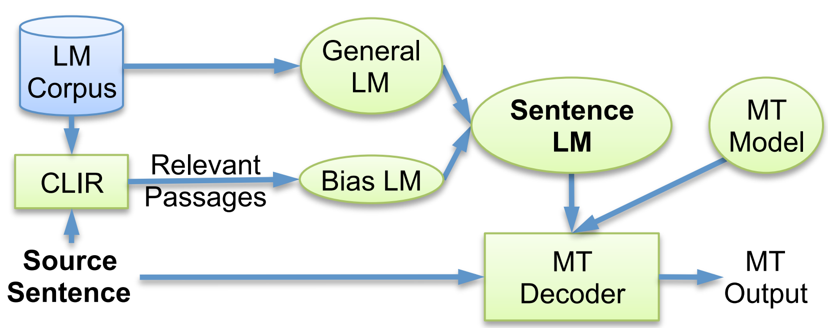
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)



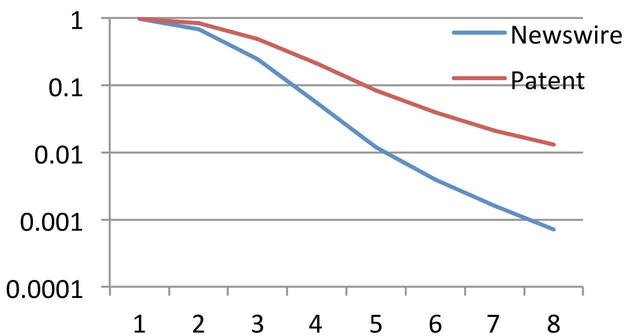
Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus



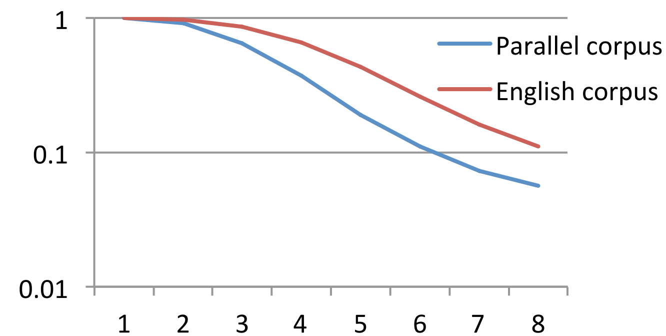
System	**Test**
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	38.06
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	39.51
+ document-level LM adaptation	**39.94**

Sentence-level LM adaptation

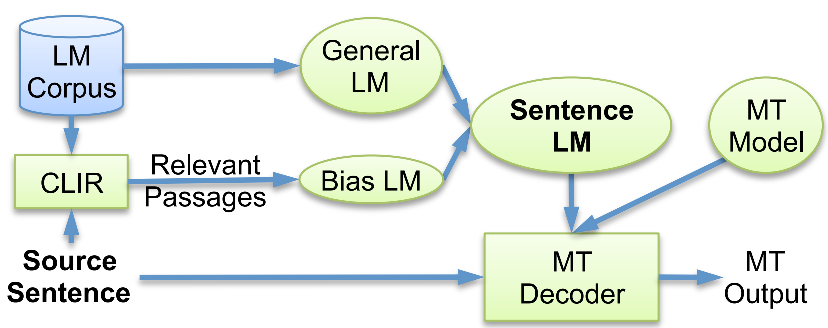
- Patent documents tend to use well-structured sentence and re-use n-grams in other patent documents



Percentage of source n-grams (tokens) in the test sentences that are observed in the parallel training for newswire (GALE) and patent (NTCIR-10)



Percentage of target n-grams (tokens) in the patent test sentences that are also observed in the patent parallel corpus and the monolingual English patent corpus



System	Test
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	38.06
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	39.51
~~+ document-level LM adaptation~~	~~39.94~~
+ sentence-level LM adaptation	**40.95**

Robust Context-Dependent Modeling

Raytheon
BBN Technologies

- High order context-dependent translation models may be very sparse

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- High order context-dependent translation models may be very sparse

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

- Common solution
  - First, apply the chain rule

$$\mathrm{P}\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = \mathrm{P}\left(t_{s_i} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Back-off each probability independent

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- High order context-dependent translation models may be very sparse

$$P\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

- Common solution
  - First, apply the chain rule

$$P\left(t_{s_i}, t_{s_{i-1}} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = P\left(t_{s_i} \mid s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right)$$

  - Back-off each probability independent
- But, unlike LM, there is no clear back-off ordering
  - Is $P\left(t_{s_{i-1}} \mid t_{s_i}, s_{i-1}\right)$ "better" than $P\left(t_{s_{i-1}} \mid s_i, s_{i-1}\right)$?

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- Our solution: interpolate all possible back-off components
  - Sparse context types can be added independently of one another

$$\mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = \omega_0 \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) + \omega_1 \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}\right) + \cdots + \omega_{30} \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}\right)$$

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- Our solution: interpolate all possible back-off components
  - Sparse context types can be added independently of one another

$$\mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = \omega_0 \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) + \omega_1 \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}\right) + \cdots + \omega_{30} \mathrm{P}\left(t_{s_{i-1}} \mid t_{s_i}\right)$$

- Each weight ω is a function of the marginal count

$$\omega_j \mathrm{P}\left(t_{s_i} \mid s_i, s_{i-1}\right) = \frac{1}{Z} \alpha_j \log\left(C\left(s_i, s_{i-1}\right)\right) \frac{C\left(t_{s_i}, s_i, s_{i-1}\right)}{C\left(s_i, s_{i-1}\right)}$$

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- Our solution: interpolate all possible back-off components
  - Sparse context types can be added independently of one another

$$P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = \omega_0 P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) + \omega_1 P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}\right) + \cdots + \omega_{30} P\left(t_{s_{i-1}} \mid t_{s_i}\right)$$

- Each weight ω is a function of the marginal count

$$\omega_j P\left(t_{s_i} \mid s_i, s_{i-1}\right) = \frac{1}{Z} \alpha_j \log\left(C\left(s_i, s_{i-1}\right)\right) \frac{C\left(t_{s_i}, s_i, s_{i-1}\right)}{C\left(s_i, s_{i-1}\right)}$$

- Weights α are optimized to maximize likelihood on a held-out set
  - Least useful components are thrown out for efficiency

Robust Context-Dependent Modeling

Raytheon BBN Technologies

- Our solution: interpolate all possible back-off components
  - Sparse context types can be added independently of one another

$$P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) = \omega_0 P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}, s_{i-2}\right) + \omega_1 P\left(t_{s_{i-1}} \mid t_{s_i}, s_i, s_{i-1}, s_{i+1}\right) + \cdots + \omega_{30} P\left(t_{s_{i-1}} \mid t_{s_i}\right)$$

- Each weight ω is a function of the marginal count

$$\omega_j P\left(t_{s_i} \mid s_i, s_{i-1}\right) = \frac{1}{Z} \alpha_j \log\left(C\left(s_i, s_{i-1}\right)\right) \frac{C\left(t_{s_i}, s_i, s_{i-1}\right)}{C\left(s_i, s_{i-1}\right)}$$

- Weights α are optimized to maximize likelihood on a held-out set
  - Least useful components are thrown out for efficiency

System	Test
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	38.06
+ robust context dependent translation	**38.72**
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	39.51
+ sentence-level LM adaptation	40.95
+ robust context dependent translation	**41.09**

Neural Net LM

Raytheon
BBN Technologies

- Trained a recurrent neural net LM for rescoring
  - Mikolov's toolkit: http://www.fit.vutbr.cz/~imikolov/rnnlm/
  - Interpolated with 5-gram KN Smoothing LM

Neural Net LM

Raytheon BBN Technologies

- Trained a recurrent neural net LM for rescoring
  - Mikolov's toolkit: http://www.fit.vutbr.cz/~imikolov/rnnlm/
  - Interpolated with 5-gram KN Smoothing LM

System	Test
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	38.06
+ robust context dependent translation	38.72
+ recurrent neural network LM	**39.35**
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	39.51
+ document-level LM adaptation	39.94
+ sentence-level LM adaptation	40.95
+ robust context dependent translation	41.09
+ recurrent neural network LM	**41.43**

Translation-based Caser

- Treats casing as a translation problem
  - Similar to (Hassan, et al. 2006)'s MaTrEx system
  - Trained on 45M LM training data
  - Use rule probabilities, case LM probability, and sparse features, e.g., *Is the target word upper cased and does it follow a period? Is the target word upper cased and a proper noun?*

Translation-based Caser

Raytheon
BBN Technologies

- Treats casing as a translation problem
  - Similar to (Hassan, et al. 2006)'s MaTrEx system
  - Trained on 45M LM training data
  - Use rule probabilities, case LM probability, and sparse features, e.g., *Is the target word upper cased and does it follow a period? Is the target word upper cased and a proper noun?*

System	Test
NTCIR-9 system with 45M LM	37.71
+ miscellaneous features	38.06
+ robust context dependent translation	38.72
+ recurrent neural network LM	39.35
+ translation-based caser	**40.02**
NTCIR-9 system with 14B LM	39.14
+ miscellaneous features	39.51
+ document-level LM adaptation	39.94
+ sentence-level LM adaptation	40.95
+ robust context dependent translation	41.09
+ recurrent neural network LM	41.43
+ translation-based caser	**42.13**

Part III:
Official Evaluation Results

Official Automatic (BLEU) Results

Raytheon
BBN Technologies

- The two BBN systems
  - BBN-1 : the primary system, trained on 45M parallel corpus plus 14B English patent corpus
  - BBN-2: the secondary system, trained on 45M parallel corpus only
- NCTIR Official Baseline systems
  - Baseline1– Moses phrase-based hierarchical SMT system
  - Baseline2– Moses phrase-based SMT system

Official Automatic (BLEU) Results

Raytheon BBN Technologies

- The two BBN systems
  - BBN-1 : the primary system, trained on 45M parallel corpus plus 14B English patent corpus
  - BBN-2: the secondary system, trained on 45M parallel corpus only
- NCTIR Official Baseline systems
  - Baseline1– Moses phrase-based hierarchical SMT system
  - Baseline2– Moses phrase-based SMT system

System	Intrinsic evaluation	Chronological evaluation	Multilingual evaluation
BBN-1	42.68	39.44 ⟶ 41.09	27.62
BBN-2	39.98	36.69 ⟶ 38.93	N/A
Baseline1	32.52	30.74	17.96
Baseline2	31.34	29.34	18.05

* ⟶ indicates the change in BLEU from NTCIR-9 evaluation to NTCIR-10 evaluation

Official Manual Evaluation Results

Raytheon BBN Technologies

- Adequacy: scores from 5 (best) to 1 (worst)

System	Average adequacy	Allocation of scores				
		5	4	3	2	1
BBN-1	42.68	156	66	44	34	0
Baseline1	32.52	46	73	91	84	6
Baseline2	31.34	38	34	75	141	12

Official Manual Evaluation Results

Raytheon
BBN Technologies

- Adequacy: scores from 5 (best) to 1 (worst)

System	Average adequacy	Allocation of scores				
		5	4	3	2	1
BBN-1	42.68	156	66	44	34	0
Baseline1	32.52	46	73	91	84	6
Baseline2	31.34	38	34	75	141	12

- Acceptability: scores in AA (best), A, B, C, and F (worst)
- Pairwise acceptability: percentage of wins and ties when comparing acceptability score with other submissions

System	Pairwise score	Allocation of scores				
		AA	A	B	C	F
BBN-1	0.69	81	36	50	35	98

Official Manual Evaluation Results

Raytheon BBN Technologies

- Adequacy: scores from 5 (best) to 1 (worst)

System	Average adequacy	Allocation of scores 5	4	3	2	1
BBN-1	42.68	156	66	44	34	0
Baseline1	32.52	46	73	91	84	6
Baseline2	31.34	38	34	75	141	12

- Acceptability: scores in AA (best), A, B, C, and F (worst)
- Pairwise acceptability: percentage of wins and ties when comparing acceptability score with other submissions

System	Pairwise score	Allocation of scores AA	A	B	C	F
BBN-1	0.69	81	36	50	35	98

- Patent examination evaluation: scores in S (perfect), A, B, C, D, and F (worst)

System	Allocation of scores S	A	B	C	D	F
BBN-1	6	19.5	3.5	0	0	0

Translation Examples

Raytheon
BBN Technologies

Source: 对于每一像素，着色引擎210使用在以上等式(2)-(4)中陈述的边等式来确定所述像素是否在三角形中。

MT output: For each pixel, the rendering engine 210 uses the edge equation set forth in equations (2) to (4) above to determine whether the pixels in a triangle.

Reference: For each pixel, the shading engine 210 determines whether the pixel is in the triangle using the edge equations set forth in equations (2) - (4) above.

Source: 上述说明书全面描述了根据本发明原理的改进型可穿透膜片的成分、制造和用途。

MT output: The above description fully describes the composition, manufacture and use of improved penetrable diaphragm in accordance with the principles of the present invention.

Reference: The above specification provides a complete description of the composition, manufacture and use of the improved penetrable membrane in accordance with the principles of the present invention.

Summary

Raytheon
BBN Technologies

- It was relatively straightforward to port BBN's MT system to work on patents
  - 4-5 weeks of efforts in NTCIR-9 evaluation
  - 3-4 weeks of efforts in NTCIR-10 evaluation
  - All techniques initially developed for other domains work well on patents
- Special attention to patents helps
  - Better tokenization, special token sharing, optimizing word segmentation
  - Sentence-level LM adaptation
  - Further improvement is possible by exploring special properties of patents
- Lots of potential
  - Patents are easier to translate
  - State-of-the-art accuracies in both automatic and manual evaluations
  - Helpful in real patent examination and possibly other tasks

Related MT Research at BBN

Raytheon
BBN Technologies

Leading performer in DARPA's MT programs

- Text-to-text translation (GALE, BOLT)
  - Arabic and Chinese to English. newswire, weblogs, web forums, SMS/chat
- Speech-to-text translation (GALE)
  - Arabic and Chinese to English. broadcast news and broadcast conversation
- Speech-to-speech translation (TransTac, BOLT)
  - English to/from Iraqi Arabic, Farsi, Dari, Pashto, Malay, and Spanish
  - TransTalk: portable (Android), two-way translation device; deployed by US Army
- Image to text translation (MADCAT)
  - Foreign text (Arabic, Chinese and Korean) in images (through OCR) to English
- Multilingual broadcast/web monitoring
  - Continuous searchable archive of international television broadcasts and web sites
  - Automatic translation to English for deep analysis

Contact: schwartz@bbn.com