

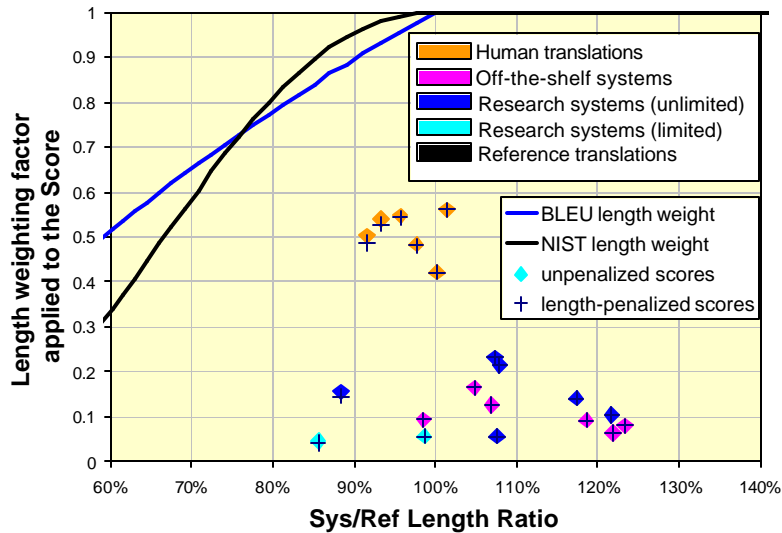
Automatic Evaluation of Language Translation using N-gram Co-occurrence Statistics

- MTEval – the NIST utility
- Results on the 2001 dry run
- Comparison with human judgments

Primary Differences between the NIST utility and BLEU

- Text preprocessing
 - Primarily reduced segmentation of numbers. For example: “**\$1,234.56**” remains one word rather than being split into six words (**\$** – **1** – **,** – **234** – **.** – **56**)
- “Gentler” length penalty
 - Quadratic penalty from mean, rather than nearest neighbor linear penalty.
- Information-weighted N-gram counts
 - Using the reference data to compute N-gram likelihoods
- Selective use of N-grams
 - Trigrams only, rather than a sum over multiple values of N

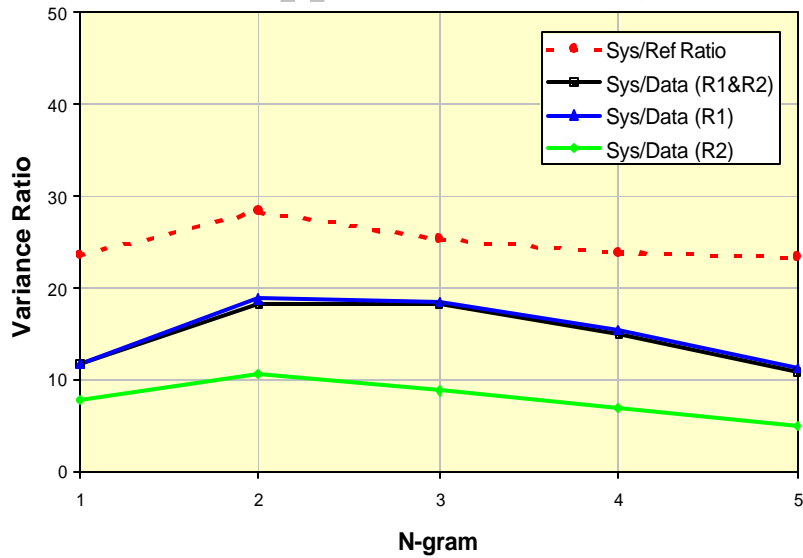
The NIST length penalty compared with BLEU



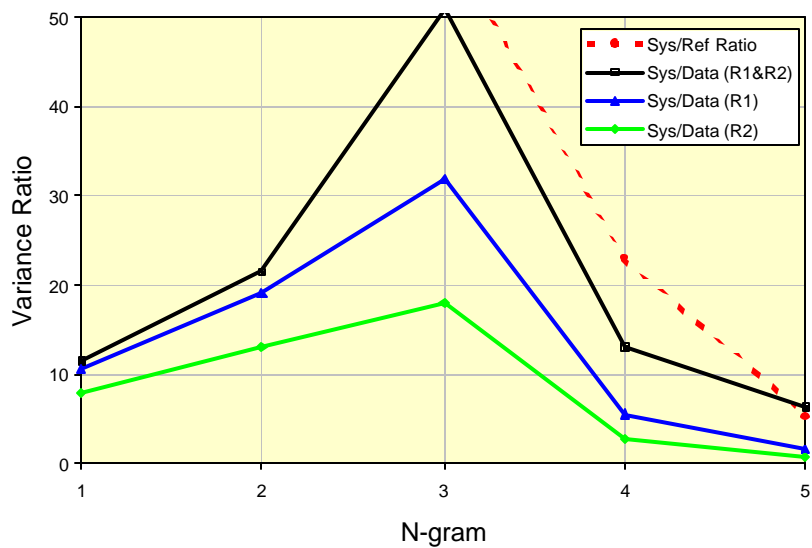
Measuring the Statistical Sensitivity of the Evaluation Tool

- Use the F-ratio of the performance measure:
 - The ratio of between-system variance (desired) to within-system variance (undesired)
- Sources of undesired variance:
 - Variance with changing source data
 - Variance with changing reference data
- Use the F-ratio to optimize performance:
 - Algorithm and feature selection
 - Parameter tuning

Statistical Sensitivity of BLEU (as supplied to NIST)



Statistical Sensitivity of the NIST MT evaluation utility (mteval-v06)



MT2001 Dry Run Source Data

Xinhua		Zaobao	Voice of America
Dev Set 25 documents 206 segments	Eval Set 27 documents 234 segments	Eval Set 27 documents 321 segments	Eval Set 26 documents 232 segments
<div style="text-align: center; font-weight: bold; font-size: 1.2em;">← The Evaluation Data Set →</div>			

Example Reference Translation Xinhua

<DOC doc_ID="chtb_050" sys_ID="12">

-Headline-

Foreign Investment in Jiangsu's Agriculture on the Increase

-Paragraph-

Xinhua News Agency, Nanjing, Dec. 16 (Reporter Zhou Fang) - Foreign investment in Jiangsu's agriculture is on the rise.

At present, there are as many as 800 agricultural projects in Jiangsu that receive investment directly from overseas, with a total amount of over 800 million US dollars.

-Paragraph-

According to officials from the Provincial Department of Agriculture and Forestry of Jiangsu, the "Three-Capital" ventures approved by agencies within the agricultural system of Jiangsu Province since 1994 have numbered more than 500 and have utilized over 700 million US dollars worth of foreign capital, respectively three times and seven times more than in 1993.

Foreign capital from the United States, Japan and Singapore has shown more increase. Foreign investment is taking a bigger and bigger share in new projects while solely foreign-owned enterprises are obviously on the rise.

Example Reference Translation

Zaobao

<DOC doc_ID="ZBN19981002_0117" sys_ID="17">

-Headline-

Due to his report of a woman **was** violated during riot in May, a correspondent was subpoenaed for interrogation by Indonesia police.

-Paragraph-

(Agence Francaise de Press, Jakarta) - An international organization, **who** safeguards freedom of the press, "Journalist Protection Committee" expressed their astonishment on Indonesian police's subpoena and interrogation of a correspondent because his report of **a Chinese origin woman was** violated during riot in May.

-Paragraph-

Journalist Protection Committee, with its office in New York, said that a correspondent named Gunawan who works for **ij**Jakarta-Jakarta*i*, magazine was subpoenaed **to a police headquarter** for interrogation on Thursday.

-Paragraph-

This committee said in one of Gunawan's reports published in July issue, he quoted a speech made by a raped woman.

In a report published in **ij**New York Times*i*, on June 10, the speech of the same victim was also quoted.

Example Reference Translation

Voice of America

<DOC doc_ID="VOM19981003_0800_2179" sys_ID="12">

-Paragraph-

Tang Jiaxuan, foreign minister of China who is visiting the United States, held a joint press conference on the 29th in Washington with Albright, U.S. Secretary of State and made a speech on the issue of the Sino-U.S. relations.

Tang Jiaxuan also mentioned that President Clinton' sex scandal is an internal affair of the United States and China will never forget the contributions President Clinton has made in promoting the Sino-U.S. relations.

The following is a spot coverage by VOA correspondent Zhou Youkang at the conference.

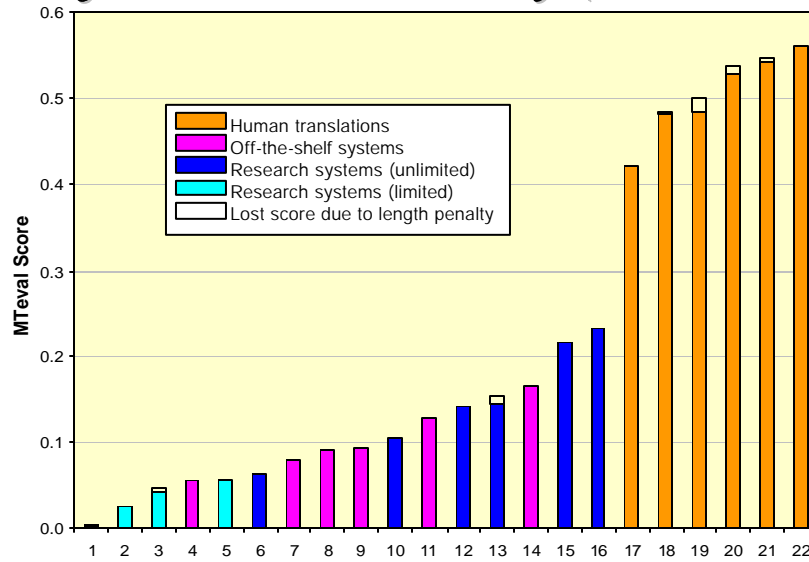
-Paragraph-

Dear listeners, at 11:30 a.m. on the 29th in the White House, **the** U.S. President Clinton met with Tang Jiaxuan, foreign minister of China who is on a visiting tour in Washington.

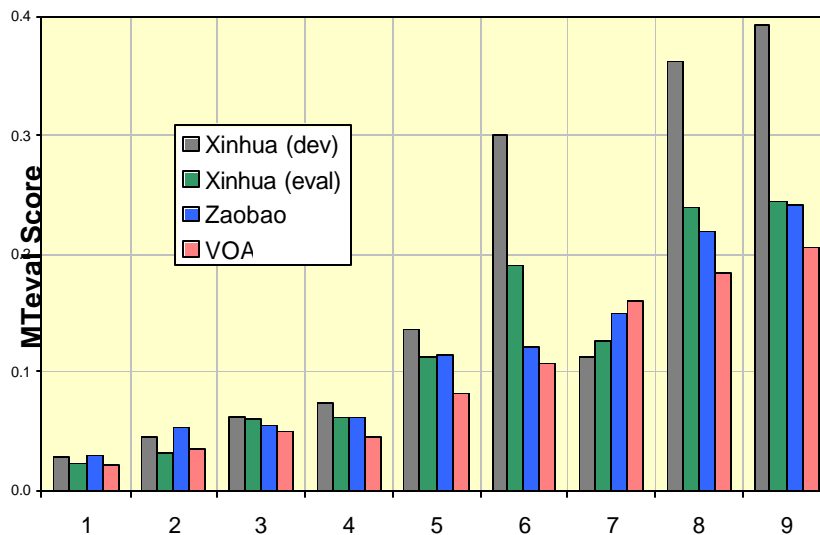
At noon on the 29th, at a joint press conference by the State Department and Foreign Minister Tang Jiaxuan, Albright, U.S. Secretary of State, announced the news that President Clinton met with Tang Jiaxuan.

Albright said:

MT2001 Dry Run Performance Summary (eval data)



MT2001 Dry Run Performance versus Source

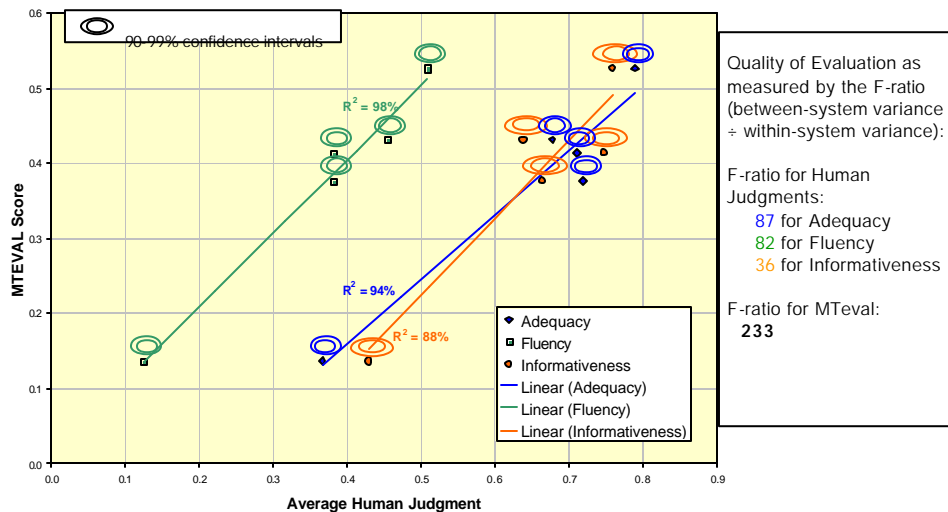


MTEval versus Human Judgment

- An MT evaluation using human judgments was conducted by DARPA in 1994
- The translation tasks were:
 - French-to-English
 - Spanish-to-English
 - Japanese-to-English
- The evaluation measures were:
 - **Adequacy** – Is the information adequately represented?
 - **Fluency** – How fluent is the English translation?
 - **Informativeness** – A question answering exercise.

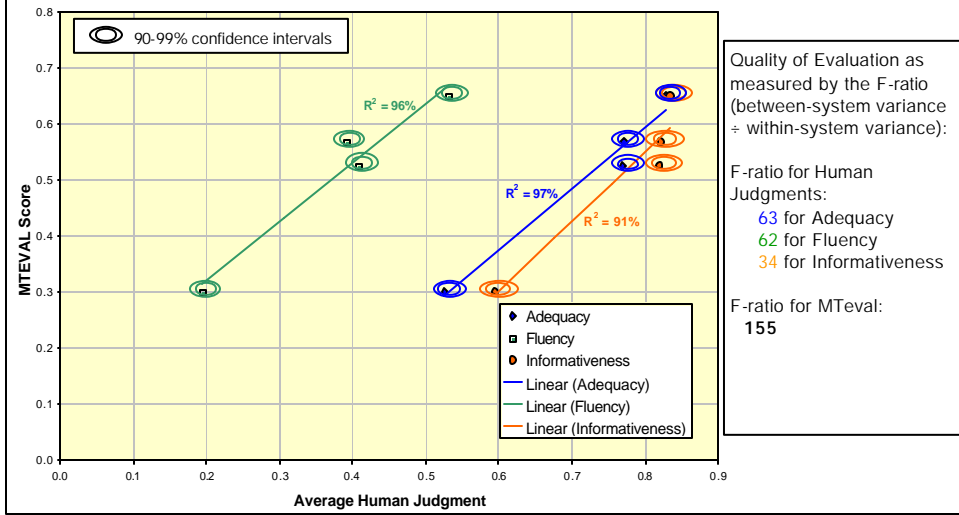
MTEval versus Human Judgment

Correlation between Human Judgments and MTEval scores
on the DARPA-94 **French** Translations



MTEval versus Human Judgment

Correlation between Human Judgments and MTEval scores
on the DARPA-94 Spanish Translations



Correlation between Performance Measures (DARPA-94 – French)

