

---

# Empirical Methods in Natural Language Processing

## Lecture 14

### Machine translation (I): Introduction

Philipp Koehn

21 February 2008



## Machine translation

- Task: make sense of foreign text like

### 毒品

本册子為家長們提供實際和有用的關於毒品的信息，包括如何減少使用非法毒品的危險。它有助於您和您的家人討論有關毒品的問題。這本小册子的主要内容已錄在磁帶上。如果您想索取一盒免費的磁帶(中文)，請在下面的

- One of the oldest problems in Artificial Intelligence
- AI-hard: reasoning and world knowledge required

## The Rosetta stone



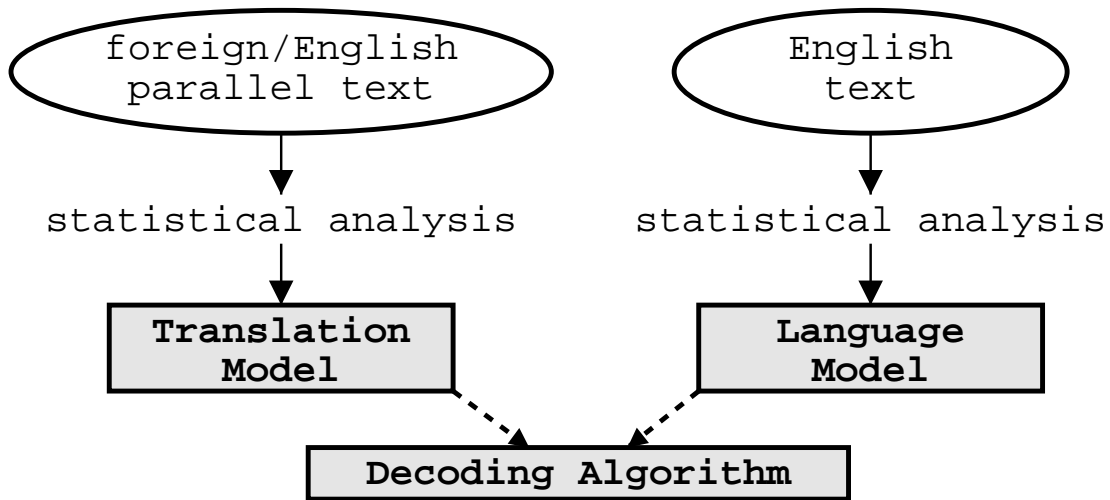
- Egyptian language was a mystery for centuries
  - 1799 a stone with Egyptian text and its translation into Greek was found
- ⇒ Humans *could learn* how to translated Egyptian

## Parallel data

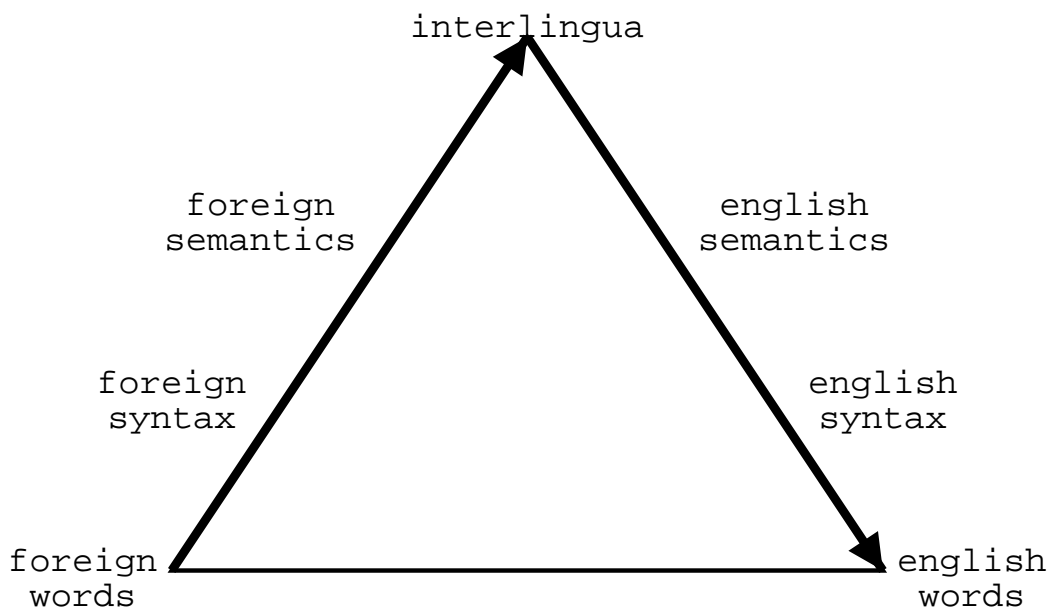
- Lots of translated text available: 100s of million words of translated text for some language pairs
    - a book has a few 100,000s words
    - an educated person may read 10,000 words a day
    - 3.5 million words a year
    - *300 million a lifetime*
    - soon computers will be able to see more translated text than humans read in a lifetime
- ⇒ Machine *can learn* how to translated foreign languages

## Statistical machine translation

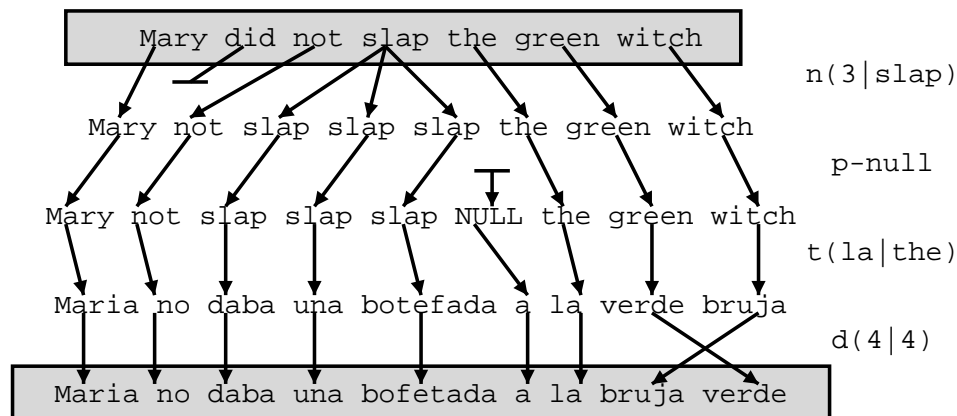
- Components: **Translation model**, **language model**, **decoder**



## The machine translation pyramid



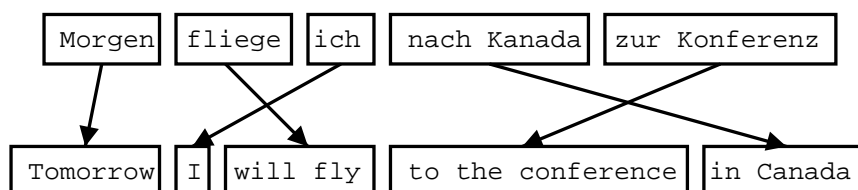
## Word-based models



[from Knight, 1997]

- Translation process is *decomposed into smaller steps*, each is tied to words
- Original models for statistical machine translation [Brown et al., 1993]

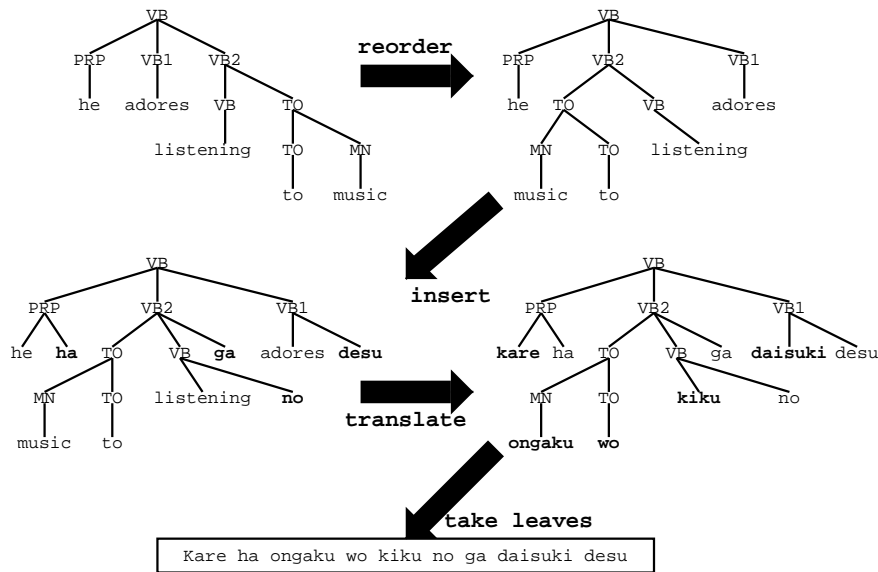
## Phrase-based models



[from Koehn et al., 2003, NAACL]

- Foreign input is segmented in **phrases**
  - *any sequence of words*, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

## Syntax-based models



[from Yamada and Knight, 2001]

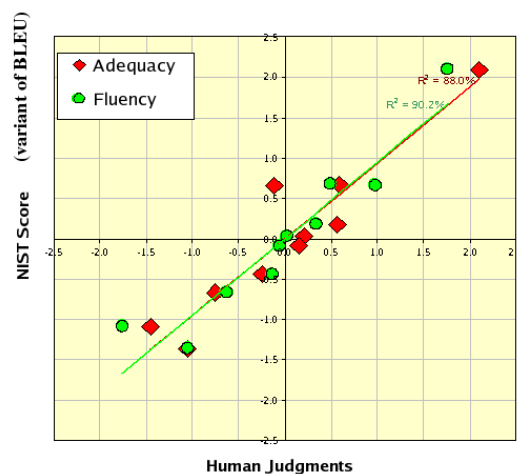
## Automatic evaluation

- Why **automatic evaluation** metrics?
  - Manual evaluation is *too slow*
  - Evaluation on large test sets *reveals minor improvements*
  - **Automatic tuning** to improve machine translation performance
- History
  - Word Error Rate
  - **BLEU** since 2002
- BLEU in short: *Overlap with reference* translations

## Automatic evaluation

- Reference Translation
  - the gunman was shot to death by the police .
- System Translations
  - the gunman was police kill .
  - wounded police jaya of
  - the gunman was shot dead by the police .
  - the gunman arrested by police kill .
  - the gunmen were killed .
  - the gunman was shot to death by the police .
  - gunmen were killed by police ?SUB>0 ?SUB>0
  - al by the police .
  - the ringer is killed by the police .
  - police killed the gunman .
- Matches
  - green = 4 gram match (good!)
  - red = word not matched (bad!)

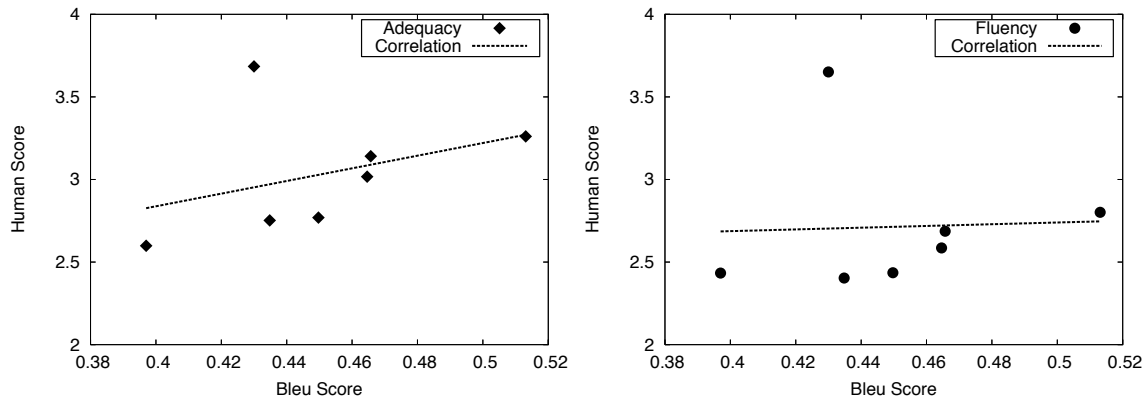
## Automatic evaluation



[from George Doddington, NIST]

- BLEU **correlates** with human judgement
  - **multiple reference translations** may be used

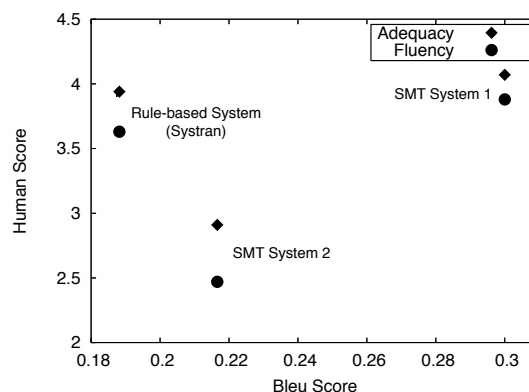
## Correlation? [Callison-Burch et al., 2006]



[from Callison-Burch et al., 2006, EACL]

- DARPA/NIST MT Eval 2005
  - Mostly statistical systems (all but one in graphs)
  - One submission **manual post-edit** of statistical system's output
  - Good adequacy/fluency scores *not reflected* by BLEU

## Correlation? [Callison-Burch et al., 2006]



[from Callison-Burch et al., 2006, EACL]

- Comparison of
  - *good statistical* system: **high** BLEU, **high** adequacy/fluency
  - *bad statistical* sys. (trained on less data): **low** BLEU, **low** adequacy/fluency
  - *Systran*: **lowest** BLEU score, but **high** adequacy/fluency

## Automatic evaluation: outlook

- Research questions
  - why does BLEU *fail* Systran and manual post-edits?
  - how can this *overcome* with novel evaluation metrics?
- Future of automatic methods
  - automatic metrics too *useful* to be abandoned
  - evidence still supports that during *system development*, a better BLEU indicates a better system
  - *final assessment* has to be human judgement

## Competitions

- Progress driven by **MT Competitions**
  - **NIST/DARPA**: Yearly campaigns for Arabic-English, Chinese-English, newstexts, since 2001
  - **IWSLT**: Yearly competitions for Asian languages and Arabic into English, speech travel domain, since 2003
  - **WPT/WMT**: Yearly competitions for European languages, European Parliament proceedings, since 2005
- Increasing number of statistical MT groups participate
- Competitions won by statistical systems



## Euromatrix

- Proceedings of the European Parliament
  - translated into *11 official languages*
  - entry of new members in May 2004: more to come...
- Europarl corpus
  - collected 20-30 million words per language
  - *110 language pairs*
- 110 Translation systems
  - 3 weeks on 16-node cluster computer
  - *110 translation systems*

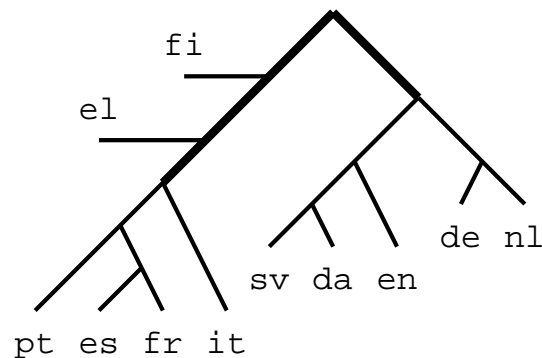
## Quality of translation systems

- *Scores* for all 110 systems

|    | da   | de   | el   | en   | es   | fr   | fi   | it   | nl   | pt   | sv   |
|----|------|------|------|------|------|------|------|------|------|------|------|
| da | -    | 18.4 | 21.1 | 28.5 | 26.4 | 28.7 | 14.2 | 22.2 | 21.4 | 24.3 | 28.3 |
| de | 22.3 | -    | 20.7 | 25.3 | 25.4 | 27.7 | 11.8 | 21.3 | 23.4 | 23.2 | 20.5 |
| el | 22.7 | 17.4 | -    | 27.2 | 31.2 | 32.1 | 11.4 | 26.8 | 20.0 | 27.6 | 21.2 |
| en | 25.2 | 17.6 | 23.2 | -    | 30.1 | 31.1 | 13.0 | 25.3 | 21.0 | 27.1 | 24.8 |
| es | 24.1 | 18.2 | 28.3 | 30.5 | -    | 40.2 | 12.5 | 32.3 | 21.4 | 35.9 | 23.9 |
| fr | 23.7 | 18.5 | 26.1 | 30.0 | 38.4 | -    | 12.6 | 32.4 | 21.1 | 35.3 | 22.6 |
| fi | 20.0 | 14.5 | 18.2 | 21.8 | 21.1 | 22.4 | -    | 18.3 | 17.0 | 19.1 | 18.8 |
| it | 21.4 | 16.9 | 24.8 | 27.8 | 34.0 | 36.0 | 11.0 | -    | 20.0 | 31.2 | 20.2 |
| nl | 20.5 | 18.3 | 17.4 | 23.0 | 22.9 | 24.6 | 10.3 | 20.0 | -    | 20.7 | 19.0 |
| pt | 23.2 | 18.2 | 26.4 | 30.1 | 37.9 | 39.0 | 11.9 | 32.0 | 20.2 | -    | 21.9 |
| sv | 30.3 | 18.9 | 22.8 | 30.2 | 28.6 | 29.7 | 15.3 | 23.9 | 21.9 | 25.9 | -    |

[from Koehn, 2005: Europarl]

## Clustering languages



[from Koehn, 2005, MT Summit]

- **Clustering** languages based on how easy they translate into each other

⇒ Approximation of language families

## Translate into vs. out of a language

- Some languages are *easier* to translate into than out of

| Language  | From        | Into        | Diff        |
|-----------|-------------|-------------|-------------|
| da        | 23.4        | 23.3        | 0.0         |
| <b>de</b> | <b>22.2</b> | <b>17.7</b> | <b>-4.5</b> |
| el        | 23.8        | 22.9        | -0.9        |
| <b>en</b> | <b>23.8</b> | <b>27.4</b> | <b>+3.6</b> |
| es        | 26.7        | 29.6        | +2.9        |
| fr        | 26.1        | 31.1        | +5.1        |
| fi        | 19.1        | 12.4        | -6.7        |
| it        | 24.3        | 25.4        | +1.1        |
| nl        | 19.7        | 20.7        | +1.1        |
| pt        | 26.1        | 27.0        | +0.9        |
| sv        | 24.8        | 22.1        | -2.6        |

[from Koehn, 2005: Europarl]

- *Morphologically rich languages* harder to generate (German, Finnish)

## Backtranslations

- Checking translation quality by **back-translation**
- *The spirit is willing, but the flesh is weak*
- English → Russian → English
- *The vodka is good but the meat is rotten*

## Backtranslations II

- *Does not correlate* with unidirectional performance

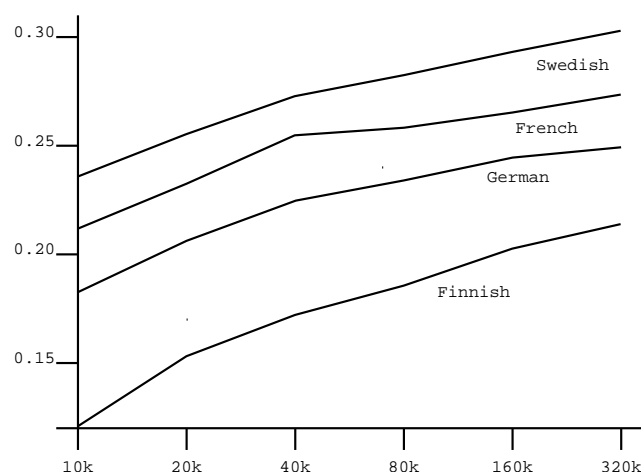
| Language | From | Into | Back        |
|----------|------|------|-------------|
| da       | 28.5 | 25.2 | 56.6        |
| de       | 25.3 | 17.6 | 48.8        |
| el       | 27.2 | 23.2 | <b>56.5</b> |
| es       | 30.5 | 30.1 | 52.6        |
| fi       | 21.8 | 13.0 | 44.4        |
| it       | 27.8 | 25.3 | 49.9        |
| nl       | 23.0 | 21.0 | 46.0        |
| pt       | 30.1 | 27.1 | <b>53.6</b> |
| sv       | 30.2 | 24.8 | 54.4        |

[from Koehn, 2005: Europarl]

## Available data

- Available *parallel text*
  - **Europarl**: *30 million words* in 11 languages <http://www.statmt.org/europarl/>
  - **Acquis Communautaire**: *8-50 million words* in 20 EU languages
  - **Canadian Hansards**: *20 million words* from Ulrich Germann, ISI
  - Chinese/Arabic to English: *over 100 million words* from **LDC**
  - lots more French/English, Spanish/French/English from **LDC**
- Available monolingual text (for language modeling)
  - *2.8 billion words* of English from **LDC**
  - *100s of billions, trillions* on the web

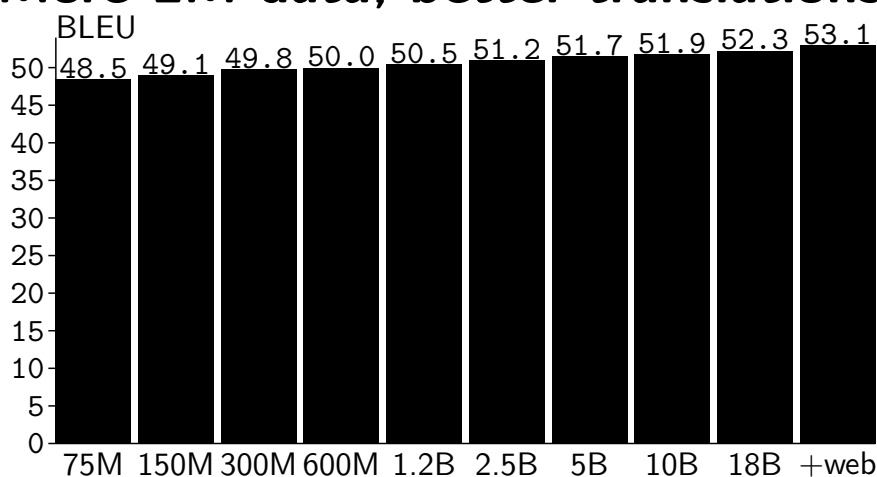
## More data, better translations



[from Koehn, 2003: Europarl]

- **Log-scale improvements** on BLEU:
  - Doubling the training data gives constant improvement (*+1 %BLEU*)

## More LM data, better translations



[from Och, 2005: MT Eval presentation]

- Also **log-scale improvements** on BLEU:  
doubling the training data gives constant improvement (+0.5 %BLEU)  
(last addition is 218 billion words out-of-domain web data)

## Output of Chinese-English system

### In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion us dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

### In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.

## Partially excellent translations

### In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion US dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

### In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.

## Mangled grammar

### In the First Two Months Guangdong's Export of High-Tech Products 3.76 Billion US Dollars

Xinhua News Agency, Guangzhou, March 16 (Reporter Chen Jizhong) - The latest statistics show that between January and February this year, Guangdong's export of high-tech products 3.76 billion US dollars, with a growth of 34.8% and accounted for the province's total export value of 25.5%. The export of high-tech products bright spots frequently now, the Guangdong provincial foreign trade and economic growth has made important contributions. Last year, Guangdong's export of high-tech products 22.294 billion US dollars, with a growth of 31 percent, an increase higher than the province's total export growth rate of 27.2 percent; exports of high-tech products net increase 5.270 billion us dollars, up for the traditional labor-intensive products as a result of prices to drop from the value of domestic exports decreased.

### In the Suicide explosion in Jerusalem

Xinhua News Agency, Jerusalem, March 17 (Reporter bell tsui flower nie Xiaoyang) - A man on the afternoon of 17 in Jerusalem in the northern part of the residents of rammed a bus near ignition of carry bomb, the wrongdoers in red-handed was killed and another nine people were slightly injured and sent to hospital for medical treatment.