Developing a Specialized Corpus for ATA Translator Certification Examinations

Geoffrey S. Koby
Kent State University, Ohio, USA
American Translators Association
Translator Certification Exam

• Examinations currently handwritten
  (transitioning to computerized)
• Data starting with 2006 testing year
• Design of certification exam determines corpus design
Exam Design

- 23 language pairs
- *Language > English or English > language*
- Three source texts - about 250 words
- Candidates translate two passages:
  - Mandatory general passage (labeled “A”)
  - Either a semi-technical passage (“B”) or a commercial/legal passage (“C”)
Annual Statistics
- In 2006, 535 candidates
- Spanish>English/English>Spanish about 55%
- 1070 passages (x two graders)
- 2140 marked passages
- Estimated 535,000 words for 2006
- 7,585 examinations 2001-2011
- Corpus grows 690,000 words per year
American Translators Association
Translator Certification Exam

• Graders mark passages
• Error codes indicate category/severity
• Target text shows **scope, type and severity of various translation errors**
AMERICAN TRANSLATORS ASSOCIATION
Accreditation Examination

German into English

Passage A – General

Instructions: The following text is taken from a guidebook on the city of Salzburg addressing foreign visitors. Translate it for an American readership.

Wenn sich heute eine Salzburger Familie eine Muschel-Suppe wünscht, ist der Weg vom Wunsch zur Erfüllung kurz und unkompliziert. Die Muscheln finden sich in gleich mehreren Fischläden, die übrigen Zutaten sind in jedem Delikatessengeschäft zu bekommen.
Unique features of the corpus

- Multidirectional parallel corpus
- All exam texts selected according to same criteria
- Common language is English
- One-to-many alignment between one source and multiple target texts
- Range of quality
- Explicit error markings (scope, severity, category)
- Designing a database tool:
These have discrepancies.
Enter New Source

<table>
<thead>
<tr>
<th>Year</th>
<th>Source Language</th>
<th>Target Language</th>
<th>Set Number</th>
<th>Passage Letter</th>
<th>Version Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Arabic</td>
<td>Arabic</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Submit Source  

Reset
This area will be the queue for new source texts needing reviewed before being locked to the database.

- eng_spa_2005_1_A_1 ----- Importer #1: root
**Final Review Of New Source Text**

<table>
<thead>
<tr>
<th>Year</th>
<th>Source Language</th>
<th>Target Language</th>
<th>Set Number</th>
<th>Passage Letter</th>
<th>Version Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>English</td>
<td>Spanish</td>
<td>1</td>
<td>A</td>
<td>1</td>
</tr>
</tbody>
</table>

**Source Passage Input**

Las Palmas de Gran Canaria is a city in the island of Gran Canaria, in Spain. It is the co-capital (jointly with Santa Cruz) and the most populous city in the autonomous community of the Canary Islands, and the ninth largest city in Spain, with a population of 383,568 in 2010. Nearly half (45.3%) of the people of the island and 18.36% of all inhabitants of the Canary Islands live in the city. It is also the fifth most populous urban area in Spain with a population exceeding 700,000 and (depending on sources) ninth or tenth most populous metropolitan area in Spain with a population of between 625,000 and 750,000. Las Palmas is the largest city of the European Union lying outside the European Continent. It is located in the northeast part of the Spanish island of Gran Canaria, about 150 kilometres (93 miles) off the northwestern coast of Africa within the Atlantic Ocean.

**Problem Comments**

Submit Source  
Reset
New Exam Record

- Year
- Exam Number
- Source Language
  - Arabic
- Target Language
  - Arabic
- Second Passage
  - B
- Source Text 1
  - eng_spa_2010_1_B_1
- Source Text 2
  - eng_spa_2010_1_B_1
- Grader #1
  - 123
- Grader #2
  - 123
- Grader #3
  - 123
- Grader #4
  - 123

Drop files to upload
(or click)

Reset Form
Enter New Exam
### Target Text Queue

The image shows a webpage with a table titled "Exams to be inputed." The table contains two columns: "Exam Input" and "Target Text." The row for Exam # 11234 is highlighted. The webpage also contains a dropdown menu labeled "Legacy Exam" and tabs for "Source Text," "Analysis," "Source Queue," and "Source Disc."
Enter a Legacy Graded Exam

inputs will be drop down selects with options pulled from database table

<table>
<thead>
<tr>
<th>Exam Year</th>
<th>Source Language</th>
<th>Target Language</th>
<th>Set Number</th>
<th>Passage Letter</th>
<th>Version Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>eng</td>
<td>spa</td>
<td>1</td>
<td>A</td>
<td>1</td>
</tr>
</tbody>
</table>

Target Passage Input

[Submit Target]  [Reset]  [View PDF Scan]
Target Text Queue

Target texts that need verified.

12346
12346
12346
12346
12346
12348
11234
12348
Las Palmas de Gran Canaria es una ciudad y municipio español situada en el noreste de la isla de Gran Canaria. La ciudad, la más grande y poblada de Canarias con 332.266 habitantes, es la capital de Gran Canaria y de la provincia de Las Palmas. Además, comparte la capitalidad de la Comunidad Autónoma de Canarias con la ciudad de Santa Cruz de Tenerife.

La ciudad fue fundada en 1478, siendo considerada la única capital (de facto) del archipiélago canario hasta el siglo XVI. Es sede de la Delegación del Gobierno de España, de la presidencia del Gobierno de Canarias en períodos legislativos alternos con Santa Cruz de Tenerife, de la Presidencia del Tribunal Superior de Justicia de Canarias, de la Delegación de Canarias, del Consejo Económico y Social de Canarias, así como otras instituciones de diversa importancia como la Casa África. El Carnaval de Las Palmas de Gran Canaria es uno de los eventos más importantes de Canarias y goza de una importante proyección nacional e internacional.

Los municipios adyacentes a la ciudad forman un área metropolitana de más de 600.000 habitantes, constituyendo el área metropolitana más grande de Canarias y noveno de España. El municipio tiene una extensión de 100.445 km² (ISTAC, 2003). Su altitud es de 8 metros sobre el nivel del mar (en la parte más meridional). El clima es de escasas precipitaciones, con una temperatura media de unos 22 °C.
Grading Queue
Grading Screen

Exam # 16621 -- Passage
eng -> ger

Source Text

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long withstand the conflict.

Target Text

Vier Kerbe und sieben Jahren unser Väter auf diesem Kontinent eine neue Nation, in Freiheit empfangen, und widmet sich der These, dass alle Menschen gleich geschaffen brachte. Jetzt sind wir in einem großen Bürgerkrieg, erproben, ob diese Nation oder irgendeine Nation so konzipiert...
Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battlefield of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

Vier Kerbe und sieben Jahren unser Väter auf diesem Kontinent eine neue Nation, in Freiheit empfangen und widmet sich der These, dass alle Menschen gleich geschaffen brachte. Jetzt sind wir in einem großen Bürgerkrieg, erproben, ob diese Nation oder irgendeine Nation so konzipiert werden kann, dass sie lange bestehen kann. Wir sind hier auf einem großen Feldzug der Kriege. Wir haben hier einen Teil des Feldes zu erweihen, als endgültiges Verbotungsplatz für diejenigen, die hier ihr Leben opferten, dass die Nation leben kann. Es ist ganz angemessen und korrekt, dass wir dies tun.
Enter Error Severity

Exam # 16621 -- Passage
eng -> ger

Source Text

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Target Text

Vier Kerbe und sieben Jahren unser Väter auf diesem Kontinent eine neue Nation, in Freiheit empfangen, und widmet sich der These, dass alle Menschen gleich geschaffen sind.

Grading Functions
TOTAL POINTS DEDUCTED: 8

Error Code: UNF  Assign Value: 1  Error Type: Normal

Error Comments:

1 2 4 8 16

Cancel  Reset Form  Submit
FLOWCHART FOR ERROR POINT DECISIONS

DRAFT REVISION November 2005

Is the meaning in context affected?

Yes

Is the effect on the meaning negligible?

Yes

0 points

No

Is the error merely a slight imprecision?

Yes

1 point

No

Could a monolingual copy editor correct the error?

Yes

2 points

No

Interference with the reader’s understanding of the passage minimal / limited in scope?

Yes

4 points

No

Could the reader work out the overall meaning of the passage despite this serious error?

Yes

8 points

No

16 points

0 points

Error apparent to a monolingual copy editor?

Yes

1 point

No

Error apparent to a typical target reader of this translation?

Yes

2 points

No

Error disruptive to a typical target reader of this translation?

Yes

4 points

No


Normal/Disjunct Error
These have discrepancies.
### Graded Exams

<table>
<thead>
<tr>
<th>Exam #</th>
<th>Passage Letter</th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam# 1111</td>
<td>Passage A</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Exam# 1111</td>
<td>Passage B</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Exam# 12347</td>
<td>Passage A</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Exam# 12347</td>
<td>Passage B</td>
<td>English</td>
<td>Spanish</td>
</tr>
</tbody>
</table>

**List of Graded Exams**

1. Exam# 1111: Passage A, English to Spanish
2. Exam# 1111: Passage B, English to Spanish
3. Exam# 12347: Passage A, English to Spanish
4. Exam# 12347: Passage B, English to Spanish
Viewing Graded Exam

Exam #12347 -- Passage

Source Text

Long ago, in the great ocean that washes the shores of Peru no fish. There corals, sponges, jellyfish, snails and other animals, but no fish you swam in the blue waters of the ocean. They lived only rivers, lakes and streams of Peru. but they were so few that not fished. One day these fish migrated to the ocean, and there multiplied. Here's how it happened: I was living in Peru a young prince, beautiful and graceful. It was very powerful and he knew magic arts. Reached a gesture to his hills and mountains aplanasen and transform themselves into green and fertile plains meadows, dipping a reed in a river, and at the same instant the waters increased, overflowing and irrigating the fields; pronounced magical formulas, and when were dried muddy swamps and lagoons, the basins were transformed into fruitful banana plantations. This prince was called Conyara, and being a right that he liked to do good, often disguised as a beggar to mingle with the poor people and learn about their needs and desires man. Often accompanied the llama herders who roamed the rugged trails of the Andes. Entered the miserable huts and saw how hard- milled corn achieved with dedication on the rocky slopes. Nobody recognized

Target Text

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum. It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem ipsum as their default model text, and a search for 'lorem ipsum' will uncover many web sites still in its infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like). Contrary to popular belief, Lorem Ipsum is not simply

Grading Functions

TOTAL POINTS DEDUCTED: 24

See Grading Flowchart  View PDF Scan  Show All Errors
### Source Text

Long ago, in the great ocean, there was no fish. There were sponges, jellyfish, sharks, and other animals, but no fish you swim in the blue waters of the ocean. They lived only rivers, lakes, and streams of Peru, but they were so few that not even the下令. One day those fish migrated to the ocean, and there multiplied. Here's how it happened: I was living in Peru's young prince, beautiful and graceful. It was very powerful and he knew magic arts. Reached a gesture to his hills and mountains apalanca them into green and fertile plains meadows, dumping a reed in a river, and at the same instant the waters increased, overflowing and irrigating the fields; pronounced magical formulas, and when they were dried muddy swamps and lagoons, the basins were transformed into fruitful banana plantations. This prince was called Coni, and being a right that he liked to do good, often disguised as a beggar to mingle with the poor people and learn about their needs and desires. Man. Often accompanied the llama herdsmen who roamed the rugged trails of the Andes, entered the miserable huts and saw how hard milled corn achieved with dedication on the rocky slopes. Nobody recognized

### Target Text

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum. It is a long established fact that a reader will be distracted by the readable content of a page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less normal distribution of letters, as opposed to using 'Content here, content here', making it look like readable English. Many desktop publishing packages and web page editors now use Lorem Ipsum as their default model text, and a search for 'Lorem ipsum' will uncover many web sites still in their infancy. Various versions have evolved over the years, sometimes by accident, sometimes on purpose (injected humour and the like). Contrary to popular belief, Lorem Ipsum is not simply

### Grading Functions

TOTAL POINTS DEDUCTED: 24

<table>
<thead>
<tr>
<th>Error #</th>
<th>Error Code</th>
<th>Comment</th>
<th>Source Phrase</th>
<th>Target Phrase</th>
<th>Points Deducted</th>
<th>Error Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Comments</td>
<td>Long ago, in the great ocean</td>
<td>Lorem Ipsum is simply dummy text of the printing and typesetting industry.</td>
<td>8</td>
<td>NORMAL_ERROR</td>
</tr>
<tr>
<td>2</td>
<td>AMB</td>
<td>Fail</td>
<td>nd they multiplied&gt;They lived only rivers</td>
<td>essentially unchanged&gt;It has survived not</td>
<td>16</td>
<td>DISJUNCT_ERROR</td>
</tr>
</tbody>
</table>
Already coded queries
User-guided queries
Statistical analyses
Keyword in Context
### Analysis Options - Preliminary

#### Analysis Queries

<table>
<thead>
<tr>
<th>Main Analysis</th>
<th>User Analysis</th>
<th>Exam Analysis</th>
<th>Keyword In Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Language*</td>
<td>Target Language*</td>
<td>Years*</td>
<td>Exams*</td>
</tr>
<tr>
<td>Select Language</td>
<td>Select Language</td>
<td>Select Years</td>
<td>Select Exams</td>
</tr>
</tbody>
</table>

*required

- **Pass/fail analysis**
- **Average score analysis**
- **Error analysis**
## Analysis Queries

### Main Analysis

<table>
<thead>
<tr>
<th>Source Language*</th>
<th>Target Language*</th>
<th>Years*</th>
<th>Exams*</th>
<th>Passages*</th>
<th>Graders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select Language</td>
<td>Select Language</td>
<td>Select Years</td>
<td>Select Exams</td>
<td>Select Passages</td>
<td>Select Graders</td>
</tr>
</tbody>
</table>

### User Analysis

- Avg. Score Analysis
- Error Analysis

### Exam Analysis

- Select Language
- Select Years
- Select Exams
- Select Passages

### Keyword in Context

- Select Language
- Select Years
- Select Exams
- Select Passages

---

**Select Source Language**

- Arabic
- Chinese
- Dutch
- English
- Finnish
- French
- German
- Croatian
- Hungarian
- Italian
- Japanese
- Polish
- Portuguese
- Russian
- Spanish
Select Target Language(s)
### Analysis Queries

<table>
<thead>
<tr>
<th>Main Analysis</th>
<th>User Analysis</th>
<th>Exam Analysis</th>
<th>Keyword In Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Language*</td>
<td>Target Language*</td>
<td>Years*</td>
<td>Exams*</td>
</tr>
<tr>
<td>English</td>
<td>-</td>
<td>1 selected</td>
<td>Select Years</td>
</tr>
<tr>
<td><strong>Pass/Fail Analysis</strong></td>
<td><strong>Avg. Score Analysis</strong></td>
<td><strong>Error Analysis</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Graders</strong></td>
<td><strong>Passages</strong></td>
<td><strong>Select Graders</strong></td>
<td></td>
</tr>
</tbody>
</table>

*required

**Pass Rate %**

<table>
<thead>
<tr>
<th>% of Passing Exams</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>null</td>
</tr>
</tbody>
</table>

Exams

Pass Rate: 64
Analysis Queries

Main Analysis | User Analysis | Exam Analysis | Keyword In Context
---|---|---|---
Source Language* | Target Language* | Years* | Exams* | Passages* | Graders
English | 1 selected | Select Years | Select Exams | Select Passages | Select Graders

*required

Pass/Fail Analysis | Avg. Score Analysis | Error Analysis

Avg. Score

Score

67

Exams
Error Analysis by Part
Errors by Quarter
Analysis Queries

Search Keywords

test

Search

Not yet implemented
 trade, clearing and settlement is a very welcome trend and makes economic sense to both.

GID024:1A

trading, clearing, and settlement is very much to be greeted and economically makes sense.

GID024:4MT-GID156:4MT

trading, clearing and settlement is quite welcome and makes economic sense to large investors.

GID023:MT4-GID165:4MT

trading, clearing and settlement are very welcome and make economic sense so that further.

GID023:2MT/MU/G-GID024:1G

trading, clearing and settlement is very welcome and economically sensible, which means that further.

GID023:MT4-GID165:4MT

trading, clearing and settlement are very welcome and make economic sense so that.

GID023:2MT/MU/G-GID024:1G

trade, clearing and settlement is very desirable and makes economic sense for both.

GID 023:MT8 Candidate lost his place, destroyed sense of text. – GID165: Mixed up reading of source text 4 MU

commerce, clearing and settlement is highly appreciated and economically useful, so that.

GID156:1C-GID200:1C

GID156:2MT-GID200:1U+2MT

GID200:1L
Summary and Outlook

- Multidirectional parallel corpus of translation exams
- One-to-many alignment between one source and multiple target texts
- Future research possibilities
Future Research Possibilities 1

- Sorting/analysis of translation errors as marked by the graders.
  - error type, error severity, language pair, etc.
- Patterns of errors, types of errors, etc.
- Analyze terms by Part of Speech
- Search, view, analyze by various tags
- Pass/fail rates & score ranges by language pair and corpus as a whole
Future Research Possibilities 2

• Clustering of score ranges above and below the pass/fail boundary
• Inter-grader/intra-grader consistency and variability in error discovery/assessment
• Congruity in discovery of translation errors
  - (i.e., the ratio of errors marked by both graders to those marked by only one of the two)
• Investigation of congruity in marking of severity and category of translation errors
Thank You for your Attention!

- Dr. Geoff Koby - gkoby@kent.edu
- For a complete description of the ATA Examination Program, see:
  Koby, Geoffrey S. and Gertrud G. Champe.
  “Welcome to the Real World. Professional-Level Translator Certification.”