ATA Certification Exam Database (ACED)
Project Specifications

Introduction

The data for this project comes from translator certification examinations administered by the American Translators Association. This data consists of two core sets of data along with metadata. In addition, we must store color scans of the original handwritten exams. We are currently working with data from 2006, but data from future years is already available and the system must be flexible to accommodate changes over time.

1. The first core component is the examination record itself, that is, a set of source texts (ST) – the original texts that were assigned to be translated on the exam – together with the target texts (TT) produced by the candidates. This is textual data, where the source texts must be aligned with the target texts in specific ways. Exams in 2006 are administered in 23 different language pairs. English is always one part of the pair, either the source text or the target text. Each source text will be in a one-to-many relationship with the target texts produced by the various candidates.
   a. An ATA examination consists of three separate source passages of approximately 250 words each. Each candidate is required to translate the general passage (labeled “A”), and must choose between a semi-technical passage (“B”) or a commercial or legal passage (“C”).
   b. Each exam package thus consists of two source texts (A/B or A/C) and the corresponding two target texts, each of which is marked by two graders for a total of four marked target text passages (AA/BB or AA/CC).
   c. It will be necessary to explicitly align the data by sentence between the source text and each target text.

2. The second core component, related to the first, is a set of error markings made on the target texts by graders. These markings (circles, underlines in a different color pen) indicate specific sections of the target texts that are errors. These errors receive a code for type and a number for severity of the error. A key component of the database will be the explicit alignment (by trained linguists) of the segments marked as errors in the target texts with the corresponding segment of the source text. This aligned error must also store the error code and error severity.
   a. Graders can also assign up to three quality points for good solutions.

3. The metadata that accompanies each exam is used for analysis purposes. This metadata includes the testing year, language pair, the exam number, the grader number, and the scores and pass or fail result recorded by the graders. In addition, the system should calculate scores based on the errors input.

All of the data is confidential, so the system must have appropriate security measures in place. A superuser should be able to set permissions for read and/or write access to various modules within the system.

1 The examination system allows for “reviews” (where a candidate contests the score) and “appeals” (where the candidate contests the markings. A review will be graded by a third grader, and an appeal by a fourth grader.
I. Data Entry

The system must allow for easy data entry and validation. The data entry system should be usable both for entering legacy data from handwritten exams as well as live grading of new exams. Comment fields should be available in a variety of places.

A. Legacy Data

The legacy data is currently handwritten on paper. The following steps allow data entry from a scan of the original documents. The system should have fields to track who did what data entry – there should be some method of validation of data at each stage – i.e., one person enters, a second person approves, then record is locked from modification.

1. Create and validate the “source text” records for each language pair using the following information:

   **Source Text Codes**
   - Source Language: SSS
   - Target Language: TTT
   - Exam Year: YYYY
   - Set Number: 1
   - Passage Letter: A
   - Version Number: 1

   **Keyword**
   - Keyword: “text”

   **Source Text**
   - Text: Unicode text storage of a passage of around 250 words or the equivalent in non-Latin scripts.

   a. Template of current source text code: SSS_TTT_YYYY_1_A_1
   b. Example of current source text code: eng_spa_2009_1_A_0 – English into Spanish, 2009, Set 1, Passage A, Version 1.
   c. Source Language and Target Language data must use the official three-letter language codes from the ISO 639-2 list (all lower-case – at [http://www.loc.gov/standards/iso639-2/php/English_list.php](http://www.loc.gov/standards/iso639-2/php/English_list.php)). The codes for current languages are:

        - ara  Arabic
        - chi  Chinese
        - dut  Dutch
        - eng  English
        - fin  Finnish
        - fre  French
ger  German
hrv  Croatian
hun  Hungarian
ita  Italian
jpn  Japanese
pol  Polish
por  Portuguese
rus  Russian
spa  Spanish

Note 1: Most ISO 639-2 codes are the first three letters of the language in English, but not always.
Note 2: It must be possible to modify the list. Some language pairs were available in 2006 and have been suspended, and new language pairs will be added in the future.

d. Year is a four-digit integer
e. Set Number is a single-digit integer. Normally, up to two sets of source texts may be active, but the examination program has up to five sets at any time.
f. Passage Letter is A, B, or C.
g. Version Number is normally “1” but may be higher if changes are made to a source text during an exam year.
h. Keyword provides a more "natural" way of referring to individual passages. A passage can more conveniently be referred to as "pension plans," for example, rather than the database designation eng_spa_2009_1_A_0.
i. The Text field will contain the core data of the source text of each passage in each language pair. This data will be reused repeatedly for explicit alignment in individual exams. The data entry person will enter the source text either from a file or from the paper copy in the exams, and then re-use it.

2. Create and validate a new record for each exam using the following fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>four-digit integer</td>
</tr>
<tr>
<td>Source Language</td>
<td>three-letter code</td>
</tr>
<tr>
<td>Target Language</td>
<td>three-letter code</td>
</tr>
<tr>
<td>Exam Number²</td>
<td>five-digit integer</td>
</tr>
<tr>
<td>Second Passage</td>
<td>“B” or “C” (the first passage is always “A”)</td>
</tr>
<tr>
<td>A passage markings</td>
<td>Yes/No – does the A passage ST have candidate markings on it?</td>
</tr>
<tr>
<td>2nd passage markings</td>
<td>Yes/No – does the second passage ST have candidate markings?</td>
</tr>
<tr>
<td>Grader No. 1³</td>
<td>a three-digit integer, allow for leading zeros</td>
</tr>
</tbody>
</table>

² The current information on each candidate is only the exam number. The system should allow for additional fields to be added later to capture information on candidate demographics if it becomes available.

³ The current information on the graders is simply a three-digit integer (001-999).
Grader No. 2  a three-digit integer, allow for leading zeros
Grader No. 3 a three-digit integer, allow for leading zeros
Grader No. 4  a three-digit integer, allow for leading zeros
TT text record 1 stores text of TT A
TT 1 error score G1-G4 Four integer fields to store the scores recorded by the graders on Passage A – for comparison with the calculated scores. Graders make mistakes!
TT text record 2 stores text of TT B or C
TT 2 error score G1-G4 Four integer fields to store the scores recorded by the graders on Passage B/C – for comparison with the calculated scores. Graders make mistakes!
“46+ stop” P1 Yes/No, was the score on passage A over 46 (stop-limit) and grader stopped grading?
“46+ stop” P2 Yes/No, was the score on passage B/C over 46 (stop-limit) and grader stopped grading?
“Not graded” P1 Yes/No, was passage A not graded for some reason?
“Not graded” P2 Yes/No, was passage B/C not graded for some reason?
“Incomplete” P1 Yes/No, was passage A incomplete?
“Incomplete” P2 Yes/No, was passage B/C incomplete?
Global QP integer in the range of 0 to 3 for Quality Points assigned to an entire passage
Exam reviewed? Yes/No
Exam appealed? Yes/No
Passage A reviewed Yes/No
Second passage reviewed Yes/No

3. Upload and attach a color PDF scan of the original document. Most of the PDFs are running between about 25 and 75 MB each, depending on the number of pages. In the further data entry process, it will be necessary to view the relevant PDF page-by-page, and to zoom in on details as necessary.

4. Input and validate texts of candidate’s two exam passages into appropriate text fields.

5. Using a dual display of ST and TT, align each paragraph and each sentence of the ST with the TT. Ideally, this would be highlighting the ST sentence, highlighting the TT sentence, and

- System should allow setup of graders so that only numbers assigned to a given language pair are available.
- System should have a field for Grader E-Mail to allow automatic sending of messages in New Data module.
- System should allow for additional demographic information fields to be added in the future if data becomes available.

4 Graders 3 and 4 will normally be hidden unless needed. Most exams have two graders.

5 Each exam and each passage and each error must have a comment field linked to it.

6 “Yes” on Exam Reviewed will activate Grader No. 3 and the fields for which passage was reviewed.

7 “Yes” on Exam Appealed will activate Grader No. 4.
storing. Should be done on the text itself using tags. The standard marks for “paragraph” and “sentence” are `<p>` and `<s>`. Alignment should be validated.

6. Using the dual display of ST and TT, mark each error in the TT. Ideally, this would be highlighting the TT error, highlighting the corresponding ST text, and assigning the category and error points marked on the exam. Note that an error may be text (one or more words, up to whole sentences), punctuation marks, or spaces (for Omission errors). It must be possible to record disjunctive errors (i.e. two words with intervening non-marked words).

Markup requires the following fields:

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error type</td>
<td>a code from a prescribed list (or Quality Points)</td>
</tr>
<tr>
<td>Error severity</td>
<td>0, 1, 2, 4, 8, or 16 points</td>
</tr>
<tr>
<td>Error modification</td>
<td>Yes/No – does the error appear to have been changed, modified, or overwritten on the handwritten text?</td>
</tr>
<tr>
<td>Allow error violation</td>
<td>Yes/No – If an error violates the rules of the grading system, is this actually on the text?</td>
</tr>
<tr>
<td>Grader markings</td>
<td>Text field to record any comments by a grader on a particular error.</td>
</tr>
</tbody>
</table>

7. Once all errors are marked, review and validate.

8. System will calculate scores on each passage, and Pass or Fail result. System should include reports on:
   a. Distance between scores on each passage
   b. Where grader markings disagree on Pass/Fail result on a passage

8 The system should allow export of the tagged text to a plain Unicode text file for use in corpus analysis software (Wordsmith).

9 The system should also allow for the possibility of later Part of Speech (POS) tagging for further research. See http://www.natcorp.ox.ac.uk/docs/c5spec.html for the basic British National Corpus (BNC) tag set for English, and http://www.natcorp.ox.ac.uk/docs/c7spec.html for the enriched set.

10 Error types and error severities should be customizable somehow, so that the system can be used with other testing plans or allow for new categories of errors, elimination of categories, different severities, etc. (Some systems work only with Critical, Major, Minor errors.) We also want to be able to specify “error groups”, which are sets of error categories, e.g., Transfer errors, Mechanical errors, Formal errors, and be able to customize which categories go in which groups.

11 A maximum of three Quality Points can be assigned per passage. This is captured like an error – with highlighting – but a quality point subtracts from an error point. It must be possible to record where the QP was assigned, and track the number of QPs assigned.

12 Graders sometimes mark an error, then assign 0 points – we want to capture that information. We also want to be able to record an error that is marked (circled, underlined) but where either no category is assigned, no code is assigned, or both (just a marking).

13 For ATA grading, mechanical errors are not allowed to be assigned more than four points; SP (Spelling) is not allowed more than two points. Graders sometimes mark such errors. We must be able to capture these errors in the system, but explicitly record that there is a violation.
B. New Data

The entry module for new data will work with current keyboarded translation exams, where the input data is already provided. ATA will be providing information on the format of the incoming data, which will provide Source Language, Target Language, Exam No., Passage A text, Second Passage text.

The New Data module should take the records supplied by the exam system, and allow a testing coordinator to assign the exam to two graders in that language pair. System should automatically generate an e-mail to the assigned graders informing them that an exam is ready to grade.

Graders should be able to log in and see a list of exams assigned to them and ready for grading. When graders do grading, they should be able to show or hide the ST, and use the same functionality as above #6 to assign errors and Quality Points.

- Graders should have a grader comment field on each error and each passage.
- System should only allow graders to assign correct codes and points
- System should allow graders to assign up to three Quality points.
- System should automatically tally points and show a running total.

When grading is complete, grader will review and validate own work, then submit.

Once both graders have submitted, system will check scores following ATA rules. The two graders must consult if there is:

- a questionable error or point marking
- a point difference of 15 points or more
- a Pass/Fail discrepancy between two graders
- an error-point score in the borderline range, i.e. within the 15–25 point range.

If no consultation is necessary, the system records this fact and does the following:

- Reports the exam as completed to the certification program manager
- Generates a Pass letter or a Fail letter to be sent out to the candidate, reporting score ranges

If consultation is necessary, the system should put the exam back in both graders’ workflow for review. The two graders should be able to see some kind of display showing both graders’ error markings, clearly labeled, for the review.

Graders should be able to change their own error category and point assignments at that time. The system should record both the original reported errors and the modified errors and allow comparison of the changes.

Once both graders have completed consultation, the exam is resubmitted as complete.

The system should generate reports by year, language pair, etc., showing scores, average scores, pass/fail rates, etc.
Data Analysis

Once the data has been entered, the database becomes a valuable resource for linguistic analysis of translation strategies, translation errors, and successful translation solutions, as well as for analysis of grader markings. The system should allow investigation of the following:

- Pass/fail rates and score ranges for each language pair and for the corpus as a whole
- Investigation of clustering of score ranges above and below the pass/fail boundary, by language pair and for the corpus as a whole
- Investigation of inter-grader and intra-grader consistency and variability in discovering and assessing errors for each exam and for each grader
- Investigation of 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 the marking of severity and category of translation errors
  - Ideally, this would be a screen where the errors marked by both graders (or three or four) are marked in some way on the target text – color coded, for instance, as to whether identical or separate, same code or different, etc.
  - Concordancing - searching for/sorting by linguistic forms
    - By primary form
    - Within primary form by secondary/tertiary words
    - Reverse alphabetical sorting
    - One-to-many (source text-target text) alignment by any linguistic criterion or by tag
- Keyword in Context (KWIC) display for each language pair, displaying ST at the top and multiple TTs below, with some highlighting and sort facility for errors. Allow for expanded view of context
  - Ability to analyze patterns of errors, types of errors, etc.
  - Ability to analyze terms by Part of Speech
  - Ability to search for, view, analyze by various tags
### Framework for Standardized Error Marking

**Revised May 2006**

<table>
<thead>
<tr>
<th>Code</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNF</td>
<td>Unfinished</td>
</tr>
<tr>
<td>ILL</td>
<td>Illegible</td>
</tr>
<tr>
<td>OTH</td>
<td>Other</td>
</tr>
<tr>
<td>MT</td>
<td>Mistranslation into target language</td>
</tr>
<tr>
<td>MU</td>
<td>Misunderstanding of source text</td>
</tr>
<tr>
<td>A</td>
<td>--Addition</td>
</tr>
<tr>
<td>O</td>
<td>--Omission</td>
</tr>
<tr>
<td>T</td>
<td>--Terminology, word choice</td>
</tr>
<tr>
<td>R</td>
<td>--Register</td>
</tr>
<tr>
<td>F</td>
<td>--Too freely translated</td>
</tr>
<tr>
<td>L</td>
<td>--Too literal, word-for-word translation</td>
</tr>
<tr>
<td>FC</td>
<td>--False cognate</td>
</tr>
<tr>
<td>IND</td>
<td>--Indecision, gave more than one option</td>
</tr>
<tr>
<td>I</td>
<td>--Inconsistency</td>
</tr>
<tr>
<td>AMB</td>
<td>--Ambiguity</td>
</tr>
<tr>
<td>G</td>
<td>Grammar</td>
</tr>
<tr>
<td>SY</td>
<td>--Syntax (phrase / clause / sentence structure)</td>
</tr>
<tr>
<td>P</td>
<td>Punctuation</td>
</tr>
<tr>
<td>SP</td>
<td>Spelling</td>
</tr>
<tr>
<td>D</td>
<td>--Accents, other diacritical marks</td>
</tr>
<tr>
<td>C</td>
<td>--Case (upper/lower)</td>
</tr>
<tr>
<td>WF</td>
<td>Word form</td>
</tr>
<tr>
<td>U</td>
<td>Usage</td>
</tr>
<tr>
<td>ST</td>
<td>Style</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Max 2 pts for mechanical errors</th>
<th>Max 4 points for mechanical errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Column Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total error points</td>
</tr>
<tr>
<td>Quality points (maximum 3)</td>
</tr>
</tbody>
</table>

A grader may stop marking errors when the score reaches 46 error points.
A grader may award a quality point for each of up to three specific instances of exceptional translation.
Quality points are subtracted from the error point total to yield a final score. A passage with a score of 18 or more points receives a grade of Fail.

<table>
<thead>
<tr>
<th>Total error points</th>
<th>Quality points (maximum 3)</th>
<th>Final passage score subtract quality points from error points:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
FLOWCHART FOR ERROR POINT DECISIONS

Is the meaning in context affected?

yes → Is the effect on the meaning negligible?

no → Error apparent to a monolingual copy editor?

yes → 0 points

no → 1 point

0 points

Error apparent to a typical target reader of this translation?

yes → 4 points

no → 2 points

1 point

Error disruptive to a typical target reader of this translation?

yes → 16 points

no → 8 points

2 points

Is the error merely a slight imprecision?

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

DRAFT REVISION November 2005
Mientras se llevaban a cabo las negociaciones que habrían de culminar en el Tratado de París de 1899, se presentaron varias resoluciones en las cámaras del Congreso de EE.UU. cuyo propósito era el de proclamar la política que ese país haría de desarrollar con respecto a las Filipinas. Debe observarse que en la mayoría de estas resoluciones la incorporación de las Filipinas en la Unión había sido rechazada de plano y en forma clara y precisa. De hecho, en varias resoluciones se traslucía la intención, por parte de EE.UU., de conceder a las Filipinas su independencia tan pronto estuviese ese archipiélago preparado para la vida independiente. Más aún, en algunas de esas propuestas se adelantaba el punto de vista de que Estados Unidos carecía de autoridad constitucional para adquirir territorios a menos que éstos no estuviesen destinados a ingresar finalmente como estados en la Unión.

Esta perspectiva se basaba en la opinión escrita por el juez presidente Roger B. Taney del Tribunal Supremo de los Estados Unidos en el famoso caso de “Dred Scott”, que argumentaba precisamente esta tesis.

Aunque es cierto que varios senadores sustentaban este argumento, no es menos cierto que resultaba ya tardío en la historia de EE.UU. para cuestionar seriamente el derecho a adquirir territorios por conquista o cesión. Obviamente, EE.UU. había ejercido esa autoridad en repetidas ocasiones desde la adquisición de Luisiana en 1803: adquirió a Florida en 1819, el territorio mejicano en 1848, Alaska en 1867 y finalmente, efectuó la anexión de las Islas Hawaianas en 1898. (Debe recordarse que muy pocas personas entonces se hubieran aventurado a predecir que Alaska y, particularmente, Hawai estaban destinados a convertirse en estados de la Unión.)
While negotiations were taking place that would have eventually resulted in the Paris Treaty of 1899, various resolutions were submitted to the United States Congress.

The Legislative Body, whose purpose was to establish the policies that it should develop, noted that in the majority of these resolutions, the admission of the Philippines as part of the Union had been rejected clearly and decisively.

As a result of the numerous resolutions on the part of the U. S. A., the intent was plan to
The objective was to grant to the Philippines its independence, as soon as its archipelago was ready for an independent status. Furthermore, in some of these proposals, the point of view was fostered that the U.S.A. lacked the constitutional authority to acquire territories unless it was done with the ultimate goal of adding them to the States of the Union. This point of view was based on the opinion written by Judge President Roger.

B. Taney of the United States Supreme Court.
in the famous "Dead Scott," who precisely argued that these.

Even though it is true that a number of senators supported that argument, it is also true that it happened too late in U.S. history of the United States to seriously question the right to acquire territories by way of conquest orcession. Obviously, the U.S. had exercised its authority on a number of occasions from the acquisition of Louisiana in 1803:

- Florida in 1819
- Mexican Territory
- Alaska in 1867

and at finally earned
out the annexation of the Hawaiian Islands. It should be noted that very few people, at that time, would have gone out on the limb to predict that Alaska and in particular, Hawaii, would be destined to be part of the States of the Union. 