Error categories in English to Japanese translations

Chiho Toyoshima†, Kikuko Tanabe†, Anthony Hartley♮ and Kyo Kageura‡

† Kobe College, ♮ Tokyo University of Foreign Studies, ‡ University of Tokyo

1 Introduction

Assessing and assuring translation quality is one of the main concerns for translation services, translation teaching institutions and the MT industry. An inherent problem concerning translation quality assessment is that it inevitably involves human judgments and thus can be subjective.[1] As White (2003) notes, "[T]here are many ways to translate the same thing, and reasonable translators will disagree about which is best." Several error classification schemes have been proposed, such as that of the American Translators Association, SAE J 2450 and MeLLANGE error types (Secară (2005) gives a compact introduction to these schemes), aiming to guide human assessments.

We are currently developing an online system for translation training (Minna no Hon'yaku for Translation Training: MNH-TT), in which modified and simplified MeLLANGE error types are used (Babych, et. al. 2012; Utiyama, et. al. 2014). MeLLANGE error types were developed as an language-independent scheme and were proved applicable to translations among English, German, Spanish, French and Italian. They have not, however, been validated for translations between languages whose structures are radically different. Also, the applicability of these error types to translations made by less advanced learners (such as undergraduate students) has not been examined fully.

We carried out an experiment to check whether the MeLLANGE scheme is useful for diagnosing translation errors in English-to-Japanese translations made by students with different learning levels. Here we present part of the results, focusing mainly on the following two questions:

- Are there any errors identified by translation teachers that resist plausible classification within the scheme? - the coverage question
- Does the distribution of errors across the scheme suggest that it is reasonably discriminant? - the granularity question

2 Setup of translation experiment

In our experiment, 19 students (11 graduates and 8 undergraduates)[2] translated 12 English texts which deal with current affairs. The texts were all taken from the same independent online broadcasting site, Democracy Now![3], to keep the level of the texts approximately the same. Some texts were translated by more than one student. The undergraduates are identified as A to H, and graduates as I to S. The translations were done between July and September 2014. Students were asked to translate within the MNH or MNH-TT environment, which provide basically the same working conditions including lookup of dictionaries (Abekawa & Kageura 2009).

3 Assigning error types

The translations of these texts were checked and error types were assigned to each error instance by the first and the second author of this paper, using the adapted version of the MeLLANGE scheme (Secară 2005) shown in Table 1.

The assignment of error types was carried out as follows:

1. The first author assigned error types to each error instance, while at the same time recording why the instance met the criteria for classification as such.
2. The first and the second authors checked together the assignment of error types, while at

[1] While automated metrics for MT quality evaluation are often presented as objective, many, including BLEU, rely on comparison with one or more human reference translations whose quality is simply assumed but not independently validated.

[2] We took some background information about students, including their TOEIC scores, so that we can analyze the result of translation errors in terms of the students' competencies. We will report this aspect in the future.

[3] http://www.democracynow.org/

Table 1. Modified MeLLANGE error types

	Group	Type
X1	content (C)	omission
X2		addition
X3		distortion
X4		sl-intrusion
X5		tl-intrusion
X6		indecision
X7	lexis (L)	incorrect-term
X8		inappropriate-collocation
X9	grammar (G)	syntax
X10		preposition/particle
X11		inflection
X12		spelling
X13		punctuation
X14	text (T)	tl-inappropriate-register
X15		awkward-style
X16		cohesion

the same time reviewing the validity of the criteria introduced from the first step.

Note that we did not concern ourselves here with inter-assigner agreement, as at this stage what matters is to validate the coverage and granularity of the error scheme and to establish guidelines for assigning error types to error instances. A total of 709 errors were identified in all the translations. Among them, 3 error tokens were categorized as undecidable. The other 706 errors were classified under one of the error types. Below are the three undecidable errors.

Case 1:

(ST) When Americans truly know about how much pain and suffering the US air strikes have caused ... they will reject this devastating targeted killing program.

(TT) 米国は国がした空襲によってどのくらいの苦痛をもたらしたかということをしっかり分かったら、悲惨な標的殺害計画を終わらせるだろう。

The literal translation of "Americans" is "アメリカ人." However, it is not easy to judge whether this should be classified as an error within this particular translation. If it is, then where is it classified?

Case 2:

(ST) The reality is, racial discrimination and segregation go hand in hand.

(TT) 現実は，人種差別と人種隔離が表裏一体であることだ。

This TT uses "da/dearu" while the remainder of the translation text basically adopts "desu/masu" style. It is clear that this error falls under "text" related errors, but we need to decide which error type should be used.

Case 3:

(ST) When she met Nelson Mandela after his release from prison, ...

(TT) 彼女はネルソン・マンデラが出所した後出会った際に、…

The Japanese "出所" has a negative connotation, and thus may not be good to use for the case of Nelson Mandela, as he was unjustly and wrongly imprisoned. For now we cannot decide sensibly which error type should be assigned to this case (though it may well be a case of content distortion - X3).

These case are left pending. Separately, it should be noted that some of the error type assignments can only be done with reference to "project-dependent" resources such as glossaries specifically made for that particular translation tasks.

All in all, we can give a positive answer to the first question we raised in the Introduction and conclude that the modified MeLLANGE error scheme is adequate for capturing all errors in English-to-Japanese translations, although in-depth understanding of the error types may well be required to use the error scheme consistently. The fact that error types could be successfully assigned to error instances in the translations made by both graduate and undergraduate students shows that it can be used in a variety of situations in translation and translation training. As such, our experiment partly validates the original aim of the MeLLANGE error scheme, which aims "to create a language independent product which will fit to as many translation institutions as possible" (Secară 2005).

4 Distribution of error types

Here we observe the quantitative nature of error groups and error types, in order to answer the second question raised in the Introduction. While the differences in errors made by different groups is not our main concern here, we sometimes refer to the difference between translations made by graduate and undergraduate students[4].

4.1 Error groups

Table 2 shows the number of errors as seen from the four error groups of content (C), lexis (L), grammar (G) and text (T). "UD" stands for errors whose types were undecidable. While errors related to content (C) account for almost half of all errors (49 per cent), other error groups are also observed, with lexis-

[4]There is no guarantee that each of these groups constitutes a coherent group from the point of view of competence in translation or language skills, but from the educational point of view they do, by definition.

Table 2. Number of errors (per 1,000 words) by error groups

	C	L	G	T	UD
A	45.45	26.17	20.66	15.11	0
B	31.9	23.31	4.91	3.68	0
C	52.08	20.83	14.32	0	1.3
D	41.42	4.73	7.1	1.18	0
E	14.3	11.7	3.9	1.3	0
F	53.61	15.15	8.16	4.66	0
G	14.36	2.61	13.05	2.61	0
H	33.25	25.86	14.78	3.69	0
I	3.17	0	0	6.35	0
J	43.8	26.76	9.73	14.6	0
K	38.19	14.32	9.55	0	0
L	32.83	15.15	7.58	5.05	0
M	19.52	22.3	7.43	4.65	0.93
N	11.85	11.85	2.37	0	0
O	14.25	48.43	2.85	8.55	0
P	31.33	7.23	16.87	4.82	0
Q	6.78	1.36	0	0	1.36
R	27.2	10.36	10.36	2.59	0
S	12.72	5.78	6.94	2.31	0
Mean	27.79	15.49	8.45	4.27	0.19
(%)	49.46	27.57	15.04	7.6	0.33
Sdev	15.13	11.49	5.49	4.29	0.44
Cval	0.54	0.74	0.65	1	2.32

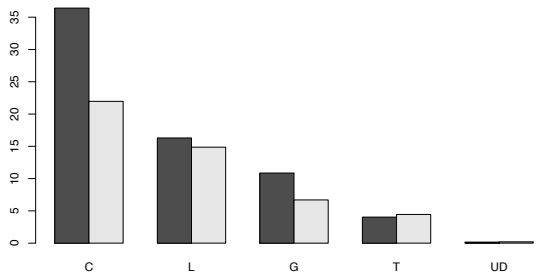

Figure 1. Mean errors by error groups for undergraduate (black) and graduate (gray) students

related errors accounting for 28 per cent, grammar-related errors 15 per cent, and text-related errors 7 per cent. The coefficient of variation (Cval in Table 3) indicates that the number of text-related errors varies widely among translators while content-related errors vary much less (assuming that we can take the range of random variation as constant).

Figure 1 shows the number of errors by error groups for translations made by graduate and undergraduate students. Although the difference between graduate and undergraduate students is not our main concern in this paper, it is interesting to note that the average numbers of content-related and grammar-related errors are substantially higher for undergraduate students than for graduate students.

4.2 Error types

Error types, rather than the error groups we just saw, constitute the most important level, as it is these which are assigned to individual error instances. Table 3 shows the normalised quantity of the sixteen error types and undecidable errors. Figure 2 shows the mean number of errors for all, undergraduate and graduate translators.

Slightly more than a quarter of all errors are classified as content distortion (X3), which most probably reflects the misunderstanding of source texts. This is followed by incorrect usage of terms (X7: 16 per cent), inappropriate collocations (X8: 13 per cent) and intrusion of source language in content (X4: 11 per cent). They together constitute two thirds of all errors. Content error by target language intrusion (X5) was not observed at all, while content indecision (X6) was observed only in a very small number.

All in all, while a relatively small number of error types account for a large portion of actual errors, other less-frequently observed error types were clearly identified. We can thus reasonably conclude that the modified MeLLANGE error types can sufficiently capture errors for English-to-Japanese translations.

Figure 2 shows some interesting points. While the 3.99 overall mean error for undergraduates is much higher than the 2.79 for graduates (the overall mean is 3.29), graduate translators tend to make certain types of errors more frequently than do undergraduates, i.e. content addition (X2), collocation (X8), inappropriate register of TL text (X14), and awkward style (X15). This may reflect the fact that, in general, graduate students tend to have less difficulty in understanding SL texts, while tending to introduce inappropriate expressions in the process of producing TL texts. Although full investigation of the factors leading to these quantitative tendencies remains one of our future tasks, the fact that the quantitative observation of error types enables us to raise this kind of questions indicates that the error types are very useful for diagnosing translations.

5 Conclusions and outlook

This paper reports the result of our experiment in assigning modified MeLLANGE error types to Japanese translations of English texts. In relation to the main questions we raised in the Introduction, we found that errors can be assigned to error types almost exhaustively. Moreover, although we can observe a few error types which account for a large portion of errors, the scheme appears useful for discriminating between a range of translation errors in En-

Table 3. Number of errors (per 1,000 words) by error types

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	UD
A	9.64	0	24.79	11.02	0	0	20.66	5.51	1.38	1.38	2.75	6.89	8.26	0	0	15.15	0
B	4.91	1.23	20.86	9.82	0	0	4.91	18.4	0	2.45	0	0	2.45	0	0	3.68	0
C	5.21	1.3	22.14	23.44	0	0	18.23	2.6	5.21	1.3	2.6	1.3	3.91	0	0	0	1.3
D	5.92	1.18	27.22	7.1	0	0	3.55	1.18	0	1.18	0	1.18	4.73	0	0	1.18	0
E	0	1.3	9.1	3.9	0	0	6.5	5.2	0	0	0	2.6	1.3	0	0	1.3	0
F	20.97	1.17	22.14	9.32	0	0	12.82	2.33	2.33	0	1.17	1.17	3.5	2.33	1.17	1.17	0
G	5.22	0	6.53	2.61	0	0	2.61	0	1.31	3.92	1.31	0	6.53	0	0	2.61	0
H	7.39	0	18.47	7.39	0	0	17.24	8.62	1.23	4.93	0	7.39	1.23	1.23	1.23	1.23	0
I	0	0	3.17	0	0	0	0	0	0	0	0	0	0	0	0	6.35	0
J	2.43	0	21.9	19.46	0	0	12.17	14.6	2.43	0	0	2.43	4.87	7.3	2.43	4.87	0
K	4.77	0	28.64	4.77	0	0	4.77	9.55	0	0	0	2.39	7.16	0	0	0	0
L	0	5.05	27.78	0	0	0	10.1	5.05	2.53	0	2.53	0	2.53	2.53	0	2.53	0
M	3.72	0	11.15	4.65	0	0	9.29	13.01	1.86	1.86	0.93	0.93	1.86	1.86	0	2.79	0.93
N	4.74	2.37	2.37	2.37	0	0	4.74	7.11	0	0	0	2.37	0	0	0	0	0
O	0	5.7	5.7	0	0	2.85	25.64	22.79	0	0	0	0	2.85	0	2.85	5.7	0
P	2.41	2.41	26.51	0	0	0	0	7.23	2.41	0	2.41	4.82	7.23	2.41	0	2.41	0
Q	1.36	1.36	4.07	0	0	0	1.36	0	0	0	0	0	0	0	0	0	0
R	3.89	1.3	11.66	10.36	0	0	6.48	3.89	0	2.59	0	1.3	6.48	2.59	0	0	1.3
S	0	0	6.94	5.78	0	0	3.47	2.31	3.47	0	1.16	0	2.31	1.6	0	1.16	0
Mean	4.35	1.28	15.85	6.42	0	0.15	8.66	6.81	1.27	1.03	0.78	1.83	3.54	1.15	0.4	2.74	0.19
(%)	7.71	2.27	28.08	11.37	0	0.27	15.34	12.06	2.25	1.82	1.38	3.24	6.27	2.04	0.71	4.85	0.34
Sdev	4.76	1.61	9.21	6.36	0	0.64	7.17	6.26	1.46	1.46	1.03	2.21	2.56	1.78	0.85	3.48	0.43

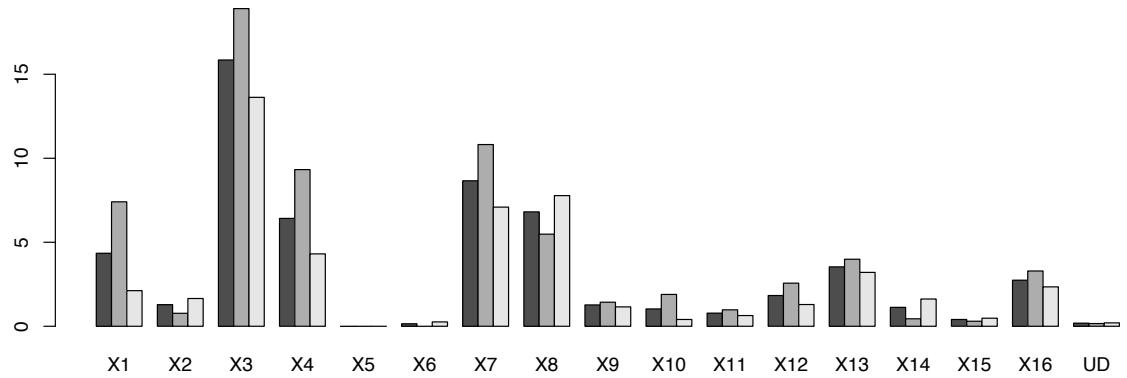

Figure 2. Mean errors by error types for all (black), undergraduates (dark gray) and graduates (light gray)

glish to Japanese translation. We are currently producing a guideline which uses examples to explain how to assign error types, to be published as one part of the MNH-TT user guide. At the same time, we are experimenting with the applicability of the error types to other language pairs such as Japanese and German.

Acknowledgement

This study is partly supported by JSPS Grant-in-Aid (A) 25240051 "Archiving and Using Translation Knowledge to Construct Collaborative Translation Training Aid System".

References

Abekara, T. and Kageura, K. (2007) "A translation aid system with a stratified lookup interface," *Proc. 45th ACL Poster and Demo*. pp. 5–8.

Babych, B. et. al. (2012) "Scaffolding, capturing, preserving interactions in collaborative translation," *LTTC International Conference*.

Secară, A. (2005) "Translation evaluation: A state of the art survey," *Proc. eCoLoRe/MeLLANGE Workshop*. pp. 39–44.

Utiyama, M. et. al. (2009) "Minna no Hon'yaku: A website for hosting, archiving, and promoting translations," *31th Translating and the Computer*.

White, J. (2003) "How to evaluate machine translation," in H. Somers (ed.) *Computers and Translation*, Amsterdam: Benjamins.