English-to-Japanese Translation vs. Dictation vs. Post-editing: Comparing Translation Modes in a Multilingual Setting

Michael Carl[1,2], Akiko Aizawa[2,3] and Masaru Yamada[4]
[1]CRITT, Copenhagen Business School, Copenhagen, Denmark
[2]NII, National Institute of Informatics, Tokyo, Japan
[3]Tokyo University, Tokyo, Japan
[4]Kansai University, Osaka, Japan
E-mail: mc.ibc@cbs.dk, aizawa@nii.ac.jp, yamada@apple-eye.com

Abstract

Speech-enabled interfaces have the potential to become one of the most efficient and ergonomic environments for human-computer interaction and for text production. However, not much research has been carried out to investigate in detail the processes and strategies involved in the different modes of text production.
This paper introduces and evaluates a corpus of more than 55 hours of English-to-Japanese user activity data that were collected within the ENJA15 project, in which translators were observed while writing and speaking translations (translation dictation) and during machine translation post-editing. The transcription of the spoken data, keyboard logging and eye-tracking data were recorded with Translog-II, post-processed and integrated into the CRITT TPR-DB[1], which is publicly available under a creative commons license. The paper presents the ENJA15 data as part of a large multilingual Chinese, Danish, German, Hindi and Spanish translation process data collection of more than 760 translation sessions. It compares the ENJA15 data with the other language pairs and reviews some of its particularities.

Keywords: translation dictation, translation process research, multi-lingual translation corpora

1. Introduction

Translation dictation is a mode of translation by which a translator reads a source text and speaks out its translation in the target language, rather than typing it. Translation dictation is thus a method of translation situated in between interpretation, where the interpreter hears a text and speaks out the translation (e.g., during conference interpreting) and conventional translation by which a written source text is translated mainly using the keyboard. It is close to sight translation, but while sight translation is usually done in the moment; there are – in principle - no time constraints in translation dictation. Translation dictation was used in some translation bureaus in the 1960s and 1970s (Gingold, 1978) but it has been used less frequently since the mid-80s, as professional translators started using micro-computers (Zapata and Kirkedal, 2015). Already the ALPAC report (Pierce et al., 1966) mentioned that "productivity of human translators might be as much as four times higher when dictating" as compared to writing. Others (e.g. Reddy and Rose, 2010, Rodriguez et al., 2012) are less optimistic about the time efficiency of dictation, but with increasing quality of voice recognition this mode of translation is becoming a valid alternative to 'conventional' translation typing (Ciobanu, 2014) and even to machine translation post-editing. See also Martinez et al, (2014) who experiment with integrating speech recognition into an online CAT system.
While translation dictation in the 1960s and 1970s was spoken on tape and transcribed by a (monolingual) typist, the usage of Automatic Speech Recognition (ASR) systems provides today an efficient means to produce texts directly without the need for an extra typist. Our experiments suggest that for some translators and types of texts it might become even more efficient than post-editing of machine translation.
In this paper we describe the ENJA15 translation study and a corpus of translation process data. The corpus was collected as a collaborative effort by CRITT and NII and is part of a bigger multilingual TPR data set which enables us to compare human translation production processes across different languages and different translation modes, including from-scratch translation, machine translation post-editing and translation dictation. We will first present the ENJA15 study and then compare the English-to-Japanese translation with other language pairs that are part of the multilingual TPR-DB subset.

2. The ENJA15 study

The ENJA15 study is part of a larger multilingual translation corpus in which six short English texts are translated under various different conditions into a number of different target languages. Each of the six English source texts has approximately 110 -160 words. Four of the texts are from a news domain and two from a sociology encyclopedia. The user activity data was recorded (keystrokes, gaze data, spoken translation) with Translog-II (Carl, 2012), and with an SMI eyetracker at 60 and 120Hz. The collected data was anonymized and post-processed as described in Carl et al. (2016), and is publicly available under a creative commons license in the CRITT TPR-DB.
The ENJA15 translation study extends the multilingual translation corpus, adding data for the language pair English → Japanese. The ENJA15 translation experiment consists of three different conditions:

1. from-scratch translation (T)
2. translation dictation (D)
3. MT post-editing (P)

Participants translated two texts in each of these conditions, in the order of the list above. For machine translation post-editing, we used Google translate (from August 2015) and pasted the MT output into the Translog target window. For translation dictation, translators used an automatic speech recognition (ASR) system, Nuance Naturally Speaking.

[1] See CRITT homepage at https://sites.google.com/site/centretranslationinnovation/

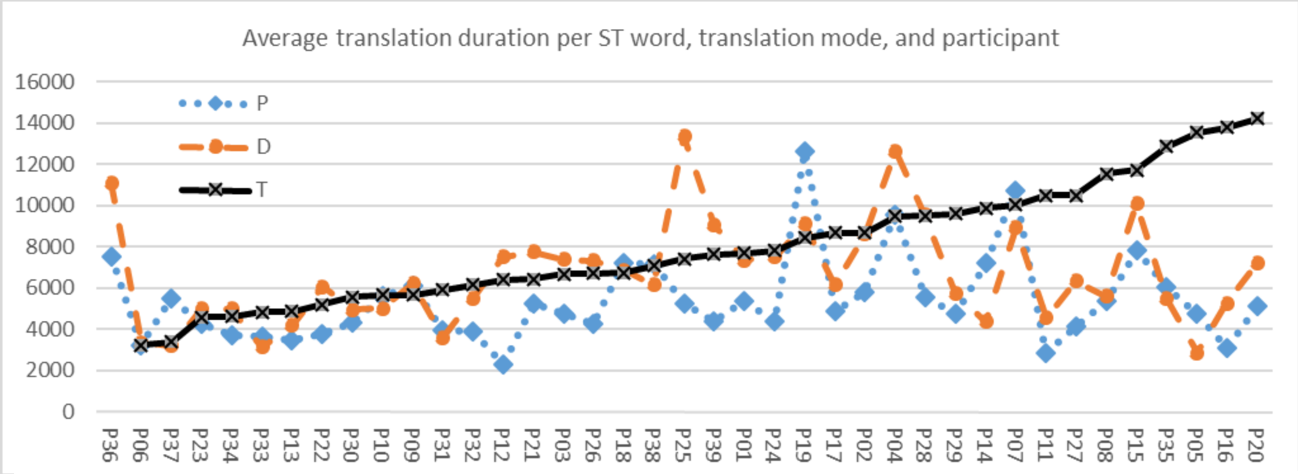

Figure 1: translation durations in the ENJA 15 experiment

The ASR system was trained for each translator prior to first translation dictation task. Training took approx. 10 minutes.

Translators were advised to produce a 'good enough' translation for publication without spending too much time on terminological or stylistic subtleties (Mesa-Lao, 2014). Translators were told not to use external help (lexica, concordance tool, etc.) during their translations and instead to concentrate only on the screen, since otherwise we would lose track of their gaze. Translators were also asked to fill out a meta-data form to keep track of their translation experiences (years of formal training, years as active translator, attitude and experience in post-editing, etc.).

The order of the translation modes remained identical, but the texts were permuted, with the goal of obtaining an equal number of translations for each text in each translation mode. The time needed to complete the translation of six texts was not restricted but usually took between 2 to 3 hours. Participants were remunerated between 4000 and 6000 yen (approx. 30€ and 45€), depending on their experience. Participants were made familiar with the goals of the translation experiment, and they signed a form in which they agreed that their translation data would be made publicly available under a creative commons license. They also out filled two questionnaires, one before starting the translation session and another after having finished. Questionnaire 1 contained questions concerning expertise of the participant, years of translation experience, post-editing experience and experience with speech recognition, etc. Questionnaire 2 was to be filled after the experiment and contained questions concerning satisfaction with the three translation modes, and an estimation of the effort used in each of the translation modes.

3. Participants

39 translators participated in the ENJA15 study. All participants had Japanese as their first and English as their second language and reported between 0 and 22 years of translator experience. According to the information given, 14 of the translators had 10 years or more translation experience and 17 translation students had a year or less experience. 20 translators reported to have no experience with post-editing, two translators said to use it every day, and another 12 at least once a month, with a level of satisfaction on a 5-point Likert scale ranging from “highly dissatisfied” to “highly satisfied”. Only two participants reported to have previously used voice recognition for translation while six used it in another context.

4. Word production times for EN → JA

An evaluation of the data was conducted with respect to the productivity of the three translation modes (T,P,D) where we found that, in most cases, translation dictation and post-editing are quicker than from-scratch translation. Figure 1 plots the average word production durations for the three translation modes and all 39 participants, sorted by the average translation duration. Average translation times span from approximately 3.5 sec (3500ms) per ST word to more than 14 sec (14000ms). The graph suggests that translators who translate more quickly (on the left) also tend to post-edit and dictate more quickly. Some translators seemed to have particularly strong problems with the ASR system (P36, P25, P04) and needed more than 10sec and 12sec. respectively per word. Others seem to struggle more with post-editing (e.g. P19), and show much worse performance than in the translation mode. However, it seems that the more time a translator needs for translation, the more likely he or she will be quicker with post-editing and dictation.

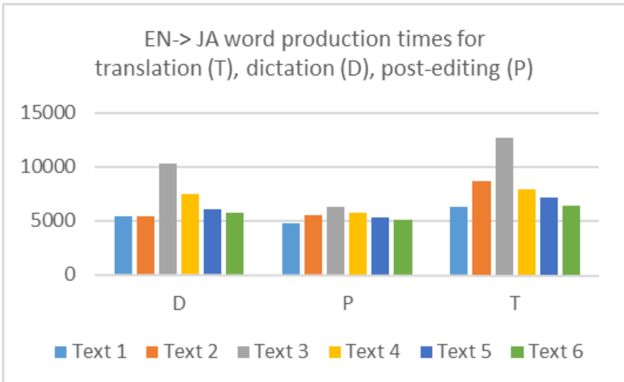

Figure 2: Production durations for different translation modes

Figure 2 shows the average translation durations per text for the three translation modes. The graph shows that different texts require – on average – different translation times. Most time consuming (per word) is Text 3, interestingly so in all three translation modes. That is, taking the per word translation time as an indicator for translation difficulty, the graph suggests that relative difficulty remains (more or less) similar across the different translation modes.

	Translation (T)						Post-editing (P)						Monling. Editing (E)						.
Text	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	Tot
da	24	23	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	69
de	7	8	8	8	8	8	8	7	8	8	7	8	7	8	8	7	8	8	139
es	11	11	12	10	12	8	10	12	10	12	8	12	10	9	10	10	10	11	188
zh	-	3	3	3	3	3	3	5	3	3	3	2	-	2	1	4	2	2	45
hi	7	7	6	7	6	6	8	12	8	10	12	11	Dictation (D)						100
ja	13	13	13	13	13	13	13	12	14	12	12	12	14	14	13	12	13	14	233
Total	62	65	64	41	42	38	42	48	43	45	42	45	31	33	32	33	33	35	775

Table 1: Number of alternative translations for each of the six English different source texts translated into six different target languages under different translation modes.

5. ENJA15 within the TPR-DB multilingual translation corpus

The ENJA15 study is part of a larger multilingual translation corpus in which six short English texts are translated under various different conditions into a number of different target languages, which so far also include Chinese, Danish, German, Hindi, Spanish, and now also Japanese. The goal of the multilingual translation corpus is to gather translators' activity data (text perception and production behaviour, as recorded by keystroke loggers, eye-trackers, etc.) in order to investigate variations in the human translation process across different translator profiles, translation modes and different target languages.
To date, experimental data has been collected from more than 150 different translators in more than 760 translation sessions, which accumulate to more than 140 hours of translation data, together 108053 ST tokens and 122323 TT tokens. Some knowledge has already been generated from this corpus which is, among other outlets, reported in an edited volume (Carl et al., 2016).
Table 1 gives an overview over the number of translation sessions collected in the different languages. The data is collected from various translations experiments between 2008 and 2016. All sessions were recorded with Translog (Jacobsen, 1999, Carl, 2012). Note that not all texts are translated under all conditions into every target language. For instance, for Danish (da) only three texts (Texts 1, 2 and 3) were from scratch translated (henceforth simply translated) by 24 different translators. For Hindi (hi) there is only translaton and post-editing data, and for Japanese (ja), we have translation, post-editing and dictation data. Also, not all translations are kept in the TPR-DB, for various reasons, mostly due to incomplete translations or to logging errors. In the monolingual editing condition[2] (E), machine translation was shown to post-editors without access to the original source text. These data were collected for German (de), Spanish (es) and Chinese (zh); the Japanese

6. A Comparative Cross Lingual Analysis

6.1 Translation durations

A comparative analysis of translation durations for the six texts shows quite different translation times for the six different languages. Figures 3 and 4 plot average production durations per word for translation and post-editing respectively. The languages are ordered according to the average translation duration for the texts. The figures show that:

- Provided that the average word translation time represents translation difficulty, we may assume that the translation into different languages correlates with quite different degrees of translation difficulty. According to our data, easiest (i.e. quickest) is the translation from English into Danish, followed by Spanish and German. More difficult are the non-European languages, Japanese, Chinese and Hindi. Translations of the same English texts into Hindi require approximately 4 times longer than into Danish or Spanish.
- On average, post-editing is quicker than translation, for all languages. With the exception of Chinese, the order of the languages (viz. the degree of the average word translation duration) is identical for translation and for post-editing. The variance in the amount of time needed to post-edit is smaller than for translation.
- The translation difficulty of a text seems to be independent from the target language: average word translation duration for text 3 is highest in the different target languages during translation and in most cases also during post-editing.

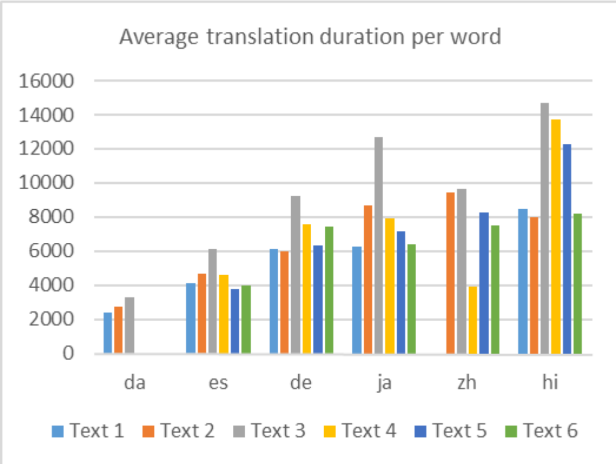

Figure 3: Word production durations in ms for different texts and different languages

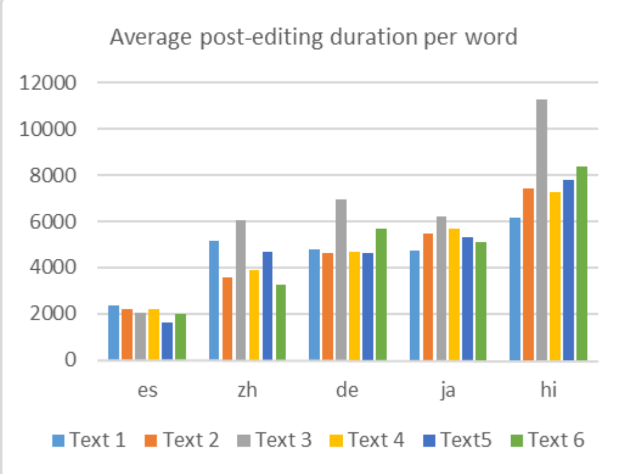

Figure 4: Word post-editing durations for different languages

[2] In this study we will not consider monolingual editing data, which amounts to 16285 ST and 17538 TT tokens

6.2 Gaze data

Previous findings showed that during translations total reading times in the target text are longer than in the source texts (Balling and Carl 2014). This finding could only partially be reproduced with the language pairs in our data. Figure 5 shows that average gaze durations during translation were longer on the ST[3] for the English → Japanese and English → Chinese data.

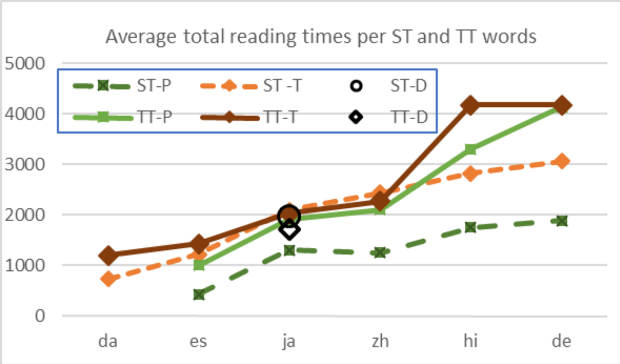

Figure 5: Gazing behavior on source and target words

Gaze durations during post-editing are 1. shorter than during translation and 2. much longer in the target text (TT) than in the source text (ST). When dictating, the gaze seems mostly fixated on the source text while during post-editing it is more often on the target text.

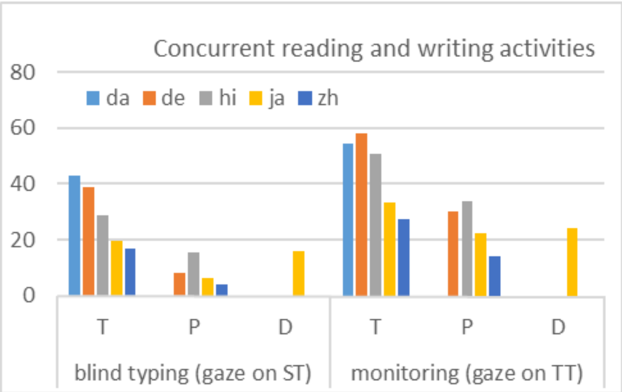

Figure 6: Concurrent reading and writing activities

6.3 Concurrent reading and writing activities

Concurrent reading and writing activities during translation production is a skill that develops over time and is a strong indicator for the translator's expertise (Martínez-Gómez et al. 2014). From the 91768 ST tokens and 104785 TT tokens in the T,P and D translation sessions under investigation only 13110 (14%) showed simultaneous gazing and typing activities. While typing activities contribute to the production of the target text, the gaze may hover over the source- and/or[4] the target text. The former activity could be referred to as blind typing while the latter one amounts to text production monitoring. Figure 6 shows different distribution of blind typing and monitoring for the different languages and production modes. It shows the percentage of words in which concurrent reading and writing activities was detected:

- Monitoring activities are more frequent than blind typing (scanning the source text while typing the translation)
- We observe a higher percentage of monitoring and blind typing during translation than during post-editing
- Highest amount of blind typing and monitoring is observed for Danish and German, the least amount for Chinese and Japanese[5]
- There is a substantially larger amount of 'blind typing' during Japanese translation dictation (16%) than during post-editing (6%). However, Japanese 'dictators' also seem to frequently monitor the emerging spoken word transcriptions, 23% of the cases, presumably to control the accuracy of the speech recognition.

6.4 Translation pauses

Pauses during the translation production have been associated with cognitive effort (e.g., Schilperoord, 1996, O'Brien, 2006). Typing pauses are defined by a lag of time beyond a given threshold which occurs between two successive keystrokes. Lacruz & Shreve (2014) introduced the pause to word ratio (PWR) as a metric for cognitive effort in post-editing and suggest a pause threshold of 300ms to correlate well with other measures of cognitive effort. The PWR analysis in Figure 7 is based on an inter-key threshold of 300ms. The higher the PWR value, the more cognitive effort is suspected to be spent. Note that the ordering of the languages in Figure 7 is similar to the previous Figures 3 and 4[6]. While for Danish (da) approximately only 1 pause > 300ms occurs per word, there are on average more than 4 such pauses for Hindi (hi). Values for PWR are lower for post-editing than for translation, for all target languages. Note also that PWR is very similar in the post-editing and dictation mode for Japanese translation.

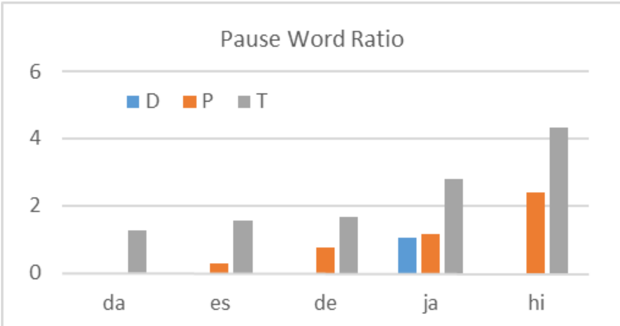

Figure 7: Comparison of PWD across different languages

[3] Spanish gaze data for participants P12-P32 was taken out due to problems in the eye-tracker logging data

[4] During the production of one target word, the gaze can go back and forth between the source and the target text.

[5] This may be partially due to problems of gaze-to-word when using IME input, since the gaze location is at a different location than the text that appears in the editor.

[6] Chinese has been taken out due to problems for computing PWR based on the IME input

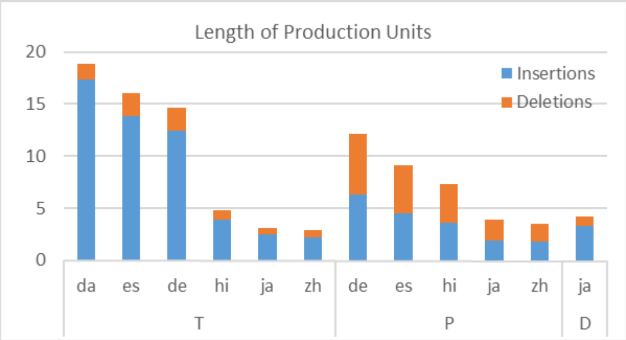

Figure 8: Length of production units in different languages

6.5 Production units

In addition to analyzing the pausing structure, we also examine length of coherent text production activities. There is some discussion how to define the length of inter keystroke pauses which separates successive typing bursts; we take it – with O'Brien (2006) - that "1 second is appropriate for observing delays in a text production event", which is also the measure adopted in the TPR-DB for the definition of production units (Carl & Kay, 2011). Production units (PUs), defined in this manner, consist of one or more keystroke. Longer production units indicate more fluent typing activity, while shorter units suggest a disruptive writing process, which may be due to difficulties in rendering the content (i.e. more meta-cognition) or less developed typing skills. Figure 8 shows length of PUs in terms of inserted and deleted characters for the six languages and three translation modes.

There is a large difference in the length of the PUs for the European languages (da, es, de) and the non-European ones (hi,ja,zh) in translation and to a lesser extent in post-editing[7]. While the percentage of deletions in post-editing is much higher than in translation for all languages, interestingly, there is not a big difference in the length of PUs for the non-European languages between post-editing and translation.

	Ins	Del
T	2.45	0.62
P	1.90	1.97
D	3.25	0.97

Table 2: Number of text modifications per Japanese Pus

The properties for Japanese PUs are reproduced in Table 2, which shows that the average number of insertions is highest during dictation. This is in line with previous study of Mees et al. (2015) who find that speaking "translations will encourage [students] to deal with larger units, and thus translate the overall meaning instead of individual words". Some participants reported that they translated longer chunks in the dictation mode than during from scratch translation, which most found an interesting effect but also cognitively more effortful. In a discussion after the experiment, one translator said:

"my brain seems to work in a different mode during translation dictation. I have the feeling I would need to better understand the source text before starting dictation so as to produce an 80% correct translation, whereas when typing I can already read ahead in the source text and delete or rearrange the translation more easily. In this sense I find translation dictation more effortful than from-scratch translation".

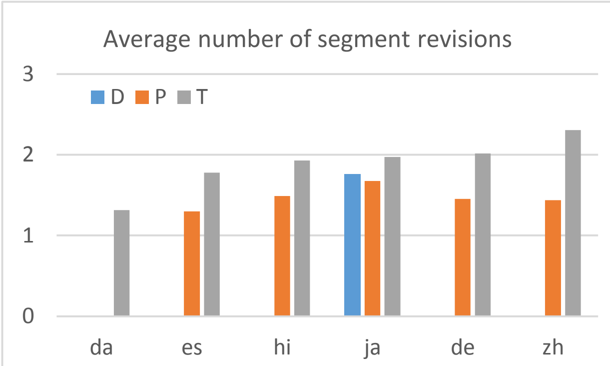

Figure 9: Revision of segments

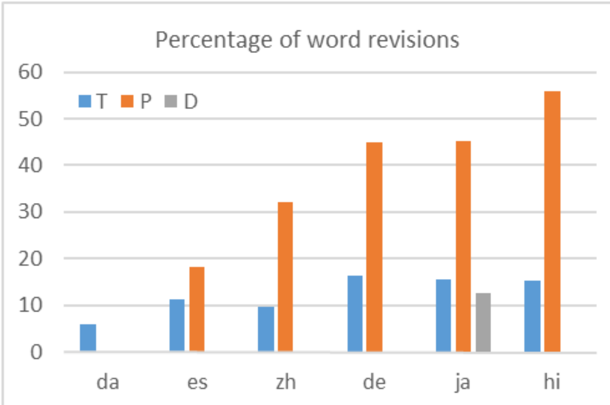

Figure 10: Revision of words

6.6 Translation revision

The number of revisions may also be considered an indicator for translation difficulty. The more translators come back to a word or a segment to revise it, the more one can expect the word or passage to be difficult and cognitively effortful to translate. Figure 9 plots the average number of revisions per segment for the six language pairs, where a segment revision is defined as a segment modification after another segment has been edited, i.e. the translator comes back to the same segment at a later point in time. Translation, drafting is considered to be a (the first) revision, so that each translated segment has at least one revision. In the case of post-editing, unmodified segments count a zero revisions. Under this definition, translations have, on average, slightly more segment revisions than post-edited data.

Interestingly, this seems to be very different when considering the percentage of word revisions, where translation drafting is not counted as revision. A much

[7] Note that a 'word' in Chinese (and also in Japanese) has, on average, less characters than in the European languages, and therefore 'meaning units' will be shorter.

higher percentage of words is revised during post-editing than during translation. Figure 10 suggests that more than 50% of the MT output is modified during Hindi post-editing while for Spanish this figure amounts to less than 20%. The percentage of word revisions in translation is between 6% for Danish and 16% for German. While these numbers might be explained by the relatively good MT quality for Spanish and quite bad MT output for Hindi[8], Japanese or even German, they also suggest that translators more often revise segments, while post-editors revise words (i.e. they don't frequently come back to previous segments). Note that during Japanese dictation even fewer words are revised than during Japanese translation, and also the number of segment revisions is lower. This might be related to and explained by the relatively longer production units that are generated during translation dictation, as mentioned above, and the felt need to "better understand the source text" before translation production. However, more research is required to arrive at a better understanding of the factors that might have an impact on these observations.

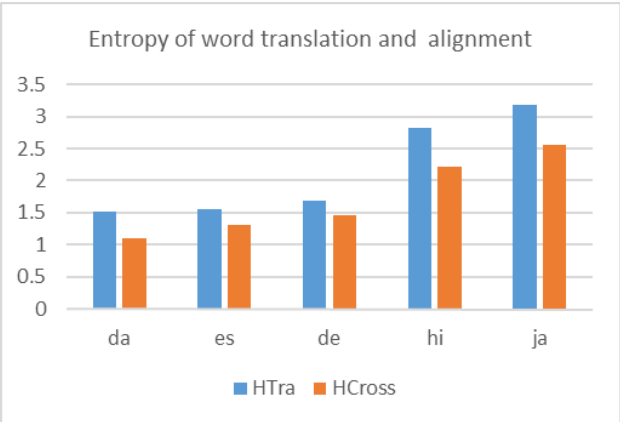

Figure 11: Literal translation

6.7 Literality of translation

Translations are considered more literal if the target text is more similar to the source with respect to the conceptual and syntactic structure. A 'free' translation deviates substantially from the structure of the source text but it may be easier to comprehend by the target audience. In a recent attempt to formalize translation literality, these components have been operationalized in terms of word translation entropy and the amount of crossing alignments (Carl et al 2016). Word translation entropy (*HTra*) measures the number of different translation realizations for a word in its context. HTra exploits the fact that all the translations in the data collection were manually word-aligned. As the data contains for each of the six English source texts up to 39 different translations, the distribution of word translation probabilities and their entropy can be computed for each source text word and every language pair. A word translation entropy of zero would mean that all translators choose the same translation for a word (and words are also aligned as such). The higher the word entropy, the more have translators produced different solutions and/or words were aligned in different way.

Crossing alignments measure the relative similarity of the word order in the source and the target text, by following the word alignment links. However, instead of measuring the distance of the relative ST-TT reordering, as suggested in (Carl et al 2016), here we measure the entropy of word re-ordering, (HCross). The value of *HCross* is zero if all translators choose the same word order in the translation, irrespectively of whether or not their order corresponds to the source text word order. HCross becomes bigger as different translators choose a different target language word order for the same sentence. While, in principle, the choice of a word could be independent from how they are ordered in the target language, our data suggests a strong correlation between HTra and HCross.

Figure 11 shows that target languages closer to English have lower HTra and HCross values (below 2) than Hindi and Japanese which are very different in structure and organization from the English source language. That is, translator seem to have more choice as how to render the target text (conceptually and procedurally) for more remote languages, as compared to the three European languages in our corpus, which are closer in terms of language and culture.

7. Conclusion and discussion

The paper presents a corpus of English-to-Japanese translation process data (ENJA) and compares it with a multi-lingual data collection within the CRITT Translation Process Research-Database (TPR-DB). Rather than an in-depth analysis, the paper provides a summary evaluation of the ENJA data by relating properties of English to Japanese translations with a corpus of alternative Chinese, Danish, German, Hindi and Spanish translations of the same English source texts under different translation conditions (from-scratch translation and MT post-editing).

The results of our study show that translating into different languages takes substantially different amount of time, implies different sizes of translation units and pausing structure, different number of word and segment revisions, different degree of concurrent reading and writing activities and different amount of variability (word translation choices) in the translation product. In many of these parameters, a sharp contrast can be observed between translation from English into the European languages (Danish, German, Spanish) and into the non-European languages (Chinese, Japanese, Hindi). However, this contrast seems to be weakened in the MT post-editing mode, where the MT quality may play an important role for the post-editing behaviour.

Particular emphasis is put on the evaluation of translation dictation, as one of the translation modes in the ENJA15 study. We find that translation dictation can be - for some texts and translators - as efficient as machine translation post-editing. Less word and segment revisions are observed during translation dictation than during from-scratch translation, the PWR metric indicates less cognitive effort for dictation than for post-editing, and we measure more gaze activity on the source and emerging target texts than during post-editing, indicating less distraction from the translation activity. Our findings confirm those of a

[8] MT output for the Hindi PEMT task was a combination of a google translate (2012) and output from Anglabharti (also from 2012).

previous study (Mees et al., 2015) who find that speaking "translations will encourage [students] to deal with larger units, and thus translate the overall meaning instead of individual words". While this is believed to lead to better translations, some translators find translation dictation more effortful than typing – presumably for the same reason. This might also explain an observation of Ciobanu (2014), who reports that "less experienced translators tend to stay away from ASR at the beginning of their careers", while "within the professional experience groups … ASR does have a positive impact on productivity": Translation students struggle often more with source text comprehension than expert translators, which may make it more difficult for them to overcome a word-by-word translation mode and to produce longer target text sequences, which are at the same time also crucial to reduce the ASR error rate and to enhance word recognition.
However, as outlined above, a number of parameters may play a role in a better acceptance and usage of the ASR technology in the translation community. As Ciobanu (2014) mentions, an important point is the proper integration of speech recognition into available CAT workbenches, the integration of dictionaries and other forms of interactivity. A beginning has been made by Martinez et al (2014) who combine speech recognition and cloud-based MT post-editing tools. We also agree with Ciobanu, in that "it is surprising that the research world has not invested much effort in investigating the impact that ASR can have on the quality of human translation output" and that ASR systems "have without doubt significant potential to make a positive contribution to the quality and ergonomics of the work of professional translators".
Since by far most of our translators in the ENJA15 study did not have any experience with translation dictation, we currently conduct a follow up study to investigate how and whether translators develop different behavioral patterns as they get more familiar with translation dictation. We therefore re-invited 7 translators (3 beginners and 4 more experienced translators who also participated in the ENJA15 study) to dictate 12 short English texts on six different days. We will record their behavioral data, the spoken language transcripts, keyboard interactions and gazing behavior and will eventually disseminate the data through the TPR-DB. We hypothesize that, within six days of using ASR for one hour each day, translators will develop particular patterns which allow them to use the tool more efficiently. We will assess and quantify to what extent this can be observed in the differently experienced user groups.
From a cognitive point of view, a comparative study of translation, dictation, and post-editing may give us a differentiated picture into the diversity of human translation processes. Given that more of less automatized or conscious mental translation processes are activated at different points in time (Schaeffer and Carl, 2013), and based on the different behavioural patterns for the different languages that we report in this paper, we take it that translation into linguistically and culturally more remote languages involves different mental processes than translating into closer languages. Similarly, from-scratch translation involves different sets of translation processes than MT post-editing or translation dictation. The nature of these processes, the extent to which they overlap, their conditioning factors and behavioural consequences still need to be explored, as well as appropriate working environments and combination possibilities of the different translation modalities.

8. References

Balling, Laura Winther; Carl, Michael. 2014. Production Time Across Languages and Tasks: A Large-Scale Analysis Using the Critt Translation Process Database. The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science. ed. / John W. Schwieter; Aline Ferreira. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014. p. 239-268

Carl, Michael. 2012. Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research, Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)}, 2012, Istanbul, Turkey, Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Thierry Declerck and Mehmet Uğur Doğan and Bente Maegaard and Joseph Mariani and Asuncion Moreno and Jan Odijk and Stelios Piperidis, European Language Resources Association (ELRA)

Carl, Michael; Kay, Martin. 2011. Gazing and Typing Activities during Translation: A Comparative Study of Translation Units of Professional and Student Translators. In: Meta, Vol. 56, No. 4, 2011, p. 952-975.

Carl, Michael; Moritz Schaeffer; Srinivas Bangalore. 2016. The CRITT Translation Process Research Database. In: New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB. . ed. /Michael Carl; Srinivas Bangalore; Moritz Schaeffer. Cham : Springer, p. 13-54

Ciobanu, Dragoş. 2014, Of Dragons and Speech Recognition Wizards and Apprentices, Número 12, Traducció i qualitat, Revista Tradumàtica: tecnologies de la traducció. desembre 2014 . ISSN: 1578-7559 https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12p524.pdf

Gingold, Kurt. 1978. "The Use of Dictation Equipment in Translation." In La traduction, une profession. Actes du VIIIe Congrès mondial de la fédération internationale des traducteurs, edited by Paul A. Horguelin, 444–448. Ottawa: Conseil des traducteurs et interprètes du Canada

John R. Pierce, John B. Carroll, et al., Language and Machines — Computers in Translation and Linguistics. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC, 1966.

Lacruz, I., and Shreve, G. M. 2014. Pauses and Cognitive Effort in Post-Editing. In O'Brien, S., Balling, L. W., Carl, M., Simard, M., and Specia, L. (Eds.). Post-editing of Machine Translation: Processes and Applications. Cam-bridge Scholars Publishing

Martinez, Mercedes Garcia; Singla, Karan; Tammewar, Aniruddha; Mesa-Lao, Bartolomé ; Thakur, Ankita; Anusuya, M. A.; Srinivas, Banglore; Carl, Michael. 2014. SEECAT: ASR & Eye-tracking Enabled Computer Assisted Translation. Proceedings of the 17th Annual Conference of the European Association for Machine Translation: EAMT2014. ed. / Marko Tadić; Philipp

Koehn; Johann Roturier; Andy Way. Basel: European Association for Machine Translation, 2014. p. 81-88.

Martínez-Gómez, Pascual; Minocha, Akshay; Huang, Jin; Carl, Michael; Bangalore, Srinivas; Aizawa, Akiko. 2014. Recognition of Translator Expertise using Sequences of Fixations and Keystrokes. Procedings of Symposium on Eye Tracking Research and Applications. ed. / Pernilla Qvarfordt; Dan Witzner Hansen. New York : Association for Computing Machinery, 2014. p. 299-302.

Mees, Inger, Barbara Dragsted, Inge Gorm Hansen, and Arnt Lykke Jakobsen. Sound effects in Translation. In Interdisciplinarity in interpretation and translation Process Research, Ehrensberger-Dow, Goepferich and O`Brien (eds). 2015, Benjamins Current Topics,

Mesa-Lao, Bartolomé. 2014. Gaze Behaviour on Source Texts: An Exploratory Study comparing Translation and Post-editing. In Post-editing of Machine Translation: Processes and Applications Edited by Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia, Cambridge Scholars Publishing pp 219 - 246

O'Brien, S. 2006. Pauses as Indicators of Cognitive Effort in Post-editing Machine Translation Output. Across Languages and Cultures, 7(1), 1-21.

Reddy, A and R. C. Rose. "Integration of Statistical Models for Dictation of Document Translations in a Machine Aided Human Translation Task". IEEE Transactions on Audio, Speech, and Language Processing, 18(8): 2015-2027. 2010.

Rodríguez, L, A. Reddy, R. Rose. "Efficient integration of translation and speech models in dictation based machine aided human translation". Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. pp. 4949 - 4952. 2012

Schaeffer, Moritz and Carl, Michael. 2013. Shared Representations and the Translation Process: A Recursive Model. In: Translation and Interpreting Studies, Vol. 8, No. 2, 2013, p. 169–190

Schilperoord, J. 1996. It's About Time: Temporal Aspects of Cognitive Processes in Text Production. Amsterdam: Rodopi.

Zapata, J & Kirkedal, AS 2015, 'Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows'. in B Megyesi (ed.), *Proceedings of the 20th Nordic Conference of Computational Linguistics.* Linköping University Electronic Press, Linköping, pp. 201-210. NEALT (Northern European Association of Language Technology) Proceedings Series, vol. 23