

Andrew Innes

Mukogawa Women`s University, Japan
andrew6_@hotmail.com
<https://orcid.org/0000-0002-6080-0566>

Differentiating between machine translation and student translation: red flags and salient lexicogrammatical features

ABSTRACT

Machine translation enables students to produce work in the target L2 which may be superior to that which they could produce otherwise. The present study examines whether the use of machine translation can be detected by teachers. Seventeen native teachers compared and assessed the authorship of five human translations (HT) and five machine translations (MT) of Japanese news stories. Findings suggest that native teachers were able to detect which essays had been written using HT and which had been written using MT.

Keywords: SFL, machine translation.

Introduction

Negative sentiments of machine translation (MT) are borne out by Van Praag` who view it as a `challenge to [their] knowledge and expertise`, and a `nuisance and distraction`. Simply banning MT has been found to be ineffective as students will use it regardless (Kazemzadeh & Kashani, 2014). While exact figures pertaining to student use of MT will perhaps be confounded by worries around personal disclosure, a study at Duke University found that more than 88% of L2 students admitted to having used it, with 77% of instructors being opposed to its use (Clifford, Merschel, & Reisinger, 2013, p. 44). Elsewhere, Briggs (2018, p. 13) found that 57.5% of Korean students strongly agreed with the statement, `I do not need to learn to write in English because [online] translators can do the work for me`. Indeed, the notion that using MT even constitutes cheating or plagiarism appears to be up for debate. Using a five-point Likert Scale ranging from strongly disagree to strongly agree, White and Heidrich (2013) received a mean of 3.59 when asking eighteen German students the degree to which they agreed with the statement, "I feel like I might have cheated" (p. 241).

The underlying epistemology for negative attitudes towards MT is perhaps better understood when viewed through the lens of the sociocultural paradigm espoused by Vygotsky (1978). Scaffolding refers to bridging the gap between what students can do on their own, and with the help of a more knowledgeable other. This largely goes against the cognitive apprenticeship model as elucidated by Gibbons (2008) which, “is particularly concerned with making thinking and the implicit processes of problem solving visible” (p. 169). Gibbons points out the importance of being “treated as apprentices in a disciplinary community, rather than as passive receivers of knowledge” (p. 170). It could be argued that the problem for educators wishing to utilize a Vygotskyian approach in the classroom is that MT turns students into the kinds of passive recipients which Gibbons alludes to. This is compounded by the fact that essays written in the student’s native language and subsequently machine translated, lack the cognitive engagement which takes place when the student does the translation by themselves.

The idea that MT has been five years away from being perfect for the last fifty years is often used to dismiss it (see for example, Lommel, 2019). News stories about botched machine translations serve to maintain the notion that MT is awkward, not to be trusted, and the source of much embarrassment (see for example, Sugiyama, 2019). Indeed, such criticisms are not limited to the lack of cognitive engagement, transference of skills to verbal communication, and sole use of the L1. While accepting that there are exceptions to the rule, Hall (1976) categorizes Japanese as a high context culture, suggesting that its message is often more implicit and less direct. This contrasts with the way in which trade languages, of which English is one, are required to be more specific and explicit due to the low degree of shared understanding (Hall, 1976). Given the status of English’s as a lingua franca, which has been estimated to be spoken by more non-native speakers than native speakers at a ratio of three to one, this is understandable (Crystal, 2003, p. 69). Hall’s concept dovetails with Davies and Ikeno (2002) discussion of the Japanese concept of ‘*aimai*’ (曖昧), which can be translated as ‘vague’, and is viewed as a concept which is not only tolerated, but also seen as a virtue of Japanese culture. While not explicitly using the term geographical determinism, Davis and Ikeno (2002) argue that the mountainous terrain of Japan formed communities into tight knit groups leading to a fear of ostracization, and a consequent hesitance to be too direct with one’s words.

Does MT have a place in the classroom?

Analysis has revealed that MT helps students with regard to tense choice, prepositions and “false friends” (Ebbert-Hübner & Maas, 2017). Garcia and Pena (2011) found that the lower a student’s ability was the higher their recourse to MT over writing words directly in the L2. The same study noted that blind marking indicated better results when MT was used despite a lower level of cognitive engagement as measured

through screen recordings. Elsewhere, Groves & Mundt (2015) demonstrated that the degree of MT accuracy is getting close to the minimum standard required for entry to many universities when measured using international testing standards, and able to produce work of similar accuracy as a mid-level L2 student.

This call for pragmatism over cognitive engagement is lent further support by Benda (2013) who points out how hiring decisions in Taiwan are not necessarily made on the basis of English ability, but on mere performance in exams. Benda suggests that we rethink the goals of English learning merely in terms of its social and cognitive elements, and reconsider them in light of the fact that many students may simply want to convey their point in the clearest and fit-for-purpose way possible. In a similar vein, White and Heidrich (2013) espouse The American Council on the Teaching of Foreign Languages (ACTFL) definition of technologically literate students in the 21st century as being, “productive global citizens [who] use appropriate technologies when interpreting messages, interacting with others, and producing written, oral, and visual messages” (p. 230). What is apparent therefore, is that there are perhaps two sides to the story of MT; those who view it as cheating and leading to a lack of cognitive engagement with the L2, and those who view it as a short-cut approach to getting the job done.

Purposes of the present study

The purpose of the present study is to investigate the extent to which native speakers of English can differentiate between a work produced by students, and a work produced via Google Translate.

Research questions

1. Can teachers differentiate between student translation and machine translation?
2. What linguistic evidence is this based on?

Literature review

While literature investigating the central theme of this investigation would appear to be a hitherto relatively uncharted territory, there is a sense that the task of differentiating between MT and student work is rather ambiguous, and points towards a current lack of reliable ways to identify translation plagiarism (see for example, Roberts, 2019). When comparing Google’s Neural Machine Translation (GNMT) system with human translators, Wu, Schuster, Chen, Le, & Norouzi (2016) found a wide distribution of ratings and cases of near identical phrasing and 60% fewer translation errors when compared with the previous phrase-based system human. The study’s conclusion indicated that raters had trouble distinguishing MT from HT. It should be noted however, that the source of some of this ambiguity was down to the differences in the ability of the translators

to completely understand the original language of the articles used in the study, leading to a degree of subjectivity. Such a view is supported by the sentiment that humans are still currently the best at detecting when MT has been used, and that, 'as of now, there are no reliable methods for spotting translation plagiarism' (Upwork.com). Conversely, Aharoni, Koppel, & Goldberg (2014) point out how distinguishing between MT and human translation is possible when examined at the level of such features as n-grams, function words, and the frequency of certain parts of speech. While this would appear to have some merit, it is important to point out that this particular study was carried out before the advent of neural machine translation in 2016 and compounded by the authors' caveat that differentiating MT from HT will become more difficult as MT itself becomes more sophisticated.

Indeed, the idea that MT could be considered 'sophisticated' does bear some weight. Even an early study by Lee and Liao (2011) found that MT helped reduce student errors when making translations and narrowed the gap between student proficiency. More recently, research by Google found that native speakers rated its translations at an average of 5.43 on a scale from 0 to 6 (McGuire, 2018). The launch of Google's Neural Machine Translation in September 2016 has heralded in a new era of accuracy which will arguably only get better with time (Schuster, Johnson, & Thorat, 2016). Based on deep learning through example-based machine translation, the system uses an artificial neural network which improves on the former system. Indeed, even without the improved system, early studies indicate that differentiating texts where MT has been used is not a straightforward task. In one study, 20 English texts were machine translated into Turkish before being edited, and then compared with 20 direct translations of the same text by professional translators. Using blind marking administered by four assistants, the texts were rated as being about the same in terms of overall acceptability (Çakır, 2013).

Given that both MT and L2 students will invariably make mistakes, it would appear that rationalizing the claim that a student has used MT may merely be down to a hunch. With regard to specifics, Somers et al., (2006), Williams (2006) and Niño (2009), have outlined nine areas of reported weakness with regard to MT such as: "grammatical inaccuracies", "literal translation", "difficulty with some idioms", and the rather vague, "errors that humans do not commit" (as cit. in Correa, 2014). For teachers all too familiar with the disparate range of essays turned in by students, the problem with this list should be quite clear; it could just as easily apply to students as it could MT.

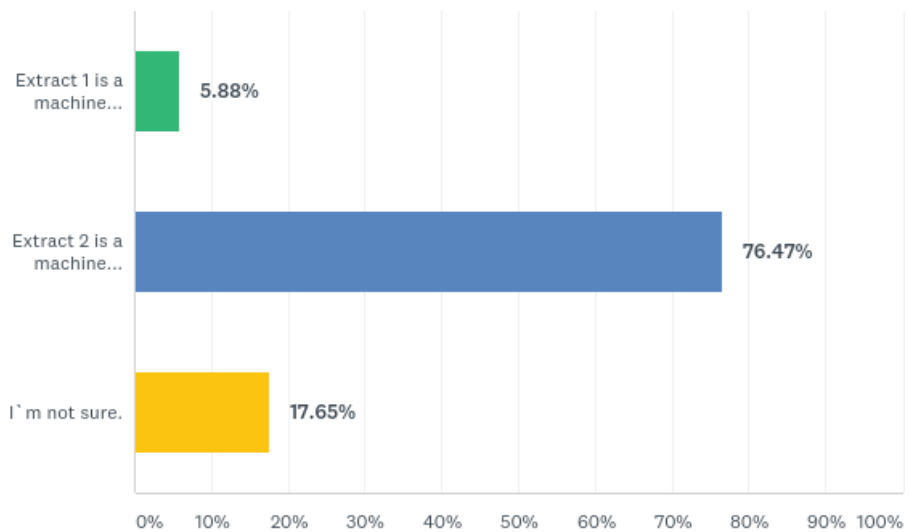
Method

Five L2 speakers of English from various work backgrounds whose native language is Japanese gave their consent to take part in the study. They were informed that their names would remain anonymous and that they were free to

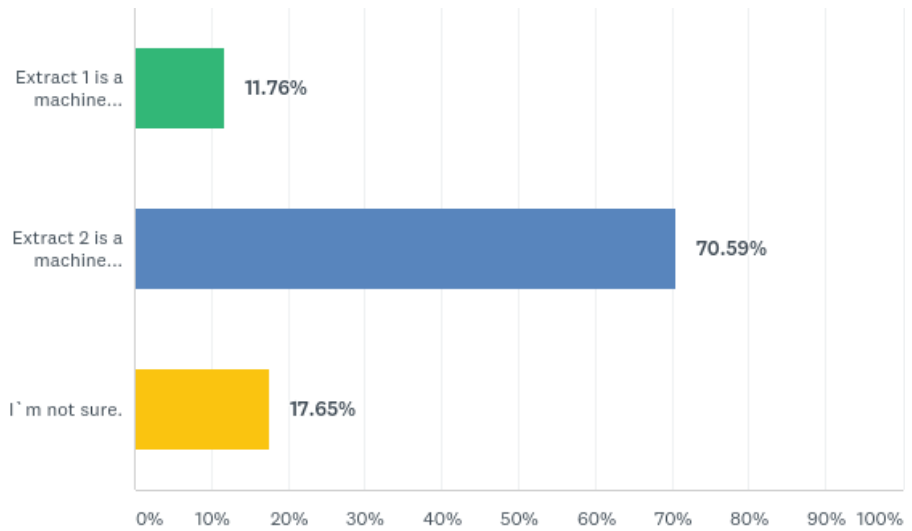
withdraw at any time. The ethical code of conduct outlined in BERA was adhered to throughout the study (BERA, 2011). The volunteers were tasked with translating one article each from the NHK website `優しい日本語で書いたニュース` while a translated version of the same story was also created using Google Translate (News Web Easy, 2019). The volunteers respective TOEIC scores for the extracts were: extract 1: 900, extract 2: 855, extract 3: 650, extract 4: 740, extract 5: 620. Using the website SurveyMonkey.com, a questionnaire asking people to choose which article they felt had been written using MT and why was sent out to native English teachers in Japan through social media and yielded seventeen responses (see appendices). Details of the survey can be found at this web link <https://www.surveymonkey.com/stories/SM-M78HX7WL/>.

Results and discussion

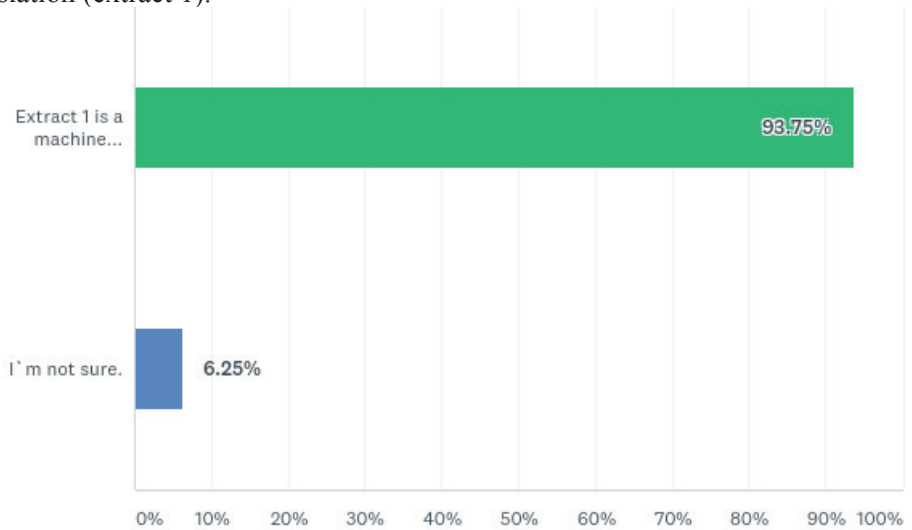
Extract one: percentage of respondents who correctly identified the machine translation (extract 2).



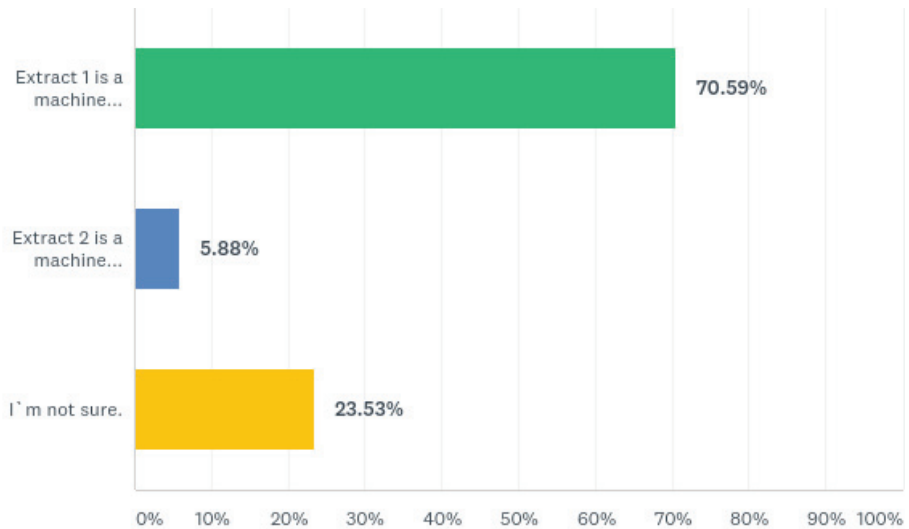
Extract two: percentage of respondents who correctly identified the machine translation (extract 2).



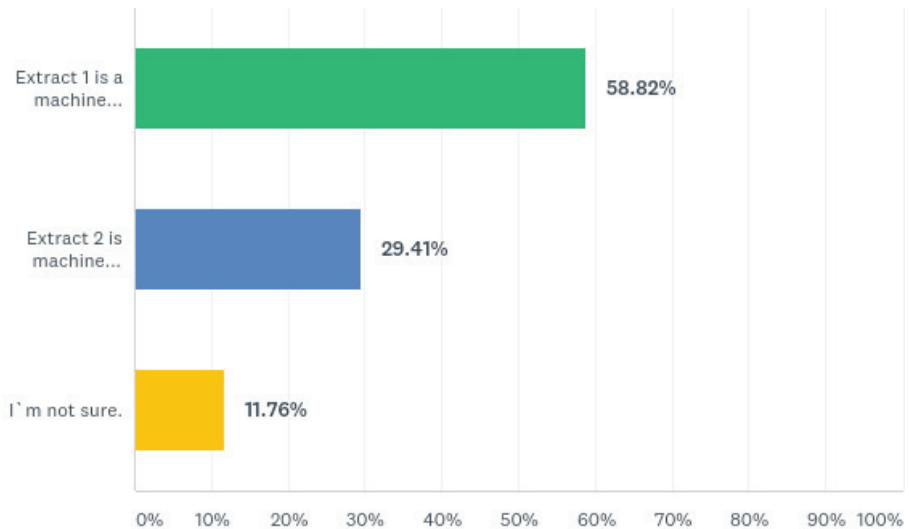
Extract three: percentage of respondents who correctly identified the machine translation (extract 1).



Extract four: percentage of respondents who correctly identified the machine translation (extract 1).



Extract five: percentage of respondents who correctly identified the machine translation (extract 1).



As can be seen from the above results, the majority of respondents were able to ascertain which extract had been written using MT, with the mean score for the five articles of 74.04% demonstrating a reasonable level of certainty. What is interesting to note is the disparity of 34.93% between the high detection rate of 93.75% in extract three, and the lower detection rate of 58.82% in extract five. While there may be various reasons for this, one possible explanation may simply

be the (understandable) matter of respondent fatigue or apathy towards the end of the task leading to a lower degree of decision-making prudence. An alternative explanation for this outlier may be the spelling mistake “students” in the MT version which contradicts points three and nine in the table below; perhaps a case of an MT red herring. In order to dig down into the reasons given for choosing one extract over the other, respondents were invited to leave comments on any salient lexical features, some of which have been highlighted verbatim as follows.

Fig 1.

Lexical features suggesting human translation.	Lexical features suggesting machine translation.
1. Common grammar mistakes made by Japanese speakers.	13. Technically correct but awkward sentences that sound like direct translations.
2. Simple sentence structure typical of students.	14. Advanced phrases not usually used by students.
3. I don't believe the machine would have gotten the spelling wrong (two similarly worded responses).	15. Inappropriate use of passive tense.
4. Seemed more typical of the Japanese learner of English.	16. Translating potential as passive.
5. I am pretty sure Google translate would not use mum' for 'mother'.	17. Unnatural use of “by all means”, suggest machine translation (two similarly worded responses).
6. Extract 1 has a suspicious ‘AI’s technology’ that is quite common in Japanese English dialect.	18. Random capital letter in first sentene (sic).
7. Typical learner errors. E.g., “in the same time”.	19. Inconsistency, “23th” and “famale”.
8. 1 contains more possible Japanese student errors.	20. The panda names change (sic).
9. Extract 2 has a spelling error that I believe Google cannot make.	21. I can't believe a human would write the sentence in extract 2 that starts “when I was born”.
10. The contraction also seems unlikely for Google.	22. Seems unlikely a student would mix up the subject here, Google Translate often does.
11. Typical learner errors.	23. Wrong pronouns.
12. Seemed to share a number of features with those typically made by my own students.	24. Inappropriate choice of subject pronouns.
	25. Unnatural use of personal pronouns.
	26. I suspect that Google can translate conditionals much more smoothly than students.

The comments above were highlighted as they were judged to be more specific than comments which could quite easily apply to either extract such as, “some of the verbs were wrong”. To a degree, these comments point to the idea of the native speaker “I know it when I see it” hunch, underpinned by the suggestion that Japanese L2 speakers of English make “typical” mistakes expressed through a kind of Japanese English “dialect” – see points 1,2,4,6,7,8, 11 and 12. With regard to the epistemological justifications used for determining

that MT was used, many respondents hedged their responses with phrases such as “I suspect”, “seems unlikely”, and “sound like” (sic) suggesting a degree of vacillation perhaps reflected less in the mean 14.12% of times “I’m not sure” was selected. Elsewhere, a few interesting red flags were pulled up such as the unusual capitalization of “Helping” (point 18), which a university level student could reasonably be expected to know not to do. A final point of interest (point 19), is the curious way in which the name of the panda in extract three changes names three times from “Aihama”, to “Ayahama”, and finally to, “Saihama”. One possible explanation for this is that – like many so called “kira kira” names which use rare kanji characters – the panda’s Chinese characters are also open to many interpretations, and Google Translate appears to have chosen them at random. It could reasonably be expected that a student would be consistent with participant names.

Passive voice.

A salient point drawn from the above comments, is the use of passive voice pointed out in points 15 and 16. Passive voice is defined by yourdictionary.com as a clause where the subject is acted on by the verb. Following an SFL approach, it would be more appropriate to say that the Goal (the participant receiving the action) comes before the process (verb), and the Actor (subject) comes after the process (Young & Fitzgerald, 2006). For example – ||the man (Goal) was bitten (material process) by the dog|| (Actor). While the question of whether active or

Fig 2.

	Use of passive voice in human translation.	Use of passive voice in machine translation.
Extract 1.	1. Kamaboko has to be place in refrigerator.	2. It could not be carried for a long time. 3. Kamaboko must be stored in a refrigerator. 4. A new fancy can be made. 5. It was made. 6. This kamaboko is sold at souvenir shops.
Extract 2.	7. He was saved. 8. [He] was rescued by a helicopter.	9. The man was helped by a helicopter. 10. After being suspended in the sea
Extract 3.		
Extract 4.	11. 10 languages can be translated.	12. The souvenir shop explanation was written in English. 13. The description of the Chinese restaurant was written in Japanese.
Extract 5.		14. Children can be contacted.

passive voice is more appropriate is not the focus of this paper, it has been marked up as a signifier of MT which warrants investigation. Looking at the following table, there is a ratio of 1 to 2.5 passive voice errors when comparing HT with MT respectively. A degree of leniency was necessary in order to make judgment calls on which clauses could fully warrant being called passive voice due to the inevitable nature of the un-grammatical clauses. For example, while the material process in the clause ||kamaboko has to be place in the refrigerator|| omits the necessary past participle, it was judged to be sufficiently different from the active process form ||you have to place the kamaboko in the refrigerator|| to warrant status as a passive voice error.

Inappropriate pronouns

The second salient point which emerged from the teacher comments was the inappropriate usage of pronouns (see points 19–25). While it is generally possible to rearrange the participants (who is involved), processes (verbs), and circumstances (where the action takes place) in a clause; the theme – or what comes at the start of the clause complex is less malleable. This is drawn out by Coffin, Donohue and North (2009) who point out the lack of opportunity to use voice, gesture or context to supplement meaning in written text, or for the interlocutor to interrupt or press for clarification as significant differences vis-à-vis spoken discourse. Theme is defined as extending to and including the first ideational element in a clause – in other words, the first process, participant or circumstance (ibid). As can be seen from the table below, the latter of these three – the participant (highlighted in italics), has been flagged up as inappropriate and confounding the speaker's message. A total of thirteen inappropriate participants were identified vis-à-vis two for the human translation.

While a university level student might also be expected to confuse pronouns in their writing, the above analysis shows that – in this case – Google Translate does so with more frequency. The text's lexical coherency is confused further through the way the participant in the non-finite dependent clauses change as follows, ||A man on the sailboat threw a lifejacket|| that allowed me to float in the water|| but the men couldn't catch it|| (points 2 and 3).

Discussion

Today's so-called digital immigrants (Prensky, 2001) may be entering a workforce where pragmatism in communication through apps and algorithms takes precedence over the ability to converse without recourse to the digital world. For classes focusing on higher order thinking skills however, the wholesale outsourcing of mental effort to computer software would appear to defeat the object of student's ability to think in English. For educators looking for red flags which suggest the use of MT in student work, the present study found some evidence that MT and HT differ in certain

Fig 3.

	Inappropriate participant use in human translation.	Inappropriate participant use in machine translation.
Extract 1.		1. It is OK without putting <i>it</i> in the refrigerator.
Extract 2.		2. [A man on the sailboat threw a lifejacket] that allowed <i>me</i> to float in the water. 3. The <i>men</i> couldn't catch it. 4. <i>Men</i> immediately took off the jeans. 5. <i>I</i> put air in my jeans. 6. <i>I</i> knew that I could use my pants instead of a life jacket. 7. <i>I</i> thought that I could not help if I did not have jeans. 8. <i>I'm</i> glad I had no holes in my jeans.
Extract 3.	9. Then showed <i>us</i> in a outside playground.	10. We have shown <i>it</i> at the playground outside. 11. <i>I</i> was born. 12. <i>I</i> grew up to 12kg. 13. I was very happy to see Saihama playing with <i>my mother</i> .
Extract 4.		
Extract 5.	14. <i>The idea of Osaka</i> made a decision.	15. <i>The country</i> could also bring a mobile phone to school.

respects. Usage of passive clause constructions was found to be used at a ratio of 1 to 2.5, and inappropriate participant choice at a ratio of 1 to 6.5 when comparing HT with MT respectively. MT was able to translate conditionals (point 21) much more smoothly than students, and use some more sophisticated phrasing which may be out of place in a lower level student's essay (point 14). Until software is developed which can distinguish HT from MT, continued identification of red flags will enable educators to have greater confidence with regard to the corporeality of the author.

References

- Aharoni, R., Koppel, M., & Goldberg, Y. (2014). Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 289–295, Baltimore, Maryland, June 23–25, 2014. Association for Computational Linguistics.
- Benda, J. (2013). Google Translate in the EFL Classroom Taboo or Teaching Tool? *Writing & pedagogy*. Retrieved March 14, 2019 from <https://journals.equinoxpub.com/WAP/article/view/19968>. DOI: 10.1558/wap.v5i2.317.
- BERA (2011). Retrieved April 5, 2019, from <https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf?noredirect=1>.
- Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use, perceptions, and analyses. *Jalt call journal*, 14(1), 3–24.

- Çakır, S. (2013). A Study on the efficiency of the Google Translate translation program. In I. Özyıldırım, S. N. Büyükkantarçioğlu, E. Yazar, & E. Alpaslan (Eds.), *Kırkıncı Yıl Yazıları* (pp. 75–83). Ankara: Hacettepe Üniversitesi Yayınları.
- Clifford, J., Merschel, L., & Reisinger, D. (2013). Meeting the challenges of machine translation. *The Language Educator*, 8, 44–47.
- Coffin, C., Donohue, J., & North, S. (2009). *Exploring English Grammar From Formal to Functional*. London: Routledge.
- Correa, M. (2014). Leaving the “peer” out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content and Language Integrated Learning*, 7(2), 1–20 DOI:10.5294/laclil.2014.7.1.1.
- Crystal, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.
- Davies, R. J., & Ikeno, O. (2002). *The Japanese Mind: Understanding Contemporary Japanese Culture*. Clarendon, VT: Tuttle Publishing.
- Ebbert-Hübner, C. & Maas, C. (2017). Can Translation Improve EFL Students’ Grammatical Accuracy? *International Journal of English Language & Translation Studies*, 5(4), 191–202.
- Garcia, I., & Pena, M. I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 24(5), 471–487. DOI: 10.1080/09588221.2011.582687.
- Gibbons, P. (2008). “It was taught good and I learned a lot”: Intellectual practices and ESL learners in the middle years. *Australian Journal of Language and Literacy*, 31(2), 155–173.
- Groves, M., & Mundt, K. (2015). *Friend or foe? Google Translate in language for academic purposes. English for Specific Purposes*, 37, 112–121. Retrieved April 5, 2019, from <https://www.sciencedirect.com/science/article/pii/S088949061400060X#>.
- Hall, E. T. (1976). *Beyond culture*. New York: Doubleday.
- Kazemzadeh, A., & Kashani, A. (2014). The effect of computer-assisted translation on L2 learners’ mastery of writing. *International Journal of Research Studies in Language Learning*, 3, 29–44.
- Lee J., & Liao, P. (2011). A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation and Translation Review*, 4(2), 105–149.
- Lommel, A. (2019). *tcworld.info – translation and localization*. *Tcworld.info*. Retrieved March, 30, 2019 from <http://www.tcworld.info/e-magazine/translation-and-localization/article/neural-machine-translation-offers-significant-advances-with-remaining-challenges>.
- McGuire, N. (2018). *How accurate is Google Translate in 2018? ARGO Translation*. Retrieved March 12, 2019, from <https://www.argotrans.com/blog/accurate-google-translate-2018>.
- News Web Easy* (2019). Retrieved April, 15, 2019, from <https://www3.nhk.or.jp/news/easy/k10011858001000/k10011858001000>.
- Prensky, M. (2001). Digital Natives Digital Immigrants.. Part 1. *On the Horizon*, 9(5), 1–6.
- Schuster, M., Johnson, M., & Thorat, N. (2016). *Zero-Shot Translation with Google’s Multilingual Neural Machine Translation System*. Retrieved March 3, 2019, from <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>.
- Sugiyama, S. (2019). *All aboard the Sakai Muscle Line? Osaka Metro axes foreign language website after botched translation. From Japan*. Retrieved February 10, 2019, from <https://www.fromjapan.co.uk/news/all-aboard-the-sakai-muscle-line-osaka-metro-axes-foreign-language-website-after-botched-translation/>.
- SurveyMonkey.com*. (2019). SurveyMonkey – Free online survey software and questionnaire tool. Retrieved April 15, 2019, from <https://www.surveymonkey.com>.
- Upwork.com*. *What is Translation Plagiarism and How do You Detect it?* Retrieved March 13, 2019 from <https://www.upwork.com/hiring/for-clients/translation-plagiarism-detection/>.
- White, K. D., & Heidrich, E. (2013). Our Policies, Their Text: German Language Students’ Strategies with and Beliefs about Web-Based Machine Translation. *Die Unterrichtspraxis/Teaching German*, 46(2), 230–250.

- Van Praag, B., & Sanchez, H.S. (2015). Mobile technology in second language classrooms: Insights into its uses, pedagogical implications, and teacher beliefs. *ReCALL*, 27(3), 288–303. DOI: 10.2017/S0958344015000075
- Vygotsky, I. (1978). *Mind in society: Development of higher psychological processes*. Cambridge: Harvard University Press.
- Wu, Y., Schuster, M., Chen, Z., Le, Q., & Norouzi, M. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. Retrieved March 18, 2019, from <https://arxiv.org/abs/1609.08144>.
- Young, L., & Fitzgerald, B. (2006). *The Power of Language How Discourse Influences Society*. London: Equinox.